AMBISONICS SYMPOSIUM 2009
June 25-27, Graz

EVOLVING VIEWS ON HOA: FROM TECHNOLOGICAL TO PRAGMATIC CONCERNS

Jérôme Daniel[1]

[1] Orange Labs/TECH/OPERA, jerome.daniel @ orange-ftgroup.com

Abstract: *From first fundamental studies in the mid 90's to today, Higher Order Ambisonics (HOA) keeps being investigated through various view angles, highlighting increasingly nice and numerous features. From the beginning, the relevance and flexibility of acoustic field reconstruction and representation format introduced HOA as a powerful approach benefiting to a large scope of contexts. Nevertheless, if one may dream of it as the format and/or generating approach for future immersive contents, making this dream a reality requires further steps: integration of content generating tools; use, adaptation and assessment by content creators / sound engineers; consumer equipment/acceptability; coding scheme development, format standardisation and compatibility. In this talk/paper, the author will outline contributions, position and expectations of Orange Labs regarding these aspects. As a turning point has been the design of HOA microphones, opening an exciting experimental field in terms of real 3D sound field recording, the author will also share experience and thoughts from experiments involving (or not): video capture, other sound recording approaches, live or post-produced diffusion, etc. In particular, the quite pragmatic, but difficult exercise of targeting standard 5.0 format and ITU setup raises itself essential issues. Indeed, the potential spatial instability and incompleteness due to this unbalanced setup may force to: reconsider decoding strategies, mistrust audio monitoring over other setups, go against the spatial fidelity principle and rather construct a twisted projection of the recorded sound space.*

Key words: High Order Ambisonics, content generation tools, 3D microphone, standard multi-channel formats, sound engineer practice

1 INTRODUCTION

Since the mid 90's, Higher Order Ambisonics (HOA) has gone a quite long way. Starting with theoretical studies that gave a promising extension to Ambisonics, many HOA concepts have turned into operative tools for several years now. Built around a sound field representation model based on its spherical harmonics decomposition, HOA can be simultaneously defined as a *versatile* 3D audio format, an approach for *"optimally" reconstructing a sound field*, and a technology for *high resolution and accurate capture* of 3D sound fields. Let's give a little explanation of these concepts right now.

First, the HOA representation format is generic in that it isn't dedicated to a specific rendering setup (loudspeaker or headphones), but can be *spatially decoded* for virtually any setup. It could be also labelled a "universal" or "versatile". This format is not only flexible in terms of reproduction, but also in terms of possible spatial manipulations of the represented sound field. Furthermore, the extension of Ambisonics to HOA introduces the notion of spatial scalability: indeed the additional spatial components (higher order spherical harmonics) that enhance the spatial resolution can be hierarchically conveyed or omitted, depending on transportation and/or reproduction constraints. Finally (and first of all!), a common and rational HOA spatial sound representation model applies for both virtual source spatialisation and real sound field recording using microphone arrays. "Elementary events" (single wave front propagation) as well as "macroscopic phenomena" (reverberated / diffuse field) are encoded in a rational and homogeneous way. From there, an appropriate spatial decoding is supposed to be able to reproduce in a predictable way main spatial features of the encoded sound field, such that angular localisation effect, depth (linked to direct / reverberated sound ratio), etc.

All these announced nice features let think that HOA will benefit to a large scope of application contexts. One could dream of it as "the" format for future immersive audio contents: imagine transporting one audio content in such flexible way to adapt it to various terminals, various transportation constraints (bitrates / bandwidth), and to offer new ways of content consumption (including interactivity). HOA can be used also for telecommunication, both for business (teleconference with improved intelligibility, presence, immersion) and mass market (ambience sharing; immersive "sound postcards"). In virtual/mixt reality applications, virtual 3D navigation, games, etc., the ability to rotate and angularly distort a HOA sound field is also of great interest to interactively adapt the rendering to changes of the avatar's point of view in the virtual or recomposed scene.

All these promises have motivated studies on HOA at Orange Labs and elsewhere.

Beyond such a promotion of HOA high potential and even beyond proofs of concept given by demonstrators, it appears necessary to make further steps. Especially if the goal is to use HOA as a "production/broadcast format", or even simply as a "content generating toolbox", one has to make it evolve towards a real usage by content creators (esp. sound engineers). At the same time, format standardization issues have to be addressed.

This paper outlines contributions, current position and expectations of Orange Labs (formerly France Telecom R&D) regarding HOA. It doesn't intend to show new scientific results, but rather to summarize a partial state of art and to bring complementary discussions. It has not the ambition of being exhaustive, all the more regarding the work done outside to Orange Labs. The first main part of this paper will mostly address *technological aspects as treated and developed "inside labs"* (often in an academic way, but also including the standardization issues). A focus will be given on 5.0 decoding and HOA 3D microphone systems, as they are technically involved in experiments reported in the second part. As many previous papers have thoroughly developed the mathematical aspects of the theory and technology, no or very few equations will be shown. When needed, the reader is invited to refer to the existing literature. The second main part deals with "real" HOA recording experiments done with spherical microphone arrays, with the stronger and stronger aim to test out the HOA "toolbox" with "real life" concerns. That means dealing with conditions and/or constraints like: targeting a standard 5.0 format and ITU setup; collaborating with professional sound engineers; dealing with various configurations (associated video or not) and contents. As a result, some conclusions are drawn and thoughts are shared.

2 HOA IN LABS

This part summarizes technological aspects of HOA *as seen from the Orange Labs' window*, with sometimes a focus on technical aspects involved in the next part (5.0 decoding and spherical microphones). This starts with theoretical studies (WFS *vs* HOA, NFC-HOA, HOA microphone design, etc.), soon completed by software and prototype developments and their integration into demonstrators. Orange Labs' contribution and position regarding format and standardization issues are also reported. Complementary to objective characterization of technological or physical limits, the report on formal and informal listening experiences invites to sometimes relativise the laudatory account given in introduction! At least, it yields to recommendations and tracks/guidelines for further investigations and improvements.

2.1. From Ambisonics to HOA

Although already considered by Gerzon in an earlier paper [1], the extension of Ambisonics to higher degrees of spatial accuracy has started in the mid 90's with studies by Bamford [2], Poletti [3], soon followed by Daniel [4, 5], Nicol [6], Sontacchi & Höldrich [7], Furse & Malham [8], and many more now…

The pre-existing, 1st order ambisonic encoding format consists of four spatial components associated to coincident pickup patterns: one omnidirectional (sound pressure *W*) and three bidirectional components (X, Y, Z, linked to the pressure gradient). Considered in the frequency domain, their ratio yields the so-called "*velocity vector*" [9, 10] which real part describes the phase propagation (thus the apparent wave front direction) and which imaginary part describes the sound field energy gradient (null in the case of a plane wave). Therefore explicit information of sound localization is encoded and thanks to an appropriate signal processing called "*spatial decoding*", it can be rendered as a "piece of wave front" locally synthesized at the centre of a loudspeaker array. Though being "explicit" regarding each encoded wave front separately, the directional information is nevertheless *minimal* and cannot allow to accurately separating sound sources angularly close to each other. Subjectively, this results in quite blur and unstable sound images. What acoustically happens is that the synthesized piece of wave front is quite small with respect to the wavelength, and therefore the localisation effect associated to the velocity vector is restricted to a low frequency / small area domain. Outside (above the limit frequency), the localization cues are altered in a way depending on the angular dispersion of loudspeakers sound contribution, which is rather large and explains the blurriness and the restricted "sweet-spot". The localization effect related to these high frequency cues is predicted by the so-called "*energy vector*" [11] [5]. Its norm r_E is an indicator of spatial concentration of energy contributions (r_E=1 corresponds to a single point source), and inversely $acos(r_E)$ reflects the angular energy dispersion of sound contributions. Despite these limitations, 1st order ambisonics can claim to *objectively represent and (with some restriction) render some essential spatial features of the sound field*: 1) the directional localisation of events; 2) the distance effect, provided that direct and reverberated sound are evenly captured whatever the direction. This principle of "*spatial objectivity*", or more emphatically "*spatial fidelity*" is something that will be discussed later with HOA.

Mathematically, the extension of Ambisonics to HOA arises by considering omni and bidirectional encoding functions/patterns as a restricted subset of the spherical harmonics basis. Higher order spherical harmonics are angular functions Y_{mn}^{σ} (1) with higher angular frequencies and therefore a higher ability for angular discrimination.

$$Y_{mn}^{\sigma}(\theta,\delta)=\sqrt{(2m+1)(2-\delta_{0,n})\frac{(m-n)!}{(m+n)!}}P_{mn}(\sin\delta) \times\begin{cases}\cos n\theta & \text{if } \sigma=+1\\ \sin n\theta & \text{if } \sigma=-1 \quad (\text{ignored if } n=0)\end{cases} \tag{1}$$

Order n defines the angular frequency of the azimuth dependency; m is the degree of Legendre polynomials and functions $P_{mn}(\sin\delta)$ that describe the elevation dependencies. Spherical harmonics are grouped as functions of same degree m, in which order n varies from 0 to m.

Acoustically, they are associated to higher order pressure field derivatives around a reference point, which help to provide an approximation of the pressure field over a larger area around this point, proportionally to the wavelength. The link between angular and radial/frequential dependency is summarized by the Fourier-Bessel series (2), which highlight weighting coefficients $B_{mn}{}^{\sigma}$ that are just the so-called HOA components, expressed in the frequency domain.

$$p(kr,\theta,\delta)=\sum_{m=0}^{\infty} i^m j_m(kr)\sum_{n=0}^{m}\sum_{\sigma=\pm1}B_{mn}^{\sigma}Y_{mn}^{\sigma}(\theta,\delta) \tag{2}$$

These series confirm that additional directional components (up to a given encoding order M) make possible to approximate the sound field over a wider area, and it can be verified that acoustic reconstruction by a loudspeaker array follow the same trend. In practice this involves a spatial decoding, which basically consists in matrixing HOA signals to derive loudspeaker signals. For the convenient case of concentric, regular arrays, spatial decoding performs like an inverse, discrete spherical (or circular for horizontal array) Fourier transform, applied in the spherical harmonics domain. At least as many loudspeakers as HOA components are required. Figure 1 shows how encoding-decoding can also be interpreted as a fixed, "multi-beamforming operation", with an angular selectivity as fine as the order is high.

