KERNEL EXPANSION: A THREE-DIMENSIONAL AMBISONICS COMPOSITION ADDRESSING CONNECTED TECHNICAL, PRACTICAL AND AESTHETICAL ISSUES

Natasha Barrett

Blåsbortvn. 10, N-0873 Oslo, Norway
nlb@natashabarrett.org

ABSTRACT

Practical and accurate application of ambisonics spatialisation presents a challenge. A working method needs to support sound design or compositional ideas in terms of space, spectral morphology and sound identity, and function technically in extremely diverse playback conditions with the minimum of information loss. A solution combining different encoding and decoding techniques is explored in the composition *Kernel Expansion.* This paper presents how technical, practical and aesthetical issues are connected in the work.

1. INTRODUCTION

Kernel Expansion was commissioned by ZKM (Centre for Art and Media Karlsruhe) for the 43-speaker Klangdom [1]. The Klangdom loudspeakers (Meyersound UPJ-1, UPJ-1, CQ-1 and CQ-2) are relatively evenly spaced in a vertically compressed, rectangular domed array. Since installation the loudspeakers have shifted from their original measured locations and are frequently repositioned 'by eye'. Repeated exact location measurement is unrealistic. Such conditions are commonly encountered and for practical purposes need to be embraced. A period of comparative study, incorporating technical encoding and decoding assumptions derived from the literature [2, 3], explored encoding and decoding methods for different types of sound material, loudspeaker configurations and aesthetical intents aimed at addressing: (a) how ambisonics may realize sound design intents beyond that of other multichannel techniques; (b) a technical method functioning over set-ups with and without vertical elevation ranging in size from the Klangdom to 4.0 or 5.0 / 5.1 home listening set-ups, and allowing the composer to pursue artistic work in smaller scale studio conditions.

2. RECORDING AND ENCODING: COMBINING FIRST- AND THIRD-ORDER SOURCES

Over the past 10 years the author's experimental sound design has involved the synthesis of Higher Order Ambisonics (HOA) sound fields to create complex scenes. Real sound images are not dimensionless points (image extent in spatial experience has been investigated elsewhere [4, 5, 6, 7]). Through the history of electroacoustic composition various practical techniques to manipulate a stereo phantom image have been explored, and more recently formalized in the technical literature [8, 9]).

2.1. Third-order sources and the sounding object

For synthesized third-order material the author investigated a number of techniques to control image size. To begin, sources were recorded in a sound isolated room so as to set apart the chosen source from its complex environment. An A-format microphone (Soundfield SPS200) and a combination of two separate cardioid (Neumann KM140) or two omni-directional (DPA 4060) microphones were used, providing six recorded channels. HOA synthesis involved:

- Positioning the six untreated mono signals (four from the SPS200 without any A- to B-format conversion, and two from the additional microphones) over a specific azimuth range. From the A-format microphone, the one (or two) capsules facing the acoustic object provided the main signals and were located relatively centrally in the new synthesized image. The other recorded channels were used as auxiliary support to 'widen' the image. This technique captured one perspective on space and frequency information (example 1). Multi-band time varying temporal decorrelation applied before the spatialisation process would furthermore influence the perception of size [10].

- Use of the author's own software for stochastically controlled asynchronous granulation incorporating amplitude and air absorption coefficients, encoded into HOA [11]. Images of a certain spatial extent are easily processed, yet in the process of granular densification are in danger of losing sounding unity and extrinsic identity; in which case it is necessary to mix in some of the original untreated recordings (example 2).

- Synthesizing a point location or trajectory in HOA, then convolving the first-order components of this with a first-order measured impulse response. The third-order (dry) and first-order (reverberant) layers are decoded separately. Impulse responses were recorded with the SPS200 microphone using the MLS method. Convolution was carried out using FFT with global support (example 3). For some material, early and late reflections were synthesized in second-order using Vspace [12].

2.2. The challenge of real-world features and the role of recorded sources.

A completely portable battery powered recording system currently restricts the recorded materials to first-order. Field and studio recordings were made using the Soundfield SPS200 and the Sound Devices 788T recorder.

