Natasha Barrett

Ambisonics and acousmatic space:
a composer's framework for investigating spatial ontology

Blåsbortvn. 10, N-0873 Oslo, Norway

nlb@natashabarrett.org

Abstract

This paper investigates how ambisonics, when embedded into a compositional methodology, transcends its function as a technical tool and influences our perception, interpretation and physical relationship to sound archetypes. The ideas address the core ontology of spatial sound in composition and in listening, laying a framework for the development of an acousmatic spatial ontology.

1. Background

We are far from a complete understanding of spatial ontology. Little has been theorised on an aesthetical level in terms of three-dimensional[1] sound projected over loudspeakers and removed from visual causation, nor have ideas which in the stereo field are restricted to concept been significantly explored in the reality of 3D. Kendall (2008) points out that, "conceptual terminology can be out of alignment with the technical capacities of spatialisation systems, out of alignment with the actual experience of spatial sound and therefore lead to under-utilization in electroacoustic composition". In an interview some years ago, (published in Otondo 2007), I also suggested that composers haven't caught up with what technology offers, and in aesthetical and creative respects this is still true.

1.1 A general problem in acousmatic listening: spatial tangibility

Despite everyday experiences of multi-media sources the reception of acousmatic music remains problematic. We can therefore question whether an underlying obstacle concerns the lack of visual information, nor a negative sense of disembodiment. Instead, reception can be hampered by an inability to enter acousmatic space-form as a tangible construct – a construct that for the composer, in the perfect listening position, may be natural and apparent. In works involving visual imagery the shared or shifted perceptual focus allows our senses to bypass or ignore the sound-space dilemma. In a conventional acousmatic stereo or multi-channel projection, our relationship to acousmatic space-form contradicts our understanding of the real world.

The problem of spatial tangibility is manifest to different extents for different types of work. In works intending a direct mimesis of a specific environment, the problem may involve a simple case of insufficient spatial information in terms of envelopment[2]. In acousmatic composition, normally interested in only partial mimesis of real-world features,

---

[1] In this paper I used the term '3D' to describe the creation of a spatial continuum, within which sound images exist independently from the loudspeakers as points with the capacity to mimic spatial reality. Even though most ambisonics systems are concerned with sound projected over a horizontal array, the term 3D is here used to differentiate from stereo phantom images and conventional surround sound.

[2] See section 4.3 for an explanation of envelopment.

the problems are multiplied in terms of gesture, agency, source-bonds and their abstractions, spatial archetypes and prototypes, and the physical real-time relationship between spatial sound and listener. As listeners, we need to consciously project ourselves into the material sound world before we can get to grips with the composition. 'Personal projection' is clearly the first hurdle for many listeners. In Smalley's article "Space-form and the acousmatic image" (Smalley 2007), the evocative text depiction of the Orbieu soundscape would tantalise internal images for many a reader, yet to achieve such in sonic art without 3D sound is, by virtue of the complex couplings of sound, time and space, upon which the description relies, another matter.

1.2 Assessing time and spatial perception: two temporal streams

Diverse studies in both the sciences and the humanities illustrate the immediacy of spatial information in the listening experience. To cite a few examples, we see that tests on how spatial audio influences both visual and sound experiences in video conferencing shows an increased memory, improved focal assurance and comprehension (Baldis 2001). Miranda (2010) points out that the processing of spatial location occurs early on in the auditory system before ascertaining sound identity in terms other than space. The immediacy of spatial perception can also be viewed from an ecological perspective. As discussed in Lennox (2001), Gibson proposed that the 'direct perception' of 'environmental affordances' is achieved without the need for intervening cognitive processes. The environment stimulated by sounding objects (termed 'ambience labelling') provides the means for 'direct perception'. In music theory, Smalley (2007) suggests that, "Possibly the most important strategy in arriving at an holistic view of the space-form ... is that I disregard temporal evolution: I can collapse the whole experience into a present moment, and that is largely how it rests in my memory." Smalley goes on to say that thought and experience combine to allow us to deduce the space and scene in question.

