



Audio Engineering Society Convention Paper

Presented at the 116th Convention
2004 May 8–11 Berlin, Germany

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Further Study of Sound Field Coding with Higher Order Ambisonics

Jérôme Daniel and Sébastien Moreau

France Telecom R&D, 2 avenue Pierre Marzin, 22307 Lannion Cedex, France
jerome.daniel@francetelecom.com
sebastien.moreau@francetelecom.com

ABSTRACT

Higher Order Ambisonics (HOA) provides a rational and flexible way for spatial encoding, conveying and rendering of 3D sound fields. For this reason it has known a growing interest over past years. Nevertheless, representing near field sources and recording natural sound fields has been addressed only quite recently. This raises the problem of "infinite bass-boost", which a recent approach (NFC-HOA) solves while being fully equivalent with spherical harmonics representation. To better handle problematic cases where bass-boost remains excessive, the present study discusses the actual usefulness of some spatial components depending on the area targeted for sound field reconstruction. Therefore it suggests frequency dependent restriction of spatial resolution by high-passing spatial components. As a particular result, it shows that a much moderated amplification is sufficient to efficiently model sound sources at any distance, and derives a safe and fine solution to simulate sources inside the listening area.

1. INTRODUCTION

Ambisonics were introduced decennials ago [1] as a way to represent and render 3D sound fields, that surpasses traditional two-channel stereophony and even quadrasonic systems. Early Ambisonics relies on a minimal, but sufficient, directional description: the so-called B-Format (omnidirectional (W) and bidirectional components X, Y, Z) that can be obtained by conventional recording means. It has the advantage of being adaptable to many different loudspeaker rigs (either 2D or 3D). Nevertheless its low spatial resolution (1st order) implies some limitations especially

regarding the sweet spot, and it cannot offer large scale sound field reproduction. Extension of B-format to higher spatial resolutions (Higher Order Ambisonics or HOA) has been knowing a growing interest for nearly a decennial now [2-6], and it has been shown to enlarge the area of correct acoustic reconstruction (and therefore the sweet spot). But until the last few years, it has been restricted to virtual encoding (no practical recording system) and furthermore to directional considerations (angular pan-pot with no distance coding nor control of the synthesized wave front curvature).

Among recent studies, some work [7, 8] addressed the modeling of finite distance sources and the problem of representing it with finite amplitude signals. A solution

was proposed by introducing at the very encoding stage a pre-compensation of the reproduction loudspeakers near-field, which bounds problematic bass-boost effects. Targeted applications are virtual source coding with distance control, but also the design of 3D microphone array processing. As a major result, [7] introduced Near-Field Control (or "Distance Coding") Filters that complete the virtual source spatial encoding scheme, which was restricted to purely directional coding before. Applied to relatively high order encoding and rendering systems, these filters have been shown to be capable of synthesizing curved wave fronts over large areas, especially for virtual sources that are outside the loudspeaker enclosure. The synthesis of "inside" sources is shown to be possible to a certain extent, though involving electro-acoustic bass-boost that becomes impracticable when the source distance is too small.

One characteristic of this previous work is that it aimed at providing encoding means that are fully representative (*i.e.* over the full frequency axis) of the theoretical spherical harmonics decomposition. The present paper discusses the actual usefulness of encoded spatial components as a function of the frequency, depending on their order m and the radius of a targeted reconstruction area. As it mostly focuses on radial characteristics of the sound field encoding and rendering, it lets the reader refer to previous literature (*e.g.* [7]) for more details about directional properties and conventions.

2. PREVIOUS WORK ON HOA

2.1. Higher Order Ambisonics Basics

2.1.1. Sound field decomposition

There are several ways of explaining how Higher Order Ambisonics represents the sound field. By analogy with well known signal processing concepts, one can say that it performs a kind of spherical Fourier Transform of all acoustic events (waves) coming from all around a reference point O . This yields spherical harmonics signals (spatial components B_{mn}^σ) associated to different angular frequencies. More formally, spherical harmonics decomposition of an acoustic pressure field p comes from writing wave equation $(\Delta + k^2)p = 0$ (with wave number $k = 2\pi f/c$, frequency f and sound speed c) in the spherical coordinate system, where any point \vec{r} is described by its direction (azimuth θ and elevation δ), and its distance (radius r) with regards to a reference

point O . This leads to the so-called Fourier-Bessel decomposition:

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

Resulting spatial (or ambisonic) components B_{mn}^σ are associated with angular functions $Y_{mn}^\sigma(\theta, \delta)$ called "spherical harmonics" (further defined in [7]). They form groups of $(2m+1)$ components B_{mn}^σ having the same "order" m , and are respectively associated to radial functions $j_m(kr)$, also called "spherical Bessel functions". Their curves illustrated in Figure 1 show how these groups of components contribute to the sound field as a function of the distance from center O : the higher the order m , the farther the m^{th} order components group contributes to sound field description compared with the wavelength. This also reflects ambisonic components as being related to spatial sound field derivatives: higher order derivatives help approximating the sound field over a larger neighborhood of reference point O .

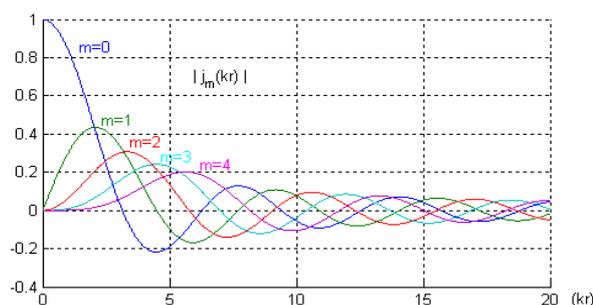


Figure 1 Spherical Bessel functions $j_m(kr)$

Of course, ambisonic systems practically rely on a finite highest order M (such that $m \leq M$), thus a finite number $K^{3D} = (M+1)^2$ of components B_{mn}^σ . Moreover, if only a 2D, horizontal rendering is addressed as often for practical reasons, then only the subset of "horizontal" components B_{mn}^σ (such that $n=m$) might be retained, which gives a total number of $K^{2D} = 2M+1$ components. Finally, let's recall that earlier B-Format is a 1st order representation, composed by omnidirectional component $W = B_{00}^{+1}$ (pressure field) and bidirectional components $X = B_{11}^{+1}$, $Y = B_{11}^{-1}$, $Z = B_{10}^{+1}$ (related to pressure gradient or acoustic velocity).

