Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format

Jérôme Daniel

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

jerome.daniel @ francetelecom.com

ABSTRACT

Higher Order Ambisonics have been increasingly investigated in the past years, and found promising as a rational, scalable and flexible way to encode, transmit and render 3D sound fields. Nevertheless, studies concerning virtual source imaging or natural 3D sound encoding mainly focussed on the directional encoding of plane waves, and neglected the near field effect of finite distance sources though it's present in any ordinary sound field.
This paper highlights that with near field, the infinite bass-boost affecting ambisonic components makes the currently accepted format unviable. By introducing from the encoding stage a near field compensation of reproduction loudspeakers, a viable, modified ambisonic format is defined, distance-coding filters are designed, and higher order ambisonic recording and synthesis become practicable.

1. INTRODUCTION

The tasks of sound spatialisation

Sound spatialisation aims at providing to the listener auditory sensations and information that are usually related to the sound propagation in environmental space. It addresses two complementary aspects. The first one is the environment acoustics ("room effect"), *i.e.* the way the waves radiated from sound sources are reflected and reverberated before reaching the listener: it provides information on the room size and the source distance, for example. The second aspect is the directional / spatial properties of such derived acoustic events (especially the first wave front and reflections): they allow the listener localising sound sources and feeling enveloped by the room effect. To reflect these aspects, a sound spatialisation system typically processes as follows.
First, from the description of a virtual sound scene (sources and environment), a virtual acoustics processor computes the signals and the positional properties associated to elementary events (first wave front and reflections), and also a signal description of macroscopic events (diffuse reverberated field). In a second step, these signals are spatially encoded, *i.e.* processed with regard to their directional or spatial properties (Figure 1). This leads to a multi-channel, 3D audio representation that can be conveyed then decoded for diffusion over loudspeakers or headphones. The present paper addresses the spatial encoding of virtual or even natural sound fields, on the basis of the ambisonic approach.

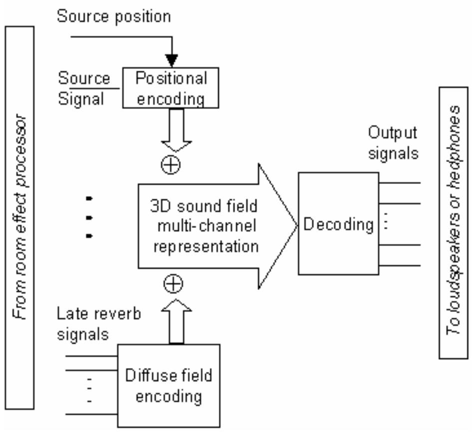

Figure 1 General spatial encoding scheme of elementary (wave fronts) or macroscopic (diffuse field) components provided by a room effect processor

Ambisonics among spatial encoding strategies

Ambisonics is a very *versatile* approach for the spatial encoding and rendering of sound fields. It has known an increasing interest during the past years thanks to studies that have extended the theory (and to a less extent, its application) from first to higher order, highlighting many advantages:

- A *rational* encoding of spatial acoustic information, and moreover independent from the reproduction layout.
- A *flexible* and *scalable* spatial sound representation: one can transform (*e.g.* rotate, see Figure 2) the sound field, and also adapt it to transmission constraints or reproduction capabilities by keeping only a subset of signals (*variable spatial resolution*).
- A *variable geometry* rendering: a decoder can be suitably designed according to the loudspeaker array geometry, and also for binaural rendering over headphones.
- A quite *optimal way to achieve "holophonic" sound field reconstruction* by means of a given number of loudspeakers, which makes Higher Order Ambisonics (HOA) comparable and even preferable to Wave Field Synthesis (WFS) in some conditions [1].

Nevertheless, application of HOA is not as spread as it deserves, yet. One reason is that practical recording systems are still restricted to 1st order microphones, while HOA being basically thought as an amplitude panning technique dedicated to virtual sound imaging. The present paper transcends this common conception of ambisonic approach and potentialities, by developing a key improvement [1] that enables ambisonics to handle realistic or natural sound fields.

From directional to positional encoding: near field effect as an essential distance feature

Ambisonic directional encoding and decoding basically assumes that virtual sources as well as reproduction loudspeakers are in far field and radiate plane waves. But in natural sound fields there are always more or less near field sources and the wave front curvature depends on their distance. This curvature allows the listener perceiving the source distance when he moves in the sound field, independently from the cues given by room effect. Even for a still listener, the near field effect of close sources is perceptible through the emphasis of ILD (Interaural Level Difference).

This paper first shows that the currently adopted HOA encoding format is unable to support near field, thus to represent natural sound fields with physically transmissible signals. Then a modified encoding scheme is introduced, which makes possible the synthesis and recording[1] of any realistic phenomenon. This leads to the detailed description of distance coding filters, the illustration of a complete positional coding and rendering scheme, and finally the specifications of a new, viable HOA format.

[1] The paper on HOA microphones announced in [1] cannot be given at the present conference.

2. SUPPORTING NEAR FIELD MODELLING WITH HIGHER ORDER AMBISONICS

2.1. Mathematical encoding formalism

Spherical harmonic decomposition

Ambisonic approach bases the sound field description on the spherical coordinate system (Figure 2). This way, it has the interesting property of providing an homogeneous description of directional information (azimuth θ and elevation δ), while separating it from the distance information (radius r).

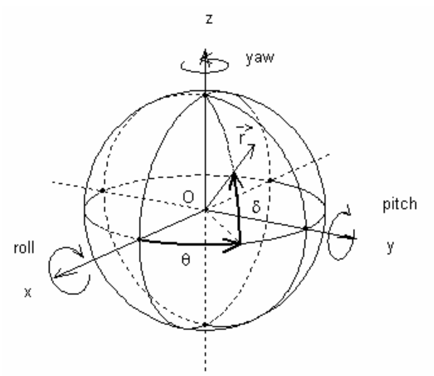

Figure 2 Spherical coordinate system, with the three elementary rotation degrees. A point $\vec{r}$ is described by radius *r*, azimuth θ and elevation δ.

The mathematical formalism comes from writing the wave equation $(\Delta+k^2)p=0$ (with the wave number $k=2\pi f/c$) in the spherical coordinate system. This leads to the Fourier-Bessel series [2]:

$$p(\vec{r}) = \sum_{m=0}^{\infty} j^m j_m(kr) \sum_{0\le n\le m, \sigma=\pm1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta,\delta), \qquad (1)$$

Each term of "order" m associates radial, spherical Bessel function $j_m(kr)$, with angular functions $Y_{mn}^{\sigma}(\theta,\delta)$ called "spherical harmonics" (Figure 3):

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

where $\delta_{q,q'}=1$ if $q=q'$ and 0 otherwise (Kronecker symbol). The P_{mn} define the associated Legendre functions of degree m and order n, and $\tilde{P}_{mn}$, their

"Schmidt semi-normalised" versions. The exponent tag $^{(N3D)}$ attached to functions $Y_{mn}{}^{\sigma}$ in (2) means that these are "3D-normalised" in the sense of a spherical scalar product [1, 3]. Other conventions with different weighting factors may also be used (see 3.1 and 4.4).

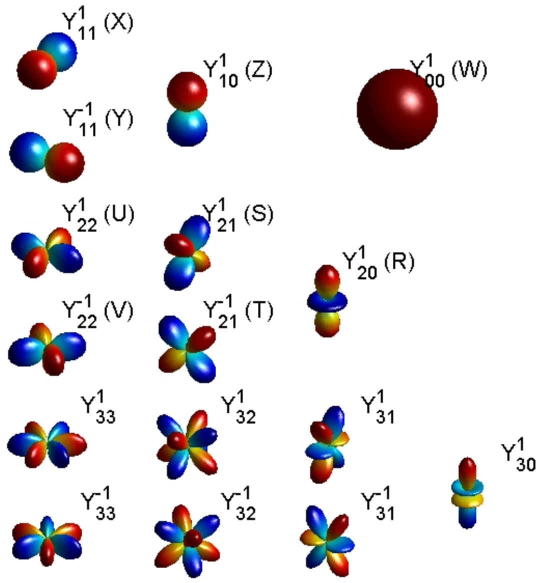

Figure 3 3D view (with respect to Figure 2) of spherical harmonics with usual designation of associated ambisonic components.

