Audio Engineering Society
Convention Paper
Presented at the 114th Convention
2003 March 22–25 Amsterdam, The Netherlands

Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging

Jérôme Daniel[1], Rozenn Nicol [1], and Sébastien Moreau[1]

[1]France Telecom R&D, 2 Avenue Pierre Marzin, 22307 Lannion Cedex, France

jerome.daniel @ francetelecom.com, rozenn.nicol @ francetelecom.com, sebastien.moreau @ francetelecom.com

ABSTRACT

Ambisonics and Wavefield Synthesis are two ways of rendering 3D audio, which both aim at physically reconstructing the sound field. Though they derive from distinct theoretical fundamentals, they have already been shown as equivalent under given assumptions. This paper further discusses their relationship by introducing new results regarding the coding and rendering of finite distance and enclosed sources. An updated view of the current knowledge is first given. A unified analysis of sound pickup and reproduction by mean of concentric transducer arrays then provides an insight into the spatial encoding and decoding properties. While merging the analysis tools of both techniques and investigating them on a common ground, general compromises are highlighted in terms of spatial aliasing, error and noise amplification.

1. INTRODUCTION

Among sound spatialisation technologies, both Wavefield Synthesis (WFS) and High Order Ambisonics (HOA) aim at physically reconstructing the sound field, though they historically belong to distinct worlds. Whereas WFS is considered as *the* solution for providing large listening areas, Ambisonics is originally known as dedicated to surround systems having a limited sweet spot. Nevertheless, the latter's extension to higher spatial resolutions (HOA) has known an increasing interest during past years, featuring scalability and flexibility properties in addition to enlarged listening areas.

The present paper provides new insights on WFS and HOA by analysing them on a common ground. It first supplies an updated state of art (theory and application). A bigger part is dedicated to HOA, to reflect recent progresses that allow it being practically compared with WFS. Then both technologies are investigated side-by-side. After completing the formal connection between their associated sound field representations, we list the reconstruction artefacts expected from practical system limitations. Next, virtual sound imaging simulations help comparing reconstruction properties, and characterising them in terms of spatial information consistency and plausible localisation effect. Enlarging considerations to "real" recording issues (*i.e.* involving microphone arrays), artefacts appear to be shared by both approaches since these obey the same practical limitations. Finally we derive preferences on encoding strategies, and compromises

on technical choices (array size).
To perform the comparison more efficiently, the scope of this paper focuses on concentric (*i.e.* circular or spherical), regular arrays. As a global result, a converging view of the two approaches is given, and a piece of "physical feeling" is offered to intuitively understand underlying phenomena.

2. WAVE FIELD SYNTHESIS (WFS)

2.1. Huygens' Principle

The Wave Field Synthesis is a concept of spatialised sound reproduction that was proposed by Berkhout in the late 80's [1, 2]. It may be identified to the acoustical equivalent to holography, and for this reason, it is sometimes referred to as "holophony" [3]. Indeed, WFS aims at reproducing sound waves (and especially the wave front curvature) by loudspeaker array. Physically, it is derived from the Huygens' Principle, and more precisely, from the idea, that a wave front may be considered as a secondary source distribution. In other words, the wave, which propagates from a given wave front, may be considered as emitted either by the original sound source (the primary source) or by a secondary source distribution along the wave front. As a consequence, the secondary source distribution may be substituted for the primary source, in order to reproduce the primary sound field.

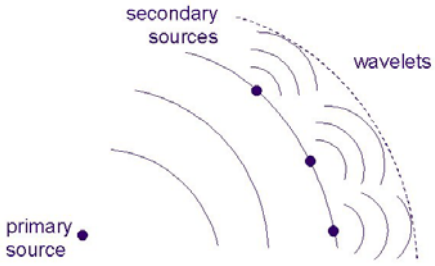

Figure 1 Illustration of the Huygen's Principle

2.2. Kirchhoff-Helmholtz Integral

The Kirchhoff-Helmholtz Integral expresses this idea in a mathematical way. The acoustical pressure p within a given area A is derived from the knowledge of the acoustical pressure p_0 and its gradient $\vec{\nabla} p_0$ over the boundary ∂A of the considered area:
$\forall \vec{r} \in A$,

$$p(\vec{r}) = \iint_{\partial A} [\vec{\nabla} p_0 . \vec{n} - \frac{\vec{R}}{R} . \vec{n} (1 + jkR) \frac{p_0}{R}] \frac{e^{-jkR}}{4\pi R} dS_0 \qquad (1)$$

with wave number k and unitary outside normal $\vec{n}$. Vector $\vec{R}$ defines the propagation path between a secondary source and the listening point.
The Kirchhoff-Helmholtz Integral may be interpreted as a continuous distribution of secondary sources. Each secondary source is composed of two elementary sources: one monopole, which is fed by the pressure gradient signal, and one dipole, which is fed by the pressure signal.
It should be noticed that the Kirchhoff-Helmholtz Integral, contrary to the Huygens' Principle, does not require that the boundary should be a wave front. The boundary may follow any geometry, which does not depend on the wave front. This remark highlights a noticeable difference between the two formulations: in the Huygens' Principle, the secondary sources are driven only by the magnitude signal, whereas in the Kirchhoff-Helmholtz Integral, they are driven both by the magnitude and the phase signal. Indeed, it should be kept in mind that in the former case, the secondary sources are distributed along a wave front, which is defined as an equal phase surface. To some extent, the Kirchhoff-Helmholtz Integral generalizes the Huygens' Principle by adding one degree of freedom for the secondary source distribution geometry, which is paid by increasing source signal complexity.

2.3. Application to spatialised sound recording and reproduction

The Kirchhoff-Helmholtz Integral gives a straightforward way of reproducing a sound field. At the recording stage, the listening area is surrounded (top of Figure 3) by a microphone array, which is composed of both pressure and velocity microphones, and which records the primary sound field due to external sources (Figure 2).

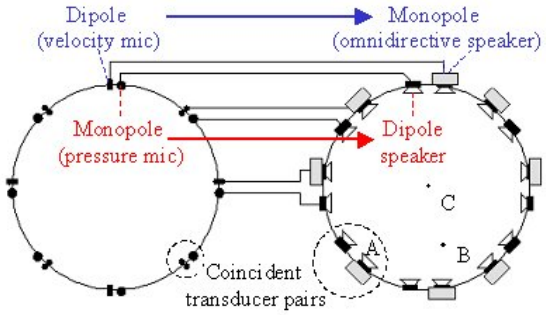

Figure 2 Application of Kirchhoff-Helmholtz Integral for holophonic sound field reconstruction.

For the reproduction stage, loudspeakers are substituted for the microphones, by replacing the

pressure microphones by dipole sources and the velocity microphones by monopole sources. Each loudspeaker is fed by the signal that was previously recorded by its associated microphone (Figure 2). It should be kept in mind that the geometry of the microphone array and the loudspeaker array should be identical. Another setup, which is exactly equivalent, consists in surrounding the primary source area by the microphone array, instead of the listening area (bottom of Figure 3).

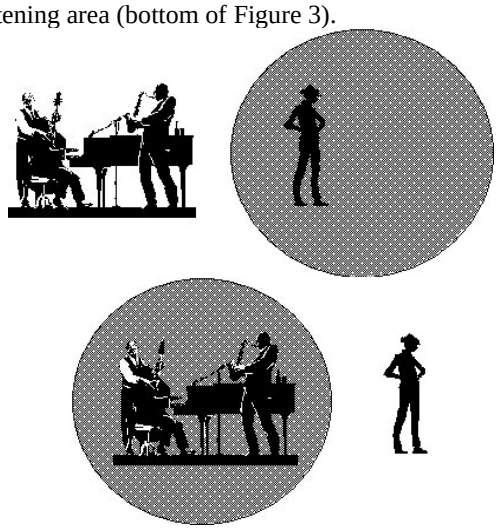

Figure 3 Two equivalent holophonic setups: by surrounding either the listener (top), or the primary sources (bottom).

The key-features of this solution of sound spatialization should now be pointed out.
Firstly, provided that the process is ideally followed (which implies for instance ideal transducers and continuous arrays), the Kirchhoff-Helmholtz Integral ensures that the sound field synthesized by the secondary sources is reproduced identical to the original one, which means that the temporal and spatial properties of the primary sound field are restored. Particularly, the localization of the sound sources is fully rendered and the listener will perceive and localize the sound sources as he would do in a real listening situation.
What's more, the sound field reproduction is valid not only at one point, but at any point within the whole area, which is delimited by the transducer array. An extended listening area is thus provided, which allows the listener to move inside the listening space and also to share this space with other listeners.
Finally, it should be mentioned that, theoretically, the process requires no signal processing between the recording and reproduction stage. The all process complexity is managed by the physics, *i.e.* the reconstruction work is handled by wave interference between the secondary sources.

2.4. Practical limitations

Though the Kirchhoff-Helmholtz Integral provides a very attractive solution of spatialised sound recording and reproduction, practical limitations are obvious.
First, it requires continuous, closed-surface transducer arrays, whereas only discontinuous arrays are available[1], which raises the problem of spatial sampling. Indeed, discrete arrays cannot correctly sample incident waves which wavelength is too small with regard to the transducer spacing $\Delta_{transducer}$. Such spatial aliasing typically occurs above the so-called "spatial aliasing frequency"[2]:

$$f_{sp} = \frac{c}{2\Delta_{\text{transducer}}}, \qquad (2)$$

Moreover linear *i.e.* not surface arrays are usually preferred in order to focus on the horizontal sound scene spatialisation.
Secondly, each secondary source is composed of two elementary transducers (both for the sound recording and reproduction), which should be coincident. This setup is also not strictly feasible.
Thirdly, the quality of the sound field reconstruction depends on the transducer characteristics, which should be the closest to the ideal one.

2.5. From holophony to WFS: approximating the Kirchhoff-Helmholtz Integral

In spite of its practical limitation, sound spatialization by holophony is not so unfeasible as it could seem at first sight. Indeed, research carried on by Berkhout & al at the TUD has shown that holophonic systems are available, provided that some approximations are applied to the Kirchhoff-Helmholtz Integral [1, 2]. These approximations define the Wave Field Synthesis concept, which has been developed by the acoustic laboratory of TUD.
Three main approximations, which are essentially based on physical feeling, have been pointed out.