2.1.2. Encoding equations (for individual sources)

Encoding equations are used to compose virtual sound scenes with a number of virtual sources positioned in the space. To define spatial encoding equations, one considers an individual sound source (which direction is described by unitary vector \vec{u}_s , or azimuth θ_s and elevation δ_s) as typically creating either a plane wave (far field source) or spherical wave (near field source at point $\vec{\rho} = \rho\vec{u}_s$, *i.e.* distance ρ). In such cases, pressure field is described by:

$$p(\vec{r}) = S e^{jk\vec{r}\cdot\vec{u}_s} \quad (\text{plane wave}) \quad (2)$$

$$p(\vec{r}) = S \frac{\rho}{|\vec{\rho} - \vec{r}|} \frac{e^{-jk|\vec{\rho} - \vec{r}|}}{e^{-jk\rho}} \quad (\text{spherical wave}) \quad (3)$$

with S describing the conveyed signal, as measured at the reference point O . Performing spherical harmonics decomposition (1) over these definitions provides the corresponding "encoding equations", *i.e.* the expression of spatial components B_{mn}^σ as a function of the virtual source position and signal.

It comes that encoding equation of a plane wave merely consists in weighting signal S by real factors depending on the wave incidence, which are nothing other than spherical harmonic functions:

$$B_{mn}^\sigma = S Y_{mn}^\sigma(\theta_s, \delta_s) \quad (4)$$

We recall that a further discussion on directional encoding functions Y_{mn}^σ is given in [7] and is not reported in the present paper, which focuses on radial characteristics of the sound field.

Compared with the plane wave case, the following spherical wave encoding equation (for a source at a distance ρ) introduces a frequency dependent factor $F_m(\omega)$ that models the near field effect and therefore the wave front curvature:

$$B_{mn}^\sigma = S F_m^{(\rho/c)}(\omega) Y_{mn}^\sigma(\theta, \delta), \quad \omega = 2\pi f = kc \quad (5)$$

$$F_m^{(\rho/c)}(\omega) = j^{-(m+1)} \frac{h_m^-(k\rho)}{h_0^-(k\rho)} = \sum_{n=0}^m \frac{(m+n)!}{(m-n)!n!} \left(\frac{-jc}{\omega\rho}\right)^n$$

where the h_m^- are the outgoing/divergent spherical Hankel functions. Near Field modeling functions $F_m^{(\rho/c)}(\omega)$ (also written $F_m(k\rho)$) are characterized by an "infinite" bass-boost (for $m \geq 1$) with a $-m \times 6\text{dB/octave}$ low frequency slope (Figure 2), and could be implemented as integrating filters. Nevertheless, these have no practical application since they are unstable by nature. Therefore mathematical encoding equation (5) is not practicable as is. A workable alternative is proposed in [7, 8] and recalled in 2.2. The present paper develops a new and even safer solution in 3.

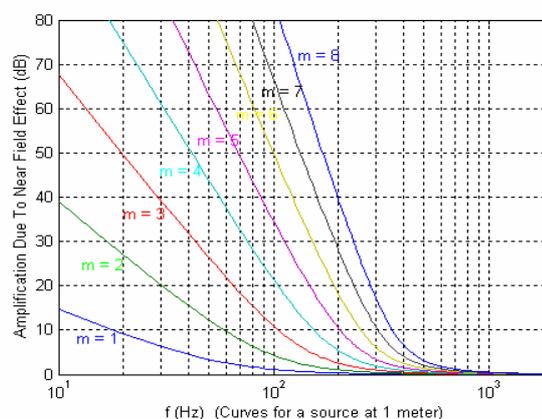


Figure 2 Low frequency infinite boost of ambisonic components due to near field effect $F_m(k\rho)$. Curves are shifted to the right (resp. the left) when the source distance ρ decreases (resp. increases).

2.1.3. Natural sound field encoding with a microphone array

In the following, we briefly recall the objectives and principle of a sound field recording system based on a discrete microphone array. For a bit more formal description, reader can refer to [8-11].

Let's first introduce the concept of spatial sampling of the sound field, which is achieved by measuring it at different points in the space by means of a microphone array. Considering that the measured sound field can be expressed in terms of spherical harmonic (regarding a reference point typically placed at the center of the array), each measured signal contains a "portion" or "sample" of these spatial components, according to its position and directivity. Therefore, defining the appropriate microphone signal processing to extract

spatial components B_{mn}^σ (HOA signals), is a matter of inverting spatial sampling equations. In the more general case, this leads to a matrix of filters [9]. When microphone capsules are distributed concentrically (*e.g.* over a sphere), the processing can be factorized into one matrix with real gains followed by a set of equalizers EQ_m [8, 12]. This is what we will further comment here. The matrix computes weighted sums and differences from the measured signals to get a rough (unequalized) estimation of spatial derivatives of different orders (*i.e.* HOA components). Then the equalizers "normalize" this estimation, depending on the order m and the array radius regarding the wave length (*i.e.* the radial position kr of measurement points on the Bessel curves of Figure 1), and also on the capsules directivity. A problem arises especially at low frequencies: information on spatial derivatives is as thin as wavelength is high with respect to the microphone array (small differences between measured signals). Since the proportion of HOA components contained in measured signals is known to be small, equalizers have to process a greater amplification to retrieve them (as shown by Figure 3).

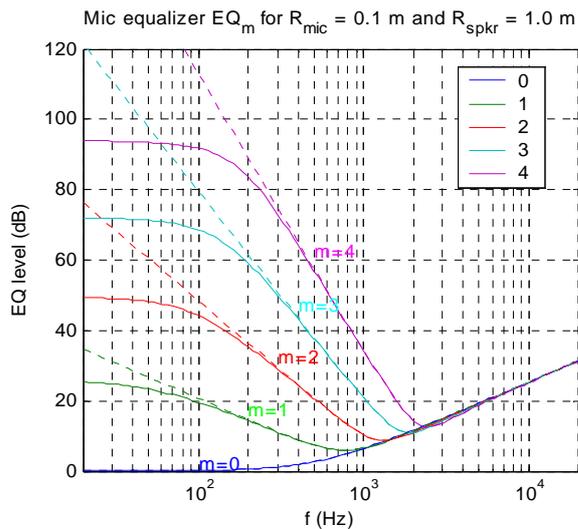


Figure 3 Equalization (EQ_m) required in microphone array processing to theoretically retrieve ambisonics components for pressure microphones distributed over a rigid sphere (radius $R_{mic}=10$ cm). Without (dotted lines) and with compensation of loudspeaker near field (cont. lines, with $R_{spr}=1$ m).