Ambisonic sound field representation

The spherical harmonic decomposition (1) exhibits frequency dependent coefficients $B_{mn}{}^{\sigma}$ that fully represent the sound field within a sphere centred on the origin *O*, provided that there is no acoustic source within this sphere. Physically, these components represent the pressure field $B_{00}{}^{+1}$ and its spatial derivatives or momentums of successive orders *m* at the reference point *O*. They also reflect the sound field propagation properties around this point [4].

Spherical harmonic factors $B_{mn}{}^{\sigma}$ are the frequency domain expression of what are called "ambisonic components". In practice, for spatial encoding, transmission and reproduction, one retains a limited set of components up to a given order *M*. Moreover, for most 2D (horizontal only) applications, this set may be restricted to "horizontal" components $B_{mn}{}^{\sigma}$ (n=m). The higher the order *M*, the larger the sound field approximation around the reference point *O* (considered as the "listener" viewpoint), as further explained in [1].

Plane wave decomposition: directional encoding

Virtual source encoding often assumes that the source is far enough, so that its contribution can be approximated by a plane wave. As shown *e.g.* in [3], the spherical harmonic decomposition of a plane wave of incidence (θ_S, δ_S) conveying a signal *S* leads to the simple expression of ambisonic components:

$$B^{\sigma}_{mn} = S.Y^{\sigma}_{mn}(\theta_S, \delta_S) \qquad (3)$$

Thus a *far field source* signal *S* is encoded by simply applying *real encoding gains*, which are the spherical harmonic functions. By the way, that means that the sound field "derivatives" properties don't vary with the frequency.
Computational details about these encoding gains are given in 3.1.

Spherical wave decomposition: near field effect

The modelling of the near field effect due to finite distance sources is rarely addressed in literature [3]. Nevertheless, it points out a fundamental issue of natural or realistic sound fields.
It is shown [2, 3] that the spherical decomposition of a spherical wave radiated by a point source at (ρ,θ,δ) leads to:

$$B^{\sigma}_{mn} = S.\Gamma_m(k\rho).Y^{\sigma}_{mn}(\theta,\delta)$$
$$\text{with} \quad \Gamma_m(k\rho) = kd_{ref}\, h^{-}_{m}(k\rho)\, j^{-(m+1)} \qquad (4)$$

where $h^{-}_{m}(kr) = j_m(kr) - jn_m(kr)$ are the divergent spherical Hankel functions, and d_{ref} is a reference distance. More conveniently, we'll consider *S* as the pressure field captured at *O*, so that the $1/\rho$ attenuation and the delay ρ/c due to finite distance propagation, which are reflected by $\Gamma_0(k\rho)$, are supposed to be already modelled. By removing the latter from (4), the encoding equations of a source at finite distance ρ become:

$$B^{\sigma}_{mn} = S.F^{(\rho/c)}_{m}(\omega) Y^{\sigma}_{mn}(\theta,\delta) \ , \qquad \omega = 2\pi f$$
$$F^{(\rho/c)}_{m}(\omega) = \frac{\Gamma_m(k\rho)}{\Gamma_0(k\rho)} = \sum_{n=0}^{m} \frac{(m+n)!}{(m-n)!n!}\left(\frac{-jc}{\omega\rho}\right)^{n} \qquad (5)$$

Such a finite distance encoding involves transfer functions $F_m(\omega)$ that affect ambisonic components especially at low frequencies, as shown by Figure 4. In other words and by comparison with the plane wave case of (3): *the near field disturbs the sound field "derivatives" as much as the source distance (i.e. the curvature radius) is small regarding the wavelength, and as the derivative order m is high.*

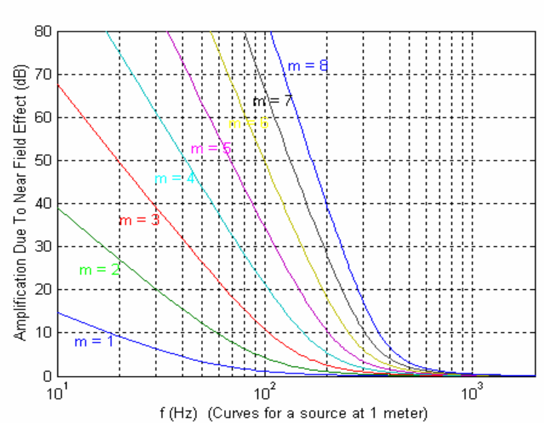

Figure 4 Low frequency infinite boost (m×6 dB/octave) of ambisonic components due to near field effect

Fundamental limitations of the encoding format

What is annoying is that transfer functions $F_m(\omega)$ typically reflect "integrating filters" (for $m \geq 1$), which are unstable by nature (infinite bass-boost shown in Figure 4). First order encoding may still remain practicable provided that every encoded signal S is centred (null mean value), but it is no longer the case for higher orders.

Not only (5) involves impracticable filters for virtual source encoding, but since it also models the physical reality, it would imply that the ambisonic representation of any natural sound field may have infinite amplitude components. This finally means that in spite of being mathematically powerful, the currently adopted HOA encoding format is unable to physically represent and convey (*i.e.* by finite amplitude signals) natural or realistic sound fields, since these always include more or less near field sources.

By addressing the decoding and reproduction issues, and introducing the loudspeaker near field modelling at this stage, the following section suggests a key to a viable encoding format.

2.2. Decoding: the need for near field compensation

Since the ambisonic components represent by themselves the sound field to be rendered, a basic goal of the decoder is to recompose or "re-encode" them at the centre of the loudspeaker array, which is the privileged listening position. As often, we'll consider concentric regular arrays in the following.

Previous decoding conception: amplitude panning

The most commonly shared conception of ambisonic decoding relies on the assumption that the loudspeakers are far field sources from the centre point of view. Therefore the decoder has to achieve sound field reconstruction by combination (interference) of presumed plane waves. This requires only combining the signals with real weighting gains, thus involving a matrix operation:

$$\mathbf{S} = \mathbf{D}.\mathbf{B}, \quad (6)$$

where $\mathrm{S} = [S_1 \dots S_N]^T$ is the vector of emitted signals, $\mathrm{B} = [B_{00}^{+1}\ B_{11}^{+1}\ B_{11}^{-1} \dots B_{mn}^{\sigma} \dots]^T$ is the vector of ambisonic components to be recomposed. The radiated signals S_i contribute to the ambisonic components re-composition according to:

$$\mathbf{B} = \mathbf{C}.\mathbf{S}, \quad (7)$$

where **C** is the so-called "re-encoding" matrix which elements are the encoding gains $Y_{mn}^{\sigma}(\theta_i, \delta_i)$ associated to the loudspeaker directions. As further detailed in [1, 3], the matrix **D** fulfil the decoding goal when being defined as the pseudo-inverse of **C**:

$$\mathbf{D} = pinv(\mathbf{C}) = \mathbf{C}^T.(\mathbf{C}.\mathbf{C}^T)^{-1}, \quad (8)$$

provided that there are at least as many loudspeakers as components to recompose.

Finally, since both encoding and decoding operations only process amplitude weightings, ambisonic sound imaging is globally a kind of amplitude pan-pot. An interesting property [1, 3, 5] is that higher orders help using loudspeakers with a finest angular selectivity around the virtual source direction, then reconstructing the sound field over a larger area, as shown by Figure 5.

For a higher frequency domain, where the reconstruction cannot be achieved at the listener scale, Gerzon [6] also introduced "psycho-acoustic" criteria in the 1st order decoding design. These have been later generalized to higher orders [3, 5]. Although this modified decoding is not detailed in the present study, it may be advantageously used in practical situation.

General comment on sound field illustrations (Figure 5 and followings): in all case, they show the ambisonic reconstruction of a single waves always coming from the same direction (but with various source distances, frequencies, system orders), by means of a 32 loudspeaker array. The instantaneous pressure field is represented in grey scale. In the case of monochromatic sound fields, blue/dark and yellow/bright contours enclose well-reconstructed areas with error tolerance of resp. 20% and 50%, and red arrows indicate the loudspeaker signal amplitudes.

Wave front curvature distortion and bass-boost effect

Figure 5 shows the case of an encoded plane wave. Its left parts, which report a traditional decoding as previously described, show that the synthetic wave has the expected propagation direction from the centred listener point of view. Nevertheless, it clearly appears that with a high (15^{th}) order rendering, this is not a plane wave that is reconstructed, but a spherical one, as being radiated by a point on the loudspeaker boundary. Therefore off-centred listeners localise the virtual source on this point and not in the direction of the original plane wave.