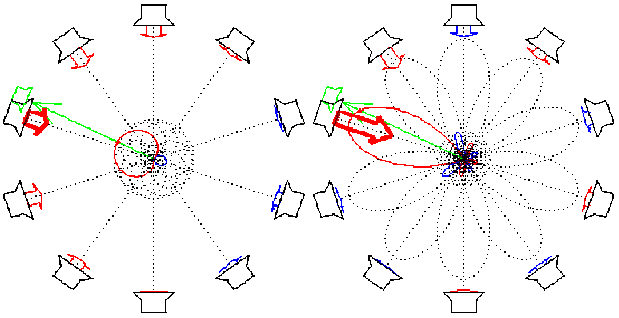

Figure 1 – Virtual pickup patterns associated to a 10-loudspeaker array, resulting from encoding and decoding combination, using 1st (left) and 4th (right) ambisonics. For the virtual source in green, equivalent panning gains are described by the large arrows lengths and colours (red/blue mean plus/minus signs).

Thus HOA has been proved to be able of "*holophony*", that means of reconstruction over a large listening area, provided that sufficiently high orders and consequently numerous loudspeakers are involved. Earliest studies [3, 5, 6] assumed that loudspeakers were far enough to consider that radiated waves were plane on the target area. Then further theory developments referred to as *NFC-HOA* [12] supported the assumption of spherical waves, modelling as "*Near Field Coding (NFC) filters*" the near field effect of the virtual source and the compensation of the loudspeakers' one. Later, Adriaensen [13] gave further improvement to NFC filters implementation. Considering that close virtual sources (inside the loudspeaker array) cause excessive bass-boost with NFC-HOA, an alternative scheme has been developed, consisting in an appropriate high-pass filtering of HOA components [14]. Note that an alternative scheme had also been proposed by Sontacchi et al [7]. NFC-HOA permitted a closer connection to *Wave Field Synthesis (WFS)*, both as holophonic approaches [15]. And more recently a very clever generalization of acoustic capture and reconstruction strategies has been offered by Fazi [16], gathering HOA, Kirschhoff-Helmholtz (*aka* WFS), and Least Mean Square approaches. As far as Orange Labs is concerned, efforts on the holophonic side have been relaxed for several years to better concentrate on 3D recording systems as shown in 2.2.

Another, though complementary branch of investigation has aimed to transpose to HOA the mathematical tools introduced by Gerzon for Ambisonics. So, the high frequency domain decoding optimization according to the energy vector has been extended to higher order [4, 5]. The same logic has applied on the basis of Malham's "in-phase" decoding style, initially dedicated to audiences with very off-centred listeners.

Table 1 summarizes some spatial quality indicators for the first few ambisonic orders M. A higher limit frequency means that natural localization cues are provided on a wider low frequency band, while dispersion angle α_E gives an idea of cues alteration above this frequency and suggests the resulting blur width.

Order M	1	2	3	4
f_{lim}	700Hz	1300 Hz	1900 Hz	2500 Hz
α_E	45°	30°	22.5°	18°

Table 1 – Limit frequencies f_{lim} of the acoustic reconstruction at a centred listener ears. Angular dispersion $\alpha_E = \text{acos}(r_E)$ of loudspeaker energetic contributions with an energy vector optimized decoding.

Benefits for unbalanced and not so generous setups

For concerns closer to mass market multi-channel standards and equipments as further addressed in section 3, it is also interesting to design HOA decoders for reproduction over the 5.0 ITU setup (more precisely referred to as ITU-R BS.775-1). Privileging the rendering of frontal scenes, the latter has a quite unbalanced geometry: 3 front loudspeakers at 0° and ±30° and 2 rear ones at about ±120° or ±110° (Figure 2). It is virtually unworkable for 1st order systems: as a matter of fact, Gerzon derived decoding 5.0 matrices for larger front angles (45° or 50°) while being forced to accommodate to either angular or level distortion of rendered sound images. The advantage of HOA here is that decoding can yield "beam patterns" (remember Figure 1) which width and shape nicely fit the angular span between loudspeakers (see Figure 2). Since the reproduction involves much fewer loudspeakers than HOA components, it doesn't permit a "perfect" reconstruction over a large area; nevertheless it improves the image robustness and offers an enlarged "sweet-area". Decoding optimization may rely on a combination of various criteria such as recommended by Gerzon for Ambisonics. There is generally a trade-off between the satisfactions of these criteria, tuning their associated weight depending on the importance that one gives to them.

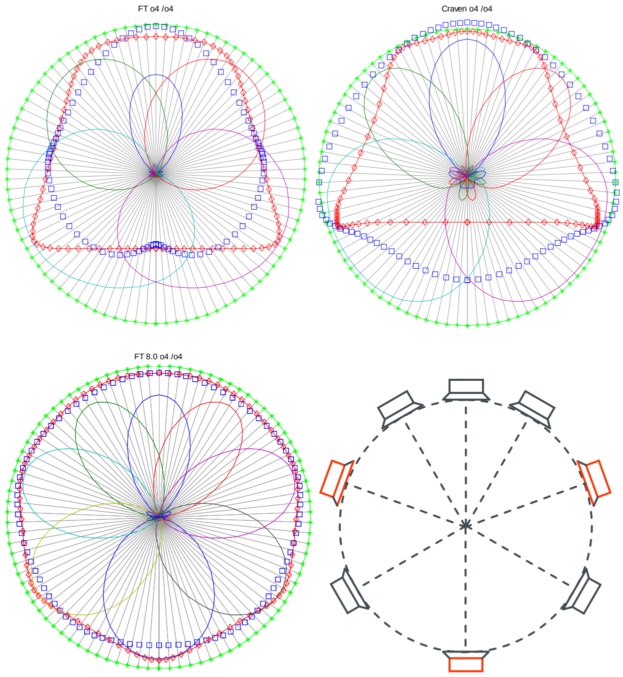

Figure 2 – Comparison between 4th order, 5.0 decoders (Top left: "FT" energy vector optimization and energy preservation; top right: "Craven") and with a 4th order 8.0 decoder (bottom left). 5.0 setup is shown bottom right, with 3 additional loudspeakers (in red) for the 8.0 setup. Equivalent pickup directivities associated to each loudspeaker feed. Velocity vectors (blue squares) and energy vectors (red diamonds) have to be compared with target unitary vectors (green crosses).

Figure 2 shows two 4th order decodings optimized in different ways. The first one has been designed to optimize the energy vector both in terms of direction and norm r_E, and to preserve the global sound energy constant for all direction. Energy vector (red crosses) actually varies in a regular and smooth way, and nearly reaches its physical bound (*i.e.* a polygon that would link the loudspeaker positions on the unitary circle), except around azimuth 0°. No care has been taken about the velocity vector and therefore about optimizing the localization for privileged centred listener, assuming that the decoder would be dedicated to a large group of listeners. That explains that r_V is often significantly less than 1. Let's recall that in the case of real panning gains, the velocity vector is real, and its norm r_V acts as a weighting factor applied to the ITD (interaural time difference), in comparison with a natural plane wave from the same direction. Therefore $r_V<1$ means that sources are perceived less lateral than expected. The second decoder has been optimized by Craven [17] (NB: rear positions at 110° and not 120°) according to various Gerzon's criteria [9], including but not restricted to velocity and energy vectors. Figure 2 shows a much more present centre channel contribution as well as separation between front left and right channels, compared with the first decoder. Little negative side lobes help reconstructing a velocity vector with a norm closer to unity, except for back virtual source directions. And finally the energy vector has a norm close to unity at all loudspeaker directions, but it also tends to be distorted and attracted towards the loudspeakers, which could cause a little "detent effect". Many other decoders can be designed (some others have been tried), but those two ones have been more systematically experienced and compared in the experiments reported in section 3. For comparison, Figure 2 also shows an 8.0 decoding for a setup including 3 additional loudspeakers (at ±70° and 180°), which we used to experienced in informal listening sessions. The 8.0 rendering shows a much better spatial homogeneity regarding both velocity and energy vectors, which perceptively brings better sound image consistency and robustness.

Spatial transformations of the sound field

The probably most common transformation of the sound field is its rotation, which would also correspond to the effect a change of the viewpoint orientation in the sound field. This transformation is "ambisonically valid" in the sense it preserves the encoding model of plane waves. The definition of 1st order rotation matrices is trivial. For higher orders there exists a recursive computation algorithm [18], which is used in the rotation effect of the VST plug-in suite listed in 2.3.

Another kind of effect can be applied to mimic an angular distortion effect occurring with a change of the viewpoint location within the sound field. Gerzon presented a "valid" 1st order transformation, called forward dominance and referring to the "Lorentz Transform", which preserves the plane wave encoding model. The associated angular distortion law cannot be extended to higher orders while preserving the plane wave encoding model. Nevertheless a number of ad'hoc transformations can be designed to yield interesting effects though not being strictly "ambisonically valid".

HOA binaural rendering

A basic method to render HOA over headphones is to combine a decoding operation for virtual loudspeakers, and their binaural simulation *via* HRTF filtering [5, 19]. The loudspeakers positions are chosen so as to match a set or a subset of an available HRTF database. Technically these two linear operations can be combined into one single "HOA to binaural decoder" that consists of 2K filters (usually FIR): one per HOA signal and ear. By assuming head symmetry, even only K filters are required.

Nevertheless such an approach appears to be perceptively suboptimal. Indeed, in binaural simulation, ears are at fixed positions at the centre of the virtual loudspeakers array. On the other hand, decoding for loudspeakers doesn't focus on fixed ears position, but is rather optimized for a centred head whatever its orientation. Binaural decoding optimization can take place in a more general framework where "binaural decoding filters" follow a multi-channel encoding using spatial functions (e.g. HOA, or VBAP, etc.). When the encoding functions are fixed such as with HOA, the decoding filters optimization aims at reconstructing the best as possible the HRTF cues. LMS optimization based on a distance between "original" and "reconstructed" HRTF provide reasonable results. Nevertheless one knows that acoustic reconstruction at the listener's scale is no longer possible above a frequency that depends on the encoding order (Table 1). It has therefore no use to try to reconstruct complex HRTF above this frequency. On the other hand, optimisation can advantageously focus on energy spectrum reconstruction, while relaxing phase reconstruction constraints makes gain several degrees of freedom. This scheme has been elaborated and implemented at Orange Labs for a few years [20].

Of course when possible, head-tracking combines very nicely with binaural rendering. The principle is simple: head-tracker orientation signals drive a sound field rotation processor placed just before the decoder so as to apply rotation angles oppositely to head-orientation angles.

2.2. HOA microphone / recording systems

Building real 3D sound field recording systems has probably been a major step for the demonstration of HOA potential and for advancing towards its exploitation. The present section aims at summarizing the actual properties and also the limits of designed microphone arrays.

Rigid spherical arrays

Technically, the goal of microphone array processing here is to transform captured signals that differ in terms of time and level as a function of the wave incidence, into a set of "HOA" signals that have amplitude relationships according the spherical harmonic functions.