Considering that, assuming correct decoding, higher order leads to a more accurate sound field reproduction, inclusion of recorded first-order materials may appear strange. However, synthesizing a scene of similar complexity and balance as found in real-world environments is a challenge. Rather than the complexity of the composed mix per se, a key issue concerns the listeners' *understanding* of spatial information. This understanding goes beyond three-dimensional source location, source extent, and the sense of environment or scene within which the sound source is located. The role of sound identity (either in terms of direct identity or as a signifier), and the relationship between sound sources and the scene in which they are embedded, play important parts. For example, we hear information *implying* distance and size in a recorded source. If a sound is recorded at distance 'x' and we attempt to project that sound at distance 'y', we resolve the problem by hearing either: (a) that the sound is still located at distance 'x'; or (b) the sound's identity has changed by virtue of its new relationship to other materials (example 4). The location of the microphone strongly influences the distance at which the listeners *feel* they are located regardless of the accuracy (or lack of) real near field sound wave reproduction (example 5). Further, our perception that a scene appears real concerns the *behavior* of objects within their environments (whether real or invented). Although problematic to test under control, the way in which behavior and spatial relationships interact has been suggested in musicology texts, such as concepts of 'local' and 'field' [13] and theories of social distance contrasted to physical distance [14].

2.3. The challenge of first-order recorded sources

Recorded ambisonic sources address some challenges concerning complex real-world features, yet introduce their own set of problems: normal B-format transformations (zoom, mirror, rotation, "tape" transposition) are insufficient, and other modifications risk losing channel coherence even though some A-format transformations have been suggested to preserve spatial impression [15]. Often it is desirable to focus on space, spectrum, morphology or extrinsic identity of specific elements of a complex scene. In this case it is more effective to isolate the element (or record a similar source out of context) and resynthesize the spatial location, preferably in HOA. Therefore, even before decoding issues enter into the process it is necessary to combine recorded sources with HOA synthesized materials. Example 6 illustrates the discussion from sections 2.2 and 2.3. Furthermore, mixing two or more Soundfield recordings, containing complex environmental information where the microphone location has changed, is often problematic; the recordings contain different spatial pictures that when mixed tend to cancel out.

3. DECODING

First-order decodings were tested using max r_E, max r_E / max r_V with shelving filters, an in-phase and a standard weights decoder (using software [16, 17, 18]). As expected, use of speaker layouts other than the 2M+2 rule performed badly, yet even for the 2M+2 layouts the sweet-spot was too small for practical concert use. Decoding with Harpex [19] over all available loudspeakers performed considerably better, creating a larger sweet spot and imagery closer to that obtained under optimal studio monitoring conditions. Third-order materials were tested in a similar way.

3.1. Vertical information

In tests in the Klangdom, the vertical dimension was problematic. The volume of the elevated source appeared to fluctuate depending on spectral content and image size, regardless of loudspeaker distance compensation, and often the spatial image appeared to collapse. It can be speculated that the loudspeaker locations were problematic for the decoding matrix and that spatial distortion was enhanced by strong floor reflections (which vary depending on audience and seating arrangements). Therefore, instead of full 3D, the encoding involved horizontal layers of first-order and third-order material, each decoded for one of three vertically displaced loudspeaker subgroups (3D information in the Soundfield materials were preserved for encoding to HRTF's using Harpex). Although not the same as a full 3D decoding this was found to be an acceptable compromise.

3.2. Vertical distribution of material

High frequency material, when distributed on the elevated layers, as expected served to enhance the perception of height [9]. Also materials containing less direction information are used as 'fill' in the upper layers. Materials containing sizeable motion were distributed in the lower layers where the loudspeakers are spaced further apart and where the audience is closer to the horizontal sound field. Also material requiring a sense of 'stage' would be located in the lower layer (example 7). Besides avoiding repeated decoding problems in other spaces this method supported the following practical considerations:

- Decoding is made in advance; fine-tuning the volume of each layer is carried out onsite 'by ear'. Although delay and volume can compensate for variable loudspeaker distance, spatial distortions are expected over the extreme variations found in different loudspeaker set-ups. Comparing the systems at SARC [20] and ZKM [1] are illustrative.

- Controlling the relative volume of the elevated layers is particularly useful in real concert situations where frequency absorption varies with audience and room acoustics.

- In horizontal only playback situations, important information biased in an elevated layer is simply mixed into the single horizontal arrangement, which although suboptimal spatially, is necessary to avoid loss of other important sound information.