Stationary phase approximation

First, the stationary phase approximation allows reducing the ideally surface transducer array to a linear horizontal "slice", in order to keep only the most useful secondary sources, according to the primary source and the listener positions.

[1] Nevertheless, the new technology of DML (Distributed Mode Loudspeaker), which are based on large vibrating panel, offers a promising answer to this issue for holophony and WFS [4].
[2] To be exact, this frequency depends also on the wave incidence with regard to the array.

Single directivity transducer array

Secondly, it has been remarked that the two elementary transducers (monopole and dipole) of the secondary source are highly redundant, so that only one of them is necessary. Thus the WFS concept practically uses monopole loudspeaker array fed by figure-of-eight or even cardioid microphones (Figure 4). Nevertheless, real-life loudspeaker or microphone directivity is neither ideal monopole, nor ideal dipole, but rather cardioid for the medium frequencies with lower directivity at low frequencies and higher directivity at high frequencies.

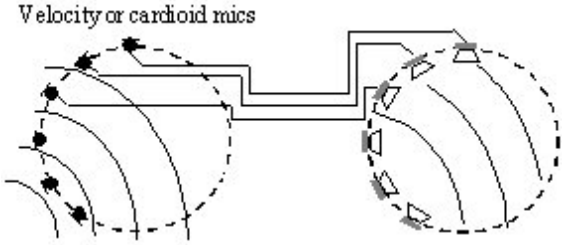

Figure 4 Restriction to single directivity transducer arrays for practical holophonic system.

Notional source encoding

Thirdly, most often, the recording is not made by microphone array, but by close microphones, which pick up the direct sound of each primary source. The microphone signal is then propagated to the virtual microphone array by applying amplitude weight and time delay, as suggested by Figure 4. Each microphone signal is thus identified to one individual primary source and may be considered as a virtual substitute for this source, *i.e.* a notional source. Such "virtual recording" allows also windowing secondary source amplitudes [5], *i.e.* using "unreal" microphone directivities, if needed for a better final rendering. Close miking provides several other advantages. In most cases, the number of microphones (and consequently the number of recorded signals) is greatly reduced in comparison with microphone array. Moreover, for up-mixing purpose, standard monophonic recordings may feed WFS rendering without extra signal processing.

Example of application

As an example, WFS has been implemented and is being experimented at the France Telecom R&D Labs, over either square, polygonal, or circular arrays composed of 48 loudspeakers (Figure 5). Especially, subjective experiments are driven in the context of the Carrouso project, with other European partners (http://www.emt.iis.fhg.de/projects/carrouso/).

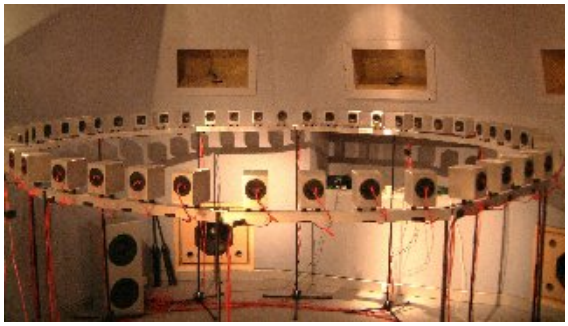

Figure 5 Quasi-circular 48-speaker array for WFS and HOA rendering experiments at the France Telecom R&D Labs

2.6. Separating the microphone and the loudspeaker arrays

It was previously pointed out that the microphone array and the loudspeaker array must be identical, mainly in terms of geometry and number of transducers. With the notional source concept, this property is already invalidated. As a matter of fact, it can be further stated that it is always possible to circumvent this constraint and to fully dissociate the microphone array from the loudspeaker array. As for the notional source concept, the two transducer arrays can be dissociated by simulating the acoustic propagation between the actual microphone array and the actual loudspeaker array, where should be the theoretical microphone array. This is done by interfacing an extrapolation matrix between the microphones and the loudspeakers [1].

2.7. Synthesizing enclosed sound source

Another constraint of the Kirchhoff-Helmholtz Integral may be overcome. Concerning the position of the primary sources, it should be realised that in theory, the listening area, which is delimited inside the loudspeaker array (Figure 3), should be free from any primary source. In other words, the loudspeaker array is able to synthesize only sound sources, which are outside the loudspeaker array. However, early developments of the WFS concept has pointed out that it is also possible to synthesize sound sources inside the loudspeaker array, merely by inverting the phase of the secondary sources, so that the last fed loudspeaker becomes the first fed and vice versa, in order that a concave wave front, instead of a convex one, is reconstructed [6].

This process may be compared with the time reversal approach applied to sound focusing [7, 8]. Indeed sound focusing by WFS may be identified to time reversal restricted to the direct sound.

3. HIGHER ORDER AMBISONICS (HOA)

Ambisonics was developed several decades ago, mostly by Gerzon, as a spatial sound encoding approach dedicated to surround (2D) and periphonic (3D) multi-speaker systems [9] [10]. For about eight years, its extension to higher spatial resolution systems has been the object of increasingly numerous studies, which promising features are becoming practicable. The following state of art includes recent and relevant progress in this field.

3.1. Mathematic fundamentals

Ambisonic representation is based on the spherical harmonic decomposition of the sound field, which comes from writing the wave equation in the spherical coordinate system (Figure 6) where a point $\vec{r}$ is described by a radius *r*, an azimuth θ and an elevation δ.

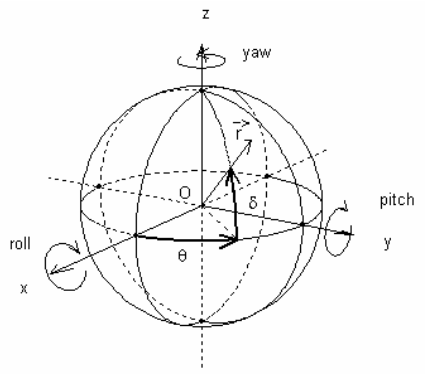

Figure 6 Spherical coordinate system, with the three elementary rotation degrees

Therefore the pressure field can be written as the Fourier-Bessel series (3), which terms are the weighted products of directional functions $Y_{mn}^{\sigma}(\theta,\delta)$ (called "spherical harmonics") and radial functions:

$$p(\vec{r}) = \sum_{m=0}^{\infty} j^m j_m(kr) \sum_{0\le n\le m,\sigma=\pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta,\delta) + \sum_{m=0}^{\infty} j^m h_m^-(kr) \sum_{0\le n\le m,\sigma=\pm 1} A_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta,\delta) \quad (3)$$

with the wave number $k=2\pi f/c$.

This is the general equation for the case of a sphere layer ($R_1 \le r < R_2$) that is free from sources (Figure 7). The weighting coefficients B_{mn}^{σ} associated with the spherical Bessel functions $j_m(kr)$ (first series) describe the "through-going" field (due to outside sources), whereas the coefficients A_{mn}^{σ} associated with the divergent spherical Hankel functions $h_m^-(kr) = j_m(kr) - jn_m(kr)$, describe the "outgoing" field (due to inside sources)[3].

As a sound spatialization approach, Ambisonics basically assumes a centered point of view, thus a centered listening area that is free of virtual sources. Thus only a "through-going" field, as represented by the coefficients B_{mn}^{σ}, is considered, the outgoing field being null ($A_{mn}^{\sigma} = 0$). Components B_{mn}^{σ} are the expression in the frequency domain of what we'll call "ambisonic" signals.

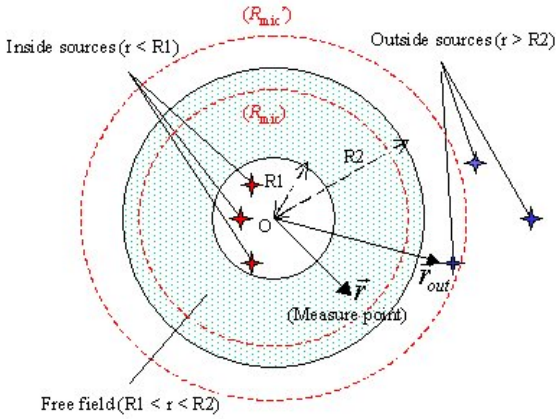

Figure 7 Inter-sphere free field volume where spherical harmonic representation (3) applies

The spherical harmonic functions

The spherical harmonic functions $Y_{mn}^{\sigma}(\theta,\delta)$ exhibited in (3) are defined as following:

$$Y_{mn}^{\sigma\,(\mathrm{N3D})}(\theta,\delta) = \sqrt{2m+1}\sqrt{(2-\delta_{0,n})\frac{(m-n)!}{(m+n)!}} P_{mn}(\sin\delta) \times \begin{cases} \cos n\theta & \text{if } \sigma = +1 \\ \sin n\theta & \text{if } \sigma = -1 \quad \text{(ignored if } n=0) \end{cases} \quad (4)$$

with the $P_{mn}(\zeta)$ being the associated Legendre functions[4] of degree *m* and order *n*, and where δ_{pq} equals to 1 if *p*=*q* and 0 otherwise (Kronecker symbol). They form an orthonormal base, *i.e.* $\left\langle Y_{mn}^{\sigma} \middle| Y_{m'n'}^{\sigma'} \right\rangle_{4\pi} = \delta_{mm'}\delta_{nn'}\delta_{\sigma\sigma'}$, in the sense of the spherical scalar product $\left\langle F \middle| G \right\rangle_{4\pi} = \frac{1}{4\pi}\oiint F(\theta,\delta)G(\theta,\delta)d\Omega$.

[3] Note that Hulsebos [11] combines outgoing field and ingoing (rather than "through-going") field.

[4] For computational application, the values of these functions can be derived using recurrence relations (see appendix A.2.2 of [12]).

Figure 8 provides a 3D view of spherical harmonics. There are $(2m+1)$ components, including 2 horizontal components (those with $n=m$), per order $m \geq 1$.