This wave curvature distortion seems to have little impact on the directional effect for a centred listener. Nevertheless, even for this position and depending on the actual array radius, the difference with a true plane may be audible as the so-called "bass-boost effect" already mentioned by Gerzon [7], and also as an emphasised Interaural Level Difference (ILD).

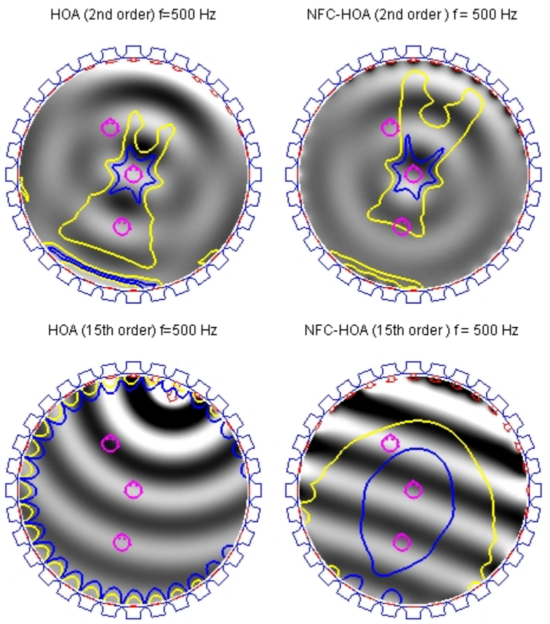

Figure 5 Reproduction of an encoded plane wave without (left) and with (right) loudspeaker near field compensation (NFC). 2^{nd} order (top) and 15^{th} order (bottom) ambisonics.

Compensating for the loudspeaker near field

In the context of earlier first ambisonic systems, Gerzon recommended to compensate for the bass-boost effect due to the finite distance of loudspeakers. Considering higher orders and with the more general aim to preserve the original curvature of the encoded wave fronts, it is now suggested to introduce the loudspeaker near field modelling into the re-encoding equation (7). That means that the elements $Y_{mn}^{\sigma}(\theta_i,\delta_i)$ of the "re-encoding" matrix **C** would have to be "multiplied" by the near field transfer functions $F_m^{(R/c)}(\omega)$ of same order. Finally, this leads to the following decoding operation [1, 3]:

$$S = D.\mathrm{Diag}\left(\left[\cdots \quad \frac{1}{F_m^{(R/c)}(\omega)} \quad \cdots\right]\right).B \qquad (9)$$

where the decoding matrix is the same as defined by (8). Thus, this new decoding consists in applying a near field compensation $1/F_m^{(R/c)}(\omega)$ to the ambisonic components B_{mn}^{σ} before decoding them classically. Unlike the near field modelling transfer functions $F_m^{(R/c)}(\omega)$, the filters $1/F_m^{(R/c)}(\omega)$ are practicable and stable.

As a result, the plane wave is actually reconstructed without curvature distortion, which is clearly illustrated for the 15th order by Figure 5 (right-bottom part).

Note that in a higher frequency domain, near field compensation is no longer effective (consider the inversed curves of Figure 4), and at the same time, the reconstruction area progressively narrows. It is still appropriate to use high frequency optimised decoding solutions mentioned above.

2.3. Distance coding, viable format: the key

Compensating for near field from the encoding stage

At this point, we have proved that for a proper sound field reconstruction, one has to compensate for the loudspeaker near field effect anyway. Why not introducing this near field compensation from the encoding stage? As a matter of fact, it rapidly appears that combining it to the near field modelling of the virtual source leads to apply finite amplitude transfer functions.

Distance Coding / Near Field Control Filters

The combination of near field effect (for a source distance ρ) and compensation (for a loudspeaker distance R) leads to the following transfer functions:

$$H_m^{\mathrm{NFC}(\rho/c,R/c)}(\omega) = \frac{F_m^{(\rho/c)}(\omega)}{F_m^{(R/c)}(\omega)} \qquad (10)$$

Figure 6 shows that they cause a finite, low frequency amplification $m\mathrm{x}20\log_{10}(R/\rho)$ (in dB), which is positive for enclosed sources ($\rho<R$) and negative for outside sources ($\rho>R$).

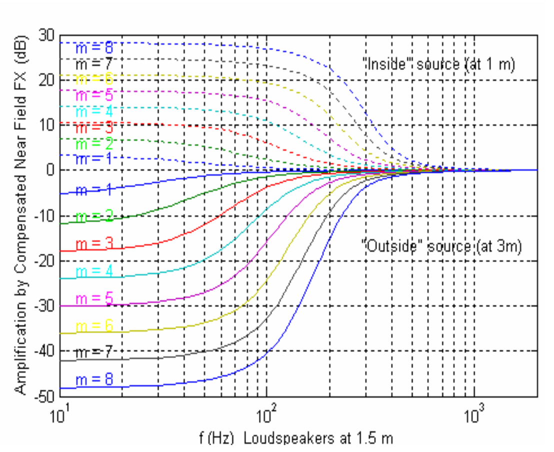

Figure 6 NFC filters frequency responses: finite amplification of ambisonic components from pre-compensated Near Field Effect (dashed lines: ρ/R=2/3; cont. lines: ρ/R=2).

They can be practically implemented as stable filters (as detailed in 3.2), which we will call "Near Field Coding" or "Control" filters, or simply "NFC filters". Now, encoding equations (5) are replaced by following the positional encoding equation:

$$\breve{B}_{mn}^{\sigma\ \mathrm{NFC}(R/c)} = S.H_m^{\mathrm{NFC}(\rho/c,R/c)}(\omega).Y_{mn}^{\sigma}(\theta,\delta) \qquad (11)$$

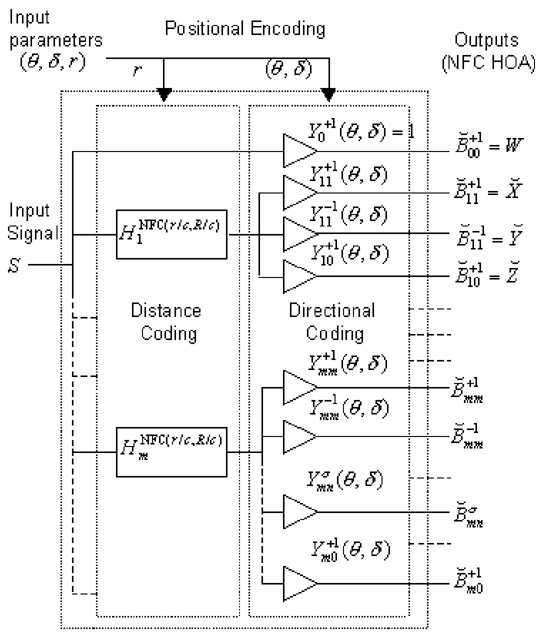

Figure 7 NFC-HOA positional encoding of a virtual sound source: a distance-coding unit (NFC filter bank) completes the directional encoding.

This new positional encoding scheme completes the earlier, purely directional one by introducing a distance-coding module (Figure 7). The latter consists of a NFC filter bank, which is preferably placed before the directional gain control in order to factorise the filtering of each group of same order components. It's worth recalling that with such an encoding scheme, the encoded sound field only requires an "ordinary" matrix decoding (6).

A viable, new ambisonic format

At the same time, it is noticeable that a new encoding format derives from the virtual source encoding scheme (11). It is more generally related to the previous higher order ambisonic (HOA) format by:

$$\breve{B}_{mn}^{\sigma\ \mathrm{NFC}(R/c)} = \frac{1}{F_m^{(R/c)}(\omega)} B_{mn}^{\sigma} \qquad (12)$$

It will be called NFC HOA, for "Near Field Compensated Higher Order Ambisonics". The advantage of this encoding format is not restricted to virtual sound encoding: it makes also possible the representation and the recording of any natural sound field. Indeed, it is shown that the equalisation filters involved in the signal processing of HOA microphone arrays become feasible when introducing the near field pre-compensation in them [1].