Rigid spherical arrays will be mostly discussed here since they have been quite extensively studied and several prototypes have been designed and experimented. A spherical distribution of sensors provides a very convenient sampling of the sound field as expressed in terms of the spherical harmonics. Indeed it permits to easily separate the angular dependency and the radial/frequency dependency. Therefore the array processing consists of a real gain matrixing which yields a set of directivities having the shape of spherical harmonics (but not the right scale and phase), followed by an EQ filtering that aims at restoring correct phase and level relationships between resulting HOA signals. For more details, refer *e.g.* to [21-23]. There are two main limitations. First, due to the finite number Q of sensors, spatial information becomes inconsistent when wavelength is less than about twice the spacing between them: this is the "spatial aliasing" occurring at high frequencies. For a 75mm-∅, 32-sensor array, the spatial aliasing frequency is about 7 kHz (but mostly annoying above about 10 kHz). The second limitation occurs thus at low frequencies. Indeed the limited size of the array implies phase differences (between sensors signals) that become very small for large wavelength. Since spherical harmonics are related to the sound field spatial derivatives of different orders [12], their estimation requires amplifying these small differences as much as the wavelength is large regarding the array and as the order is high. Therefore the theoretical EQ filters should act as bass-boost with an infinite slope of -m×6 dB/oct for each order m. Such curves are shown in Figure 3.

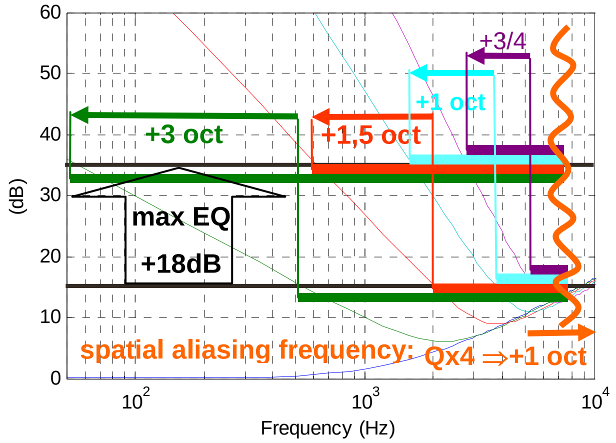

Figure 3 – Theoretical EQ curves involved in the processing of a 7cm-∅, 32-sensor spherical array (order/colour: 0=blue, 1=green, 2=red, 3=cyan, 4=magenta). Spatial aliasing limit suggested in orange. Effect to the level limitation of the EQ on the bandwidth of estimated HOA signals: suggested as horizontal bars with the same colour code as curves.

Of course such filters are impracticable for matters of stability. Thus one limits the EQ to a maximum level, which determines the low frequency bounds of the correct estimation of HOA signals (found when crossing the theoretical curves in Figure 3). This maximum level can be different for each order m. In [14] it was suggested to adjust it as a function of a target size of the sound field reconstruction area, which generally implies higher level limits for higher orders. In practice one chooses about the same max level for all the orders (or even a bit lower for higher order), and adjust it according to a tolerance to the resulting background noise amplification and to the potential calibration errors.

Improving HOA encoding quality with spherical arrays is constrained by several trade-offs. First notice that the bandwidth of each correctly estimated HOA component is fixed (in terms of octaves) by the number of sensors Q and the EQ level limit. Choosing a greater array radius will be better for low frequency capture but worse for high frequency, and inversely with a smaller radius. Otherwise, you can increase the maximum EQ level, provided that you don't fear noise amplification at this level and that the sensors are calibrated well enough. As illustrated by Figure 3, this enlarges the HOA bandwidths towards low frequencies. Nevertheless, since this extension is inversely proportional to the EQ slope (m×6dB/oct), the benefit gets lower as higher orders are concerned. With an increase of +18 dB, one gains resp. 3, 1.5, 1 and ¾ octave for 1st, 2nd, 3rd and 4th order components. Finally you may increase the number Q of sensors, which has two benefits. Firstly, it pushes spatial aliasing towards higher frequencies since spacing between sensors decreases (for the same array radius): quadrupling Q makes approximatively gain 1 octave (Figure 3). Secondly, since signal redundancies naturally improve the SNR (by $10\,log_{10}(Q)$ dB, *e.g.* 15dB for Q=32) and make spatial information more robust to sensors calibration and positioning errors, this permits to increase the maximum EQ level and therefore to further extend the estimation towards low frequencies.

To summarize, the benefit of more numerous sensors in terms of higher order capture has to be relativised since high orders are actually encoded over a reduced bandwidth. The benefit has to be shared with the extension of the encoding quality (i.e. the "bandwidth") for "not so high orders".

And what about focussing on 2D sensors arrays, for horizontal only reproduction? The idea of using less sensors and placing them over a horizontal circle (even around a sphere) wouldn't be bad… if recorded sound fields were themselves restricted to horizontal wave propagation. Nevertheless a number of waves have a vertical component, which yields spatial aliasing artefacts on estimated horizontal components because the circular array cannot separate them correctly. Beyond a poor localization effect, non-horizontal contributions may suffer from undesirable spectral coloration.

Alternative microphone array structures

Alternative array structures have been proposed to improve the capture (*i.e.* the spatial encoding) without necessarily increasing the number of sensors. More specifically, Epain and Daniel [24] investigated structures that are more diffracting than a simple sphere, thanks to cavities hosting the sensors at their ends (this idea can be also found in [25]). The main idea is to acoustically enhance the directivity of each sensor so as to capture more substantially the desired higher order spherical harmonic components. Another interpretation is that spatial aliasing at high frequency is lowered by diminishing the gap between apertures, while the global structure keeps a size favourable to relatively low frequencies. This kind of structure might also be useful to reduce the spatial aliasing annoyance from vertical onto horizontal components in the case of horizontal-only arrays. As a proof of concept, the work of Epain concluded by the design and prototyping of an 8-cavity/sensor structure (see "The Sceptre" in Figure 4) and associated array processing for 2D, 3rd order encoding. Although not subjectively evaluated in a formal way, it has been successfully experimented in the contexts of immersive audio conference as well as music performance recording (see Section 3).

There are other kinds of arrays that aim at improving the spatial encoding, such as multi-radius concentric arrays (Ward, Abhayapala, Jin et al), "shell arrays" (Rafaely), sensors distributed within a volume (Laborie et al), etc. Since no practical experience has to be reported by the author, these are not further described in the present paper.

Experimented prototypes

Several spherical microphone systems have been built and/or experimented since 2004, with a step by step progress in terms of quality as well as facilities. Most of Orange Labs' prototypes were made with a 7cm-∅ plastic ball (the last ones with DPA4060 microphones), with variant versions comprising either 8, 12, 16, 20, 24, or 32 sensors. Except for the 8-sensor array, they generally required a bulky and heavy rack case of preamps and converters, big cables and a desktop PC. Finally since mid 2008, a number of experimental recordings have been made with the famous EigenMike™ ("em32") purchased to mh-acoustics [26], which has virtually the same geometrical configuration as our previous 32-sensor prototypes, except a slightly larger diameter (8cm). The EigenMike "just" offers much higher use convenience thanks to the integration of preamp and ADC inside the sphere, and a potentially long connection (via an Ethernet cable) to a FireWire interface box that makes the system handled as a 32-IO ASIO device. Processing is still done with Orange Labs' HOA VST plug-in suite (on the laptop or desktop). These different prototypes have been used in various recording opportunities which are reported in section 3. The compactness and inconspicuousness of such systems appears to be a very attractive feature and is an additional criterion for the potential adoption of HOA microphones in audio or audiovisual production.

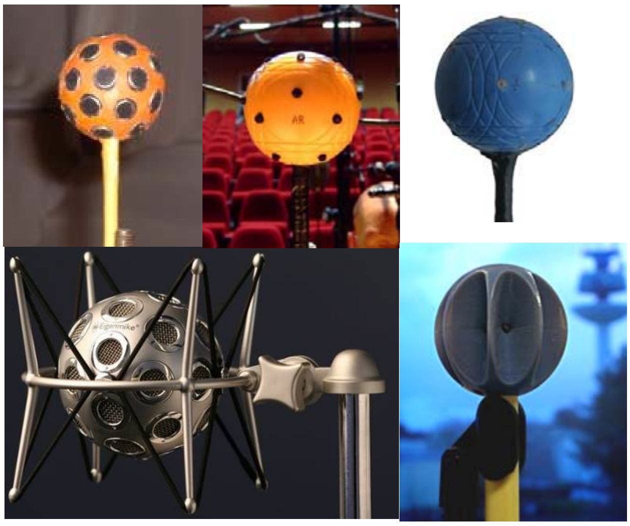

Figure 4 – Some microphone arrays experimented, built by Orange Labs (top, plus "The Sceptre" at bottom-right) except the EigenMike™ from mh-acoustics (bottom-left). Other prototypes are shown in [27]

Impacts of imperfect encoding with microphone arrays

Developing a small-size, compact HOA microphone system is a nice thing for sound engineers or many other application concerns. Nevertheless Figure 3 and the accompanying discussion clearly show that the spatial encoding of the recorded sound field is far from being "ideal" compared with the theoretical encoding functions (1). Furthermore, 5th or higher order encoding with a single sphere array is virtually unworkable.

As a consequence, recordings with such arrays are not really compliant with "holophonic reconstruction" over large areas. Firstly, because of the limited encoding order; secondly, because the lower frequency bounds of HOA estimation are generally not low enough to contribute to acoustics reconstruction over the expected area size. To understand it, one simply has to wonder if it seems reasonable to extrapolate the sound field captured by a small sphere over a much larger area. Furthermore, Near Field Coding will mostly not be within reach of microphone array encoding, except regarding sources very close to the sphere.

Another consequence is that even for "non holophonic reconstruction", a decoder optimized for *e.g.* a 4th order encoding won't remain strictly optimal for a recording from a 4th order HOA microphone, since encoding accuracy actually decreases from 4th to 1st order as the frequency decreases to low frequency (cf Figure 3). Nevertheless the loss of optimality has to be relativised depending on the actual decoding. For instance, Figure 5 shows that the 4th order "Craven" decoder seems to resist rather well to reduced orders in terms of the velocity vector though not in terms of the energy vector, which can be considered as less important at low frequency. Regarding another category of decoders, some informal trials to adapt the decoding as a function of the frequency didn't show a great subjective improvement for the listener.