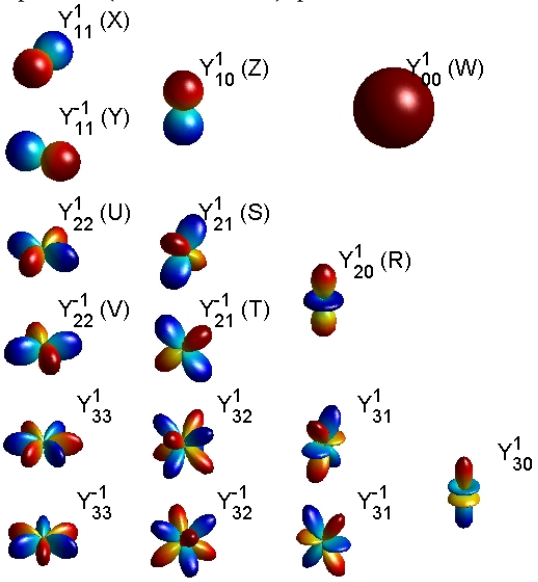

Figure 8 3D view (with respect to Figure 6) of spherical harmonic functions with usual designation of associated ambisonic components.

Interpretation: directional information and radial approximation

Spherical harmonic components B_{mn}^{σ} are closely related to the pressure field and its derivatives (or momentums) of successive orders around the origin *O*. The first four components are well known: $B_{00}^{+1} = W$ is the pressure, and $B_{11}^{+1} = X$, $B_{11}^{-1} = Y$, $B_{10}^{+1} = Z$ are related to its gradient or also the acoustic velocity. Each additional group of higher order components or momentums provides an approximation of the sound field over a larger neighbourhood of the origin with regard to the wavelength (Figure 10).

In practice, only a finite number of components (up to a given order *M*) can be transmitted and exploited, and even estimated. Thus the represented field is typically approximated by the truncated series:

$$p(\vec{r}) = \sum_{m=0}^{M} j^m j_m(kr) \sum_{0 \leq n \leq m, \sigma = \pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \delta), \qquad (5)$$

involving $K_{3D} = (M+1)^2$ components. An interesting interpretation comes from commenting Figure 8 and Figure 9 with respect to each other: harmonics $Y_{mn}^{\sigma}(\theta, \delta)$ with higher angular variability are associated with radial functions $j_m(kr)$ which first maximum occur at larger distances *kr*. To summarize: *higher directional resolution goes with greater radial expansion* and *vice-versa*.

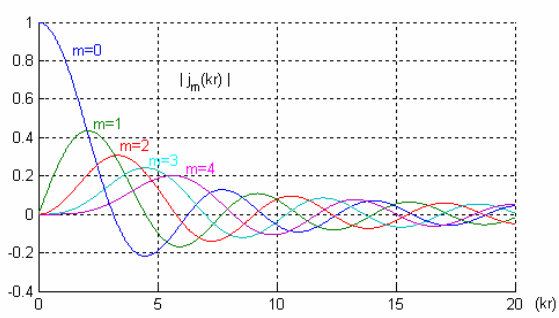

Figure 9 Spherical Bessel functions $j_m(kr)$

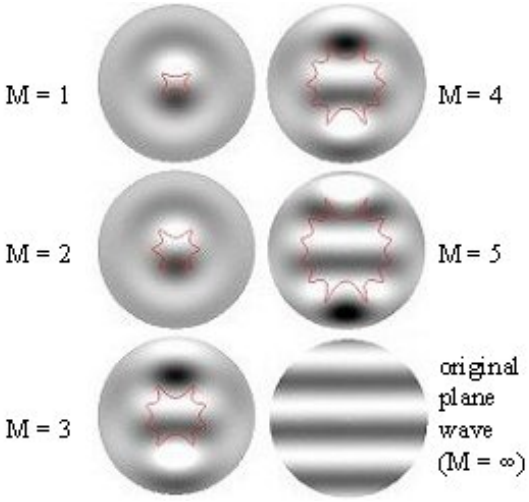

Figure 10 Monochromatic plane wave sound field and its approximation by the truncated Fourier-Bessel series (5) for several orders *M*

The plane wave case: directional encoding equations

The spherical harmonic decomposition of a plane wave of incidence (θ_S, δ_S) conveying a signal *S* leads to the simple expression of the ambisonic component.

$$B_{mn}^{\sigma} = S.Y_{mn}^{\sigma}(\theta_S, \delta_S) \qquad (6)$$

Thus a far field source is encoded by simply applying *real gains* to the received pressure signal *S*.

2D-restricted formalism: cylindrical decomposition

The cylindrical coordinate system has often been used in the literature when dealing with horizontal-only reproduction system and virtual sources [5, 12-14]. This leads to the Fourier-Bessel series:

$$p(r,\theta) = B_{00}^{+1(\mathrm{N2D})} J_0(kr) + \sum_{m=1}^{\infty} \left(B_{mm}^{+1\ (\mathrm{N2D})} \underbrace{\sqrt{2}\cos m\theta}_{Y_{mm}^{+1\ (\mathrm{N2D})}(\theta,0)} + B_{mm}^{-1\ (\mathrm{N2D})} \underbrace{\sqrt{2}\sin m\theta}_{Y_{mm}^{-1\ (\mathrm{N2D})}(\theta,0)} \right) J_m(kr) \qquad (7)$$

This way the 2D (horizontal) ambisonic components B_{mm}^{σ} derive from a kind of circular Fourier Transform of the sound field involving angular functions $Y_{mm}^{\sigma\ (\mathrm{N2D})}(\theta,0)$. They form an orthonormal base in the sense of the circular scalar product: $\langle F|G\rangle_{2\pi} = \frac{1}{2\pi}\int_0^{2\pi} F(\theta)G(\theta)d\theta$

One can unify the two formalisms by considering the circular (thus horizontal) harmonics as a subset of the spherical ones (4), modulo a weighting factor [12]:

$$Y_{mm}^{\sigma\ (\mathrm{N2D})}(\theta,\delta) = \sqrt{\frac{2^{2m}m!^2}{(2m+1)!}}\,Y_{mm}^{\sigma\ (\mathrm{N3D})}(\theta,\delta) \qquad (8)$$

The cylindrical formalism, which encoding functions (7) are simpler and less numerous ($K_{2D}=2M+1$) than the spherical ones, is useful to design the decoding for horizontal-only loudspeaker arrays (see next section).

3.2. The reproduction step: decoding design

The re-encoding principle

The design of ambisonic decoding basically relies on what could be called "the re-encoding principle" [12, 15]: the aim is to acoustically recompose encoded ambisonic components (pressure field and its "derivatives") $\tilde{B}_{mn}^{\sigma}$ (9) at the centre of the array. Assuming that loudspeakers are far enough from the listening centre point, their signals S_i are encoded as plane waves with coefficient vectors $\mathbf{c}_i$:

$$\mathrm{c}_i = \begin{bmatrix} Y_{00}^{+1}(\theta_i,\delta_i) \\ Y_{11}^{+1}(\theta_i,\delta_i) \\ Y_{11}^{-1}(\theta_i,\delta_i) \\ \cdots \\ Y_{mn}^{\sigma}(\theta_i,\delta_i) \\ \cdots \end{bmatrix} \quad \tilde{\mathrm{B}} = \begin{bmatrix} \tilde{B}_{00}^{+1} \\ \tilde{B}_{11}^{+1} \\ \tilde{B}_{11}^{-1} \\ \cdots \\ \tilde{B}_{mn}^{\sigma} \\ \cdots \end{bmatrix} \quad \mathrm{S} = \begin{bmatrix} \mathrm{S}_1 \\ \mathrm{S}_2 \\ \cdots \\ \mathrm{S}_N \end{bmatrix} \qquad (9)$$

Thus the re-encoding principle can be written in the matrix form (10), with $\mathbf{C}=[\mathbf{c}_1\ \ldots\ \mathbf{c}_N]$ being the "re-encoding matrix":

$$\tilde{\mathbf{B}} = \mathbf{C}.\mathbf{S}, \qquad (10)$$

The decoding operation aims at deriving signals **S** from matrixing original ambisonic signals **B**:

$$\mathbf{S} = \mathbf{D}.\mathbf{B} \qquad (11)$$

To ensure $\tilde{\mathrm{B}} = \mathrm{B}$, system (10) must be inverted. Therefore decoding matrix **D** is typically defined as:

$$\mathbf{D} = pinv(\mathbf{C}) = \mathbf{C}^T.(\mathbf{C}.\mathbf{C}^T)^{-1}, \qquad (12)$$

provided that there are enough loudspeakers: *i.e.* $N \geq K_{2D}$ or $N \geq K_{3D}$. The case of regular layouts[7] simplifies the expression of the decoding matrix **D**. Indeed, by choosing the appropriate normalized encoding convention (either (4) for full-3D or (8) for horizontal-only reproduction), one easily shows [12, 15] that:

$$\mathrm{D} = \frac{1}{N}\mathrm{C}^T \qquad (13)$$

Outside the reconstructed domain (HF/off-center)

The reconstruction over a given listening area is achieved only up a frequency that depends on the area size. Above this frequency, other decoding criteria and solutions (called "max r_E" and "in-phase") are preferably applied to optimize the perceived spatial rendering [12, 15]. Such decoding optimization is simply done by applying gains g_m on the appropriate frequency-bands to the components B_{mn}^{σ} before processing the "basic" decoding:

$$\mathrm{D} = \frac{1}{N}\mathrm{C}^T.\mathrm{Diag}\left(\begin{bmatrix}\cdots & g_m & \cdots\end{bmatrix}\right) \qquad (14)$$

Generic formulae for g_m are fully defined in [12].

Equivalent panning functions and directivities

By combining encoding equation (6) and decoding matrix (14), one derives an equivalent panning function G(γ) [12, 15]:

$$G(\gamma) = \frac{1}{N}\left(g_0 + 2\sum_{m=1}^{M} g_m \cos(m\gamma)\right) \quad \text{for 2D,} \qquad (15)$$

$$G(\gamma) = \frac{1}{N}\sum_{m=0}^{M}(2m+1)g_m P_m(\cos\gamma) \quad \text{for 3D,} \qquad (16)$$

such that loudspeaker i is fed with $S_i = S.G(\gamma)$, with $\gamma = \arccos(\vec{u}_i.\vec{u}_S)$ being the angle between the speaker and source directions $\vec{u}_i$ and $\vec{u}_S$. Figure 11 shows that higher orders help using loudspeakers with a finest angular selectivity for sound imaging, thus benefiting of their angular density.