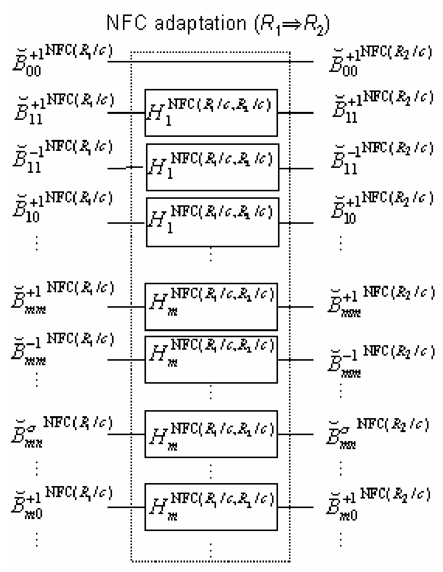

Figure 8 Adaptation of the near field compensation to a loudspeaker distance different from the reference one

Finally, the NFC-HOA format comprises a reference distance[2] R with an implicit parameter, which corresponds to the radius of the reproduction loudspeaker array. Nevertheless, this shall not be

[2] As a matter of fact, the implicit parameter is rather a reference delay $\tau=R/c$.

thought as a constraint. Indeed, the NFC filters may also be used for adapting one NFC-HOA representation to a different reference distance or loudspeaker radius (Figure 8):

$$\breve{B}_{mn}^{\sigma\ \mathrm{NFC}(R_2/c)} = H_m^{\mathrm{NFC}(R_1/c,R_2/c)}(\omega).\breve{B}_{mn}^{\sigma\ \mathrm{NFC}(R_1/c)} \qquad (13)$$

Note that this also applies to the earlier, "un-compensated" HOA format as a particular case for which the reference distance is R=∞.

Illustrated example of close source synthesis

Figure 9 illustrates the ability of simulating a finite distance source, and takes as an example the more critical case of a source inside the loudspeaker array (ρ=1m<R=1.5m). As expected from Figure 6, a quite important amount of energy is involved at low frequencies (long red arrows). One notices that the reconstructed area is constrained by the same validity conditions as the theoretical representation: it is a disk that excludes the virtual source (see section 2.1). A further discussion can be found in [1], and the later section 4.1 completes the rendering analysis with time domain considerations.

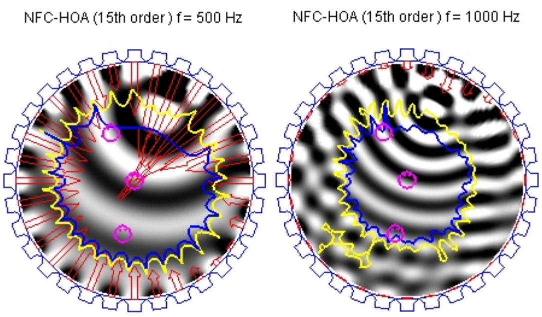

Figure 9 Spherical wave reconstruction (inside source) at 500Hz and 1kHz, with NFC-HOA.

3. DIRECTIONAL AND DISTANCE CODING: IMPLEMENTATION ASPECTS

3.1. Computing directional encoding gains

For a generic, recursive definition

Up to now, most of people interested in higher order ambisonics for virtual sound space encoding (*e.g.* for music), use explicit encoding equations such as provided by Malham and Furse [8], with a restriction to the 2nd or 3rd order. This Furse-Malham Harmonics set (FMH) is characterised by the fact that each function (excepted Y_{00}^{+1}) reaches a maximal value of 1.

On the other hand, the computation of encoding gains without order restriction may rely on the following few lines of matlab code:

```
pm = legendre(m, sin(elev), 'sch');
ymn_p = sqrt(2*m+1)*cos(n*azim).*pm(n+1);
ymn_m = sqrt(2*m+1)*sin(n*azim).*pm(n+1);
```

which provide values that conform to the "3D-Normalised" (N3D) encoding convention (2), the first line computing the values pm of functions $\tilde{P}_{mn}$.

Nevertheless, such a code cannot be used in a practical, matlab-independent DSP platform.

That's why it is useful to describe a generic algorithm for the computation of spherical harmonic encoding gains of any order. The algorithm detailed below relies on the recursive definition [2] of the Legendre polynomials and associated functions P_{mn}, and of the cosine/sine functions (see appendix A.2.2 of [3]). Thus, it globally process a recursive computation of directional encoding gains y_{mn}^{σ}:

$$y_{mn}^{\sigma} = Y_{mn}^{\sigma}(\theta,\delta) = Y_{mn}^{\sigma}(\vec{u}), \qquad (14)$$

starting from the cartesian coordinates u_x, u_y, u_z of the unitary incidence vector $\vec{u}$, or, which is equivalent, from its azimuth θ and elevation δ.

Step one: initialisation

$$\begin{cases} \bar{r} = \cos\delta,\ u_z = \sin\delta \\ c_1 = \cos\theta,\ s_1 = \sin\theta \end{cases} \text{ or } \begin{cases} \bar{r} = \sqrt{1-u_z^2} \\ c_1 = u_x/\bar{r},\ s_1 = u_x/\bar{r} \end{cases} \qquad (15)$$

If $\bar{r}=0$ (purely vertical incidence), there's no azimuth dependence, then c_1 and s_1 can be set to arbitrary values. Let's recall that: $u_x^2+u_y^2+u_z^2=1$.

Cross-recurrence for azimuth/horizontal dependent terms c_n=cos(nθ) and s_n=sin(nθ)

$$\begin{aligned} c_{n+1} &= 2c_1c_n - c_{n-1} \\ s_{n+1} &= c_1s_n + s_1c_n \end{aligned}, \text{ for } n=1 \text{ to } M-1. \qquad (16)$$

Double-recurrence for elevation/vertical dependent terms p_{mn}=$P_{mn}(u_z)$= P_{mn}(sinδ)

$$\begin{aligned} p_{m+1,0} &= \frac{2m+1}{m+1}u_z\,p_{m,0} - \frac{m}{m+1}p_{m-1,0}, \\ p_{m+1,n+1} &= p_{m-1,n+1} + (2m+1)\,\bar{r}\ p_{m,n} \end{aligned} \qquad (17)$$

applied for n=0 to m-1, for each m=1 to M-1, with $p_{m'n'}$=0 when m'<n'.

And finally:

$$\begin{aligned} y_{mn}^{+1} &= \beta_{mn}.p_{mn}.c_n \\ y_{mn}^{-1} &= \beta_{mn}.p_{mn}.s_n \quad (\text{for } n>0) \end{aligned} \qquad (18)$$

with β_{mn} being coefficients depending on the encoding convention used. These ones lead to (SN3D)-compliant encoding gains:

$$\beta_{mn}^{(\mathrm{SN3D})} = \sqrt{(2-\delta_{0,n})\frac{(m-n)!}{(m+n)!}}, \qquad (19)$$

whereas the following ones are (N3D)-compliant (2):

$$\beta_{mn}^{(\text{N3D})} = \sqrt{2m+1}\,\beta_{mn}^{(\text{SN3D})} \quad (20)$$

It is clear that such coefficients are recursively computable. In practice, they would even be tabulated.

Horizontal only encoding and components

For the case of a completely "horizontal" restriction (components with n=m and sources with δ=0), recurrence (17) is useless. Considering either "2D semi-normalised" (SN2D) or "2D normalised" (N2D) encoding convention, directional gains are just:

$$\begin{cases} y_{mn}^{+1\,(\text{SN2D})} = c_n \\ y_{mn}^{-1\,(\text{SN2D})} = s_n \end{cases} \text{and} \begin{cases} y_{mn}^{+1\,(\text{N2D})} = \sqrt{2}\;c_n \\ y_{mn}^{-1\,(\text{N2D})} = \sqrt{2}\;s_n \end{cases} (\text{for } m>0) \quad (21)$$

Note that y_{00}^{+1}=1 for any convention. More generally, the relation between (N2D) and (N3D) conventions is given by [1, 3]:

$$\beta_{mm}^{(\text{N2D})} = \sqrt{\frac{2^{2m}\,m!^2}{(2m+1)!}}\;\beta_{mm}^{(\text{N3D})} \quad (22)$$

3.2. Design of distance coding filters

Design strategy for parametric low cost digital filters

One basic method for deriving digital filters from their analytic, frequency responses is to process an inverse Fourier transform of these responses. This leads to a Finite Impulse Response (FIR) model. This approach is actually not very interesting for several reasons: it has to be computed for each new distance parameter, its processing may be expensive, according to the FIR length, and artefacts called "Gibbs oscillations" are always present at the FIR extremities, and they are progressively smoothed only by enlarging the FIR.