Theoretically, equalizers applied to respectively m^{th} order components would cause an infinite bass-boost (with generally $-m \times 6$ dB/octave as a low-frequency slope, shown as dotted lines in Figure 3). Solution

developed in 2.2 introduces finite low frequency amplification.

2.1.4. Decoding and rendering

Since it isn't the topic of this paper, we will simply make a brief summary of the principle of spatial decoding over loudspeakers. Further considerations can be found *e.g.* in [3, 6, 8]. Considered loudspeakers arrays are typically concentric (either circular or spherical), centered on the reference point O , and with preferably a regular angular distribution. That's the case we'll use in this paper for some illustrations. Earlier literature considered loudspeaker as emitting plane waves (*i.e.* in far field) from the center point of view. Under this assumption, the decoding process applied to the set spatial components B_{mn}^σ to get loudspeaker signals merely consists in a matrix \mathbf{D} with real gain factors, the aim being to recombine the encoded sound field at the center O by applying (4) to each loudspeaker contribution. It has more recently highlighted [6] that the finite distance of loudspeakers should be considered, which requires compensating their near field effect, as developed in next section.

Notice that spatial decoding can be also defined for non concentric array shapes at the expense of a higher computational cost. For convenience, sound field reconstruction will be illustrated considering a regular, circular loudspeaker array in the rest of the paper.

2.2. Near-Field Compensated HOA

2.2.1. Introducing Near-Field pre-Compensation

When reproducing a sound field by means of loudspeakers, their finite distance causes a near field effect that has to be compensated, so that encoded wave fronts are accurately synthesized with their proper curvature, instead of being distorted by the loudspeakers waves' curvature. Now, combining the near field effect $F_m(k\rho)$ of virtual sources (Figure 2) with the inverse of the loudspeakers one $F_m(kR)$ implies that the $-m \times 6$ dB/oct low frequency slope is stopped (at a given frequency) by a slope with opposite value. The same property arises with equalization filters EQ_m . Since compensation of loudspeakers near field is required for a correct rendering, [7] suggests introducing it from the encoding stage. This yields a variant definition of HOA encoding, called Near Field Compensated (NFC) HOA:

$$\tilde{B}_{mn}^{\sigma \text{ NFC}(R/c)} = \frac{1}{F_m^{(R/c)}(\omega)} B_{mn}^{\sigma} \quad (6)$$

This ensures that any near field sources and therefore any natural sound field are modeled and represented by finite amplitude signals $\tilde{B}_{mn}^{\sigma \text{ NFC}(R/c)}$. Note that such spatial description remains fully equivalent to the original spherical harmonics decomposition. Moreover, even if it involves a reference distance R as an intrinsic parameter (representative to the loudspeaker array size), this is not a restriction. Indeed, adaptation from distance R_1 to R_2 is simply performed by:

$$\tilde{B}_{mn}^{\sigma \text{ NFC}(R_2/c)} = \frac{F_m^{(R_1/c)}(\omega)}{F_m^{(R_2/c)}(\omega)} \tilde{B}_{mn}^{\sigma \text{ NFC}(R_1/c)} \quad (7)$$

The great advantage of NFC-HOA is that encoding tools associated to this new description now involve stable filters.

2.2.2. Distance coding filters

Reporting new sound field description (6) in spherical wave encoding equation (5) yields the following encoding formula for a finite distance source:

$$\tilde{B}_{mn}^{\sigma \text{ NFC}(R/c)} = S.H_m^{\text{NFC}(\rho/c, R/c)}(\omega) Y_{mn}^{\sigma}(\theta, \delta), \quad (8)$$

featuring distance coding transfer functions:

$$H_m^{\text{NFC}(\rho/c, R/c)}(\omega) = \frac{F_m^{(\rho/c)}(\omega)}{F_m^{(R/c)}(\omega)}, \quad (9)$$

which combine the near field effect of the virtual source and the compensation of the reproduction loudspeakers' one, as previously explained. These transfer functions can be realized as stable filters with finite LF amplification. Previous paper [7] further describes how to design and implement them as parametric, minimal cost IIR filters.

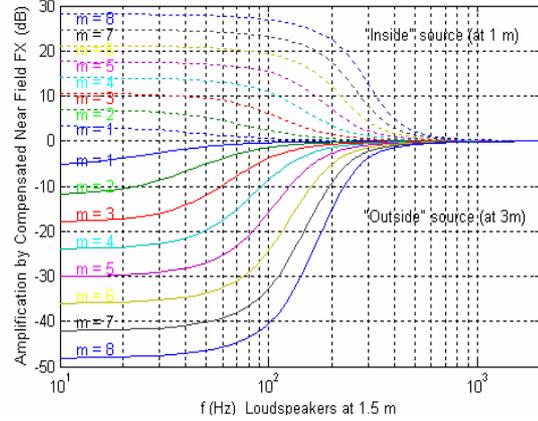


Figure 4 NFC filters frequency responses: finite amplification of ambisonic components from pre-compensated Near Field Effect (dashed lines: $\rho/R=2/3$; cont. lines: $\rho/R=2$).

Low frequency amplification is proportional to order m and depends on distance ratio R/ρ : more precisely it equals $m \times 20 \log_{10}(R/\rho)$ (in dB). Therefore distance coding filters cause an attenuation (at low frequencies) when the virtual source is beyond the loudspeaker array, and an amplification when the virtual source is "inside", as shown by Figure 4.

2.2.3. Equalizers for microphone processing

Like near field modeling, theoretical equalization involved in microphone array processing would also present a $-m \times 6 \text{ dB/oct}$ low frequency slope for each order m . Near Field Compensation $1/F_m(kR)$ helps stopping it thanks to an opposite slope, as shown by Figure 3. Note that the illustrated case (Figure 3) involves a microphone radius $R_{\text{mic}}=10 \text{ cm}$ that is small compared with the loudspeakers array radius $R_{\text{spkr}}=1 \text{ m}$.