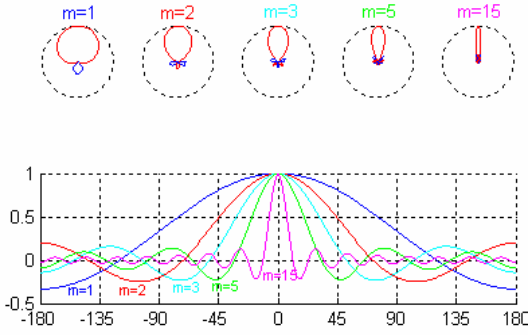

Figure 11 Equivalent directivities and panning laws associated to basic 2D decoding with various orders *m* (normalized regarding their max values).

This will be a helpful tool for interpreting rendering properties (section 4.2). Another helpful interpretation can be derived in terms of equivalent

recording setup: the polar diagram of a given order m (top of Figure 11) describes the directivity pattern of coincident microphones pointing to the loudspeakers that they would respectively feed.

3.3. Recent progress: supporting near field

Previous literature only rarely addressed the modelling of spherical waves, radiated by finite distance sources [12]. Nevertheless, correct encoding and reconstruction of realistic sound fields require it, and couldn't satisfy themselves with the usual plane wave approximation.

Spherical wave encoding (finite distance sources):

From the decomposition of a spherical wave given in [16], one derive [12] the formulae (17) (18) describing source encoding at a finite distance ρ:

$$B_{mn}^{\sigma} = S.F_m^{(\rho/c)}(\omega)Y_{mn}^{\sigma}(\theta,\delta) \qquad (17)$$

$$F_m^{(\rho/c)}(\omega) = \sum_{n=0}^{m} \frac{(m+n)!}{(m-n)!n!}\left(\frac{-jc}{\omega\rho}\right)^n \text{, with } \omega=2\pi f \qquad (18)$$

Note that equation (17) involves the pressure field S captured at O, assuming that $1/\rho$ attenuation and delay ρ/c due to finite distance propagation are already modeled.

Filters shown in (18) are typically "integrating filters", which are unstable by nature (infinite bass-boost, see Figure 12). This means also that the currently adopted HOA encoding format is unable to physically represent (*i.e.* by finite amplitude signals) natural or realistic sound fields, since these always include more or less near field sources.

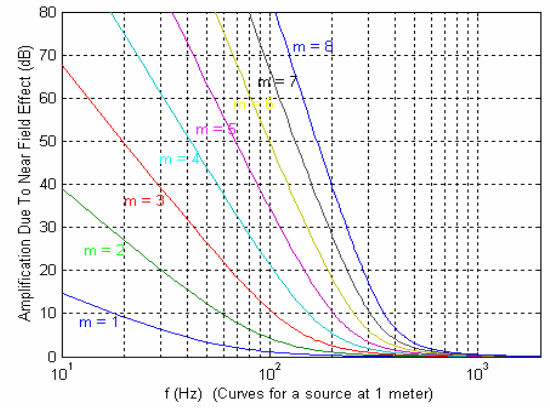

Figure 12 Low frequency infinite boost (m×6 dB/octave) of ambisonic components due to near field effect

Corrected decoding for a proper reconstruction

In order to truly satisfy the "re-encoding principle" and recompose the encoded sound field, the near field effect of the loudspeakers has to be considered, *i.e.* compensated, as already illustrated in [12]. The corrected decoding operation is thus:

$$S = D.\mathrm{Diag}\left(\left[\cdots \quad \frac{1}{F_m^{(R/c)}(\omega)} \quad \cdots\right]\right).B \qquad (19)$$

This decoding is practicable since inverse filters $1/F_m^{(R/c)}(\omega)$ are stable, and actually manages to preserve wave fronts original shape[5] (Figure 17).

Distance coding filters

The solution for practicable, finite distance source encoding, is to introduce the near field compensation (19) from the encoding stage and no longer at the decoding. Its combination with the virtual source near field effect (17) leads to the definition of stable "Distance (or Near Field) Coding filters"[6]:

$$H_m^{\mathrm{NFC}(\rho/c,R/c)}(\omega) = \frac{F_m^{(\rho/c)}(\omega)}{F_m^{(R/c)}(\omega)} \qquad (20)$$

They are characterized by a finite, low frequency amplification $m \times 20\log_{10}(R/\rho)$ (in dB), which is positive for enclosed sources ($\rho<R$) and negative for outside sources ($\rho>R$), as shown Figure 13.

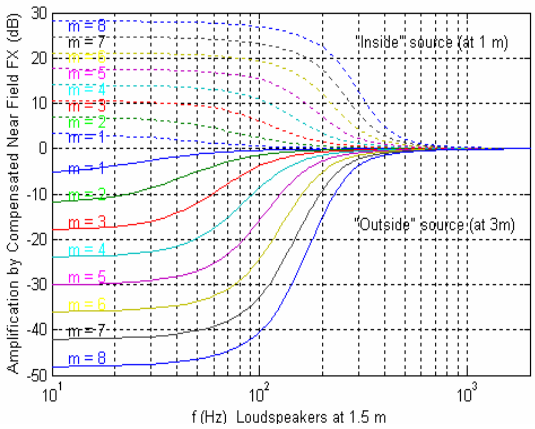

Figure 13 Finite amplification of ambisonic components from pre-compensated Near Field Effect (dashed lines: ρ/R=2/3; cont. lines: ρ/R=2).

A viable, new ambisonic format

Thus filters (20) advantageously replace filters (18) in encoding equation (17). At the same time, we have to consider a new encoding format called "Near Field Compensated Higher Order Ambisonics" format (NFC-HOA), and defined by the relation:

[5] Without near field compensation, an encoded plane wave is reconstructed as a spherical one coming from the loudspeaker array.

[6] Efficient, parametric digital filters (for practical use) are described in [17].

$$\breve{B}_{mn}^{\sigma\ \mathrm{NFC}(R/c)} = \frac{1}{F_m^{(R/c)}(\omega)} B_{mn}^{\sigma} \qquad (21)$$

It can represent any realistic sound field by finite amplitude signals and only requires the decoding operation (13), while implying a reference parameter: the reproduction loudspeaker distance R. Nevertheless, adaptation to any other array radius R' is possible by applying filters defined in (20) (replace ρ by R and R by R') before decoding. Finally, finite distance source encoding equation (17) becomes:

$$\breve{B}_{mn}^{\sigma\ \mathrm{NFC}(R/c)} = S.H_m^{\mathrm{NFC}(\rho/\mathrm{c,R/c})}(\omega) Y_{mn}^{\sigma}(\theta,\delta) \qquad (22)$$

This direct and rational way of encoding distance is an advantageous alternative to the WFS+HOA coupling scheme suggested in [18].

Equivalent pickup directivity or panning law

The same way as in 3.2, an equivalent panning law or pickup directivity can be derived, from replacing or multiplying gains g_m in (15) (for 2D case) by the frequency dependent complex gains (20):

$$G^{\mathrm{NFC}(R,c)}(\rho,\gamma,\omega) = \frac{1}{N}\left(g_0 + 2\sum_{m=1}^{M} g_m.H_m^{NFC(\rho/c,R/c)}(\omega).\cos(m\gamma) \right) \qquad (23)$$

Figure 14 shows the case of far virtual source (plane wave) at different frequencies and for g_m=1 (basic decoding).

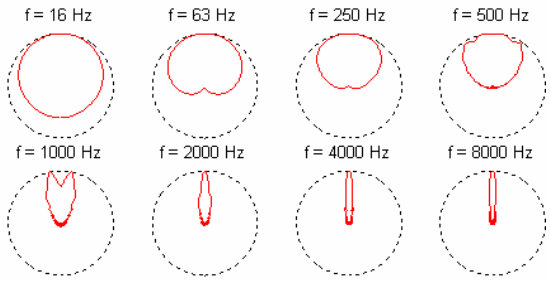

Figure 14 Equivalent pickup directivities or panning laws (normalized absolute values) for ρ=∞ (plane wave) and R_{spk}=1.5m, and with "basic" (g_m=1), 2D, 15th order rendering with Near Field pre-Compensation.

It's worth noticing that high frequency equivalent directivity tends to be the same, thus as high, as without near field coding (Figure 11, m=15), whereas a lower directivity (down to cardio or even omni) is observed at low frequencies.

3.4. Natural sound field recording

Up to this section, only virtual source encoding has been addressed. But natural sound field recording is also a possible and important feature of Higher Order Ambisonics, although up to very recently, Ambisonics recording possibilities have been restricted to the 1st order "Soundfield" microphone [19]. Indeed, theoretical studies regarding higher orders addressed recording systems based horizontal circular microphone arrays [5, 14] and more recently spherical arrays [12, 20-22].

The reader interested in further issues of ambisonic recording systems will find a full study in [23]. The following, lighter description aims at bringing out basic issues of practical systems, which are *spatial aliasing* and *noise amplification*.

Basic principle

The basic idea is to process a discrete spherical Fourier Transform of the sound field, based on its spherical sampling. For this purpose, one considers an array of N microphones distributed over a sphere (or an horizontal circle, in more restricted systems) of radius R_{mic} and centre O, and positioned and oriented according to the directions $\vec{u}_i$. From these we can process a directional sampling of the spherical harmonics: $\mathbf{y}_{mn}^{\sigma} = [Y_{mn}^{\sigma}(\vec{u}_1) \quad Y_{mn}^{\sigma}(\vec{u}_2) \quad \cdots \quad Y_{mn}^{\sigma}(\vec{u}_N)]$.