So far we can see that science and aesthetics agree that our understanding of spatial information is primarily immediate and less concerned with temporal processes. Yet the second part of Smalley's discussion presents a contradiction in pointing out that identity unfolds through time and deduction is a slow process contingent on, rather than disregarding time. Likewise, Emmerson's concepts of 'space frames' (Emmerson 2007 pg 92) are described as both specific time-independent frames (event and stage, arena and landscape), yet malleable within musical discourse.

If we are to address spatial ontology we therefore need to take a broader view that divides spatial perception two temporal streams: (a) the immediate effect on the listener, which is both instant and timeless, preceding 'transmodal' deductions – for example 'prototype' spaces, and (b) that which develops, unfolds and provokes temporal engagement. In sound example 1a we hear a Soundfield recording from a small, inner city park. This sound space can be heard as a prototype image where, in everyday experience, many encounters of this familiar image collapse into one timeless prototype. In sound example 1b, we hear a Soundfield recording from the exact same location but with an unusual addition: a bicycle with rattling ornaments is pushed across the scene. The additional information provokes temporal engagement of the new sound object and the changing space-environmental cues stimulated by the moving sound object.

On hearing a new work listeners often experience not being able to keep up with the real-time flow of information, but are able to grasp the gist of what they hear. Timeless spatial perception may therefore be that most memorable, and in this regard space itself is not an 'event', but a 'frame' to group events. An understanding of temporal spatial development on a musical structural level is more likely learnt through experienced listening.

2. A method that allows us to investigate spatial ontology: The 3D projection of spatial images.

Let us first consider the projection of spatial information to the listener without complicating the discussion in terms of sound identity or source bonding. Through the history of electroacoustic composition various practical techniques to manipulate the implied size, depth and distance of a stereo phantom image have been explored. These images are then interpreted in sound diffusion, which although advantageous in terms of performance practice fail to satisfy the dimensions of spatial reality in which we are interested[3]. If we wish to use spatial reality as a way to engage the listener we are interested in communicating the size, shape, distance, action and interaction of sounding objects, and the context of environmental cues. Only then can we address real spatial archetypes and subsequently our physical relationship to the sound-space. In other words we need control over the 3D sound field.

The investigation of three-dimensional sound has intensified in recent years. Ambisonics and wave field synthesis are the two most used systems. In contrast to wave field synthesis, ambisonics allows for a realistic number of loudspeakers and also provides a portable encoded format (wave field synthesis relies on real-time spatial control and projection where encoding could feasibly involve storing mono sources and spatial meta data). The advantages of the encoded format become apparent with use; by combining the possibilities of both real-time and non-real-time image manipulation, it yields greater control over sound image and spatial depth, the composer can encode spatial information to be reproduced without the need for performance interpretation and carry out the creative work monitoring over a smaller studio set-up. Although both wave field synthesis and ambisonics allow for real recorded scenes, only ambisonics B-format Soundfield recordings are currently practical for field recording. Ambisonics is therefore a favourable tool. In the artistic milieu, ambisonics is mainly understood in terms of these technical solutions to the spatial limitations of stereo or conventional surround and multi-channel.

2.1 A hybrid approach to ambisonics

For any single approach to ambisonics there exist limits to what is realistic in both technical and artistic realms. For the purpose of investigating spatial ontology, we are interested in the success of the recorded or synthesised source in capturing the intended information and the success of the decoding method in recreating that information. A hybrid approach that combines features of different encoding and decoding techniques overcomes many of these problems.

Without entering too deeply into a technical description, ambisonics is a system of sound reproduction based on the spherical harmonics of the sound field. The more spherical harmonics, or in other words the higher the order of representation (high order ambisonics, or HOA), the greater the spatial resolution over a larger listening area. Basic ambisonic encoding (rather than recording) involves the directional encoding of plane waves and neglects the near field effect of finite distance sources. In the real world, when a sound is located in near-field the wave-front curvature depends on the sound's distance. The curvature effects how the listener perceives parallax, relative and absolute distance. Soundfield recordings capture the real sound field, but with a low spatial resolution and a small sweet spot. HOA allows some control over these acoustic (rather than extrinsic) distance cues, and combined with the spatial