2.3. Problematic cases in terms of bass-boost

Near-Field pre-Compensated HOA has solved the problem of representing the sound field with finite amplitude (ambisonic) signals and at the same time, the problem of involving signal processing with finite amplification. Nevertheless, even being now limited, low frequency amplification appears to be excessive in some particular cases for practical applications.

2.3.1. Sound field recording with small size microphone array

Here we consider the problem of sound field recording with a microphone array of relatively small size with regard to typical loudspeaker distances. Even with NFC, equalization filters involved for higher order components present great low-frequency amplification (Figure 3). There is a simple interpretation of this excessive estimation effort: at low frequencies and/or for higher order, the system tries to catch spatial information that is very thin at the measurement points and is substantial only at a distance from the microphone array.

As a result, the system mostly amplifies microphone background noise, calibration errors, capsules positioning error, and more generally any deviation from the theoretical model of sound field encoding by capsules (including directivity model). Recent studies have introduced sensible criteria to reduce this amplification: in [10] equalization filters are limited according to a "maximal white noise gain" criterion; in [9] a "regularization parameter" λ is introduced in the "filtering matrix" computation (when inverting the spatial sampling equations as shortly explained in 2.1.3), as a compromise between signal SNR and spatial SNR. The method proposed in the present paper (in 3.2) and further developed in [11], involves another criterion that predicts the quality of sound field reconstruction.

2.3.2. Encoding of virtual source "inside" the loudspeakers enclosure

For virtual source encoding and rendering, NFC-HOA as previously introduced is very effective for sources that are beyond the loudspeaker array, but problems progressively occur with sources "inside" the loudspeaker enclosure. As a matter of fact, even if encoded ambisonic signals of great amplitude can be digitally conveyed using *e.g.* with floating point representation, the later diffusion of decoded loudspeakers signals become prohibitive regarding electro-acoustic considerations. Their amplitude is indeed given by the pan-pot law (for an N-loudspeaker regular and circular array):

$$G^{\text{NFC}(R,c)}(\rho, \gamma, \omega) = \frac{1}{N} \left(1 + 2 \sum_{m=1}^M H_m^{\text{NFC}(\rho/c, R/c)}(\omega) \cdot \cos(m\gamma) \right) \quad (10)$$

where γ is the angle between virtual source and each considered loudspeaker. It shows how the bass-boost of HOA components (see top curves of Figure 4) is reported in loudspeakers signals.

Figure 5 (time domain simulation) shows that huge amplitude interfering waves are present between the disk excluding the virtual source and the loudspeaker array. This excessive energy disappears as loudspeakers waves have constructively combined to each others to recompose expected spherical wave.

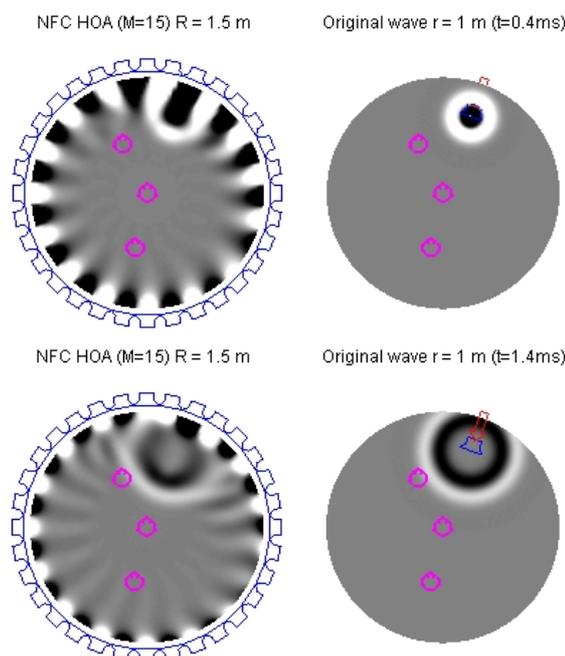


Figure 5 Two snapshots (time domain) of spherical wave synthesis with NFC-HOA for an enclosed virtual source ($r=1\text{m} < R=1.5\text{m}$). Left: 15th order rendering over a 32-loudspeaker circular array. Right: reference wave from a single source.

3. A SAFE ENCODING STRATEGY

Problems of excessive amplification discussed above arise from seeking for a sound field representation that is strictly equivalent to mathematical definition (1) of spherical harmonics decomposition. By "strictly" we mean: "over the full frequency scale". In this part of the paper, we further discuss the usefulness of spatial (HOA) components as a function of their respective order and the frequency, considering the size of a

targeted reconstruction area. Then we derive solutions for workable microphone array processing and virtual source encoding. Particularly interesting results arise for the simulation of sources inside the loudspeaker array enclosure, and even inside the listening area.

3.1. Characterizing representation usefulness

3.1.1. Quality of sound field approximation for a given radial extent

To address the quality of the sound field reconstruction as a function of the representation order M ("spatial resolution"), we introduce separately "2D" and "3D" approximations, as being representative to what a horizontal (*e.g.* circular) or respectively a 3D (*e.g.* spherical) loudspeaker array would achieve. For simplicity, we consider that such sound field approximation is sensibly expressed¹ by respectively a cylindrical (11) or a spherical (12) harmonics expansion series truncated to order M . Applied to both cases of a plane wave (2) and a spherical wave (3), this yields:

$$\hat{p}_M^{(2D)}(\vec{r}) = S \left(J_0(kr) + 2 \sum_{m=1}^M j^m J_m(kr) F_m(k\rho) \cos m\gamma \right) \quad (11)$$

$$\hat{p}_M^{(3D)}(\vec{r}) = S \sum_{m=0}^M (2m+1) j^m j_m(kr) F_m(k\rho) P_m(\cos \gamma) \quad (12)$$

where γ is the angle between the measurement point $\vec{r} = r\vec{u}_r$ and the wave incidence \vec{u}_s ($\cos \gamma = \vec{u}_s \cdot \vec{u}_r$). For a plane wave, one set the source distance to $\rho = \infty$, thus $F_m(k\rho) = 1$ for any k and m .