3.3. Use of multi-channel panning

In contexts of both commercial (e.g. cinema) and art-music (e.g. the spatialisation of stereo sources over loudspeaker orchestras) the focused 'punch', where a source is rapidly thrown to a single or closely located pair of loudspeakers, is often used for dynamic emphasis. To achieve similar results with ambisonics would demand extreme high order and accurate near-field encoding. Alternative focused decoding solutions such as Harpex would require loudspeakers to be exactly located at the mean spatial location of the encoded sound. At present, practical application of the 'punch' effect requires a

departure into conventional panning techniques. Amongst the available options the standard 8-ch panner in Nuendo was chosen (multi-channel equal power panning). This layer is added after the ambisonics decoding, mapped onto the given loudspeaker layout manually or by using VBAP [21] such that each of the eight channels may be located on or between loudspeakers as appropriate for the room (example 8).

4. ENVELOPMENT AND IMMERSION

Listener envelopment has been discussed at length in connection to room acoustics and laterally reflected sound. Berg and Rumsey [6] make an appropriate 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 (example 9). The nature of ambisonics allows our perception of envelopment to go one stage further into the sensation of being 'inside' or 'immersed' by the sound.

4.1. Experiments with immersion

Although spaciousness and envelopment have been studied in terms of low inter-aural cross correlation coefficients [22], tests in the current project indicate immersion to be connected to the sound capture and decoding method, and the spectral temporal content of the source.

- Sound capture method: first-order recordings were made with the SPS200 microphone placed inside resonating bodies (a drum and a metal bathtub) capturing the enclosed sound field in close-up.

- Directionality: The sensation of immersion is directionally undefined. Clear directionality (or even worse the appearance of sound originating from a single loudspeaker) creates sensations of 'looking to' rather than 'being inside' the sound. For example, an environmental recording containing a multitude of related yet unique sound sources will surround, envelop and even enclose a listener, but may not lead to the sensation of immersion.

- Frequency: Low frequency sounds, being more difficult to localize, should enhance the sense of immersion. However, the chosen sources involve mid-range frequencies and a complex spatial temporal frequency distribution that, by removal through filtering and "tape" transposition, were found to be important to the sense of immersion. This seems in line with theory showing that a degree of spectral content is necessary for us to distinguish sounds from the front and rear hemisphere [9]. It was not possible to achieve the same effect with W panning where the source radiates the same signal uniformly [23].

- Decoder: Decoding to quad using shelving filters was most immersive. This may be because the dominating lower frequencies are decoded optimizing for directional cues using phase, where spatial artifacts outside the central listening position serve to confuse direction information and in fact enhance the sense of immersion. In larger loudspeaker setups it was necessary to decode these first-order sources using Harpex, for which the sense of immersion was less evident but still present.

Examples 10 to 12 serve as illustrations.

5. SOUND EXAMPLES

The following sound examples serve to illustrate the text. Where appropriate they include reference to the full context in Kernel Expansion (KE). Examples are downloadable from www.natashabarrett.org/KE-examples.html in various formats, including HRTF decoding.

Example 1 – Creating an image by positioning multi-channel recordings over a specific azimuth range.
Example 1a: source mono channels from SPS200 (LF, RF, LB, RB) and two channels from DPA4060.
Example 1b: sources from example 1a located at: -10, 10, -29, 29, -35, 35 degrees respectively, synthesized in third-order horizontal.
Example 1c: sources from example 1a located at: -25, 25, -60, 60, -90, 90 degrees respectively, synthesized in third-order horizontal.
Example 1d: a new source, synthesized in third-order horizontal, 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).

Example 2 - Granulation [11] to control image size, in this instance creating a 'denser' but smaller image (KE 5'19-5'40).
Example 2a: Original first-order recording from the SPS200 inside metal bowl with rolling marble.
Example 2b: One mono channel from example 2a.
Example 2c: Third-order granulation of example 2b, 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 in third-order.

Example 3 - Convolution (KE 3'39-4'02).
Example 3a: Original sound spatialised in third-order.
Example 3b: First-order recorded impulse response.
Example 3c: Convolved result.
Example 3d: Dry and convolved sources.

Example 4 - The sound's identity changes due to a new relationship to other materials.
Example 4a: Original first-order recording.
Example 4b: Isolating part of the sound from its environment (at the end of the extract) allows a synthesized spatial motion, changes the perceived distance and suggests a change in identity.
Example 4c: Full context from KE 0'00-0'08.

Example 5: The location of the microphone influencing the perception of distance, in this example the contrast between close up studio recordings and environmental recordings (KE 9'51- 10'01).

Example 6 - Combines ideas from section 2.2 and 2.3 (KE 0'40-1'15).
Example 6a: Recorded first-order source.
Example 6b: Acceleration of example 6a changes both 'behavior' and spatial cues, effecting our perception of distance.