[3] In a conventional approach to acousmatic composition listeners are often assumed to invent a place that may be richer than the sounding information. (For an interesting embellishment see Kim 2010). 3D sound provides the composer with a choice: to create a complete set of real spatial cues, or to reduce spatial information to be invented in the listening process.

resolution over a large listening area we may ask why not work in as high order as possible? Artistically and technically there are trade-offs:

Technical tradeoffs: There remain practical problems in synthesising the near-field information. Besides needing a very high-order representation for the effect to be significant, extreme energy in the low frequencies is necessary to recreate the sound field. Although the bass cancels out towards the centre of the array, on the edges remains problematic. (For a technical summary see Daniel et. al. 2003 and 2004). Also, room reflections will serve to enhance the bass and confuse the wave reconstruction process.[4] Even if we avoid encoding the near-field information (most ambisonics encoders avoid this) and just take advantage of the spatial resolution and larger sweet spot, the encoded format must be decoded over a minimum number of matched loudspeakers, evenly spaced on the circumference of a circle. For a horizontal array, this number is two times the order plus two ($2m+2$). So for 3rd-order ambisonics we already require a minimum of eight identical, suitably located loudspeakers. Higher order formats also consume exponentially increasing computer resources. We can however encode in as high order as we wish: a convenient feature of ambisonics means, that if we lack enough loudspeakers, we can decode just the lower order components by using the correct decoder.

Artistic trade-offs: HOA requires the synthesis of spatial information[5] and the simplest approach involves spatialising point sources by hand or automated with control data. Sound example 2 is an extract from "Tensile Elements" (Hofmann, 2000) created from the multiple trajectories of point sources. The intention here is not to create real sound images of spatial extent, rather to project behaviour and gestural trajectories within a space defined by motion rather than environment. Although this approach can be successful, as is clear in the sound example, *real* sound images are rarely dimensionless points (image extent in spatial experience has been investigated in many sources such as Martens 2003, Rumsey 2001, Berg 2003, Ford 2003). Creating an image of size and shape presents a challenge. Positioning the two channels of a stereo recording within the ambisonics sound field is insufficient: the image tends to disassociate. More recently various solutions have been explored. Some documented approaches involve multi-band time varying temporal decorrelation applied before the spatialisation process (e.g. Potard 2004), along with methods for simulating the diffuse field and controlling a 1st-order image (such as experimental approaches described in Anderson 2009 and commercial room acoustics modelling software). My own techniques are threefold (detailed in Barrett 2010): recording with microphone arrays and synthesising the image from these signals, stochastically controlled asynchronous granulation incorporating amplitude and air absorption coefficients, and ambisonics impulse response convolution. Sound examples 3-5 serve as illustrations (refer to the examples appendix for content).

Recorded 1st-order sources, although of a lower spatial resolution, capture 'reality information' from the original source, including proximity, image size, the real balance between sound events in the original scene and a unified 3D image of fore- mid- and background – or in other words the sense of environment and perspective, which although in reality is complex and precise, for our perception involves a blended field of sound. In theory we can therefore combine 1st-order and HOA, taking advantage of their features:[6] Soundfield recordings capture spatial reality at a low resolution while HOA adds precise spatial control and articulation for perceptual and musical ends. The auditory contrast between 1st-order recording and HOA synthesis also highlights size and shape, influencing the way in which

---

[4] The author is currently testing various solutions as well as investigating the degree to which acoustic proximity information is important over extrinsic proximity information.

[5] Recorded sources are currently restricted to a 1st-order representation even though higher-order prototype microphones exist.

[6] Different decoding questions are required for different mathematical orders. 1st-order materials can be decoded over larger numbers of loudspeakers via newly developed non-linear decoding methods. 'Up-mixing' from a 1st to HOA can reduce the need to mix decoders, but current methods produce audible artefacts.

spatial behaviors and relationships interact. In example 6 we can hear that even for a first-order representation, the synthesis of point sources is in clear contrast to the Soundfield recorded materials. Furthermore, the perceptual significance of the distorted image from off-centred listening in a 1st-order representation is reduced when anchored by the HOA's larger sweet spot.[7]

3. Addressing spatial ontology in hybrid ambisonics

3.1 Distance, proximity and the listener

In a traditional stereo or multi-channel image, a clear source-bond combined with the correct psychoacoustic cues is vital in understanding distance, yet at best can only *imply* distance. For ambisonics, proximity in acoustic terms can be combined with, and enhanced by, psychoacoustic cues: the relationship between source identity, spectrum, volume, image size and realistic motion relating the sound image to its environment. Proximity cues are also embedded in Soundfield recordings such that a straight-forward combination of close microphone and ambient recordings will yield distance differences to the listener. In using these distance cues in conjunction with sound identity we can influence a listeners' understanding of where they are located with respect to the source.