For further comments and illustrations, let's concentrate on 2D approximation (11), which suits to the most usual application cases. The series involves radial dependent terms $J_m(kr)$ (for plane wave) or $J_m(kr)F_m(k\rho)$ (spherical wave) that reflect the contribution of the m^{th} order group of spatial components $\{B_{mm}^\sigma\}$ to the sound field. These contributions are shown as functions of the radius r : in Figure 6 for a relatively low frequency $f = 100$ Hz, and in Figure 7 for a higher frequency $f = 500$ Hz.

¹ Equivalence is shown when loudspeakers are far field and numerous compared with the number of spatial components ($K=2M+1$ for 2D and $K=(M+1)^2$ for 3D).

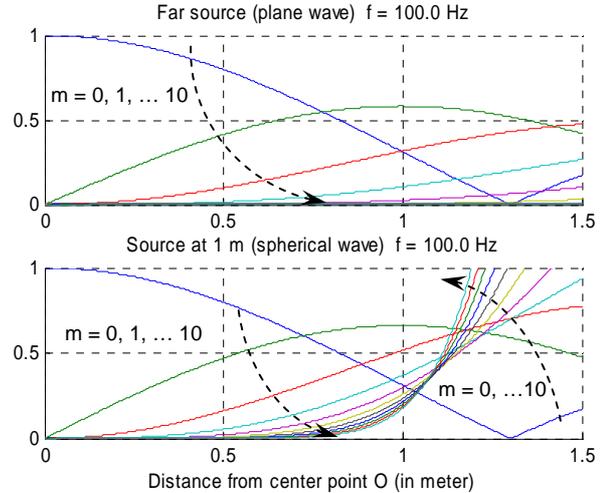


Figure 6 Radial contribution of spatial components B_{mm}^σ of orders $m=0$ to 10 and for a low frequency $f=100$ Hz in the cases of a plane wave (top) and a spherical wave (bottom, with $\rho=1$ m), as quantified by functions $|J_m(kr)|$ and $|J_m(kr)F_m(k\rho)|$ respectively.

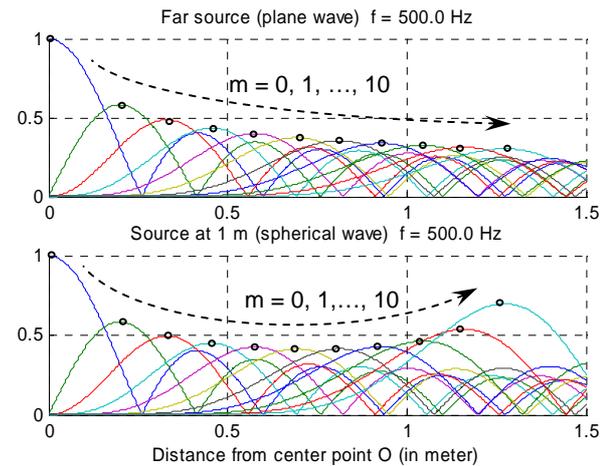


Figure 7 Same as Figure 6 but for a frequency $f=500$ Hz

3.1.2. Considerations on the plane wave case

A general property arises by considering $J_m(kr)$ curves (or even $j_m(kr)$ curves of Figure 1): it says that spatial components begin to substantially contribute to the sound field at the distance r (from O) that is as high as the order m is high for a given frequency, this distance r being furthermore proportional to the wavelength (and therefore as high as the frequency is low). Turned in another way: if one targets a reconstruction area

characterized by a radius (e.g. $r_{\text{target}}=1\text{m}$), it can be deduced that higher order components are significantly useful to the reconstruction only above a certain frequency that increases with the component order m .

This is particularly well shown in the case of a plane wave (top of figures): contributions are significant over a 1 meter radius area up to order 2 or 3 for the low frequency $f=100\text{ Hz}$ (Figure 6), and up to 9 or 10 for the higher frequency $f=500\text{ Hz}$ (Figure 7).

3.1.3. Special considerations on spherical wave

The spherical wave case (bottom of figures, with a source distance $\rho=1\text{m}$) gives rise to special comments. What is noteworthy is that each shown contribution $J_m(kr)F_m(k\rho)$ (for $m\geq 1$) is the result of two antagonist trends: when frequency decreases and/or the order m increases, $J_m(kr)$ gets smaller values over a given radius range (and "peaks" move to farther distances) while the overall contribution is increased by the near field effect $F_m(k\rho)$. A balance point seems to settle around the source distance ρ where curves seems to converge to a same amplitude for high orders m (see bottom of Figure 6). At the left of this balance point (smaller distances) amplitude of contributions is bounded, whereas it becomes huge (divergent) on the right part (farther distances), which constitutes an "energetic barrier", in a way.

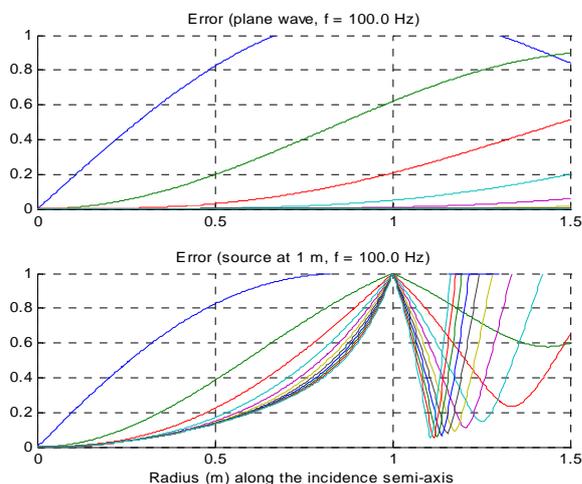


Figure 8 Relative error of sound field reconstruction along the semi-axis oriented towards the wave incidence ($\gamma=0$), in conditions similar to Figure 6.

As a consequence, reconstruction error (bottom of Figure 8 and Figure 9) is always infinite or unacceptable for radius $r\geq\rho$ whatever the truncation order M , since the accumulation of missing higher order contributions diverge. Therefore true reconstruction cannot go beyond nor even reach the source distance, what was expected from the theory: it's indeed both a physical impossibility and a limit of representation (1) which is valid only in a centered free field sphere (thus excluding any sound source). Furthermore, pressure field should be theoretically infinite at the source place.