In contrast, we preferably seek a parametric, lower cost, IIR (Infinite Impulse Response) filter implementation. It appears that the bilinear transform, which is well known in digital filter design, does perfectly the job. Let's first define the successive steps of the following design strategy. With the final aim being to describe filters with second and first order sections, we have first to find their poles and zeros. For convenience, this pole-zero extraction is preferably done directly on the analog domain filters, before applying the bilinear transform. So, let's first rewrite the near field modelling transfer function (5) as the Laplace function:

$$F_m^{(\tau)}(p) = \sum_{n=0}^{m} \frac{(m+n)!}{(m-n)!\,n!}\left(2\tau p\right)^{-n}, \quad (23)$$

with respectively $\tau=\rho/c$ or $\tau=R/c$ if the matter is to simulate the virtual source distance or to compensate for the loudspeaker near field.

Pole-zero extraction

To find the poles and zeros of filter $F_m(p)$, it is convenient to set $X=2\tau p$ and rewrite (23) as:

$$F_m(X) = X^{-m} Q_m(X)$$
$$Q_m(X) = \sum_{n=0}^{m} \frac{(m+n)!}{(m-n)!\,n!} X^{m-n} = \prod_{q=1}^{m} (X - X_{m,q}) \quad (24)$$

While the poles of $F_m(p)$ are clearly null, its zeros p_{mq} appear to be related to the complex roots $X_{mq} = 2\tau p_{mq}$ ($0\leq q\leq m$) of the polynomial $Q_m(X)$, which is a particular case of the generalized Bessel polynomials. Traditional roots extraction algorithms are stable only for limited orders. Matlab function '`roots`' provides usable approximations up to order 24, which is enough for most applications. For more precise approximations or higher orders, Pasquini [9] provides a robust method dedicated to the generalized Bessel polynomials.

Some approximated values are given in Table 1.

m	Roots X_{mq} of Q_m
1	-2
2	-3.0000±1.7321*j*
3	-3.6778±3.5088*j*; -4.6444
4	-4.2076±5.3148*j*; -5.7924±1.7345*i*
5	-4.6493±7.1420*j*; -6.7039±3.4853*j*; -7.2935
6	-5.0319±8.9853*j*; -7.4714±5.2525*j*; -8.4967±1.7350*i*

Table 1 Roots of Q_m for the first few orders m.

In the following, we consider that the roots X_{mq} are arranged in decreasing order of imaginary parts.

Applying the bilinear transform

The second step is to transpose the pole-zero filter form from the analog (Laplace) domain to the digital domain (z-transform). For this purpose, the bilinear transform consists in applying the substitution $p=2f_s(1-z^{-1})/(1+z^{-1})$:

$$F_m^{(\tau)}(z) = F_m^{(\tau)}(p)\Big|_{p=2f_s\frac{1-z^{-1}}{1+z^{-1}}} \quad (25)$$

with f_s being the sampling frequency. Therefore, it's easy to write the zeros z_{mq} of $F_m(z)$ in terms of the zeros p_{mq} of the Laplace function $F_m(p)$:

$$z_{m,q}^{-1} = \frac{1-\dfrac{p_{m,q}}{2f_s}}{1+\dfrac{p_{m,q}}{2f_s}} \quad \textit{i.e.} \quad z_{m,q}(\tau) = \frac{1+X_{m,q}/(4\tau f_s)}{1-X_{m,q}/(4\tau f_s)} \quad (26)$$

Finally, by setting $X = 2\tau p = \alpha(1-z^{-1})/(1+z^{-1})$, with $\alpha = 4f_s\tau$, the "near field compensating" digital filter can be written in the pole-zero form:

$$\frac{1}{F_m^{(\tau)}(z)} = \frac{(1-z^{-1})^m}{\prod_{q=1}^{m}\left[\left(1-\frac{X_{m,q}}{\alpha}\right)-\left(1+\frac{X_{m,q}}{\alpha}\right)z^{-1}\right]} \quad (27)$$

More generally, a near field control filter H_m is formed by the ratio of two versions of (27) with different implicit parameters τ' and τ.

Second and first order sections

Any m^{th} order IIR filter can be implemented under the *Direct Form II* (28), with m/2 second order sections (or "cells") for even m, or $(m-1)/2$ second order sections plus one first order section for odd m:

$$\begin{aligned} H_m(z) &= \prod_{q=1}^{m/2} \frac{b_0^q + b_1^q z^{-1} + b_2^q z^{-2}}{a_0^q + a_1^q z^{-1} + a_2^q z^{-2}} \times \frac{b_0^{\frac{m+1}{2}} + b_1^{\frac{m+1}{2}} z^{-1}}{a_0^{\frac{m+1}{2}} + a_1^{\frac{m+1}{2}} z^{-1}} \\ &= g \prod_{q=1}^{m/2} \frac{1 + b'^q_1 z^{-1} + b'^q_2 z^{-2}}{1 + a'^q_1 z^{-1} + a'^q_2 z^{-2}} \times \frac{1 + b'^{\frac{m+1}{2}}_1 z^{-1}}{1 + a'^{\frac{m+1}{2}}_1 z^{-1}} \end{aligned} \quad (28)$$

the right factor (first order cell) being present only for odd orders m.

In order to define the coefficients of our NFC filter, let's first consider the denominator of (28) as related to the "near field compensation" part: it equals the denominator of (27), with $\tau = R/c$ as an implicit parameter. Each second order cell denominator $a_0^q + a_1^q z^{-1} + a_2^q z^{-2}$ derives from the two 1st order cells of (27) that involve conjugate complex roots $X_{m,q}$ and $X_{m,m-q+1} = X_{m,q}^{\ *}$:

$$\begin{aligned} a_0^q &= 1 - 2\frac{\mathrm{Re}(X_{m,q})}{\alpha} + \frac{\left|X_{m,q}\right|^2}{\alpha^2} \\ a_1^q &= -2\left(1 - \frac{\left|X_{m,q}\right|^2}{\alpha^2}\right) \qquad \text{for } 1 \le q \le m/2 \\ a_2^q &= 1 + 2\frac{\mathrm{Re}(X_{m,q})}{\alpha} + \frac{\left|X_{m,q}\right|^2}{\alpha^2} \end{aligned} \quad (29)$$

For odd order filters, the coefficients of the additional first order cell merely derive from the remaining real root $X_{m,(m+1)/2}$ as follows:

$$a_0^{\frac{m+1}{2}} = 1 - \frac{X_{m,(m+1)/2}}{\alpha}, \quad a_1^{\frac{m+1}{2}} = -\left(1 + \frac{X_{m,(m+1)/2}}{\alpha}\right) \quad (30)$$

Numerator coefficients b_i^q, related to the "virtual source distance coding" part, are computed exactly the same way, but with $\tau = \rho/c$ as an implicit parameter instead of $\tau = R/c$.

The second line of (28) suggests a lower cost implementation that saves a number of multiplications. It involves filter coefficients b'_i, a'_i and g that straightforwardly derive from coefficients b_i and a_i.

Finally, for a more efficient implementation, it would be recommended to tabulate the real part and the modulus of each root X_{mq} (for $1 \le q \le (m+1)/2$) rather than the complex roots themselves.

Frequency scale distortion: practically ineffective

The bilinear method transforms the unlimited frequency axis $p = j\omega$ ($\omega \in]-\infty,+\infty[$) of the Laplace's complex plan into the unitary complex circle $z = e^{j\varpi}$, which reflects bounded frequencies $f \in]-f_s/2, +f_s/2[$. Thus theoretically, there is a frequency scale distortion between analog and digital filter responses:

$$f_{\text{analog}} = \frac{f_s}{\pi} \tan\left(\pi f_{\text{digital}} / f_s\right) \quad (31)$$

Nevertheless, this distortion is insignificant for frequencies that are small with respect to the sampling frequency f_s. Now in the present case, the filter response varies only on a low frequency domain, and no longer above a frequency that depends on the distance ratio ρ/R and the order m (Figure 6). Considering the parameters typically used in practice, one can verify that the digital filter response fits the analytic one very well. Thus such designed filters are fully satisfying in practice.