For the study needs, let's consider the simple case of cardio-like directivity: $\mathrm{G}(\theta) = \alpha + (1-\alpha)\cos(\theta)$ (far field directivity). It combines the pressure value p as expressed in (5) and its radial derivative with respective weighting factors α and (1-α). Therefore the field captured by the microphone i is:

$$p_R(\vec{u}_i) = \sum_{m=0}^{\infty} W_m(\omega) \sum_{0\le n\le m,\sigma=\pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\vec{u}_i) \qquad (24)$$

with the weighting factor:

$$\mathrm{W}_m(\omega) = j^m \left(\alpha\, j_m(kR_{\mathrm{mic}}) - j(1-\alpha)\, j_m'(kR_{\mathrm{mic}}) \right) \qquad (25)$$

Provided that the array geometry verifies some regularity conditions[7], one can estimate ambisonic components by projecting the spatially sampled sound field $\mathrm{p} = [p(\vec{u}_1) \quad \cdots \quad p(\vec{u}_N)]^T$ onto each sampled spherical harmonic y_{mn}^{σ}:

$$\hat{B}_{mn}^{\sigma} = \mathrm{EQ}_m(\omega) \left\langle \mathrm{p} \middle| \mathrm{y}_{mn}^{\sigma} \right\rangle_N , \qquad (26)$$

where the following equalization filters are also applied:

$$\mathrm{EQ}_m(\omega) = \frac{1}{\mathrm{W}_m(\omega)} \qquad (27)$$

[7] The underlying condition is that spatial sampling preserves spherical harmonic base orthonormality, *i.e.* $\left\langle \mathrm{y}_{mn}^{\sigma} \middle| \mathrm{y}_{m'n'}^{\sigma'} \right\rangle_N = \frac{1}{N} \mathrm{y}_{mn}^{\sigma}.\mathrm{y}_{m'n'}^{\sigma'\ T} = \delta_{mm'}\delta_{nn'}\delta_{\sigma\sigma'}$ (for $m,m' \le M$).

Spherical harmonic spectrum aliasing / spatial aliasing

According to [12], components B_{mn}^{σ} ($m \le M$) are estimated with the residual error:

$$\varepsilon_{mn}^{\sigma} = B_{mn}^{\sigma} - \hat{B}_{mn}^{\sigma} = \sum_{m'>M} \mathrm{EQ}_{m'}(\omega)\mathrm{W}_{m'}(\omega) \sum_{0\le n'\le m',\sigma'=\pm 1} B_{m'n'}^{\sigma'} \left\langle \mathrm{y}_{mn}^{\sigma} \middle| \mathrm{y}_{m'n'}^{\sigma'} \right\rangle_N \quad (28)$$

which exhibits the projection of higher order, insufficiently sampled components $B_{m'n'}^{\sigma'}$, onto the estimated one $\hat{B}_{mn}^{\sigma}$. This is an aliasing effect on the estimated spherical harmonic spectrum.

The "projection factor" $\left\langle \mathrm{y}_{mn}^{\sigma} \middle| \mathrm{y}_{m'n'}^{\sigma'} \right\rangle_N$ typically decreases when increasing N, thus the sensor angular density (spatial "oversampling"). The other weighting factor $\mathrm{EQ}_m(\omega)\mathrm{W}_{m'}(\omega) = \mathrm{W}_{m'}(\omega)/\mathrm{W}_m(\omega)$ is an increasing function of the frequency, and it also globally increases with the array radius R_{mic}. It appears finally that the frequency, above which spherical harmonic spectrum aliasing (28) becomes significant, decreases when the distance between acoustic sensors increases. It clearly has to be related to the "spatial aliasing frequency" (2), as introduced with WFS in 2.4! This was also pointed out in [24].

Applying near field pre-compensation (for practicable systems)

As shown in section 3.3, one has to introduce a near field pre-compensation at the encoding stage, in order to be able to represent any sound field with finite amplitude components. Therefore, instead of (27), the required equalization filters become:

$$\mathrm{EQ}_{\mathrm{m}}^{\mathrm{NFC}(R_{\mathrm{mic}}/c, R_{\mathrm{spk}}/c)}(\omega) = \frac{\mathrm{EQ}_m(\omega)}{F_m^{(R_{\mathrm{spk}}/c)}(\omega)} = \frac{1}{F_m^{(R_{\mathrm{spk}}/c)}(\omega) W_m(\omega)} \quad (29)$$

Unlike filters (27), these are now stable (finite low frequency amplification as shown Figure 15) and produce signals $\breve{B}_{mn}^{\sigma\ \mathrm{NFC}(R_{\mathrm{spk}}/c)}$ that are compliant with the "NFC-HOA" format (21). The reference distance R_{spk} is preferably chosen as the radius of a typical loudspeaker array. Figure 15 shows the case of a "reproduction distance" R_{spk}=1m much larger than the microphone radius R_{mic}=5cm.

Now it's possible to discuss to another important issue of practicable ambisonic microphones, which is the noise and error amplification problem.

Noise and error amplification

Electric signals derived from real life acoustic sensors always include noise. It's important to know what this noise becomes when computing ambisonic components and then, when decoding them and rendering the sound field. For this purpose, let's consider the signals **p** as pure, uncorrelated noise signals of equal energy $|p|^2$. One easily finds that the resulting noisy component has energy[7]:

$$\left|\breve{B}_{mn}^{\sigma\ \mathrm{NFC}(R_{\mathrm{spk}}/c)}\right|^2 = \left|\mathrm{EQ}_m^{\ \mathrm{NFC}(R_{\mathrm{spk}}/c)}(\omega)\right|^2 \frac{1}{N^2} \mathrm{y}_{mn}^{\sigma} \cdot \mathrm{y}_{mn}^{\sigma\ T} \left|p\right|^2 = \frac{1}{N}\left|\mathrm{EQ}_m^{\ \mathrm{NFC}(R_{\mathrm{spk}}/c)}(\omega)\right|^2 \left|p\right|^2 \quad (30)$$

Thus noise amplification fits equalization law (29) (Figure 15) lowered by -10.log$_{10}$(N) dB (*e.g.* –15dB for N=32). The quite high amplification (especially for low frequencies and high orders) reveals an important "effort" for extrapolating the sound field knowledge from a little radius R_{mic} to a much larger one R_{spk}.

Figure 15 shows that very low frequency amplification is about "one order lower" with ideal cardio sensors than with pressure sensors over rigid sphere. Indeed, cardio sensors already include first order directivity, but in real life they tend to be omni and/or noisy at low frequency!

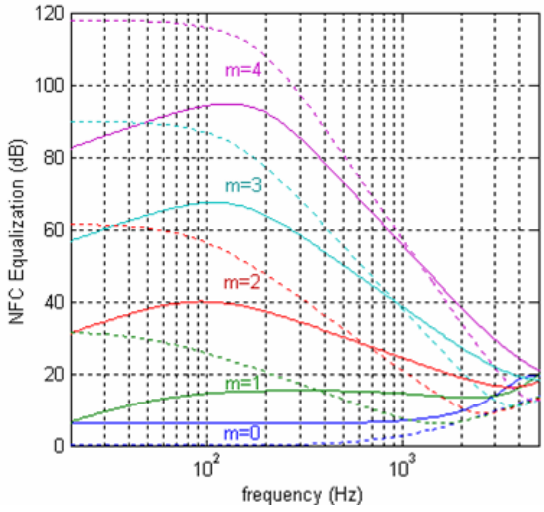

Figure 15 Near Field Compensated Equalization (29) involved in sphere microphone processing (R_{mic}=5cm, R_{spk}=1m). Cont. lines: ideal cardioid sensors; dotted lines: pressure sensors over a rigid sphere.

It should be added that even the effective captured signals don't necessarily exactly fit the theoretical modeling (24): that may be because of acoustic disturbance from the mechanical structure or because of bad sensor calibration, etc. This kind of error is also amplified during the processing.

It will be later discussed what this noise becomes after decoding and regarding the reconstructed sound field.

3.5. Applied State of Art

After mostly theoretical studies on High Order Ambisonics, their promising potentialities are becoming reality. Indeed, some essential features have been made practicable by solving the near field problem (as mentioned in sections 3.3 and 3.4). Related work done at the France Telecom R&D Labs is briefly listed below for illustration.

DSP tools (software generic implementation without limitation on system order)

- Encoding tools available are now: directional encoding functions and distance coding filters.
- Decoding tools include matrix and shelf-filters design for loudspeaker presentation. Design and processing of "Ambisonics to Ears" Transfer Functions, are also concerned, for binaural rendering (over headphones).
- Sound field transformations: addressing rotation matrix design (Figure 6 shows basic rotation angles) and focalisation.

Practical embodiments and experimentations

- 2D holophonic configurations are used (48 speaker, circular or dodecagonal arrays), also for comparison with WFS (Figure 5).
- A 4th order 3D microphone (based on 32 capsules placed over a rigid sphere) with associated DSP is being experimented [23].
- A full (or nearly-full) 3D ambisonics configuration (4th order, 32 or 21 loudspeakers) is in project: to be built in our anechoic room.

3D audio multi-channel format:

Original (1st order) "B-format" introduced by Gerzon has been recently extended to a 2nd order encoding format FMH ("Furse-Malham Harmonics": http://www2.york.ac.uk/inst/mustech/3d_audio/secondor.html), used by *e.g.* some music composers. There has been also a first attempt by Richard Dobson at handling them as sound files using an extension of the WAV format (WAVE-EX).
Since a viable, new ambisonic format is now mathematically defined (21), specifications are being discussed [17] for handling it as a multi-channel WAV-EX file, and also as a compressed multi-channel AAC stream in MPEG-4 (output document w5386 from the Awaji Meeting, December 2002).

4. COMPARING WFS AND HOA: FROM CONNECTIONS TO COMPROMISES

Up to this point, each approach has been described separately. It is now intended to further investigate them side-by-side, with the hope of offering the reader an "intuitive feeling" of the underlying physics, while making the views converge. First, a formal connection between their intrinsic representations is completed, and a list of reconstruction artefacts is drawn from examining typical departures from theoretical conditions. Then, some major artefacts are physically interpreted and characterized with the help of visualizations of simulated sound fields. Finally, recommendations and compromises are highlighted regarding virtual sound imaging and natural recording strategies, and also the encoding format.

4.1. Formal connection - Consequences of departures from theory

For a long time, WFS and HOA have been considered as two different, and even opposite, ways of sound spatialization. Nevertheless, it has been recently pointed out that they are closely connected approaches of 3D audio recording and reproduction [3, 5, 12]. Indeed, several analogies can be mentioned. Both WFS and HOA are based on sound recording and reproduction by resp. microphone and loudspeaker arrays. These respectively perform an acoustic encoding and decoding of the spatial sound information, which aim at a physical reconstruction of the primary sound field. While being based on two different representations of the sound field, (Kirchhoff-Helmholtz Integral for WFS and spherical harmonic expansion for HOA), both provide exact solution to the sound wave equation, and, for this reason, are fully equivalent. Moreover, under given assumptions[8], it has been shown how the ambisonic encoding and decoding equations may be derived from the Kirchhoff-Helmholtz Integral [3, 5, 12].
In this section, the connection between both intrinsic representations is completed and further discussed, regarding finite radius boundary (top of Figure 3). Then departures from theoretical conditions due to practical constraints are considered in terms of reconstruction error. While classifying expected resulting artefacts, difference and convergence are highlighted regarding extreme and intermediate forms of both approach implementations, especially regarding the microphone array radius R_{mic}.