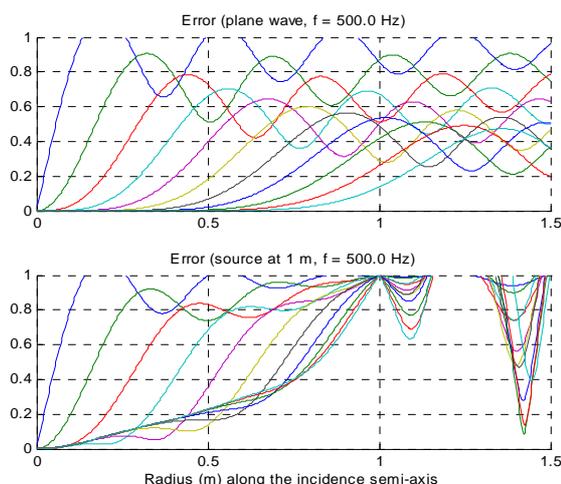


Figure 9 Same as Figure 8 for conditions similar to Figure 7 ($f=500\text{Hz}$)

To summarize, radial expansion of the sound field reconstruction is bounded by the source radius ρ (bottom of Figure 8 and Figure 9), unlike with a plane wave (top of Figure 8 and Figure 9). Thus the target "valid reconstruction area" is necessarily limited by the source distance. Introducing higher order components still improves the reconstruction over this targeted area, but not as efficiently as for a plane wave. The same kind of conclusion applies anyway: for a given target area, the highest order of significantly contributing (*i.e.* "useful") spatial components decreases when the frequency decreases.

Though the present discussion on spherical waves relies on 2D sound field approximation or reconstruction, its conclusions apply to 3D approximation as well. It's worth mentioning that reconstruction is even better (computed error is smaller) with a 3D approximation, though being similarly bounded by the source radius.

Indeed the theory foresees that a 2D array cannot properly render the $1/|\bar{r}-\bar{\rho}|$ amplitude law that is characteristics to spherical waves, but rather a $1/\sqrt{|\bar{r}-\bar{\rho}|}$ law. By the way, this is the reason for the so-called "stationary phase approximation" applied for Wave Field Synthesis [5, 13]. On the other hand, later Figure 13 shows that 2D reconstruction is able to properly render the wave front curvature (phase characteristics).

3.1.4. A safe representation: the concept of "useful high frequency bands"

The discussion above can be summarized as follows: according to a target radius for the reconstruction area, spatial (HOA) components of orders $m \geq 1$ appear to be useless in a low frequency band that increases with increasing orders m , and also with decreasing target radius. Therefore one can avoid excessive amplification effects (mentioned in 2.3) that occur in these low frequency bands when processing near field source simulation or natural sound field recording (with a microphone array) as described in previous literature [7, 8].

Typical useful high frequency bands of HOA components are shown in Figure 10. Thin leaning line draws a typical frequency limit law as a function of the order m and depending on the target radius r_{target} .

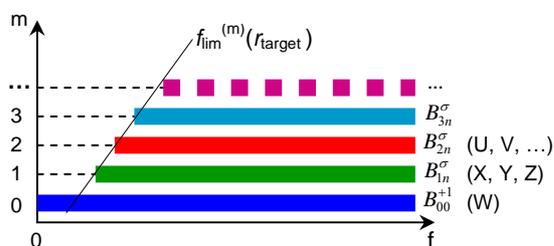


Figure 10 Schematic view of "useful" high frequency bands for the first few groups (colored horizontal bars) of spatial components (mentioned on the right part) according to their order m .

3.1.5. Frequency dependent truncation by high-passing HOA components

So we come to the following strategy for the design of safe coding tools, that is: applying a high-pass filtering to HOA components as suggested by Figure 10, which

is similar to applying a frequency dependent truncation of the spatial representation order.

For now, we can give some first requirements on the low frequency slope of high-pass filters. In cases (see 2.3) where NFC doesn't prevent from excessive bass-boost, H_m^{NFC} (or EQ_m) curves (Figure 3, Figure 4) present a $-m \times 6\text{dB/oct}$ slope between the boosted low frequency band and the high frequency band where amplification is moderated or null. To efficiently limit the bass-boost, high-pass filters shall stop this slope, thus present themselves at least an $m \times 6\text{dB/oct}$ slope with a sharp transition.

Following sections further reveal appropriate frequency limits definition with respect to each application case: microphone array processing for natural spatial sound field recording, and especially virtual source encoding with consideration to sources inside the listening area.

3.2. Application to sound field recording

In the context of recording system design, there is no prior knowledge on the sound field composition, especially about the presence of near field sources or not. Since we consider the problem of relatively small microphone arrays, we will also consider that most of real sound sources that create the sound field are in far field. Therefore we assume that measured sound field is composed with plane waves and we rely on error curves on top of Figure 8 and Figure 9 to appreciate the quality of reconstruction as a function of the target area radius and of the frequency. Quadratic integrated error curves could be also used. For a target radius and a given component order m , a simple error threshold can be used to derive the frequency limit below which it becomes useless to accurately model higher order components. Then such frequencies have to be reported on equalization curves of Figure 3, which are representative to noise amplification too, apart from a level offset related to the number of capsules [8]. Therefore a maximum useful amplification is derived with respect to the targeted area radius. If such amplification is judged still excessive, ambitions of sound field estimation can be moderated by choosing a smaller target radius. Derived limit frequencies are characteristic to "useful frequency bands" as shown by Figure 10, and used as parameters of high-pass filters as introduced in 3.1.5. This strategy and its results are further developed in [11].

3.3. Application to virtual source encoding

3.3.1. Toward a relevant criterion for truncation

For the case of simulating a finite distance source, one recalls that the possible sound field reconstruction is bounded by its distance ρ . Therefore the target radius should be fixed a bit lower than ρ : $r_{\text{target}} = \alpha \cdot \rho$ with $\alpha < 1$. But to define frequency limits of HOA components usefulness, it is no longer relevant to rely on a maximal tolerated error criterion at distance r_{target} , as used in 3.2. Indeed fixing r_{target} is somewhat arbitrary and most of all, a small radius change may imply great error changes. Moreover there is an irreducible error with 2D approximation, related to the badly rendered amplitude law discussed in 3.1.3.