Time properties: viewing impulse responses

Figure 10 exhibits some NFC filter temporal responses computed with f_s=44.1kHz and c=340m/s, for "inside" and "outside" sources. In the top case of an inside source, a kind of "Dirac" is followed by a ventral section which amplitude increases with order m. This will be further discussed in the next section, while interpreting synthetic sound field snapshots.

Let's finish by a remark regarding the use of NFC filters for adapting NFC-HOA signals from a reference distance R_1 to another one R_2 (as discussed in 2.3). It is verified that the original signals are exactly restored by backward conversion (R_2 to R_1): the impulse response of $H_m^{(R1/c,\, R2/c)}.\ H_m^{(R2/c,\, R1/c)}$ is an un-delayed Dirac.

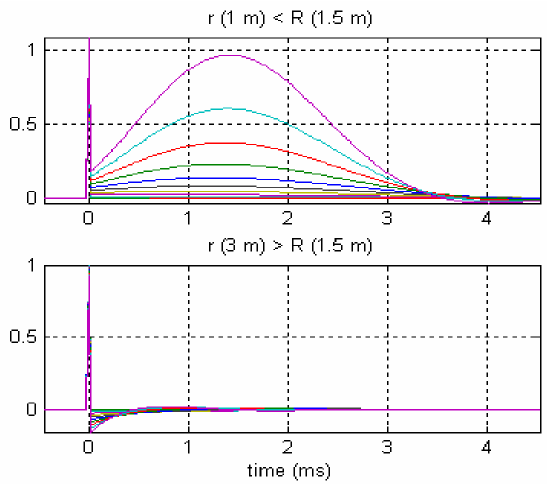

Figure 10 Impulse responses of NFC filters for a source inside (r=1m) and outside (r=3m) the loudspeaker array (R=1.5m). 1st to 11th order responses are shown with increasing amplitudes after the first "Dirac".

4. APPLICATIVE ISSUES

A first summary and comparison

The two previous sections introduced respectively theoretical and practical solutions for enabling Higher Order Ambisonics to encode and render sources at arbitrary distances, and especially near field sources.

The Near Field Control (or "Coding") filters are designed as parametric IIR digital filters that may be implemented with the lowest possible cost regarding their functionality. For these reasons among others, this encoding scheme would be preferred to the distance coding scheme recently introduced by Sontacchi and Höldrich [10], which combines WFS (Wave Field Synthesis) "notional" encoding (using a virtual circular microphone array) and HOA encoding (applying a circular Fourier transform on the simulated microphone signals). In spite of arising a quite relevant concept, this encoding scheme has a number of disadvantages in practice. First, it is computationally expensive because it involves simulating and "ambisonically" encoding a lot of virtual microphone signals. Moreover, this indirect encoding suffers from artefacts that are typical to WFS [1], such as: spatial aliasing (which is reduced only by increasing the number of virtual microphone), vertical aliasing (horizontal-only microphone array), and time-reversing for inside sources, which causes an inverted ITD (Interaural Time Delay). And finally, it constraints the loudspeaker array to a fixed radius (the same as the virtual microphone array).

The following addresses some applicative issues involving format aspects and signal processing tools described in previous sections.

4.1. Illustration of positional rendering

The NFC filters designed in 3.2 are now applied for simulating the positional rendering of virtual sources in the time domain, which completes the frequency domain simulations of 2.3. A high (15th) order, 32-speaker system is involved. This actually results in a large area, "holophonic" reconstruction. To make the sound field visualisation clearer, a gaussian pulse (a windowed single sine, with the centre frequency f_c=500Hz) is chosen as the encoded signal and conveyed by the wave fronts.

Far field virtual sources

The plane wave case shown Figure 11 implies only few comments. The reconstruction looks very good on the disk just including the three illustrated listeners. Outside this disk, some artefacts (off-centred interference "rose" patterns) appear on the synthetic wave front due to the higher frequency content of the pulse, but its spatial consistency remains on a quite large area. It is moreover verified that the reconstruction is even better with nearest, but still outside sources, since the acoustic phenomenon to be synthesised becomes closer to what the real sources (loudspeakers) can actually create.

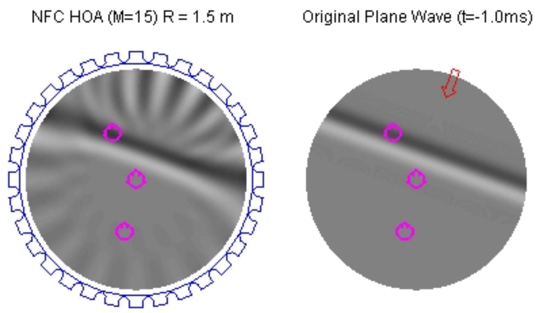

Figure 11 Time domain synthesis of a plane wave.

Near field, enclosed virtual sources

The two snapshots of Figure 12 give a time domain view of the case of an enclosed source, previously shown in the frequency domain (Figure 9). The first one (beginning of pulse emission) exhibits a strong interference pattern on the border area ("behind" the virtual source distance), caused by loud emitted signals with alternatively opposite phases. Figure 9 helps understanding that this border interference concerns the lower frequency content, which is particularly amplified by high order NFC filters (Figure 6). Interfering elementary wave fronts radiated by the loudspeakers rapidly and partially cancel each other while converging towards the centre (see the second snapshot at t=3ms), to synthesise the expected sound field. Like on Figure 9, the latter is well reconstructed on the disk excluding the virtual source.

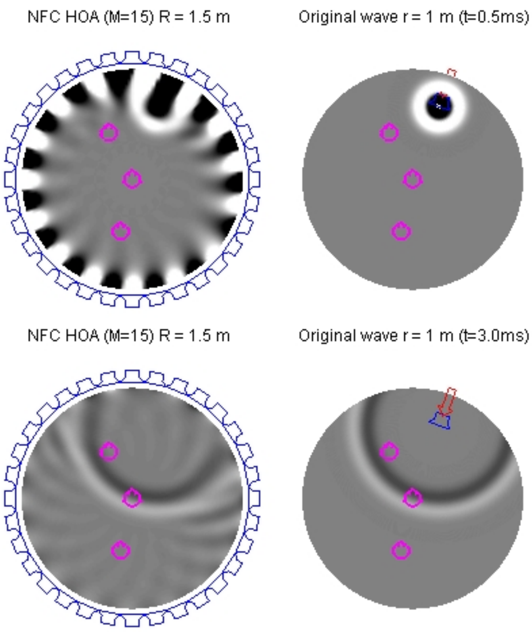

Figure 12 Two snapshots of spherical wave time domain synthesis, for an enclosed virtual source (r=1m<R=1.5m).

Another interesting observation arises from the time domain simulations. Indeed, the loudspeaker contributions have to "prepare", *i.e.* anticipate the pulse emission by the enclosed source. Moreover, It's worth noticing that for an inside source (r=1m), the ventral section of NFC impulse responses (Figure 10) tends to reach its maximum about 1.5ms, which is the time $(R\text{-}r)/c$ for the virtual source point being reached by the nearest loudspeaker wave.
Finally, it's worth recalling that the synthetic wave propagates in the correct direction, whereas Wave Field Synthesis (WFS) would synthesize a time-reversed wave front (*i.e.* converging towards the virtual source). A further comparison between WFS and NFC-HOA is given in [1].

4.2. Various uses of NFC filters

Positional encoding of elementary wave fronts

As mentioned in introduction, the positional encoding and rendering process illustrated in 4.1 may apply for any individual contribution computed by the environmental acoustics processor (Figure 1). Thus NFC filtering (Figure 7) may apply to direct sound as well as individual, discrete reflections (as emitted by "mirror sound images").

Diffuse field encoding

A general encoding scheme (Figure 1) may also process diffuse signals that typically correspond to the remaining room effect and especially the late reverberation. What's important is to actually provide the effect of a diffuse field, *i.e.* uncorrelated parts coming from surrounding directions. Thus, one can simply encode this reverb signals as plane waves, using NFC filters with the distance parameter $\rho=\infty$.

Another option consists in considering that a diffuse field is theoretically represented by uncorrelated ambisonic components of same energy (with the "N3D" encoding convention). Therefore, the "reverb" signals provided by the room effect processor can also be directly added to ambisonic signals, with an appropriate gain adaptation if the encoding convention is not "N3D" (refer to 3.1 or to [3]), and also with a NFC filtering like mentioned just above.