Completing the connection between sound field intrinsic representations

To complete the convergence of views, both descriptions have to be confronted in identical

[8] These assumptions are: plane wave (infinite distance boundary and secondary sources), in addition to continuous sources distribution.

conditions/situation of recording and reproduction, that means: with the same transducer arrays. At least and even for ambisonics, we have to examine the case of a microphone array having a non-negligible radius R_{mic}, and especially the case $R_{mic}=R_{spk}$. We'll show how these recording considerations address the issue of sound field intrinsic representation.

Let's place the microphone array in a free field sphere strata as shown in Figure 7: we chose $R_1<R_{mic}<R_2$. It's worth highlighting that both intrinsic representations (derived from spherical harmonic decomposition and Kirchhoff-Helmholtz Integral) are able to distinguish between inside and outside sources, *i.e.* between outgoing and through-going field. This appears directly regarding respectively A_{mn}^{σ} and B_{mn}^{σ} components for HOA (3), and indirectly regarding the pressure values $p(\vec{R})$ and normal velocity values $v_n(\vec{R}) = \frac{1}{j\omega R}\vec{\nabla}p_0.\vec{n} = \frac{1}{j\omega R}\frac{\partial p}{\partial r}(\vec{R})$ for WFS (1). A more explicit connection derives from applying spherical Fourier Transform onto the pressure and velocity boundary distributions:

$$\begin{aligned} P_{mn}^{\sigma} &= \frac{1}{4\pi}\oiint_{|\vec{u}|=1} p(R.\vec{u}).Y_{mn}^{\sigma}(\vec{u})d\Omega \\ &= j^m j_m(kR)B_{mn}^{\sigma} + j^m h_m^-(kR)A_{mn}^{\sigma} \\ V_{mn}^{\sigma} &= \frac{1}{4\pi}\oiint_{|\vec{u}|=1} \frac{1}{j\omega R}\frac{\partial p}{\partial r}(R.\vec{u}).Y_{mn}^{\sigma}(\vec{u})d\Omega \\ &= \frac{j^{m-1}}{cR}\left(j_m'(kR)B_{mn}^{\sigma} + h_m^-{}'(kR)A_{mn}^{\sigma}\right) \end{aligned} \quad (31)$$

where we have used the series (3) and its radial derivative to express p and $\partial p/\partial r$, and also the orthonormality of spherical harmonics Y_{mn}^{σ}. Therefore "through-going" and "outgoing" field descriptors are:

$$\begin{aligned} B_{mn}^{\sigma} &= j^{-m}\frac{h_m^-{}'(kR)P_{mn}^{\sigma} - jcRh_m^-(kR)V_{mn}^{\sigma}}{j_m(kR)h_m^-{}'(kR) - j_m'(kR)h_m^-(kR)} \\ A_{mn}^{\sigma} &= j^{-m}\frac{j_m'(kR)h_m^-{}'(kR)P_{mn}^{\sigma} - jcRj_m(kR)V_{mn}^{\sigma}}{j_m'(kR)h_m^-{}'(kR) - j_m(kR)h_m^-{}'(kR)} \end{aligned} \quad (32)$$

Inversely, one can recompose the pressure and velocity fields at the boundary using the series (3) and its radial derivative. This is the transposition, in terms of 3D representation, of the relationship established by Huselbos[3] [11] for the horizontal case. Finally, this completes the connection previously stated for an infinite boundary radius R [3, 5, 12].

It's worth recalling and highlighting here that the exact representation intrinsically depends on R_{mic}: if we move R_{mic} to a distance R_{mic}' beyond one or several "outside" sources (Figure 7), these become "inside" sources and then the spherical harmonic representation changes in terms of A_{mn}^{σ} and B_{mn}^{σ}! On the opposite, there are only "outside" sources ($A_{mn}^{\sigma}=0$) from the centre point of view, *i.e.* for $R_{mic}=0$. From similar considerations, the extrapolation of the Kirchhoff-Helmholtz Integral to $R_{mic} \neq R_{spk}$, may become mathematically invalid.

Introducing practical limitations: Classification of expected artefacts

Reconstruction errors arise from departures from theory, when obeying constraints of practical system embodiment (limited number of loudspeakers, use of a single kind of transducer directivity, restriction to 2D), or even from fundamental limitations ("inside" sources). A qualitative and restricted comparison between WFS and HOA was already given in [25], in terms of artefacts and compromises related to the reproduction constraints. The following list rationally classifies expected artefacts by referring to previous sections. Points 1, 3 and 4 address encoding issues whereas artefact 2 arise from reproduction array restrictions.

1. *Restriction to single directivity microphone arrays (section 2.4) disables inside/outside dissociation and may cause encoding confusions.*

Indeed, equation (32) clearly shows that if the captured signals are just a combination of pressure and velocity, "inside" and "outside" descriptors A_{mn}^{σ} and B_{mn}^{σ} cannot be unambiguously derived. If we assume $A_{mn}^{\sigma}=0$ (free field enclosed area) even though there are enclosed sources, then these are rendered with an inverted wave front curvature (spatial mirroring with regard to the centre *C*), in addition to the time reversal effect (see also point 5 below) as explained in 2.7.

2. *Restriction to single directivity loudspeaker arrays (section 2.4) implies a reconstruction error along the array border.*

Figure 2 helps understanding that the combination of monopole-dipole contributions (from the array point *A*) is not the same from a central viewpoint *C*, as from a viewpoint *B* along the border of the listening area, *i.e.* closer to the array. The single directivity approximation is only acceptable for the centre C, and is no longer valid along the border, which implies that reconstruction may be affected.

3. *Spatial aliasing arises from the microphone spacing.*

This artefact has been first introduced in 2.4 (with WFS) as depending on the spatial sampling of secondary source array, and then further identified in 3.4 (with HOA) as occurring at the recording stage. Therefore, it is a sound field *encoding* issue, which depends on both the array radius R_{mic} and the number of transducers *N*. That's why the extreme, "ideal" encoding form (22) of HOA (virtual source with

R_{mic}=0) doesn't suffer from spatial aliasing, as illustrated in the next section.

4. *Vertical/horizontal aliasing occurs when using circular arrays for recording.*

This is another kind of spatial sampling artefact. Even if a horizontal restricted reproduction is targeted, only a spherical microphone array allows discarding vertical sound field components from the horizontal sound field representation and prevents these unwanted components from spoiling the audio rendering as completely unrelated acoustic phenomena. This could be proved by extending the aliasing error computation (28) of section 3.4.

5. *Enclosed sources: exact sound field reproduction over the enclosed area is physically impossible.*

Nevertheless, both approaches can even do something for reproducing sources inside the reproduction area: WFS' trick consists in temporally reversing the wave front propagation (section 2.7), whereas HOA may just extrapolate the wave front description, as seen from the centre (A_{mn}^{σ} =0), up to the source distance.

4.2. Characterizing and interpreting artefacts

This section illustrates rendering properties and artefacts (as listed in 4.1) of each system, through visualizations of simulated sound field reconstruction (restricted to the horizontal plane for convenience). This will help understanding and interpreting them physically. This also leads to characterize the rendering in terms of listening area wideness, or in terms of plausibly perceived effect or annoyance.

Note that the two systems are shown in their respective basic and extreme forms, *i.e.* considering virtually R_{mic} =0 for HOA and R_{mic} = R_{spk} for WFS. Sound imaging relies respectively on "virtual source" or "notional source" encoding. For the latter, WFS involve cardioid microphones that point outwards for outside sources and inwards for inside sources. For reproduction, loudspeakers are supposed to be omnidirective (no dipole "secondary source"). They are placed at a distance R_{spk}=1.5m from the centre. We consider of course a limited number (N=32), thus a limited ambisonic order (M=15).

In all figures, instantaneous pressure amplitude in represented in grey scale. Regarding the case of monochromatic fields (which show one frequency at once), this is also the real part of the complex pressure value in the frequency domain. The length of red wide arrows represents the signal amplitude of associated loudspeakers. Reconstruction error *err* is then computed (for each position) as the absolute, normalized difference between the synthesized field p_{syn} and the reference (original) field p_{ref}:

$$err = abs\left(\frac{p_{\text{ref}} - p_{\text{syn}}}{p_{\text{ref}}}\right) \tag{33}$$

To better comment the artefacts, three listening positions are symbolized with small heads and are referred to in the following as: *position C* (at the Centre), *position U* (in the Upper half of the disk, and also "Upstream" regarding the wave propagation) and *position D* (in the lower half-disk, and also "Downstream").

Low frequency: border error

According to the discussion of section 3.1 and considering the restriction to a given order M, the radial expansion of the achievable sound field approximation is proportional to the wavelength. This is also true for the reproduction with infinite distance loudspeakers. For example, a 15th order approximation of a 200Hz plane wave would be achieved over more than a 4 m radius area. Nevertheless, reconstruction shown Figure 16 doesn't even reach the area boundary (R_{spk}=1.5m), with both HOA and WFS. It has been moreover verified that increasing the order M (and N) doesn't improve it. This illustrates the "border error" expected in point 2 of section 4.1 and explained by the restriction to a single directivity transducer array (instead of using monopole + dipole pairs).

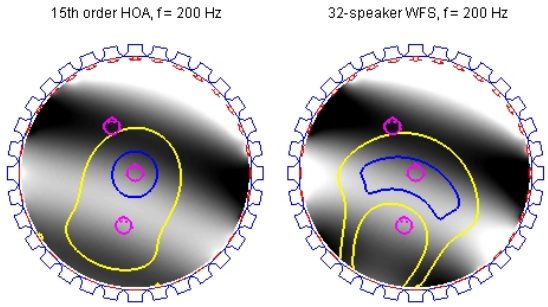

Figure 16 Reconstruction of a low frequency plane wave (f=200 Hz) with HOA and WFS (R_{spk}=1.5m). Blue/dark and yellow/bright contours enclose well-reconstructed areas with error tolerance of resp. 5% and 20%.