A more relevant criterion is found by taking care of sources simulated "inside" the loudspeaker array (at distance $\rho < R$) and possibly inside the listening area. The idea is to consider not only the correct reconstruction area (bounded by the source distance) but also sound field characteristics over the remaining area (beyond the source distance). Bottom of Figure 6 and Figure 8 shows that at a relatively low frequency, spatial components of relatively high order ($m > 2$, here) yields little improvement to the reconstruction at radius $r \leq \rho$, and on the other hand, have unfortunate consequences regarding sound field properties at radius $r > \rho$: they cause great energy levels that sound annoying and unrealistic to listeners in this area. In a few words, one has to find a good compromise between reconstruction improvement below the source distance and acceptable level properties of the sound field composed beyond.

3.3.2. Energy focalization on the virtual source location

To refine our criterion, a first approach is to point at the frequency for which spatial components of a given order m have their higher "usefulness" (or contribution) on the source location. Top of Figure 11 shows $J_m(kr) \cdot F_m(k\rho)$ curves for $\rho = 1\text{m}$ (like bottom of Figure 6 and Figure 7) and for the frequency ($f = 462.2\text{ Hz}$) such that $J_m(kr) \cdot F_m(k\rho)$ (or merely $J_m(kr)$) has its maximum at $r = \rho$ for $m = 7$. Bottom shows the absolute value of M^{th} order truncated series (11) (*i.e.* sound field "2D" approximations) on semi-axis (O, \vec{u}_ρ) passing through the virtual source ($\gamma = 0$). Incidentally, we observe that for $M = m + 1 = 8$ the curve reaches its maximum value at the virtual source location $r = \rho$.

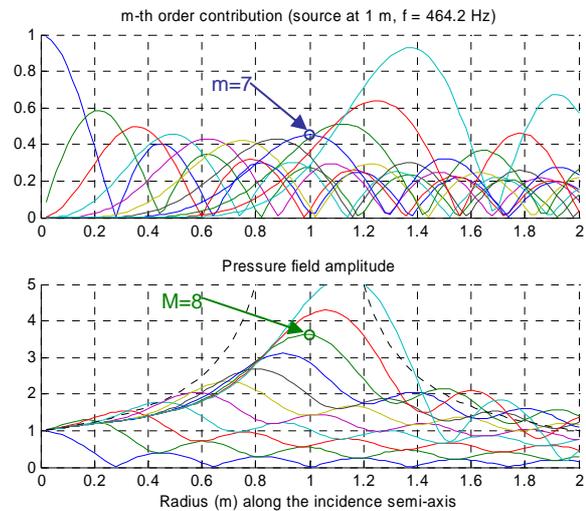


Figure 11 Top: radial contribution $J_m(kr) \cdot F_m(k\rho)$ of spatial components of various orders m for a source at distance ρ . Bottom: associated sound field approximations (truncated series in absolute value) on the semi-axis passing through the virtual source. Dotted curve indicates $\rho/|r-\rho|$ amplitude law as a reference (case of a true spherical wave).

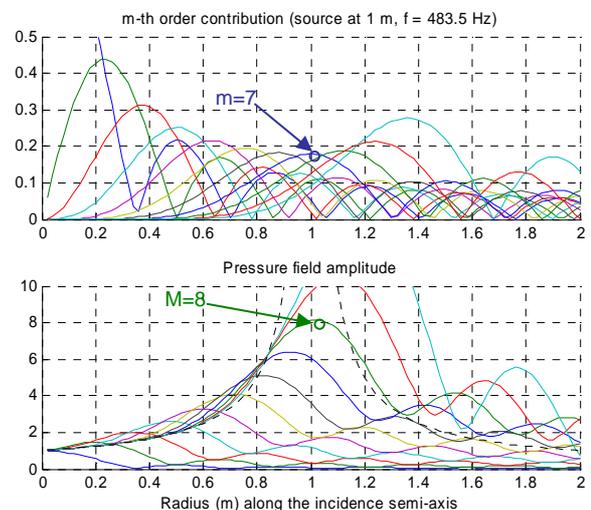


Figure 12 Same comments as for Figure 11, but with spherical Bessel functions and "3D" sound field approximation.

Similar observation applies for a "3D" approximation of the sound field, involving spherical Bessel functions $j_m(kr)$ and truncated series (12) with $\gamma=0$ (Figure 12). It is worth noticing that reproduced amplitude law looks like the theoretical one (dotted curve) much better than with "2D approximation" (Figure 11). Moreover, the amplitude peak observed at the virtual source location (for $M=8$ on the figure) is higher.

In both 2D and 3D cases, an interesting feature is that amplitude of sound field approximation (for the same truncation order $M=8$ in the figures) falls down to plausible values, though in a non-monotonic way, when moving beyond (thus away from) the virtual source. That's a property that should have appreciable auditory effects on listeners placed in such off-centered area.

Figure 13 provides 2D visualizations of approximated and synthesized sound fields. A color scale is used for amplitude field (see bottom views as a reference) while wave front are drawn as constant-phase lines, with an inter-line propagation time that corresponds to a 20cm distance for a natural wave. It confirms that focalization on virtual source (at distance $\rho=0.75\text{m}$) occurs with a $M=8^{\text{th}}$ order restriction at the expected frequency $f_{\text{lim}}^{(M=8)} = 619 \text{ Hz}$, while wave front curvature looks quite acceptable over the centered disk just excluding the source. Wave front shape is better refined using higher orders (*e.g.* $M=12$), but it is at the expense of a huge energy field at radius greater than the source distance ρ .

Finally, one notices a difference in amplitude field when comparing series truncation (left) with reconstruction by the loudspeaker array (right): for the latter, focalization spot still includes the virtual source but spreads out towards the nearest loudspeakers.

Now we have found a particularly relevant criterion, that is: focussing amplitude peak on the virtual source location for all possible frequencies. On Figure 13, the second pair of subplots from the top might be considered as optimal in the sense of this criterion.