Example 6c: The combination of recorded and synthesized materials, the illusion of proximity and how the change in behavior (from example 6b) contributes to a change in space.

Example 7: Vertical distribution of material (KE 7'03 - 7'48).
HOA Layer 1 (bottom): 'Staged' material followed by wide motion.
HOA Layer 2 (middle): Wide motion.
HOA Layer 3 (high): Vague space and higher frequency layer.
First-order layer 1 (bottom): Sparse clear articulations (with resonance to blend with upper layers).
First-order layer 2 (middle): Less directional resonant materials.
First-order layer 3 (high): Higher frequencies and general resonance.
Eight-channel pan layer: see section 3.3.

Example 8 -'Punch' effect using multi-channel panning (KE 0'15-1'29).
Example 8a: Eight-channel pan layer.
Example 8b: All layers and formats playing together.

Example 9: Synthesized 'envelopment' (KE 8'30-8'51).

Example 10: SPS200 inside the sound source where the spectral-temporal information enhances directional cues leading to close-up envelopment rather than immersion.

Example 11 – The same recoding technique as in example 10, but with a different source (containing less high frequency and less clear articulation location), suggesting immersion.
Example 11a: Untreated recording.
Example 11b: With processing, using the techniques mentioned in the text (KE 9'23-9'45).

Example 12: A mixture of layers displaying envelopment and immersion (KE 6'49-7'56) maintained by avoiding obvious spectral, spatial, temporal, and to some extent extrinsic relationship between materials.

6. REFERENCES

[1] Ramakrishnan, C., Goßmann, J., Brümmer, L. "The ZKM Klangdom". In *Proc. of the International Conference on New Interfaces for Musical Expression*, 2006.

[2] Bertet, S., Daniel, J., Parizet, E., & Warusfel, O. "Influence of Microphone and Loudspeaker Setup on Perceived Higher Order Reproduced Sound Field". In *Proc. of the 1st International Symposium on Ambisonics,* 2009. binaural synthesis". In *Acoustical Science and Technology*, vol 24 (5), 2003.

[3] http://gyronymo.free.fr/audio3D/the_experimenter_corner.html

[4] Martens, W. "Perceptual evaluation of filters controlling source direction: Customized and generalized HRTFs

[5] Rumsey, F., Berg, J. "Verification and correlation of attributes used for describing the spatial quality of reproduced sound". In *AES 19th Int. Conf.* June 2001.

[6] Berg, J., Rumsey, F. "Systematic Evaluation of Perceived Spatial Quality". In *AES 24th Int. Conf.* June 2003.

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

[8] Potard, G., Burnett, I. "A Study on Sound Source Apparent Shape and Wideness". In *Proc. of the 2003 International Conference on Auditory Display,* 2003.

[9] Blauert, J., *Spatial hearing*. MIT press, 1974, rev 2001.

[10] Potard, G., Burnett, I. "Decorrelation Techniques for the Rendering of Apparent Sound Source Width in 3D Audio Displays". In *Proc. of the 7th Int. Conference on Digital Audio Effects.* 2004

[11] Barrett, N., Hammer, Ø. "Granny: Three-dimensional granulation software". www.natashabarrett.org/granny.html, 2002.

[12] Furse, R. "Vspace". www.muse.demon.co.uk/vspace/vspace.html, 2000.

[13] Emmerson, S. *Living Electronic Music*, Ashgate, 2007. pages 89-102

[14] Blesser, B., Salter, L. Spaces Speak, Are You Listening? Experiencing Aural Architecture. MIT Press, 2007. Pages 33-36.

[15] Anderson, J. "The Ambisonics Toolkit". In Proc. of the 1st Ambisonics Symposium, 2009.

[16] Wakefield, G. "Ambi.decode" www.grahamwakefield.net/soft/ambi~/index.htm, 2006.

[17] Adriaensen, F. "AmbDec" www.kokkinizita.net/linuxaudio/, 2009.

[18] Kocher, P., Schacher, J. "ambidecode", www.icst.net, 2009.

[19] Berge, S., Barrett, N. "High Angular Resolution Planewave Expansion". In *Proc. of the 2nd International Symposium on Ambisonics,* 2010.

[20] www.sarc.qub.ac.uk/main.php?page=soniclab

[21] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *AES Journal* vol. 45 (6), 1997.

[22] Okano, T., Beranek, L., Hidaka, T. "Relations among interaural cross-correlation coefficient, lateral fraction and apparent source width in concert halls". In *J. Acoust. Soc. Am.*, vol. 104 (1),1998.

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