When including acoustic proximity, distance cues and source-bonding collaborate in both directions. In one direction, understanding the sounding material (for example metal, wood, stone) may be enough such that, if the listener has some idea even if not an exact idea of the source, loudness and image size will indicate proximity. In sound example 7a (an extract from "Metal Clouds", Hofmann, 2001), the synthetic sound sources are on the whole extremely abstract. Lacking any source bond they linger relatively close to the listener. Sometimes the material sounds similar to a high-flying jet aircraft, where sound identity adds the sense of distance even though the spectrum is essentially identical to the abstract material. Sound example 7b (an extract from "Untitled-2", Barrett, 2010) is a collage of Soundfield recordings. Distance relationships are is expansive, combining near-field effects and source bonds (identifying bees, a plane, and their appropriate behaviours as stabile aspects of a complete sound picture).

In the other direction, for highly ambiguous sources a clear sense of physical proximity will anchor the sound location, if not in absolute terms in the relationship between other spatial anchors – or in other words distance contrast - and influence how we interpret sound identity.

In being able to control acoustic and psychoacoustic proximity we can explore the trickier real-world contradictions between perceptual and physical proximity. Lennox (2007) explains how perceptual and physical "near" may differ, giving the example that a physically near item in an adjacent room is perceptually further than items in this room. Action also takes on an important roll: 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.

3.2 Gesture, agency and a projection of the self

3.2.1 'Causal agency' and 'social distance'

A sound has to be activated; causation leads to 'agency'. Combining real physical distance with causal agency in 3D space evokes 'social distance'. Blesser and Salter (2007 pg 34)

[7] Although no formal tests exist the effect is clear in personal study.

suggest that, "Understanding the social rules for acoustic arenas requires the concept of social distance..." They draw attention to Hall's (1966) idea of dividing social distance into four spheres: intimate sphere (0.5 meters), personal sphere (1 meter), conversational sphere (up to 4 meters) and public sphere (determined by the acoustic horizon). For our practical ambisonics acousmatic application, real distance on a human scale is embedded in the material and can connect directly to the definitions of social spheres. Furthermore, the strength of the ambisonics image avoids disturbance of the real "arena" (Emmerson 1998 and Smalley 2007) in all but the most reverberant spaces, and the listener more easily enters the arena of the composition.

3.2.2 Self-projection

In 'human' terms Emmerson (2007 pg 62) discusses the body in music: heart beat, breath, voice, exchange, touch, proximity and movement. To explain how 'self projection' is contingent on 3D sound we need to abstract the idea of agency in two directions.

Firstly, in the current discussion is it interesting to abstract from human bodily sounds and consider the links between sonic imagery and movement imagery. Godøy (2010, 2006) explains that human mental activity is intimately linked with sensations of movement. Mental simulation of sound-producing gesture can involve the transfer of energy from the human body to some resonating medium rather than necessarily being the action of playing a specific instrument. In other words, motor theory, linking perception to the mental simulation of sound-producing actions, can be used on a general level where sounds triggers images of actions and vice versa. The crucial point is that we have a generalised schemata for sound-perception which we use in understanding familiar and unfamiliar sounds.