Format adaptation (change of the reference distance)

For the purpose of mixing different "NFC-HOA" material (multi-channel streams) or decoding for an arbitrary loudspeaker layout, NFC filters are also used to adapt the encoded material from a reference distance to another one. Figure 8 describes this adaptation scheme.

Sound field transformations and effects

It's worth noticing that sound field transformations that process separately each group of components of same order m (like rotation) apply the same way to NFC-HOA as to HOA.

It is not the same with focalisation, which consists in combining spherical harmonics to form a directive beam, as if a highly directive microphone were pointing in a given direction of space [3]. When applied to NFC-HOA material, the focalisation effect is the most effective at a distance that is roughly the reference distance. One can control this distance by applying NFC filters before focalisation. Nevertheless, radial selectivity is much less than angular selectivity, and moreover it depends on the frequency.

Finally, NFC filters can be applied to artificially distort the wave field curvature, and at the same time the bass-boost effect.

4.3. Binaural synthesis of close sources

Virtual NFC Ambisonics (using virtual loudspeakers)

Although ambisonic approach is initially dedicated to reproduction over loudspeakers, its rendering over headphones is also possible. A basic approach for that is to combine ambisonic decoding with the so-called "virtual loudspeaker" process, which consists in the binaural simulation of each loudspeaker for a centred listener position, *i.e.* the filtering of its signal by the corresponding HRTF (Head related Transfer Function). For a more efficient implementation, one merges the decoding matrix and the HRTF filter bank associated with the loudspeaker directions. The resulting filter bank directly processes ambisonic components B_{mn}^{σ}, producing signals that are combined to form the binaural signals.

It has been shown [3, 5] that such a "virtual ambisonics" process progressively reaches the performance of a direct binaural synthesis in terms of spectral reconstruction when considering increasing ambisonic orders. Now, binaural synthesis usually addresses "far field" sources, which distance actually corresponds to the one of the loudspeaker used for the HRTF measurements. As an improvement, the new positional encoding scheme (Figure 7) associated with the "virtual ambisonics" rendering can be used to binaurally synthesise sound sources at arbitrary distances. In particular, it can render the emphasised ILD (Interaural Level Difference) of close sources.

Extension of the binaural B-format scheme

Binaural synthesis using NFC-HOA scheme can be even more efficient by focussing on each ear separately, as if each one were at the origin of the spherical coordinate system and at the centre of the loudspeaker array. Thus, one has to derive the virtual source properties in the referential of each ear (Figure 13). Delaying and weighting the original signal S (left part of Figure 14) leads to two signals S_L and S_R that are spatially encoded (middle part of Figure 14) with respective positional parameters (r_L, θ_L, δ_L) and (r_R, θ_R, δ_R). Then the resulting multi-channel streams have just to be decoded (right part of Figure 14) according to the "virtual ambisonics" scheme mentioned above, but focussed separately on each ear. The whole process can be seen as a modified version of the "binaural B-format" scheme [11], enriched by the NFC positional encoding scheme.

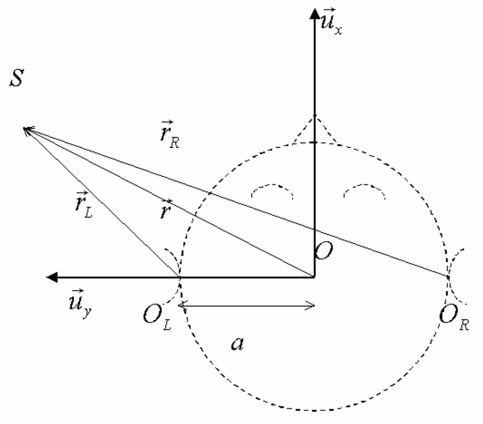

Figure 13 Deriving two coordinates systems centred on each ear (O_L, O_R) from the head centred one (origin O).

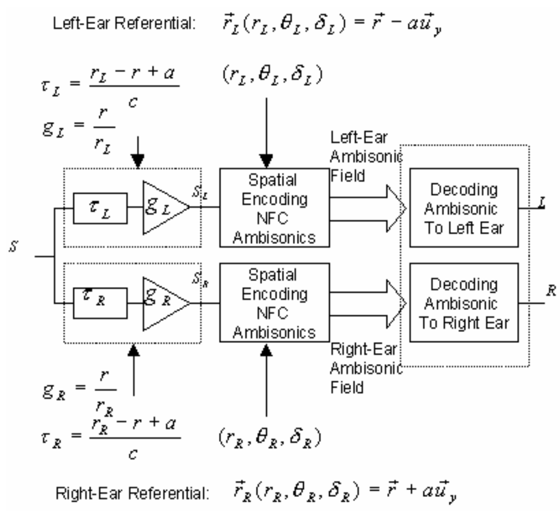

Figure 14 Application of the new positional coding (as described Figure 7) in the referential of each ear, for accurate binaural synthesis of close sources.

4.4. Generic format specifications

When conveying a 3D audio multi-channel material like 1st or higher order ambisonics, specifications must accompany the audio streams (or channels) to enable the player interpreting them and applying appropriate decoding or transform operations. The specifications listed below address multi-channel file formats like the Wave file format and its extensible definition [12], as well as multi-channel compressed audio streaming in MPEG-4 [13]. The fields that are introduced compose a structure that could extend the format chunk ("`fmt `") of the wave field header, for example[3]. Nonetheless, they may be refined and renamed, and their arrangement and size specifications still have to be discussed.

Encoding convention - conversion rules

Several encoding conventions may be possibly used [3]: "N3D" and "SN3D" derive from the most generic definition of 3D spherical harmonics (see 2.1 and 3.1); "N2D" and "SN2D" (section 3.1) initially concern a 2D-restricted formalism (cylindrical harmonics) [14] but may also apply to 3D spherical harmonics provided that clear extension rules are given; finally, Furse-Malham Harmonics ("FMH") [8] are "Max-Normalised" (MaxN) excepted for m=0 and only provided up to the 2nd or 3rd order.
It's important to specify which convention a given multi-channel material obeys. For this purpose, we introduce a field **`encodingConvention`** that takes a value among the following list: {N2D=0, SN2D=1, N3D=2, SN3D=3, MaxN=4, FMH=5, etc.}.
It's also important to now how to convert the material from such a convention ("e1") to another one ("e2"), in order to properly apply transformation, decoding or mixing operations that wouldn't be defined for the same convention. This involves the conversion rule:

$$B_{mn}^{\sigma\ (e2)} = \alpha_{mn}^{(e2\leftarrow e1)}.B_{mn}^{\sigma\ (e1)},\ \ \alpha_{mn}^{(e2\leftarrow e1)} = \frac{\beta_{mn}^{\sigma\ (e2)}}{\beta_{mn}^{\sigma\ (e1)}}, \qquad (32)$$

where $\alpha_{mn}^{(e2\leftarrow e1)}$ is the conversion factor, as defined by Table 2.

	α_{00}	α_{1n}	α_{22}	α_{21}	α_{20}	α_{mn}
FMH←SN3D	$1/\sqrt{2}$	1	$2/\sqrt{3}$		1	none for m>2
N3D←SN3D	1	$\sqrt{3}$	$\sqrt{5}$			$\sqrt{2m+1}$
N2D←SN2D	1	$\sqrt{2}$	$\sqrt{2}$			$\sqrt{2}$
N2D←N3D	1	$\sqrt{2/3}$	$\sqrt{8/15}$			$\sqrt{\frac{2^{2m}m!^2}{(2m+1)!}}$

Table 2 Conversion factors between the existing encoding conventions

Any other conversion would merely derive by using the relations:

$$\alpha_{mn}^{(e3\leftarrow e1)} = \alpha_{mn}^{(e3\leftarrow e2)}.\alpha_{mn}^{(e2\leftarrow e1)},\ \alpha_{mn}^{(e2\leftarrow e1)} = 1/\alpha_{mn}^{(e1\leftarrow e2)} \qquad (33)$$

For the conversion of a sound field transformation matrix $\mathbf{T}^{(e1)}$ (e.g. a rotation) or a decoding matrix $\mathbf{D}^{(e1)}$, initially defined for a convention (e1), one has to apply the conversion matrix $\mathbf{A}^{(e1\leftarrow e2)}$ which elements are the factors $\alpha_{mn}^{(e2\leftarrow e1)}$:

$$\mathbf{T}^{(e2)}= \mathbf{A}^{(e2\leftarrow e1)}.\ \mathbf{T}^{(e1)}.\ \mathbf{A}^{(e1\leftarrow e2)}$$
$$\mathbf{D}^{(e2)}= \mathbf{D}^{(e1)}.\ \mathbf{A}^{(e1\leftarrow e2)} \qquad (34)$$

Reference distance or reference time

Any "NFC-HOA" material comprises implicit parameters, which are a reference distance R and the sound velocity c. More synthetically, we will specify a reference delay **`nfcReferenceDelay`**, which is the ratio $\tau=R/c$. This field can take an infinite value to indicate the case of an "uncompensated" HOA material (section 2.3).