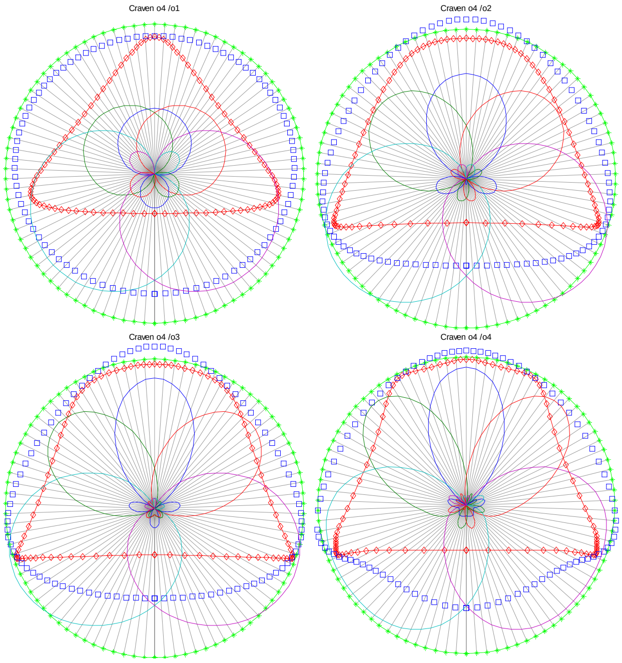

Figure 5 - Behaviour of the "Craven" decoding when actually applied to 4th and lower order encoding (top to bottom and left to right: from 1st to 4th order). Same interpretation as Figure 2

2.3. Development of HOA tools and demonstrators

Development of a HOA toolbox (VST plugins)

The HOA processing units presented above have been implemented and embedded by Orange Labs as VST plugins with GUI (compiled for Windows). They support parameters settings like HOA format, microphone or loudspeaker array configurations. Each plug-in can be instantiated with various numbers of inputs and outputs.

- HOAEncoder (virtual sources encoder): *N* monophonic signals to *K* HOA signals (supports NFC-HOA), with 3D coordinates GUI
- HOAMicProcessor: HOA microphone array processor, choice of a configuration file containing precalculated matrix and FIR coefficients within a list
- HOARotator: HOA to HOA signals, 3Degrees-of-Freedom rotation, with 3D symbolic representation of the sound field and the listener head
- HOASpkDecoder: *K* HOA to *N* loudspeaker signals, choice of a preconfigured decoder within a list
- HOABinDecoder: *K* HOA to 2 headphones signals, choice of a decoder (FIR coefficients precalculated from HRTF databases) within a list

Integration in host applications - demonstrators

All HOA VST plugins have been successfully tested in several VST host applications under Windows XP: multichannel editors like Cubase, Nuendo (Steinberg), and Podium (Zynewave); modular softwares like PlogueBidule (extensively!) and in a less extent MaxMSP. Nevertheless some editors (Cubase, Nuendo) impose a bit annoying constraints in terms of channels per bus limitations, and bus-to-bus routing to spatially encode several mono tracks using one HOAEncoder plugin instance (this is possible but requires a bit twisted trick). Despite unusual ergonomics, Podium appears to be quite well adapted, with up to 32-channel busses.

For dynamic binaural rendering, one just has to connect, prior to the HOABinDecoder, a HOARotator plugin which angle parameters are driven by a head-tracker. Besides quite expansive trackers like Polhemus, Fastrack, etc., relatively low cost trackers have been successfully experimented for demonstration purpose at Orange Labs, deriving from e.g. a gyroscopic wireless mouse, or associated with head-mounted displays.

3D audioconference using HOA microphone

The same basic HOA processing units have been embedded in a software teleconference tool developed in Orange Labs. Full duplex spatial communication could be demonstrated, with: an 8-sensor array in each place, binaural and/or loudspeaker decoding, Noise and Acoustic Echo Cancellation, compression (or not) and transmission of either HOA or decoded loudspeaker signals (the former case appearing more destructive than the latter regarding audio and spatial properties). Further work has to be made especially regarding specific echo cancellation schemes.

2.4. Format, standardization and coding issues

A highlighted in introduction, one key concept that initially motivated studies on HOA was the high versatility and therefore attractiveness of this spatial audio format. Indeed it promises new perspectives for 3D audio content creation and consumption in a flexible and generic format: one content for 2D or 3D, high or lower resolution depending on transportation constraints and reproduction facilities, static or interactive reproduction. To be able to actually share such contents, the standardization of HOA format is certainly an essential step.

Specification levels for requirements

Nevertheless it is also useful to identify the different contexts concerned by HOA format standardization, as they might specify different requirements. Let's distinguish between the following specification levels:

1. Format specifications to ensure the interoperability between HOA processing units (e.g. encoding, decoding plugins, etc.), as e.g. connected in a Digital Audio Workstation
2. Format specifications to share HOA sound files with no or few concern with data size issues (no or lossless compression)
3. Format specifications for insertion, manipulation and interaction in composite multimedia contents, for virtual 3D applications (virtual reality, games, etc.)
4. Format specifications for generalized, "mass market" exchange and/or consumption (broadcasting, etc.), with high concern with compression issues

Of course these levels require all clear specifications to make possible to decode HOA signals with no risk of misinterpretation. The first two levels may recommend certain simplicity to be easily handled. But in the author's opinion, they should also take into account the various HOA normalisation and sequencing schemes that have a legitimacy (2D, 3D, "SN2D", "N3D", etc.), while the existing HOA software tools should be able to adapt to them, either by software upgrade or additional conversion plugins or tools. An idea could be that a common API and associated code be shared for conversion between normalisation and routing between different sequencing schemes. Otherwise, let’s keep in mind that any format specification that becomes widely adopted in practice, may have an influence on the definition of future compressed/broadcast formats, and therefore should be thought so as to not burden them.

Format specification for broadcast contents raises additional issues like efficient compression schemes, which may imply reconsidering spatial representation models. This is still the object of research work.

Let's first focus on specifications that concern the rendering of potentially composite and interactive contents, as it was standardized in MPEG4-AudioBIFS-V3. They are inspired by the proposal given in [12] and comprise a few more options.

Comments on the MPEG4-AudioBIFS V3 standard

Currently in MPEG4, the specification of HOA format is restricted to the BIFS (for Binary Information for Scenes) dedicated to the insertion of media in virtual 2D or 3D scenes, and which rendering requires the "system layer". It isn't handled by the "audio layer" dedicated to coding/decoding audio streams, and actually restricted to mono, stereo and usual multichannel formats. A good summary of the "new" features (Version 3) of the standard [28] can be found in [29]. Two nodes of the AudioBIFS-V3 are related to HOA and more generally 3D audio contents. The **AudioChannelConfig** node can be considered as a "patch" to label audio flows as belonging to "spatial formats" that the *audio layer* is currently unable to describe and handle itself, so that the *renderer* (of the system layer) can correctly interpret them and apply the appropriate spatial decoding if needed. This node supports ambisonics/HOA as well as discrete multichannel with generic loudspeaker positioning specifications, and also binaural. The **SurroundingSound** node can be used to interpret a spatial audio flow as a "sound field object" that surrounds the listener / avatar in the virtual scene. This way, rotations and/or perspective distortions might be applied to the concerned sound field at the rendering stage, in accordance to changes of node parameters and/or of the listening point (orientation, location) in the virtual scene. With the benefit of hindsight, let's now bring additional comments regarding HOA features.

HOA specifications in the AudioChannelConfig node are largely inspired from [12]. Fields named ambResolution2D, ambResolution3D, and ambEncodingConvention, have the same interpretation of fields `resolution2D`, `resolution3D`, `encodingConvention` in [12]. The ratio of the fields ambNfcReferenceDistance and ambSoundSpeed corresponds to the `nfcReferenceDelay` in [12], and is used to properly handle NFC-HOA.

One retrieves the concept of a mixt sound field resolution for the case where e.g. 1st and 2nd order contents would have been mixed together (read [12] for further explanation). ambSoundfieldResolution corresponds to `lowerResolution` and `lowerResolution`. (used when `mixtResolution` is true). With hindsight, this feature will probably have a minor use.

Fields ambArrangementRule, ambRecombinationPreset, ambComponentIndex have about the same use as `orderingRule` and `componentIndex` in [12]. But they also may indicate the use of an additional field **ambBackwardMatrix** that allows retrieving HOA channels that would have been transported in a matrixed form. This allows alternative though backward compatible representations of HOA sound fields. It can also be useful for compression issues, as discussed in the next subsection.

Regarding contents that are produced with microphone arrays such as described in 2.2, a format extension could be to specify the frequency bands where HOA components are actually encoded and present (or to specify encoding characteristics more precisely), in order to enable to optimize the decoding as a function of the frequency. Nevertheless this becomes useless if the microphone array content is mixed with ideally encoded close microphone signals.

A format for transporting and coding 3D audio

Format specifications described above don't provide any compression scheme for HOA content. They just apply to audio streams that may have been coded/decoded in a more or probably less efficient way in the audio layer. If one wants to use an existing codec (therefore non HOA specific) to HOA content, it would surely be better to apply it on a matrixed form of HOA channels. Indeed, HOA spatial encoding implies many redundancies and phase opposition between channels. If HOA channels would be compressed individually (generating quantization noise that is supposed to be perceptually masked), there are many chances that the quantization noise proportionally increase and become unmasked after the final spatial decoding step, i.e. after matrixing of HOA signals. In this compression context, one would rather an alternative spatial representation basis, where spatial functions would present better spatial separation and less opposite-phase relationships. A candidate could be an approximation of a plane wave decomposition as resulting from the HOA domain by a simple matrix transformation (or also a particular spatial decoding), and from which one could go back to HOA signal by applying e.g. the **ambBackwardMatrix** mentioned above. Anyway, efficient HOA compression strategies are currently still the object of research work.

2.5. Learning from subjective evaluation studies

A quest: determine a quality-cost ratio as a function of the order and derive recommendations

HOA is claimed to provide a higher spatial resolution of the rendered sound field, and therefore more perceived satisfaction. When thinking about using HOA as an actual content format, quite important questions arise: is it worth generating and transporting many HOA signals? What is the actual improvement of perceived quality? Keep in mind that depending on the application context (entertainment, teleconferencing, etc.), perceived quality can be appreciated in terms of immersion, intelligibility, comfort, sensation of presence, ability to localize precisely. To take the problem with an approach as generic as possible, one initial aim of the PhD work of Stephanie Bertet was to find and explicit connections between objective and subjective characterization scales, trying to highlight a main parameter that would be the "spatial resolution".