Secondly, we can add the idea of 'personal agency' as enlarged by Casey (1976 pg 45), who sets the active act of 'imagining-how' against 'imagining-that'. "To imagine-how is to imagine what it would be like to do, think or feel so-and-so, or to move, behave and speak in such-and-such ways.... realised by entertaining an imagined state of affairs in which he is envisaged as himself an active and embodied participant..." In other words, there is a sense of personal agency, of the imaginer's own involvement in what is being imagined. "... the imaginer has also projected himself (or a surrogate) as an active being who is experiencing how it is to do, feel, think, move etc. in a certain manner." Casey adds a footnote: "This is not only in terms of action activity, but includes a broad spectrum of ways in which the imaginer becomes implicated via self-projection or by proxy in his own imaginative presentation". In general terms, Kendell (2010) points out that in the everyday world, listeners identify agency with the cause of the changes of state associated with the event. In the case that the agency of an event is unknown or unknowable, there are potential default values. Although Kendell goes on to say that for many listeners of electroacoustic concert music, the default agent is the composer, in the 3D context outlined so far the default *agency* behind certain 'performed' sounds may be a projection of the self (the listener).

We can conclude that the listener's connection to the sound involves not only 3D acoustic and psychoacoustic distance cues, sense of agency behind the material and social distance in relation to the arena. These features together provoke a physical projection of the 'self' in the two temporal-spatial streams outlined above: through immediate (3D scene) and time-dependent (agency and social distance) spatial constructs, as part of the listener's creative consciousness in the acousmatic ambisonics experience.

3.2.4 The need for an accurate spatial recreation

To place the above ideas in a 'real', non-conceptual realm relies on the spatial projection of suitable spatial imagery and a clear mimesis of real-world spatial archetypes (rather than necessarily a mimesis of all aspects of sound identity). Furthermore, in composition, if we are to successfully embody these ideas we need an accurate spatial recreation rather than an interpretation. To give an example, in Klang (Harrison 1982), in the opening moments of unprocessed sound and clear imagery, we understand the object's material (metal) and maybe its geometry. But further, we understand a human agent behind the articulation and a sense of 'performance gesture'. In sound diffusion, my own spatialisation interpretation attempts to project this sense of 'intimate performance' on a virtual stage. But the result is fragile and possibly only clear from where I sit. I have heard many other interpretations that consider other aspects than agency and social distance. In the projection of a stereo image, we have to accept such interpretations and the ensuing effects on the listeners' conception of the work. Encoding and then decoding spatial information is the only way to approach the truth behind the work as composed. Example 8 presents the opening of the ambisonics work "Pacific Slope" (Anderson, 2002). Here we find many similarities to the opening of Harrison's work "Klang", such as image, physical materials and gesture. Anderson's choice of ambisonics spatialisation leads to a stable reproduction of the intent behind the sound; its azimuth location, image size and psychoacoustically implied distance, implication of stage, human scale and agency in relation to the listener.

3.3 Breaking the paradox between reality and fiction: envelopment and immersion

In accurately projecting imagery and distance we can create the sensation of 'really' being in a 'real' space on a *human scale*. In 3D sound the sense of being within the sound field is called 'listener envelopment'. Listener envelopment has been discussed at length in connection to room acoustics and laterally reflected sound. Berg et. al (2003) make a distinction between 'room envelopment' – or the extent to which we feel surrounded by the reflected sound, and 'source envelopment' - or the extent to which we feel surrounded by the sound source(s). For source envelopment, the listener is aware of directional cues and is either looking out 'towards' or is 'inside' a scene, which may consist of a complex array of directional and diffuse information from sounding objects and reflections from the environment. In terms of sound field recreation we 'really' feel we are in a 'real' space. An obvious example is an untreated Soundfield recording of an outdoor environment. Example 9 presents a composed version where only the most salient features of the original recording remain, yet still concerns a combination of room and source envelopment.

'Listener immersion' is less clearly formulated. It involves the listener being convinced they are in a real space, yet the scales are *non-human and in reality impossible.* This approach is in contrast to Paine (2007) who suggests a difference between engulfment and envelopment, where he argues that engulfment is a factor of vertical information. However, his test set-up involved normal vertical-horizontal playback. Also, Emmerson (2007, pg 161) suggests that high sound levels and omni-directionality creates an immersive sound field. I would describe this effect as saturation rather than as immersion, unless significant temporal-frequency colouration was introduced by the real listening space serving to destroy the sense of omni-directionality (which is less interesting for composition due to the lack of control). In contrast to envelopment, immersion is directionally less defined. It is however dependent on spatial-temporal frequency distribution of a specific low-mid range. Excessive bass removes the effect, as does the classic method of 'zooming' in a B-format sound using W-panning (Menzies 2002). This is in line with theory suggesting variation in spectral content is necessary to spatially distinguish between front and rear hemisphere (Blauert 2001). The need for these irregularities means immersion can be creatively controlled, rather than simply an amorphous experience. In the stereo field I remember striving for immersion using countless tricks to create a fictional sea of spectral wash, all-encompassing to hearing and body, only to find the image collapsed as soon as I moved out of the perfect sweet spot. Multi-channel projection was insufficient: each loudspeaker locking the sound rather than freeing it into the space.