Hybrid resolution and channel ordering

An ambisonic material may have a higher directional resolution in the horizontal plane than in the rest of the sphere. That supposes to distinguish between an upper order M_{2D} (the field **`resolution2D`**) for 2D-components B_{mm}^{σ}, and an upper order M_{3D} (the field **`resolution3D`**) for the other 3D-components (such that $n<m$). The number of components involved is:

$$\begin{aligned} K &= 2M_{2D}+1+(M_{3D}+1)^2-(M_{2D}+1)^2 \\ &= M_{3D}{}^2 - M_{2D}{}^2 + 2M_{3D}+1 \end{aligned} \qquad (35)$$

We already suggested arranging the components of each order m by beginning with the horizontal ones, *i.e.* with n decreasing from m to 0. This choice allows designating each component by a single index SID (for "Single Index Designation"), as described in Table 3.

Name	W	X	Y	Z	U	V	…
$^{\sigma}_{mn}$	$^{+1}_{00}$	$^{+1}_{11}$	$^{-1}_{11}$	$^{+1}_{10}$	$^{+1}_{22}$	$^{-1}_{22}$	…
SID	0	1	2	3	4	5	…
No more usual names							
$^{+1}_{mm}$	…	$^{+1}_{mn}$	$^{-1}_{mn}$	…	$^{+1}_{m0}$		
m^2	…	$m^2+2(m-n)$	$m^2+2(m-n)+1$	…	$(m+1)^2-1$		

Table 3 Single Index Designation of ambisonic components B_{mn}^{σ}

According to this default "ordering rule", the case (M_{3D}=1, M_{2D}=3) would lead to the index list {0, 1, 2, 3, 4, 5, 9, 10}, for example. Note that the FMH file format, which ordering is described by the list {0, 1, 2, 3, 8, 6, 7, 4, 5}, doesn't exactly obey this rule.
For the case the component ordering would need to be defined explicitly or according to another rule, one introduces the field **`orderingRule`**. Its possible values are at least: 0 (default ordering) and 1 (explicit ordering). They could be extended to other rules,

[3] This extends an earlier mail discussion between the author, Dave Malham, Richard Furse and Richard Dobson.

such as a "lateral preference" rule that would place *Y* before *X*, and more generally σ=-1 before σ=+1.
The case of an explicit ordering implies defining another field **componentIndex** as the array of component indexes (according to Table 3).
These ordering issues are important in contexts (like MPEG-4) that enable multi-channel scalability, *i.e.* the possibility to convey only a subset of audio channels to fit transmission rate limitations.

Mixed order format

Dave Malham [8] raised the problem of mixing and conveying ambisonic materials of different original resolutions, while allowing, in the end, to optimally decode them according to their respective original resolution. Such a differentiated decoding supposes controlling the relative weightings of components of different orders, according to the resolution of the material they originally belong to. To make it possible after mixing 1st and 2nd order materials, Malham suggests conveying two versions of the W-component, one resulting from the mixing of both materials, and the other regarding only the 1st (or the 2nd) order material. The principal is also described and extended to higher resolutions in [3].
More generally, one has to foresee and allow the transmission of several versions of the same components, with the indication of the upper or/and the lower resolution of the materials they originally belong to. Let's introduce a field **mixtResolution** with the possible values: 0 (no differentiated resolutions); 1 (specification of the lower resolution of each component); 2 (specification of the upper resolution of each component); 3 (specification of both lower and upper resolutions). If specified, arrays **upperResolution** and/or **lowerResolution** are filled in with the same number of integer elements as the number of channels.

Possible future specifications

Some existing stereo or multi-channel formats like MS, UHJ, BHJ [6] or G-Format [15] are closely related to first order ambisonics. They are kinds of matrixed forms of the ambisonic components W, X, Y, (and Z). The latter are fully or partially pre-decoded so that the conveyed signals are ready to feed loudspeakers (*e.g.* stereo or "5.1" arrangements), but they can be also restored and decoded for another rendering configuration. These options, and their possible extension to higher order, could enrich the **orderingRule** list.
Finally, the presently discussed format could cover the extended binaural B-format mentioned in section 4.3, by adding optional "left" and "right" specifications to the transmitted components.

Structure for multi-channel format specifications

The following structure gathers the fields introduced above. As previously said, the arrangement and bit allocation of these fields have to be discussed. Some integer or enumeration values (int) typically required only a few bits. The values of the optional arrays could be interleaved.

encodingConvention	int (enum)
resolution2D	int
resolution3D	int
nfcReferenceDelay	float
orderingRule	int (enum)
mixtResolution	bool
componentIndex	int array
lowerResolution	int array
upperResolution	int array
…	

Table 4 Structure for NFC-HOA format specifications

5. CONCLUSION

This paper has stated that the commonly adopted higher order ambisonic (HOA) encoding format suffers from a theoretical obstacle that addresses the ability to represent near field. Indeed, the wave front curvature due to finite distance sound sources causes an emphasis of the spherical harmonic components that tends to be infinite at low frequencies. This obstacle is overcome by compensating, from the encoding stage, the near field of reproduction loudspeakers. This compensation is anyway required to preserve the original curvature of encoded wave fronts. This way a new, "Near Field Compensated" (NFC) HOA encoding format is defined, which is now viable since its components are ensured to have finite amplitude. At the same time, HOA ambisonic recording systems are made practicable since they now require only a finite equalization. Regarding virtual source encoding, distance-coding (or "Near Field Coding/Control") filters can be designed. Such NFC filters also apply for NFC-HOA format adaptation to arbitrary distances of reproduction loudspeakers.
Implementation details have been given on signal processing tools for both directional and distance encoding. For the latter, we designed parametric, low cost digital filters.
The efficiency of the resulting positional encoding scheme has been illustrated with sound field simulations of outside and inside sources. Then a particular application has been described, which is the accurate binaural synthesis of close sound

sources. Finally, we detailed a list of specifications related to the NFC-HOA format. These specifications form a structure that may extend wave file header and MPEG-4 audio stream description, in order to handle NFC-HOA as a 3D audio multi-channel format.

6. REFERENCES

[1] J. Daniel, R. Nicol, and S. Moreau, Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging, presented at the AES 114th Convention, Amsterdam, 2003 22-25 March.

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

[3] 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.

[4] J. Daniel, J.-B. Rault, and J.-D. Polack, Acoustic properties and perceptive implication of stereophonic phenomena, presented at the AES 16th International Conference, Rovaniemi, Finland, 1999 April 1999.

[5] 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.

[6] M.A. Gerzon, Ambisonics in Multichannel Broadcasting and Video, Journal of the Audio Engineering Society, 33(November), pp. 859-871, 1985.

[7] M.A. Gerzon, General Metatheory of Auditory Localisation, presented at the 92nd AES Convention, 1992.

[8] D. Malham and R. Furse, Second and Third Order Ambisonics - the Furse-Malham Set, http://www2.york.ac.uk/inst/mustech/3d_audio/secondor.html

[9] L. Pasquini, Accurate computation of the zeros of the generalized Bessel polynomials, Numerische Mathematik, 86, pp. 507-538, 2000.

[10] 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.

[11] J.-M. Jot, S. Wardle, and V. Larcher, Approaches to binaural synthesis, presented at the 105th AES Convention, San-Francisco, 1998 March.

[12] Microsoft, Multiple Channel Audio Data and WAVE Files, http://www.microsoft.com/hwdev/tech/audio/multichaud.htm

[13] MPEG, Audio Group, Output Document N5386, Awaji Island, 2002 Dec. 12, http://mpeg.telecomitalialab.com/

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

[15] R. Elen, Ambisonics for the New Millennium, http://www.ambisonic.net/gformat.html