Evolving methodological approaches for HOA evaluation

One initial assumption that sounds quite reasonable is that while increasing ambisonic order, the spatial resolution also increases, which also means that the localization blur decreased. From an objective point of view, these expectations are encouraged by acoustical and mathematical indicators, which are the reconstruction area width (or low frequency band width) and the energy vector (see Table 1). Therefore Bertet first performed a localization test. It involved several encoding systems (mostly microphone arrays) with various degrees of spatial encoding accuracy in terms of theoretical order (from 1st to 4th) and actual estimation bandwidth. The results have confirmed the expectations [30]: the higher the encoding accuracy, the lower the localization blur. Like most localization tests, sound stimuli were broadband noises and didn't compose a plausible sound scene. Thus another methodological approach has been adopted to appreciate spatial characteristics on more "realistic" sound scenes will several spatialized natural sounds (though synthesized without room effect). This was a modified MUSHRA test, reported in [27]. Results were very coherent with the system hierarchy previously highlighted by the localization test. Besides quantitative results, listeners informally reported artificial and annoying auditory sensations (coloration, “phasiness”) that appear in some configurations. Such effects had been already noticed during previous informal tests, and reported in papers like [31]. It had been hypothesized that this was due to the use of more loudspeakers than the minimum number *N* required regarding the encoding order *M*. For horizontal reproduction, a minimal regular setup involves $N=2M+2$ loudspeaker according to Gerzon ($N=2M+1$ for other people). Now, a common 12-loudspeaker setup was used for all encoding systems, from 1st to 4th order. Therefore Bertet performed another listening test to compare various combinations of reproduction setup versus encoding order. This time, ideal encoding equations were used to exclude the influence of microphone array artefacts. To get results in an efficient way, a pair-wise comparison protocol was chosen: an “AB-test” where the listener is asked to judge the subjective distance between two stimuli. One highly interested result derives from the Multidimensional scaling analysis. It highlights two main dimensions which one feels inclined to interpret as: 1) the effect of the encoding order in terms of localization accuracy; 2) the effect of excessively numerous loudspeakers (with respect to the encoding order) in terms of artificial auditory sensations (coloration, “phasiness”). Turning this interpretation into a scientific truth would require an additional test involving explicit subjective attributes and maybe a preference scale. At least, this study corroborates a objective study done by Solvang about “spectral impairments” [32].

Although the initial quest is not closed yet, one interesting point with Bertet’s work is the investigation of a palette of methodologies which use is not so usual in the field of spatial audio.

Spatial and audio features are not orthogonal: incomplete objective indicators

One should notice that for a given order and regular reproduction setup, traditional indicators like the reconstruction area width and the energy vector, remain about the same whatever the number *N* of loudspeakers, provided that $N\geq 2M+2$. Regarding Bertet and Solvang’s studies, that means that focussing only on these indicators hide an important part of both objective and subjective realities. In the first two tests made by Bertet, a 12-loudspeaker setup had been chosen as fixed whatever the encoding system so as to not introduce an additional influent factor, which turned to be a mistake. Indeed, keeping it constant while changing the encoding system order introduced itself a variation of some important perceptual attributes!

To summarize, Gerzon’s criteria based on velocity and energy vectors for decoder optimization should be considered with carefulness in a number of cases. What is questioned here is the reproduction system transparency. Additional criteria are probably necessary to design a transparent decoder.

Further comments on reproduction transparency issues

Undesirable coloration and/or phasiness effects may even occur in particular cases where there are quite few loudspeakers regarding the ambisonic order, esp. with irregular distributions like ITU 5.0 (Figure 2). Their perception can be present in a more or less amount. Artefacts are generally emphasized in presence of broadband signals, applauses, etc. It also depends on the reproduction room. And finally their detection depends on the listener expertise.

The issue can be discussed in the more general context of multi-channel panning and recording. Indeed many sound engineers and mixers already noticed that subjective coloration generally occurs when playing a coherent sound onto more than two loudspeakers on a 2D rig or more than three on a 3D rig[1]. This brings a further interpretation of the two 4th order, 5.0 decoders shown Figure 2. Indeed, the first one (“energy vector optimized”) present an important overlap of the three front pickup patterns for sources about 0°, whereas the second one (Craven decoder) offer a much higher channel separation, at the cost of more angular distortion[2]. And as a matter of fact it has been observed that the “Craven decoder” behaves better than the other one in terms of coloration and phasiness on applause contents (NB: if these are essentially captured in the horizontal plane).

Further comments on the impact of encoding artefacts

Section 2.2 has shown microphone arrays typically suffer from encoding artefacts at low frequencies (reduced actual order) and also at high frequencies, due to the spatial aliasing. To reduce the latter, alternative array structures have been studied. One may object that once localization cues are well reconstructed in a mid-low frequency range, it isn't necessarily worth reducing spatial aliasing in the high frequency range. After all, most WFS systems are reputed to work quite well even spatial aliasing is present above 1 or 2 kHz. HOA recording and listening experiences with e.g. cymbal sounds tend to prove that reducing spatial aliasing is a good thing. Indeed, since that kind of sound has a lot of energy in a high frequency domain, including above the spatial aliasing frequency (depending on the array), much spatial information becomes inconsistent, and therefore the sound might be distributed "everywhere". More generally, listening to broadband, moving sources reveals an

[1] Informally reported from the workshop entitled “Surround Sound Recording and Reproduction with Height” of the AES 120th Convention, Paris, 2006.

[2] The trade-off would be solved with 5th order systems, as Laborie et al do with their Trinnov SRP system.

amazing localization phenomenon, where the perceived sound image splits between high frequency and mi-low frequency portions: the former (affected by spatial aliasing) remaining about a "middle/nowhere position" while the latter coherently follows the expected trajectory. From informal listening, this splitting artefact (with a particular content, indeed!) mainly occurs with "ordinary" sphere microphones (especially with a reduced number of sensors, e.g. 8 sensors on the equator) whereas broadband sound image integrity better resists with the "sceptre" (cf 2.2).

2.6. Summary

HOA as sound field reproduction / spatial sound imaging approach

One of the first features highlighted while extending Ambisonics to HOA was the ability for "holophony", that means acoustic reconstruction over a wide listening area, provided that many loudspeakers and HOA components are involved (*e.g.* several dozens). Though this is still the object of further clever and elegant theoretical developments (generalization of approaches), it is likely that the physical limits have been nearly reached and that no significant subjective improvement can be expected. At least, Orange Labs' efforts have been relaxed on this topic to better concentrate on more modest setups, closer to the mass market reality. To design HOA decoders for more general setups, including standard 5.0 ITU or extensions, decoder optimization usually relies on combinations of Gerzon's criteria formerly introduced for 1st order Ambisonics. Though resulting rendering properties are globally satisfactory, objective and subjective characterizations (formal or informal) have pointed out potential auditory cues alterations that the usual objective indicators (*e.g.* the energy vector) are unable to predict. These are coloration and/or "phasiness" effects. There's probably latitude (requiring additional criteria) for decoder improvements towards a better reproduction system *transparency*. It is recommended that this research topic be supported by listening tests, considering an extended palette of evaluation methodologies as Bertet's work began to investigate.

HOA as a production approach

As Orange Labs has *a priori* less interest in *e.g.* Computer Music than in mass market and business applications (content production and delivery, teleconferences, etc.), a key and turning point was the design and experiment of HOA microphones for real sound field recordings. Many kinds of microphone arrays can be considered for HOA recording: of small or big size, using more or less sensors, transparent or diffracting structures… The spatial encoding accuracy (in terms of ambisonic order and bandwidth) of course depends on all these parameters. [Reversely the same microphone arrays could be efficiently used for direct production of loudspeaker signals without explicitly providing an intermediary HOA content.] For experimental concerns, we focussed on small size spherical arrays and one more diffracting structure, with known qualities and limitations. Spatial content generation can also take benefit from a mix with close microphone signals, spatially encoded using the HOAEncoder. A very exciting topic is now to further develop spatial audio editing tools based on 3D sound field visualisation and analysis in order to assist and partially automate the content creator tasks. Finally, it is worth mentioning that all the technological developments related to HOA can benefit to various specifically targeted multichannel setups or formats, without necessarily involving intermediary HOA signals. With such a consideration in mind, one can gain additional degrees of freedom to optimize *e.g.* the generation of loudspeakers signals directly from the microphone array signals.

HOA as an exchange and delivering format

For the exchange of HOA contents, several propositions of format specifications currently exist. It is suggested that a consensus could be adopted with the support of common conversion tools, while waiting that HOA content generation and use get a higher degree of maturity. Requirements regarding HOA compression issues are more demanding: one might consider modified representation basis (e.g. plane wave decomposition, with backward compatibility with HOA basis) to better support quantization issues. An interesting track of investigation would be also to transpose the concept of Spatial Audio Object Coding to HOA.

From labs to "real life" use

Section 3 will provide further (though mostly informal) learning from experiences of HOA recording and reproduction, especially over the standard ITU 5-speaker setup. One main concern is to adapt the "HOA exploitation strategy" according to the current standard and market reality (broadcast possibilities).

3 HOA IN LIFE

Beyond laboratory developments and experiments, an important step (with short or mid term concerns) is to try the HOA technology with the current use, equipment and delivery constraints. Considering the multichannel market in Europe, that means focussing essentially on 5.0 contents. Regarding now the content generation issue, this supposes to make the technology accessible to sound engineers, and to adapt the tools to their practice if needed. An expected benefit from such an exchange is also to improve the technology according to their demand and criteria.

3.1. Report on recording opportunities

Brief history

With the first HOA microphone prototypes, recording trials done in 2004 and 2005 were difficult due to heavy hardware and soundcard constraints (which required *a posteriori* correction of dephasing between the four 8-

channel Terratec PCI cards), and finally hardly workable. Then a new 32-sensor recording setup benefited from better hardware choices (two MOTU24IO) and “packaging” (in a rack case), and therefore could be reinstalled and operative in a few minutes. With this system, the JES’06 (Journées d’Etude sur la Spatialisation, Paris) in January 2006 have been the first public demonstration opportunity of real time 4th order 3D recording and reproduction (over 5 to 8 loudspeakers, or over headphones with head-tracked binaural). The same demonstrator has been shown a few months later at the AES 120th Convention in Paris.

From then, experimental recordings have been performed regularly, first on an in-house basis, then (since 2008) in collaboration with external sound engineering teams. Recordings have been as many opportunities of combining various kinds of configurations and constraints:

- associated video or not (constraint of putting the microphone out of the camera view fields)
- 3D/stereoscopic video (*Broadway tout Show, Don Giovanni*)
- concurrent recordings with multichannel trees (*EWO workshop*)
- collaboration with professional sound engineers
- live / direct diffusion or post-production
- various contents: music (symphonic, chamber music, voice, jazz…), nature ambience, crowds (football match)…
- with or (mostly) without public address system
- spatial organisation: frontal scene, panoramic scene, with height sources
- with additional microphones or not; with crosstalk (spot mic) or virtually not (close mic, HF mic) ; with HOA encoding or other panning laws/means (mixing desk)

Besides many informal opportunities, some recording experiments have been important milestones. They are reported in the next paragraphs.