Paradoxically, ambisonics *real recordings* can be used to create *surreal immersion*. My own method involves placing a Soundfield microphone inside a resonating body and capturing the internal sound field. This method is similar to conventional close microphone recording techniques while yielding immersion, stability and reproducibility. The decoding technique also enters into the method; regardless of whether horizontal only, by decoding the dominating lower frequencies optimizing for directional cues using phase, spatial artifacts outside the central listening position serve to confuse direction information and enhance the sense of immersion (contrary to textbook decoding practice). Immersion is a good example of where a sounding object becomes a sounding space. (Examples 10a and 10b). The ambisonics field gives the space a sense of reality and plausibility, even though impossible on a human scale.

Conclusion

A hybrid approach combining 1st-order recorded sources and HOA synthesised materials allows us to sidestep theoretical obstacles in the control of 3D space. Our future technology may hold complete solutions, but for now we can combine the features of what is currently realistic in terms of direction, distance and identity of the 3D image. Once in that position, the stability of the spatial image and its direct transmission from composer-intent to listener-reception allows us to address musical spatial ontology in new directions:

- A 3D acousmatic experience allows the listener to immediately enter the sound world as a tangible and real entity.
- Spatial perception can be evaluated in terms of immediate and temporally dependent streams. Each stream yields different spatial constructs. The immediate stream links directly to real-world space, while the temporal stream is closer to traditional acousmatic musical concepts.
- An approach that captures and projects real-world images can be extended to make 'apparently real' non-real sources through real spatial archetypes.
- Extending these archetypes further in a 3D image allows the listener to find a sense of place inside the sound world. It allows understanding of being a spectator or actor in the scene and a relationship to sound-space through physical or social distance on a human scale.
- When we compare experiences of envelopment and immersion we see that physical-spatial reality is possible for unreal sounds and unreal spaces. There are likely more ways in which spatial reality can serve surrealism and surrogacy in a compositional context beyond the stereo image.

Although outside the scope of this paper, the discussion clearly implies the need to readdress acousmatic compositional structural procedures in an ambisonics acousmatic context.

References

Anderson, J. (2009). "The Ambisonics Tool Kit". In *Proceedings of the 1st International Symposium on Ambisonics.*

Baldis, J. (2001). "Effects of spatial audio on memory, comprehension, and preference during desktop conferences". *Proceedings of the SIGCHI conference on Human factors in computing systems.* Seattle.

Barrett, N. (2010). "Kernel Expansion: A Three-Dimensional Ambisonics Composition Addressing Connected Technical, Practical and Aesthetical Issues". In *The Second International Symposium on Ambisonics and Spherical Acoustics.* Paris.

Barrett, N. (2003). "Spatio-musical composition strategies". *Organised Sound* 7(3). Cambridge University Press.

Berg, J., Rumsey, F. (2003) "Systematic Evaluation of Perceived Spatial Quality". In *AES 24th Conference.* June 2003.

Blauert, J. (1974, rev. 2001). Spatial hearing. MIT press.

Blesser, B. Salter, L. (2007). Spaces speak, are you listening? MIT Press.

Casey, E. S. (1976). Imagining: A Phenomenological Study. Bloomington: Indiana University Press.

Daniel, J. Nicol, R. Moreau, S. (2003). "Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging". *Audio Engineering Society Convention Paper 5788 Presented at the 114th Convention*

Daniel, J. Moreau, S. (2004). "Further Study of Sound Field Coding with Higher Order Ambisonics" *Audio Engineering Society Convention Paper Presented at the 116th Convention 2004 May 8–11 Berlin, Germany*

Emmerson, S. (ed) (1986). The Language of Electroacoustic Music. Palgrave Macmillan.