Nevertheless, this isn't a very damaging error (a low error tolerance is chosen for Figure 16) since it only causes a slight wave front shape distortion. Moreover, it concerns only the border where higher frequency artefacts are much more annoying, as it is illustrated below. By the way, to rightly discuss the reconstruction extent as a function of the frequency, one have to compare yellow/bright contours of Figure 16 with blue/dark ones in Figure 17 (20% error tolerance).

Higher frequency: spatial aliasing versus decreasing radial expansion

Besides the latter "border error" and according to 2.4, WFS is expected to provide a good reconstruction over the enclosed area up to the so-called "spatial aliasing frequency" f_{sp} (2). In the present case, its value is about 586Hz (let's say 600Hz). Indeed, the top of Figure 17 shows that the shape of a 600Hz plane wave is rather well preserved even outside error contours. It is noticeable that HOA provides a quite similar reconstruction quality. Moreover, in both cases the 20% error contour is about the same as for the 200Hz wave shown Figure 16.

WFS and HOA begin to distinguish from each other in terms of plane wave reproduction only above the spatial aliasing frequency, as shown Figure 17.

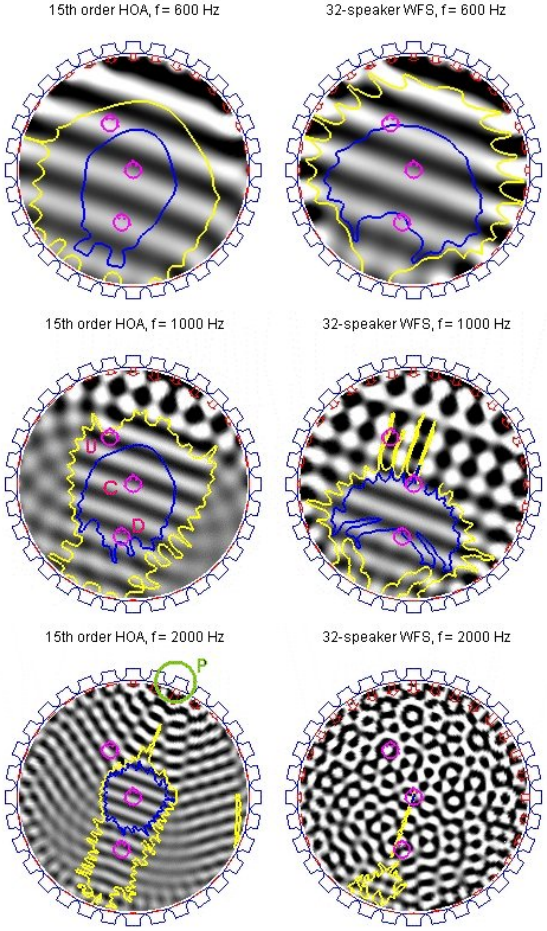

Figure 17 Reconstruction of monochromatic plane waves with HOA and WFS. Blue/dark and yellow/bright contours enclose well-reconstructed areas with error tolerance of resp. 20% and 50%.

HOA simulations clearly show that the reconstruction area progressively narrows around the centre position *C* when the frequency inversely increases, as expected from discussion of section 3.1. By the way, the centred listener *C* is especially favoured, since reconstruction will be perfect for him up to about 10kHz and for any sound incidence. But it's worth highlighting that at other listener positions, wave front shape remains quite consistent, even if their apparent origin progressively moves to a fixed point "*P*" on the loudspeaker array.

At the same time but with WFS, a strong interference effect rapidly spreads over the area as the frequency increases. Correct reconstruction is still observed over the quarter of area opposite to the virtual sound incidence for *f*=1000Hz, then not at all for *f*=2000Hz ("honeycomb" interference pattern).

To summarize: unlike with WFS, there's no spatial aliasing effect with HOA virtual imaging. A first explanation is that *spatial aliasing is related to the transducer spacing at the recording stage* (point 3 of section 4.1)… which is virtually null in the case of HOA virtual source encoding (R_{mic}=0).

Another explanation comes from interpreting HOA and WFS renderings in terms of e*quivalent panning functions or sound pickup directivity*, observed as a function of the frequency. Indeed one understands that out of exact reconstruction conditions, *wave interferences at a given position are the stronger and the more damaging, as significant contributions come from widely spread directions and therefore are contradictory*[9]. This is very well shown with WFS, which relies on a quite low (cardioid) directivity for all frequencies. The case *f*=1000Hz is especially instructive: sound field is highly disturbed at the listener position *U* (*U*pstream), which is surrounded by the most contributing loudspeakers; but there's no damaging interference effect at the "remote" listener position *D* (*D*ownstream), which "sees" the contributing loudspeakers as being less angularly spread. HOA has a fully different behaviour: Figure 14 means that loudspeaker contributions are used with a finest angular selectively around the virtual source direction as the frequency increases. As a consequence, only quite slight sound field disturbances appear off-centre and at relatively high frequencies. In the end (high frequency tendency), panning law fits the "old style" rendering (*i.e.* without Near Field Control), and the sound image tends to be "projected" over the loudspeaker array[5] (point "*P*", bottom-left of Figure 17). At intermediary frequencies (*e.g.* 1000Hz),

[9] This angular spread could be concisely characterised by the so-called "*energy vector*", this being computed locally (*i.e.* for given position and frequency). Its modulus varies from 1 (single contribution) to 0 (fully contradictory contributions).

interference patterns similar to WFS' ones appear at a distance and upstream from the centre, because of the less angular selectivity.

Consequences of artefacts in terms of audible effects

With HOA: localisation cues (especially ITD, *i.e.* Interaural Time Difference, and ILD, *i.e.* Interaural Level Difference) remain quite consistent along all the frequencies, though being progressively distorted for off-axis positions. Future listening experiments should precise the actual subjective effect.

With WFS: localisation relies essentially on cues up to the spatial frequency (thus mainly on the low frequency ITD) or a little higher, depending on where the listener is placed. Spatial information is objectively poor at higher frequency. Moreover, interference effects due to spatial aliasing are perceived as coloration effects (according to experiments done at the TUD). A slight decorrelation of loudspeaker signals, or an additional room effect, can reduce this coloration.

"Inside" (enclosed) sources

Simulating sources inside the reproduction area is a very special case of acoustic field reconstruction. Indeed, a full, true reconstruction is physically impossible in this case. Nevertheless, the following illustrates that partial (with HOA) or time inverted (with WFS) reconstruction is achievable.

Figure 18 shows the case of an inside source at a distance ρ =1m from the centre.

It is first noticeable that the spherical wave front shape seems correctly synthesized by WFS and over the whole area, up to the spatial aliasing frequency (about 600Hz). For the same frequencies, HOA only reproduces the shape over a disk of radius ρ =1m, just excluding the virtual source. This may be linked to the intrinsic limitation of HOA representation (section 3.1), which validity is limited to a free field sphere ($A_{mn}{}^{\sigma}$=0).

A further viewing reveals that the sound field phase is inverted with WFS, *i.e.* that the synthesized wave propagates *towards* the virtual source (*time-reversing*). At the same time, HOA involves *a great energy to restore the proper direction* of propagation, especially at low frequencies (see loudspeaker feedings shown by red wide arrows, top-left of Figure 18; see also the NFC-amplification shown in Figure 13). This reconstruction effort causes strong interferences in the periphery beyond the virtual source ($\rho<r<R_{spk}$): the interference angular frequency is directly linked to the highest and most amplified spherical harmonic mode (M=15 periods per 2π).

Above the spatial aliasing frequency, spatial aliasing affects WFS reconstruction once again, although a centred area is still preserved at f=1000Hz. With HOA, the reconstruction effort is progressively relaxed and peripheral interferences tend to be reorganised as wave fronts coming from the array projection point "*P*", like for the plane wave case (bottom-left of Figure 17).

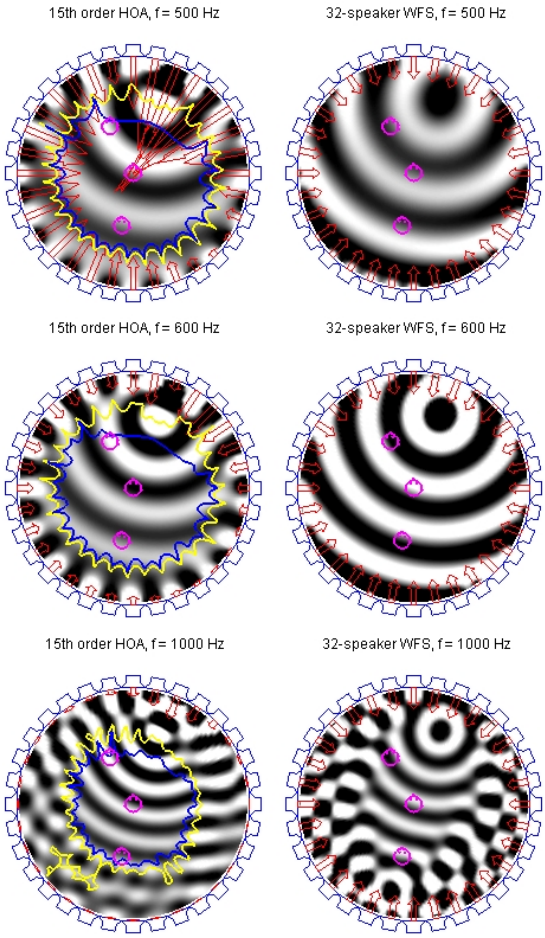

Figure 18 HOA and WFS rendering of an inside source (monochromatic spherical wave). Reconstructed spherical wave is propagating *from* the virtual source with HOA, but *towards* the same convergence point with WFS.

Consequences of artefacts in terms of audible effect

With WFS: where the time-reversed spherical wave is reconstructed, a correct ILD is expected since the spherical wave energy gradient is restored, but the ITD is inverted because of the reversed propagation. Thus these two primary localisation cues are contradictory. This leads to an amazing, but consistent effect, already noticed at the TUD.

With HOA: future listening experiments should teach us about the actual subjective effect in the presence

of the strong interference patterns, as seen beyond the source radius and for relatively low frequencies.