Joint 3D audio-video recording: "Broadway tout show"(May 2007)

For experimental purposes motivating both 3D video and 3D audio teams of Orange Labs, one capture session has been organized near Rennes two years ago, on a *musical* entitled “*Broadway tout show*”. It involved simultaneously 3 stereoscopic cameras, 4 or 5 other video cameras, 3 HOA microphones (one 32-sensor array in a centred place of the first rows of the audience, one 12-sensor array close to the stage and on its side, one 8-sensor array on the stage), 3 dummy heads for binaural capture, HF Lavallier microphones for singers, close microphones for the piano, and some microphone pairs for an alternative artistic multichannel mix. Two front loudspeakers (on both sides of the stage) were present as a public address system, and a soundtrack was played on some pieces. After recording, the audience was invited to a demo workshop at Orange Labs Rennes, where extracts were reproduced with stereoscopic projection and 8.0 / 5.0 multi-channel reproduction. The idea was that they could compare the sensations provided by the reproduction, with the recollection of the live event. For this demo, the 4th order, 32-sensor recording was decoded over 8 loudspeakers (cf Figure 2), and a little amount of close microphone signals had been panned (under Nuendo) in accordance to the spatial organization reproduced by HOA. The HOA version and the artistic mix sounded very different: the latter used couples of microphones at different locations and had applied EQ effects, whereas the former was representative to a centred point of view, with no EQ correction of the “concert hall” (which was not acoustically great). Finally, the HOA reproduction together with the stereoscopic projection sounded with a very good audiovisual coherence, not only in terms of localization but also of “room recognition”, providing a striking sensation of “being there” again.

Workshop "Ears Wide Open", Rennes, March 2008

This workshop (*EWO*) which took place in Rennes (France) was co-organized by the AES French section, the *Société Française d'Acoustique* (SFA), and Orange Labs. An evening session at *le Tambour* (University of Rennes 2) was dedicated to the multi-channel recording of a guitar quartet, a big-band and a jazz combo, a comedy with songs. Conferences was held at Orange Labs, as well as listening sessions the day following the recording session. A more complete report on this event can be found in [33], and videos of conferences and recording / listening sessions are available online on www.uhb.fr/lairedu/ (search "Ears Wide Open").

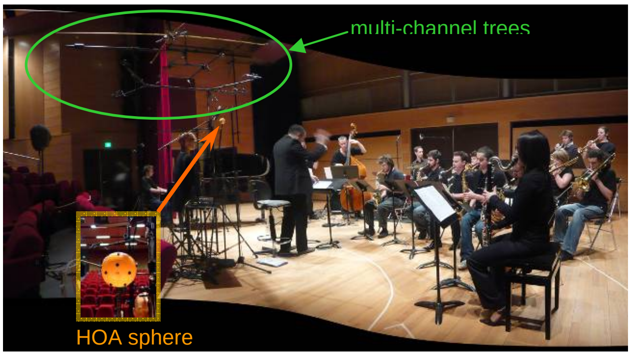

Figure 6 – Concurrent 5.0 recordings of a big-bang with several multi-channel trees and a 20-sensor HOA sphere. Additionally: artificial heads for binaural recording.

The HOA microphone was a 7cm-∅, 20-sensor sphere (with DPA4060). For reproduction 3rd order, 5.0 and 8.0 decoding matrices were optimized only according to energy vector and energy preservation criteria (see Figure 2), since we considered that listening was dedicated to a group and not a single listener at the "sweet spot".

This microphone system has been found quite impressive by its very small size (especially compared trees extended over several meters, as shown Figure 6) and its quite good results, although rendering actually suffered from 2 bugs. We report them here for people who attended the workshop: 2 capsules (over 20) were deficient and the 5.0 decoding delivered front left and front right signals in opposite phase with respect to the other channels. This explains some coloration issues and "phasiness" sensations in the comparative listening session. On the other hand, the 8.0 decoding (not bugged) has let a good impression in the listening room dedicated to HOA. On both 5.0 and 8.0 renderings, localization was found to be robust to out-of-centre listening.

A relatively "high level" has been observed in the rear channels of the HOA 5.0 (only a few dB below the front channels) especially compared with other multichannel trees. There are several possible explanations: rear/back microphones of other trees were several meters away from the front position; the sound scenes were generally quite close and therefore very large (Figure 6); perhaps early side reflections were relatively important (indeed rear signal was found unusually a bit loud for all systems); the spatial encoding was probably altered by the missing sensor signals and by possible calibration errors (preamp levels were adjusted individually and manually).

Further observations arise from subjective comparison experiments on these 5.0 contents, which have started in collaboration with the University of Brest (and which results are not published yet). After removing the "opposite phase bug", timbral features of HOA system become quite acceptable though perfectible, letting think that both encoding and decoding steps could be optimized with respect to this feature (especially in the spatial aliasing domain). Other trees, where signals captured by high quality microphones feed the loudspeakers without being matrixed, generally presented very nice timbral features. Regarding the relatively high back/front level ratio for HOA, its consequence appeared to be highly dependent on the reproduction room: a room with reflective surfaces on its back tends to alter the front image and imply front-back confusions. When one replaces reflecting ("live") surfaces by absorbing ("dead") ones, a correct perceived spatial organization is restored.

Once again, the angular span of the frontal scene is certainly an issue that one should consider with care for a comfortable ITU 5.0 rendering. It will be discussed again in 3.2.

Experimental symphonic recording with Radio-France, Paris, June 2008

The *EWO* experience has motivated Orange Labs and Radio-France teams to further experiment the HOA system on typical radiophonic contents, such as symphonic music, dramatic, public / entertainment programs. The first step has focussed on recording a working session of the Orchestre National de France conducted by Kurt Masur, playing Beethoven's 2nd Symphony at the "Studio 104" of Radio-France. Recording and mixing have been done in collaboration with sound engineer Christian Prévot and thanks to an invitation by Didier Gervais.

The recording setup involved a 4th order ambisonic microphone, a 7cm-∅, 32-sensor (DPA4060) spherical array; additionally 9 spot microphones were placed over strings, winds, and timpani. HOA and spot microphones were connected to high quality preamp and A/D converters (Radio-France) and finally processed in real-time by the HOA VST plugins (developed by Orange Labs, cf 2.3) hosted by PlogueBidule on a PC. For experimental purpose, the spatial encodings and decodings associated respectively to the HOA sphere and the spot microphones were done separately so as to provide separate 5.0 flows to the SSL mixing desk of the "Cabine 104" which hosted the monitoring. Spot mic panning (i.e. the tuning of angles in the HOA encoder) was adjusted under the instructions of Christian Prévot, by listening comparison between the sound images provided respectively by the sphere and the spot mic encoding (by switching between them).

An initial aim was to render the conductor's point of view, considering that "it also reflects the sound the composer was expecting while composing". It quickly appeared that this didn't fit the basic sound engineering criteria, especially in terms of direct/reverberated sound ratio. After some trials, the HOA microphone has finally been placed about 4 meters above the ground, nearly behind the conductor, overhanging the orchestra so as to have a good balance between front and back rows. To perfect the spatial impression, the trial of several decoding proposals has led to choose the "Craven" decoder, which brought the feeling of having more "air" in the sound image, maybe thanks the good separation between front channels and to the good velocity vector properties (see Figure 2 and associated comments).

On this configuration, a very clear “role distribution” could be noticed between front channels (orchestra sound, relatively dry), and rear channels (the reverberated sound). Rendering was judged comfortable. Indeed, Ch. Prévot told us that, considering the long and intensive listening sessions, the listening fatigue was quite low compared with what he usually experienced with other recording approaches… A sign of the quite good coherency and naturalness of reconstructed localisation cues!?

The experiment ended with listening sessions open to any sound engineer of Radio-France. A same music extract was given in three versions: the global capture by the "nude HOA sphere"; the "HOA sphere" plus the mixing of the spot microphones for the contrabasses, plus some artificial reverb; the "HOA sphere" plus all spot microphones plus artificial reverb (as it seems to be a frequent practice). A shared opinion was that adding spot microphones made the sound images less clear and localisable, although it surely reinforced timbral features of instruments. Otherwise opinions were divided, be it in terms of space interpretation or in terms of preference, which also revealed various sound recording philosophies, expertises and listener sensibilities. One listener regretted the lack of depth (foreground vs background), which could objectively be explained by the fact the HOA sphere captured the orchestra nearly "from the top". Some people preferred the "nude HOA version" and found it sufficient to greatly read and feel the space ("*what misses would be just the smell*", to quote one listener). Other people found the mix with spot microphones necessary for colour features. Highly experienced people generally had strong timbral references "in their ears", and therefore clear expectations on how instruments should sound. Regarding timbral objections, one could wonder whether the HOA encoding/decoding system has to be further optimized (maybe!), or if anyway a "faithful reproduction" of what would have be listened at the microphone place doesn't fit the usual sound engineering requirement.

« La Trahison Orale », Opera de Rennes, Feb. 2009

To prepare the 3D audiovisual capture and diffusion of Mozart's « Don Giovanni » at the *Opera de Rennes* described *infra*, trial HOA recordings have been prealably performed in the same place in order to appreciate its acoustics and to find the best sphere microphone placement for a good spatial imaging and impression. To anticipate "Don Giovanni", the sphere was constrained to be outside the cameras' field of view, that means: either below or above. Indeed, the HOA sphere (the EigenMike™ in this case) has been successively placed at the first row of the stalls (at the level of a listener ears, position "A" in Figure 7), and hanged about 8 meters above the orchestra pit (position "C" in Figure 7). Trials have been done on a works of Mauricio Kagel, "La Trahison Orale", with an interesting stage set and spatial sound sources distribution. The stage set consisted of a 9-meter height front wall with 3 floors and 3 windows at each floor (Figure 8). Therefore narrators/singers and even some instruments could take place at various elevation, depth and of course lateral positions on the stage.

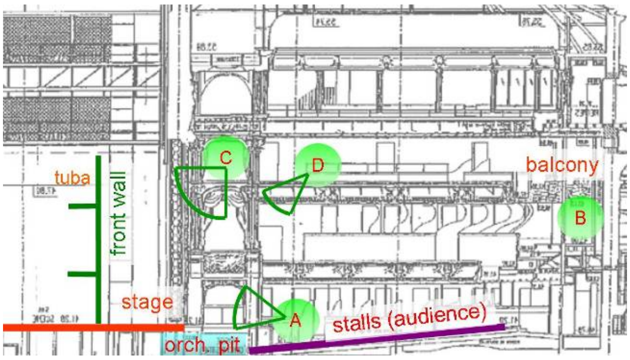

Figure 7 – Side view of the *Opera de Rennes* (stage set represented for "La Trahison Orale"), with 4 tried HOA microphone positions: A (1st row at stalls), B (balcony), C (overhanging the orchestra pit), D (overhanging the 2nd or 3rd row of the stalls, for "Don Giovanni"). "Pie chart like" sectors suggest the span of musical sources in terms of elevation angles.