Emmerson, S. (1998). "Aural landscape: musical space". *Organised Sound* 2(3). Cambridge University Press.

Emmerson (2007). Living Electronic Music. Ashgate Publishing Limited.

Ford, N; Rumsey, F; Nind, T. (2003). "Evaluating Spatial Attributes of Reproduced Audio Events Using a Graphical Assessment Language - Understanding Differences in Listener Depictions". In *AES 24th Conference.* June 2003.

Godøy, R. (2010). "Images of Sonic Objects". *Organised Sound* 15(1). Cambridge University Press.

Godøy, R., Haga, E., Jensenius, A. (2006). "Exploring Music-Related Gestures by Sound-Tracing. A Preliminary Study". In *2nd ConGAS International Symposium on Gesture Interfaces for Multimedia Systems.* Leeds.

Hall, E. (1966). The Hidden Dimension. New York: Doubleday

Harrison, J. (1982). "Klang". *empreintes* DIGITALes IMED 0052.

Kendall, G. (2010). "Meaning in Electroacoustic Music and the Everyday Mind". *Organised Sound* 15(1).

Kendall, G., Mauricia, A. (2008). "The Artistic Play of Spatial Organization: Spatial Attributes, Scene Analysis and Auditory Spatial Schemata". *Lecture Notes in Computer Science.* Springer Berlin / Heidelberg.

Kim, S-J. (2010). "Imaginal Listening: a quaternary framework for listening to electroacoustic music and phenomena of sound-images". *Organised Sound* 15(1). Cambridge University Press.

Lennox, P., Vaughan, J., Myatt, T. (2001). "3D Audio as an Information environment: Manipulating Perceptual Sigificance for Differentiation and Pre-selection". In *Proceedings of the 2001 International Conference on Auditory Display*. Espoo, Finland.

Lennox, P., Myatt, T. (2007). "Concepts of Perceptual Significance for Composition and Reproduction of Explorable Surround Sound Fields". In *Proceedings of the 13th International Conference on Auditory Display*, Montréal, Canada.

Martens, W. (2003). "Perceptual evaluation of filters controlling source direction: Customized and generalized HRTFs binaural synthesis". In *Acoustical Science and Technology*, 24(5).

Menzies, D. (2002). "W-Panning and O-Format, Tools for Object Spatialization". In *Proceedings of the International Conference on Auditory Display*.

Miranda, E. (2010). "Organised Sound, Mental Imageries and the Future of Music Technology: a neuroscience outlook". *Organised Sound* 15(1). Cambridge University Press.

Otondo, F. (2007). "Creating Sonic Spaces: An Interview with Natasha Barrett". *Computer Music Journal.* 31(2). MIT Press.

Paine, G. (2007). "Perceptual Investigation into Envelopement, Spatial Clarity, and Engulfment in Reproduced Multi-Channel Audio". In *Proceedings of the 31st AES Conference: New Directions in High Resolution Audio*.

Potard, G., Burnett, I. (2004). "Decorrelation Techniques for the Rendering of Apparent Sound Source Width in 3D Audio Displays". In *Proceedings of the 7th Conference on Digital Audio Effects*.

Rumsey, F., Berg, J. (2001). "Verification and correlation of attributes used for describing the spatial quality of reproduced sound". In *Proceedings of the 19th AES Conference.* June 2001.

Smalley (2007). "Space-form and the acousmatic image". *Organised Sound* 12(1). Cambridge University Press.

Sound examples:
Download from: http://users.notam02.no/~natashab/ems2010/

Sound examples are provided in formats appropriate to common listening environments:

- B-format (WXYZ, FuMa weighting – decode this to your loudspeaker set-up).
- Pre-decoded to an **exact square**. Channel routing: 1=front left, 2=front right, 3=rear left, 4=rear right.

- Pre-decoded to binaural stereo. The binaural is **only** useful over **headphones**, and will not make sense over stereo loudspeakers.

The use of B-format examples means that features contingent on high-order ambisonics (HOA) cannot be illustrated. Please contact the author for HOA examples (5^{th} order horizontal for a minimum of 12 matched loudspeakers in an evenly distributed circular array).