4.3. Recommendations and compromises

The previous section has begun pointing out some preferences and compromises on virtual sound imaging strategies. They are further discussed in the following, and then enlarged to natural sound field recording as well as encoding format issues.

Virtual sound imaging strategies: compromise for inside sources

Objective considerations would encourage using HOA as a preferred sound imaging strategies, at least for "outside" virtual sources. Indeed, HOA reconstruction quality is similar to WFS up to the spatial aliasing frequency f_{sp} (2), and above it, HOA is more robust and affected by less damaging artefacts.

A more intriguing question concerns *enclosed virtual sources*, especially *at low frequencies, below* f_{sp}. Here, there's a compromise to find between a reconstruction that preserves the propagation but is energy demanding and which radial extension is limited by the source (HOA), and wave shape preserving, but time reversing reconstruction over the whole area (WFS). Future subjective experiments are expected to bring some answers.

Real recording system: noise/error amplification versus spatial aliasing

Unlike mathematic encoding equations (4)(20)(22) for HOA virtual source imaging, natural sound field recording systems (section 3.4) must obey physical constraints. More particularly, estimating pressure field derivatives (*i.e.* ambisonic components) from capture points that are close to each other with respect to the wavelength, is all the harder, especially considering high order components. Therefore, a too small microphone array radius R_{mic} may imply a strong noise and error amplification, as stated in 3.4. On the opposite, a too large radius (*e.g.* $R_{mic}=R_{spk}$ in the extreme) causes spatial aliasing artefacts, as shown for WFS. One could envisage an intermediary radius ($0<R_{mic}<R_{spk}$): then the artefacts observed for WFS (Figure 17) would be rescaled down according to radius R_{mic} instead of R_{spk}. Nevertheless, if one considers spherical, instead of circular, microphone arrays (while keeping fixed number N and radius R_{mic}) in order to avoid vertical aliasing (point 4 of section 4.1), then microphone spacing increases, thus spatial aliasing frequency decreases.

Noise and error issues may also be examined regarding the reconstructed field. Figure 9 shows how ambisonic components, including their "measurement noise", participate to the sound field reconstruction as a function of kr. One notices for example that the highest order, and at the same time the noisiest components have a negligible presence at a small distance kr from the centre. From combining curves of Figure 9 and Figure 15, one could further state that the noise "recomposed" at a given listener position is directly linked to the "effort" for extrapolating the sound field knowledge from the radius R_{mic} to the listener distance $R_{listener}$. Therefore, the noise/error issues are less damageable for moderated sizes of listening area.

A last aspect is concerned with the choice of the radius R_{mic}: it is not desirable that the microphone array encloses real sources, since unlike with WFS "notional source encoding", it cannot naturally operate time reversing in order to avoid wave front curvature inversion (see 2.7, and points 1 and 5 of 4.1), and neither apply microphone inward pointing.

The compromises between noise/error amplification, spatial aliasing, and listening area won't be further and more quantitatively discussed here. At least we have highlighted how HOA and WFS approaches begin to share their originally own characteristic artefacts when dealing with practicable recording systems.

3D audio encoding format

Slightly different spatial encoding formats may derive from either HOA encoding equations or WFS-like "notional source encoding" scheme, or even from their coupling [18], which would consist of the components P_{mn}^{σ} of equation (31), but relying on a discrete spherical integration.

Until next discussions, the HOA encoding scheme (section 3.3) is preferred as being exact, efficient and scalable at once, and is further described in [17].

5. CONCLUSION

WFS and HOA approaches have been reviewed regarding their mathematic fundamentals *and* their practical application (*i.e.* usability and efficiency), on the basis of an updated state of art. Recent and relevant progresses regarding HOA have to be noticed. The first addresses near field modelling, which allows: preserving original wave front curvatures even when considering finite distance loudspeakers; deriving distance coding filters; and defining a viable "Near Field Compensated HOA" format. The second addresses feasible, higher order microphone systems.

A formal connection has been given between both intrinsic spatial sound field representations. In addition, it has been shown that when regarding practical (recording and reproduction) systems, both

approaches begin to share their own characteristic encoding and reconstruction artefacts, and especially the spatial aliasing.

Objective artefact characterisation relying on sound field simulations, has led to globally *prefer HOA as a more robust and efficient strategies for virtual sound imaging* (virtual source encoding). Indeed, it isn't affected by spatial aliasing. Not only reconstruction is achieved beyond the spatial aliasing frequency (though over a narrowing, centred area), but also off-centre distorted wave fronts (at high frequencies) keep consistent spatial information, unlike WFS "aliased" sound field.

Nevertheless *HOA and WFS meet similar limitations and compromises when dealing with real recording systems*. Indeed, both may suffer from noise/error amplification and/or spatial aliasing, which depend on the size of the microphone array, but in opposite ways. Therefore, the array size has to be defined regarding specific constraints and priorities, like the listening area extent. Finally, the established convergence of view may help refining the design of one technique by benefiting from the knowledge of the other.

Experiments are in preparation in the France Telecom R&D Labs and have to be conducted to subjectively characterise some of the discussed artefacts in terms of degree of annoyance. An interesting comparison is expected regarding the case of sources enclosed by the loudspeaker array. This may lead to further recommendations on the virtual sound imaging (WFS or HOA) in this case.

6. REFERENCES

[1] A.J. Berkhout, A Holographic Approach to Acoustic Control, J. Audio Eng. Soc., pp. 977-995, 1988.

[2] A.J. Berkhout, D.d. Vries, and P. Vogel, Acoustic Control by Wave Field Synthesis, J. Acoust. Soc. Am., vol. 93, pp. 2764-2778, 1993.

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

[4] E. Horbach, E. Corteel, and D.d. Vries, Spatial Audio Reproduction using Distributed Mode Loudspeaker Arrays, presented at the AES 21rst Int. Conference, St Petersburg, Russie, 2002.

[5] R. Nicol, Restitution Sonore Spatialisée sur une Zone Etendue : Application à la Téléprésence, Ph. D. Thesis, Université du Maine, Le Mans, France, 1999, http://gyronymo.free.fr/audio3D/Guests/RozennNicol_PhD.html.

[6] E. Verheijen, Sound Reproduction by Wave Field Synthesis, Ph. D. Thesis, Faculty of Applied Physics, Delft, The Netherlands, 1996.

[7] S. Yon, M. Tanter, and M. Fink, Sound Focusing in Rooms : The Spatio-Temporal Inverse Filter, J. Acoust. Soc. Am., 2002.

[8] S. Yon, M. Tanter, and M. Fink, Sound Focusing in Rooms: The Time Reversal Approach, J. Acoust. Soc. Am., 2002.

[9] M.A. Gerzon, Ambisonics in Multichannel Broadcasting and Video, J. Audio Eng. Soc., vol. 33(11), pp. 859-871, 1985 Nov.

[10] M.A. Gerzon, Periphony : With-Height Sound Reproduction, J. Audio Eng. Soc., vol. 21(1), pp. 2-10, 1973.

[11] E. Hulsebos, D.d. Vries, and E. Bourdillat, Improved Microphone Array Configurations for Auralization of Sound Fields by Wave Field Synthesis, preprint 5337 presented at the AES 110th Convention, Amsterdam, The Netherlands, 2001 May 12-15.

[12] J. Daniel, Représentation de Champs Acoustiques, Application à la Transmission et à la Reproduction de Scènes Sonores Complexes dans un Contexte Multimédia, Ph.D. Thesis, University of Paris 6, Paris, France, 2000, http://gyronymo.free.fr/audio3D/download_Thesis_PwPt.html.

[13] J.S. Bamford, An Analysis of Ambisonics Sound Systems of First and Second Order, M. Sc. Thesis, University of Waterloo, Waterloo, Ont., Canada, 1995.

[14] M.A. Poletti, A Unified Theory of Horizontal Holographic Sound Systems, J. Audio Eng. Soc., vol. 48(12), pp. 1155-1182, 2000 Dec.

[15] J. Daniel, J.-B. Rault, and J.-D. Polack, Ambisonic Encoding of Other Audio Formats for Multiple Listening Conditions, preprint 4795 presented at the AES 105th Convention, San-Francisco, USA, 1998 Sept.

[16] P.M. Morse and K.U. Ingard, Theoretical Acoustics, McGraw-Hill ed, 1968.

[17] J. Daniel, Spatial Sound Encoding Including Near Field Effect : Introducing Distance Coding Filters and a Viable, New Ambisonic Format, presented at the AES 23rd International Conference, 2003 23-25 May.

[18] A. Sontacchi and R. Höldrich, Further investigations on 3D sound fields using distance coding, presented at the DAFX-01, Limerick, Ireland, 2001 Dec. 6-8.

[19] P.G. Craven and M.A. Gerzon, Coincident Microphone Simulation Covering Three Dimensional Space and Yielding Various Directional Outputs, US Patent 4,042,779, filed July 7, 1975, issued Aug. 16, 1977.

[20] T.D. Abhayapala and D.B. Ward, Theory and Design of High Order Sound Field Microphones Using Spherical Microphone Array, presented at the IEEE ICASSP-02, Orlando, Florida, USA, 2002 May 13-17.

[21] J. Meyer and G. Elko, A Highly Scalable Spherical Microphone Array Based on an Orthonormal Decomposition of the Soundfield, presented at the IEEE ICASSP-02, Orlando, Florida, USA, 2002 May 13-17.

[22] P. Cotterell, On The Theory of the Second-Order Soundfield Microphone, Ph. D. thesis, University of Reading, UK, 2002, http://www.personal.rdg.ac.uk/~shr97psc/Thesis.html

[23] J. Daniel and S. Moreau, Theory and Design Refinement of High Order Ambisonic Microphones - Experiments with a 4th Order Prototype, presented at the AES 23rd International Conference, 2003 May 23-25.

[24] D. Malham, Higher order Ambisonic systems for the spatialisation of sound, presented at the ICMC99, Beijing, 1999 Oct.

[25] J. Daniel, Position Paper, presented at the ACM-SIGGRAPH and EUROGRAPHICS Campfire on Acoustic Rendering for Virtual Environments, Snowbird, Utah, USA, 2001 http://www.bell-labs.com/topic/conferences/campfire/abstracts/daniel.pdf.