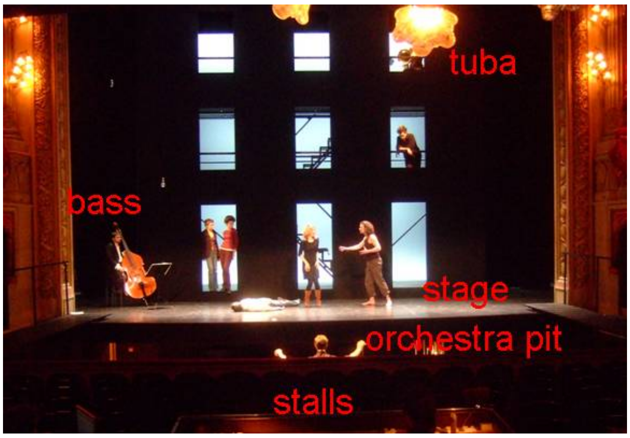

Figure 8 – Front view of the stage set for "La Trahison Orale" at the *Opera de Rennes* (during a rehearsal)

We've had to face two kinds of issues. The first issue would concern any global recording approach, which is to provide an acceptable balance between instruments, vocalists, foreground and background, direct and reverberated sound, and preferably a good spatial readability. In the case of the Opera de Rennes, which is relatively small and not acoustically very generous, we came to the conclusion – already predicted by Radio-France sound engineers – that using the HOA sphere alone, no ideal placement could be found to fit the audio production requirements. HOA capture and reproduction could be found faithful or "realistic", but artistically not sufficient. The "bottom" microphone location ("A") globally provided a very good spatial readibility in terms of angular localisation. But on the other hand, the balance between instruments and voices was not satisfactory. Furthermore the room acoustics at this place was very dry. And finally there was the risk of noise disturbance from the very close people of the audience. The "suspended" sphere location ("C") provided much more room presence (though a not very long reverb), but generally less localization accuracy: it became a bit less in the best case, and very problematic for a number of source locations. Let's mention a third positioning option (not retained): the sphere has been placed at 1st balcony level (position "B" in Figure 7), where the sits are reputed to be the best for the sound balance from the audience point of view. The problem was that the balcony reflections that made the sound better balanced, also completely destroyed the localization effect!

We come now to the second issue, related to the projection of a 3D sound field on the "2D" reproduction plane (when targeting a horizontal reproduction setup, esp. the ITU 5.0 setup). From the stalls point of view (position "A" in Figure 7), the most critical sources were at about 45° of elevation, and their 5.0 reproduction made them still localized on a front direction though with less accuracy ad stability. From the suspended point of view (position "C" in Figure 7) and thanks to a sound field rotation plug-in, the "projection plane" (or "microphone virtual orientation") could be adjusted so as to focus on the orchestra pit (projection plane oriented vertically) or the 3rd floor (nearly horizontal plane) or a intermediary part of the stage (inclined plane). Therefore in the first case, the orchestra was reproduced as being in the horizontal plane (with respect to the loudspeaker setup) whereas the tuba and sometimes narrators were at a relative elevation angle of 90°. Over a 5.0 ITU setup, the latter were mainly reproduced by rear loudspeaker, which went against the listener's expectations could be not well accepted. This spatial configuration wouldn't be such a problem with a 3D reproduction (*e.g.* hemispherical) setup.

A listening of these recording trials has been proposed to sound engineers from Radio-France to decide the HOA sphere location choice for "Don Giovanni". It resulted in a preference for the suspended sphere location, because of the room presence and the better balance between sound sources. Let's comment that the "Don Giovanni" spatial configuration wasn't as critical as the one of "La Trahison Orale" since singers stood simply on the stage ground.

Don Giovanni in 3D/HD, Opera de Rennes, June 2009

The "Don Giovanni" recording and diffusion event has been the first step of a project called 3DLive, mainly financed by the Media&Networks Cluster (www.images-et-reseaux.com). The aim was to simultaneously capture, broadcast and render the last performance of Mozart's "Don Giovanni" at the *Opera de Rennes*, in both 3D and HD versions. For further details, a more complete presentation of this operation can be found on the Media&Networks Cluster web site[3]. Orange Labs managed the 3D video capture (with stereoscopic cameras) as well as the coordination of sound capture and diffusion. For the latter, the HOA capture and 5.0 decoding setup (using the EigenMike and the Orange Labs' HOA plugins suite) took place within a much larger setup managed by a Radio-France team. This comprised between four and five dozens of additional microphones (mainly spots in the orchestra pit and at the stage border), which signals was conveyed together with the HOA 5.0 signals to the mobile unit of Radio-France. From there, sound director Cyril Bécue assisted by Paul Malinowsky, performed a "live" 5.0 mix dedicated to the "3D" audiovisual production (broadcasted by satellite towards 6 different places in France). Christian Lahondès was in charge of a stereo downmix dedicated to the HD-only and SD versions (the latter being broadcasted by Mezzo TV and TVRennes35). Mixing/panning of spot microphone signals was done in accordance to a far camera view showing the entire stage, itself quite coherent with the global HOA capture (the sphere being hanged about 8 meters above the stalls first flows, see Figure 9). The goal of spot mixing was to emphasize the timbral quality of voices and instruments. For voices, this naturally required to tune stage mic level as a function of the singers' locations. Furthermore, as the acoustics in the Opera was "not very generous", an artificial reverberation has been added (on spot mikes mix as well as on the 5.0 ambience captured by the HOA), which made feel the theatre more spacious than in reality.

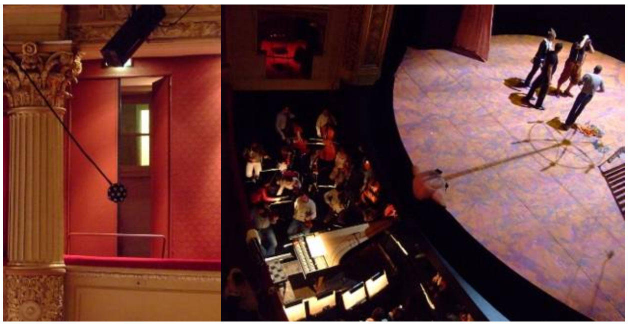

Figure 9 – The HOA microphone, overhanging the stalls (about 8 m above and 2 m behind the conductor, see also position "D" in Figure 7). View of the orchestra pit and the stage (during rehearsal).

The sound result was unanimously judged as great by listeners in the different reproduction places. Now, several technical and artistical aspects can be discussed. In the present configuration, audio and spatial fidelity with respect to the HOA sphere point of view seemed not compliant with artistic requirements. Nevertheless people familiar to the Opera de Rennes recognized something very natural and "realistic" by listening to the "nude HOA version". With the final mix version, other listeners mentioned the slight regret of no longer having sound directivity effects coherent with the actual singers' orientation. These comments make desire to still have the feeling of "being there" while improving the presence and/or timbre of sources. Such a purpose would require imagining technological improvements and new tools to assist the spot-mic mixing, maybe according to HOA encoding rules. Another debate concerns the audio mixing and panning strategy with respect to the camera viewpoints chosen in the video production. Two opposite positions can be taken: provide a constant and stable view angle of the audio scene; or make both visual and audio imaging spatially coherent, which can generate some annoyance if done in a systematic and too changing way (already experimented on a post-production by Radio-France). In the context of direct production and broadcasting, the latter option was anyway unworkable, all the more that there were simultaneously two distinct video productions (one "3D" and one simply "HD"). In the author's opinion, it would be worth working of joint audio-video panning at least for visible sound sources. Indeed, especially when both visual and audio imaging technologies provide strong localization cues, it might be relevant to ensure their coherency so as to not unsettle the spectators.

[3] See these pages: http://www.images-et-reseaux.com/en/l-actualite/fiche.php?id=322 and http://www.images-et-reseaux.com/upload/actualite/fichier/320fichier1.pdf

To conclude on this event that was a great artistic, technical and popular success, let's mention the nice communication impact for HOA. One had never heard/or read from HOA at such a wide audience level. (Just type "HOA" and "Don Giovanni" in a web search engine!)

3.2. Summary / lessons / open issues

From research to sound engineering positions

Our initial position *as R&D engineers* when starting experimental HOA recordings was to render the acoustic reality as faithfully as possible, in other words: to adopt a *spatial fidelity or objectivity principle*, as the HOA is supposed to be designed for. One lesson, by working *with or as sound engineers,* has been that this doesn't necessarily fit usual artistic production requirements (unless for *e.g.* outdoor, ambience recordings). Indeed, the acoustic configuration of the recording place and the microphone positioning constraints, make in numerous cases unreachable the quest of an ideal recording point of view. Nevertheless in many listening opportunities a very strong feeling of "being there" was reported by the listeners. This is generally emphasized when listening experience is associated with video (especially stereoscopic video).

Dealing with the ITU 5.1 setup constraints... mistrust monitoring over another setup!?

Before experimenting more closely with professional sound engineers, we (at Orange Labs) used to listen recordings on an 8.0 setup because it was simply more pleasing that a 5.0 setup. This practice is restrictive to get lessons from recording experiences and derive rules for using the HOA system while targeting a standard multichannel setup. Indeed, additional loudspeakers at ±70° facilitate the rendering of robust side/lateral images, and hide the front-back separation issue that occurs with large frontal scenes. Therefore a sound imaging that is very satisfying with 8.0 monitoring, might become problematic with a 5.0 rendering. More generally, this is the problem of retrogressing to a more severely "sampled" setup. To disambiguate the 5.0 rendering and make it more robust and transparent, alternative decoding strategies are currently studied. Similar issues occur when a 3D sound field including sound sources at high elevation angles is *projected* over a "2D"/horizontal setup, as reported in 3.1: a spatial organisation that should sound clear over a 3D setup (e.g. a spherical or hemispherical loudspeaker array) might yield unexpected, if not undesirable, localization effect. Finally, monitoring over headphones (using a binaural decoder) generally sounds good, *modulo* usual difficulties of headphone presentation to provide external and frontal sensations of localization. But one must take care that it may hide varying interference effects that could occur with loudspeaker reproduction, when the listener's head moves within a sound field synthesised by loudspeakers. To summary, what is challenged there is the transparency of the reproduction system, which motivates further studies on spatial decoding strategies. To take the problem by the other side and to adopt the position of the devil's advocate, this also questions the concept of universality of a generic 3D content such as represented in a HOA format.

Specific vs. generic thinking of microphone positioning and/or spatial organisation

When addressing a sound scene to be spatially recorded either with HOA or another system, the "global microphone" positioning has an impact and therefore should be adjusted depending on the desired width of the sound scene, the direct/reverberated sound ratio, the potential disturbance from the audience, etc. From the recording experiments reported in 3.1, it seems that even with a HOA 3D recording, the sound engineer might make choices that will be specific to the target setup, the 5.0 ITU. In this context the following question becomes crucial: "*Is the main sound scene supposed to be frontal?"* Traditionally music or theatre scenes are assumed to be frontal. In the quest of further enlarging this frontal scene (beyond the front loudspeaker angular limits), there is the risk, with a 5.0 ITU setup, that the most lateral sources be "rejected" to the rear loudspeakers, for listeners in the rear half part of the listening area. It might be less risky though apparently more audacious, to record and render a full panoramic scene since in the latter case the listeners can accept to locate sounds at rear positions.