Examples

1a: Soundfield recording from a small, inner city park. This sound space can be heard as a prototype image. In everyday experience, many encounters of this familiar image collapse into one timeless prototype.

1b: Soundfield recording from a small inner city park with unusual addition (a bicycle with rattling ornaments is pushed across the scene). This sound space is the same as in example1a, but the additional information provokes temporal engagement involving the new sound object and the changing space-environmental cues stimulated by the moving sound object.

2: "Tensile Elements" (extract 0'00-1'27. Hofmann, 2000). In this work Hofmann controls multiple trajectories of point sources. The intention here is not to create real sound images of spatial extent, rather to project behaviour and gestural trajectories within a space defined by motion rather than environment.

3: Creating an image by positioning multi-channel recordings over a specific azimuth range.
3a: Mono source channels from SPS200 (LF, RF, LB, RB) and two channels from DPA4060.
3b: Sources from 3a located at: -10, 10, -29, 29, -35, 35 degrees respectively.
3c: Sources from 3a located at: -25, 25, -60, 60, -90, 90 degrees respectively.
3d: A new source, displaying a widening of the image over 2500 ms from -10, 10, 29, -29, -35, 35 degrees to -25, 25, -60, 60, -90, 90 degrees respectively (the widening starting half way through the example).

4: Granulation to control image size, in this instance creating a 'denser' but smaller image.
4a: One mono channel.
4b: Granulation of 4a, displaying an image width transition from 180 to 45 degrees and a distance transition (with amplitude and filtering) from two to 20 meters. The mono source is also mixed centre-front to maintain focus.

5: Convolution.
5a: Original spatialisation of mono source.
5b: First-order recorded impulse response.
5c: Convolved result.
5d: Dry and convolved sources.

6: "Kernel Expansion" (extract 0'40-1'05. Barrett, 2009). This extract contrasts synthesised point sources with Soundfield recordings (note that the synthesised materials are only a first-order representation).

7a: "Metal Clouds" (extract 4'32-6'32 Hofmann, 2001). Hofmann's synthetic sound sources are on the whole are extremely abstract. Sometimes the material sounds similar to a high-flying jet aircraft. Even though appearing relatively close to the listener, when the 'plane' enters, psychoacoustic cues combined with sound identity add the sense of distance.

7b: "Untitled-2" (extract 6'04-7'07. Barrett, 2010). In this collage of Soundfield recordings the distances information is expansive due to a combination of near-field effects and sound identification (identifying the bees, the plane and their appropriate behaviours as stabile aspects of a complete sound picture).

Examples 8 to 10 are most successful over loudspeakers.

8: "Pacific Slope" (extract 0'00-1'00, Anderson, 2002). In the opening of "Pacific Slope" we find many similarities to the opening of Harrison's work "Klang", such as image, physical materials and gesture. Anderson's choice of ambisonics spatialisation leads to a stable reproduction of the intent behind the sound; its azimuth location, image size and psychoacoustically implied distance, implication of stage, human scale and agency in relation to the listener.

9: A composed 'envelopment'.

10a. Immersion effect, (sound recorded with an Soundfield SPS200 microphone placed inside the sound source).

10b: This example uses the same recording technique as 10a. In contrast to 10a, the spectral temporal information enhances directional cues, leading to close-up envelopment rather than immersion.

Works references:

Anderson, J. (2002). "Pacific Slope". Original composition in 1st-order periphonic. Released on "Epiphanie Sequence" SCD28056. London: Sargasso. www.sargasso.com/khxc/index.php?app=ccp0&ns=prodshow&ref=SCD28056

Barrett, N. (2009). "Kernel Expansion". Original composition combines 1st-order and 3rd-order encodings in vertically displaced horizontal layers. www.natashabarrett.org

Barrett, N. (2010). "Untitled-2". Original composition in 1st-order periphonic. www.natashabarrett.org

Harrison, J. (1982). "Klang". *empreintes* DIGITALes IMED 0052.

Hofmann, J J. (2000). "Tensile Elements". Original composition in 2nd-order periphonic. www.sonicarchitecture.de