Reproducing the spatial organisation in an objective and coherent way?

From an artistic point of view, in it not always required to reproduce the strict angular location of each source of the recorded sound scene, since the listener has a capacity of building an internal spatial organisation, and since anyway angular distortions occur depending on the listening point during the reproduction. Nevertheless it seems important to provide consistent localization cues so as to not imply too much interpretation effort and therefore listening fatigue. When video capture and reproduction is associated, one can debate on whether proposed audio and visual views should be coherent or not in terms of image localisation. Experiments reports in 3.1 have shortly outlined such a debate, which isn't specific to HOA by the way. We won't develop it further in this paper.

HOA format and technology can bring facilities and some answer to this issue. Indeed HOA has the potential to apply changes to the spatial organisation of the recorded sound field (angular distortion, thus go against a certain spatial fidelity) while still providing consistent localisation cues regarding each elementary contribution (*i.e.* individual wave front). Spatial editing tools would have to be further developed to offer to the sound engineer the possibility to adapt the recorded sound scene to its desires.

Mixing with close or spot microphones

In several experiments, signals from close or spot microphones have been mixed with the global HOA capture. One should highlight that this may not respect the “HOA spatial encoding model”, and therefore the spatial imaging mechanism peculiar to HOA. Of course, mixing and panning in a “discrete loudspeakers” multichannel domain (e.g. 5.0 or 8.0) is no longer “HOA” encoding. Moreover, mixing signals from spot microphones raises issues whatever the encoding/panning tool (even a “HOA encoding”): indeed, the plane wave encoding model associated to direct sounds is destroyed by the crosstalk between recorded sources and spot microphones. Finally, the only way of preserving a “valid” HOA encoding is to spatialize signals from very close microphones (with virtually no crosstalk effect, as *e.g.* with Lavallier microphone) in the HOA domain (with a HOA encoder).

Using HOA as an end-to-end format or just as a technological toolbox? Now and in the future? Make steps towards sound engineers… then reciprocally!?

As we began to work more closely with professional sound engineers on short term experiments (especially for “Don Giovanni”), our approach has been to adapt the HOA technology to their practice and work environment, which means: use HOA as a toolbox and no longer as an end-to-end approach and format. At the same time and reversely, a mid or long term approach consists in inviting sound engineers / content creators to adapt their practice and get familiar to HOA concepts and tools, and to help to make both research and sound engineering worlds converging.

Technical demands and recommendations

Very practical issues have been pointed out by sound engineers: the recording system latency should be reduced (currently about 40ms with the EigenMike, including ADC, FireWire, ASIO buffering, digital filtering, etc.); software interface could be simplified (one processing box and interface instead of splitted spatial encoding, transformation and decoding interfaces if the intermediary HOA format is not necessary); solutions for outdoor use should be adapted (appropriate windshield, check resistance to moisture, etc.).

4 CONLUSION

From the birth of HOA, development efforts and concerns have evolved from theoretical to technological issues, and finally to "real life" experiments. Developed tool boxes (*e.g.* a HOA VST plugins suite) and demonstrators have permitted a number of informal and formal evaluations that help refining design criteria. They have given convincing proofs of concept, especially when combining real sound field recording with spherical microphone arrays, and surround diffusion on more or less standard setup (5.0 ITU and extended).

Beyond these proofs of concept, recording opportunities with production concerns have multiplied, with numerous combinations of conditions and constraints. Some recent collaborative work with professional sound engineer was especially informative on the way of approaching 5.0 multichannel content generation. As a result, exchanges between the research and sound engineering worlds reveal that their convergence is "on the way" while still being a work in progress. This involves both communication and joint experimentation efforts. One may have to simplify and at least vulgarise some essential HOA concepts that are unfamiliar to the usual sound engineering world (intermediary HOA format *versus* “one microphone for one loudspeaker” traditional approach). Convergence also concerns the way of describing the auditory sensations, implying the learning of a common vocabulary. A first step has been to adapt HOA technology to the current standard constraints and practices. A further step would be that sound engineers / content creators adapt or transpose their practice to HOA-specific concepts, and to bring them to think in terms of a generic content that could be rendered on various setups. This assumes to further develop, improve and integrate HOA spatial processing tools, with care on ergonomics, provision of usual audio effects collection (including dynamic range processing, reverb, etc.). This also invites to draw up kinds of "content creation rules" including spatial organisation concerns, audiovisual coherency issues, etc.

Reported experiments mostly targeted the standard 5.0 ITU setup. It has indeed the merit of existing and enriching the listening experience compared with 2-channel stereo, even though it can be criticized because its poor ability to render lateral sound images. One shall also foresee even richer formats and setups (e.g. 22.2), which experimentation and use are currently growing in Japan (cf Hamasaki work at NHK for Super-High Vision).

Finally, interesting remaining topics on the technological side are: to improve the spatial decoding towards a better reproduction transparency on any horizontal or "3D" setups; to develop advanced spatial audio editing tools that could assist and partially automate content creator's tasks; to develop relevant compression schemes for HOA and converge towards a workable broadcast format.

REFERENCES

[1] Gerzon, M.A., *Maximum Directivity Factor of n'th-Order Transducers.* J. Acoust. Soc. Am., 1976. **60**: p. 278-280.

[2] Bamford, J.S., *An Analysis of Ambisonics Sound Systems of First and Second Order*. 1995, University of Waterloo: Waterloo, Ont., Canada.

[3] Poletti, M., *The Design of Encoding Functions for Stereophonic and Polyphonic Sound System.* J. Audio Eng. Soc., 1996. **vol. 44**(11): p. 948-963.

[4] Daniel, J., J.-B. Rault, and J.-D. Polack. *Ambisonic Encoding of Other Audio Formats for Multiple*

Listening Conditions. in *AES 105th Convention*. 1998. San-Francisco, USA.

[5] Daniel, J., *Représentation de Champs Acoustiques, Application à la Transmission et à la Reproduction de Scènes Sonores Complexes dans un Contexte Multimédia*. 2000, University of Paris 6: Paris, France.

[6] Nicol, R. and M. Emerit. *3D-Sound Reproduction over an Extensive Listening Area: A Hybrid Method Derived from Holophony and Ambisonic*. in *AES 16th Int. Conference on Spatial Sound Reproduction*. 1999. Rovaniemi, Finland.

[7] Sontacchi, A. and R. Höldrich. *Further investigations on 3D sound fields using distance coding*. in *DAFX-01*. 2001. Limerick, Ireland.

[8] Malham, D. and R. Furse, *Second and Third Order Ambisonics - the Furse-Malham Set.*

[9] Gerzon, M.A. *General Metatheory of Auditory Localisation*. in *92nd AES Convention*. 1992.

[10] Daniel, J., J.-B. Rault, and J.-D. Polack. *Acoustic properties and perceptive implication of stereophonic phenomena*. in *AES 16th International Conference*. 1999. Rovaniemi, Finland.

[11] Gerzon, M.A. *Psychoacoustic Decoders for Multispeaker Stereo*. in *AES 93rd Conv.* 1992. San Francisco, USA.

[12] Daniel, J. *Spatial Sound Encoding Including Near Field Effect : Introducing Distance Coding Filters and a Viable, New Ambisonic Format*. in *AES 23rd International Conference*. 2003.

[13] Adriaensen, F., *Near Field filters for Higher Order Ambisonics*. 2006.

[14] Daniel, J. and S. Moreau. *Further Study of Sound Field Coding with Higher Order Ambisonics*. in *AES 116th Convention*. 2004. Berlin.

[15] Daniel, J., R. Nicol, and S. Moreau. *Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging*. in *AES 114th Convention*. 2003. Amsterdam.

[16] Fazi, F. and P. Nelson. *A theoretical study of sound field reconstruction techniques*. in *19th Internation Congres on Acoustics*. 2007. Madrid.

[17] Craven, P.G. *Continuous Surround Panning for 5-speaker reproduction*. in *AES 24th International Conference on Multichannel Audio*. 2003.

[18] Choi, C.H., et al., *Rapid and stable determination of rotation matrices between spherical harmonics by direct recursion.* Journal of Chemical Physics, 1999. **111**.

[19] Noisternig, M., et al. *A 3D Ambisonic based Binaural Sound Reproduction System*. in *AES 24th International Conference*. 2003. Banff, Canada.

[20] Faure, J., J. Daniel, and M. Emerit, *Optimization of Binaural Sound Spatialization Based on Multichannel Encoding*.

[21] Moreau, S., J. Daniel, and S. Bertet. *3D Sound Field Recording with Higher Order Ambisonics – Objective Measurements and Validation of a 4th Order Spherical Microphone*. in *120th AES Convention*. 2006. Paris.

[22] Meyer, J. and T. Agnello. *Spherical microphone array for spatial sound recording*. in *AES 115th Conv*. 2003. New York.

[23] Rafaely, B., *Analysis and Design of Spherical Microphone Arrays.* IEEE Transaction on Speech and Audio Processing,, 2005.

[24] Epain, N. and J. Daniel. *Improving Spherical Microphone Arrays*. in *AES 124th Convention*. 2008. Amsterdam, The Netherlands.

[25] Dedieu, S. and P. Moquin, *Microphone array with physical beamforming using omnidirectionnal microphones*. 2007.

[26] mh-acoustics, *em32 Eigenmike® microphone array (http://www.mhacoustics.com/page/page/2949006.htm).*

[27] Bertet, S., et al. *Influence of microphone and loudspeaker setup on perceived higher order ambisonics reproduced sound field*. in *Ambisonics Symposium'09*. 2009. Graz, Austria.

[28] Schmidt, J. and O. Baum, *Text of ISO/IEC 14496-11/2003 PDAM-3*. 2003(ISO/IEC 2003: N6207).

[29] Schmidt, J. and E. Schröder. *New and Advanced Features for Audio Presentation in the MPEG-4 Standard*. in *AES 116th Convention*. 2004. Berlin, Germany.

[30] Bertet, S., et al. *Investigation of the perceived spatial resolution of higher order ambisonic sound fields : a subjective evaluation involving virtual and real 3d microphones*. in *AES 30th International Conference*. 2007. Saariselkä, Finland.

[31] Pulkki, V. and T. Hirvonen, *Localization of virtual sources in multichannel audio reproduction.* IEEE Transactions on Speech and Audio Proc., 2005. **13**(1): p. 105-119.

[32] Solvang, A., *Spectral impairment for two-dimensional higher order ambisonics.* Journal Audio Eng. Soc. 56(4): p. 265-279.

[33] Williams, M., News of the Sections / French ears wide open. Journal Audio Eng. Soc., 2008. 56(4): p. 310-311.