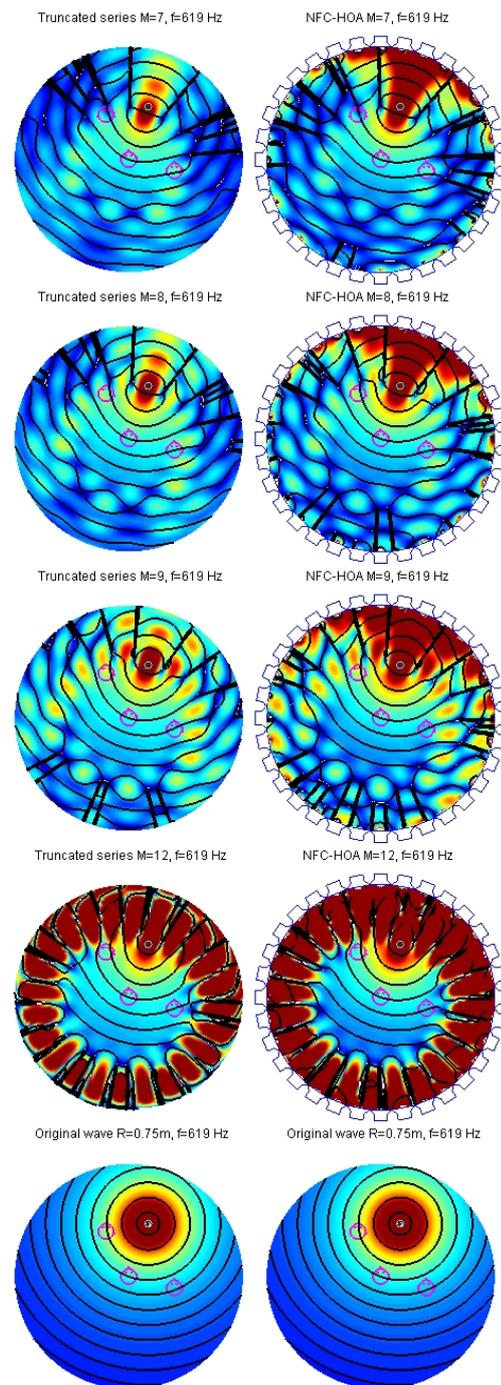


Figure 13 Top view of 2D sound field approximations (truncated Fourier-Bessel series (11): left) and reconstructions (by a 32-loudspeaker array: right), for a source simulated at 75cm from the center (bottom views). Orders $M=7, 8, 9, 12$ from top to bottom.

3.3.3. Refining high-pass filters specifications

Our "focalization" criterion is fulfilled by truncating ambisonics representation order according to the following principle: at each frequency $f_{\text{lim}}^{(M)}$ such that M^{th} order truncated series reaches its maximum on the source location, preserve all spatial components of order $m \leq M$ and attenuate (or cancel) higher order ones ($m \geq M+1$). This can be reinterpreted as follows: on each component of order m , apply a high-pass filter $H^{(m)}$ ² which pass-band low frequency limit is $f_{\text{lim}}^{(m)}$ and which rejection frequency is $f_{\text{lim}}^{(m-1)}$. An essential requirement is that high-pass filters have a null phase response in their respective pass-band. Figure 14 shows how amplitude spectra of such filters may look like.

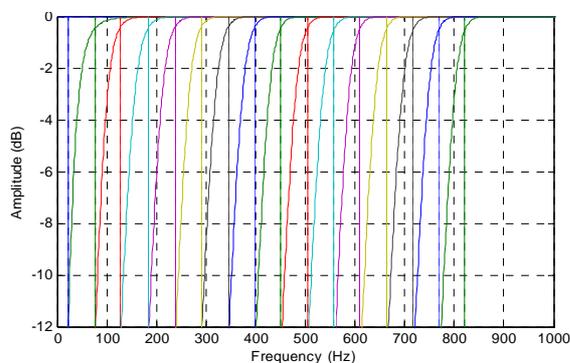


Figure 14 Spectra of high-pass filters $H^{(m)}$ designed to limit functions $F_m(k\rho)$ for orders $m=1$ to 15 (from left to right) and a source distance $\rho=1\text{m}$. For each curve (of $H^{(m)}$), pass-band frequency limit is marked by a vertical bar (with the same color as the curve). The same bar marks rejection frequency of higher order filter $H^{(m+1)}$.

Before going farther, let's further define frequencies $f_{\text{lim}}^{(M)}$. Actually, the fact $f_{\text{lim}}^{(M)}$ is also such that function $J_{m=M-1}(2\pi f_{\text{lim}}^{(M)} r/c)$ is maximal for $r=\rho$, is fairly fortuitous and is not so exact for arbitrary orders M . Therefore, defining frequency limits $f_{\text{lim}}^{(M)}$ requires a bit trickier computational work. Some approximated results are reported in Table 1. An interesting property remains that $f_{\text{lim}}^{(M)}$ values are inversely proportional to source distance ρ and derive from generic values $kr_{\text{lim}}^{(m)}$ as listed in Table 1, according to formula:

$$2\pi f_{\text{lim}}^{(m)} \rho / c = kr_{\text{lim}}^{(m)} \Rightarrow f_{\text{lim}}^{(m)} = kr_{\text{lim}}^{(m)} \cdot c / (2\pi\rho) \quad (13)$$

² Do not mix up with NFC filters $H_m^{\text{NFC}}!!!$

M (or m)	1	2	3	4	5
kr ($p_M^{(2D)}$ max)	1.405	2.355	3.417	4.429	5.393
kr ($p_M^{(3D)}$ max)	1.699	2.836	3.919	4.981	6.031
M (or m)	6	7	8	9	10
kr ($p_M^{(2D)}$ max)	6.404	7.409	8.357	9.347	10.345
kr ($p_M^{(3D)}$ max)	7.074	8.110	9.132	10.147	11.171

Table 1 kr -values related to frequencies where M^{th} order sound field approximation has their maximum at the place of the source ($r=\rho$).

Higher orders values can be derived by affine extrapolation of values of Table 1. That's why we shown frequency limits as an affine function of order m in Figure 10.

What's interesting now is to derive the "maximal useful amplification" of Near Field Functions $F_m(k\rho)$, i.e. their values at respective frequencies $f_{\text{lim}}^{(m)}$. Figure 15 shows that it is much moderated: it goes from about 2 dB ($m=1$) then slowly rises up to about 5dB for $m=14$ or 15. Moreover, it doesn't depend on source distance ρ anymore, since the latter causes the same frequency scale distortion on both $F_m(k\rho)$ and $f_{\text{lim}}^{(m)}$.

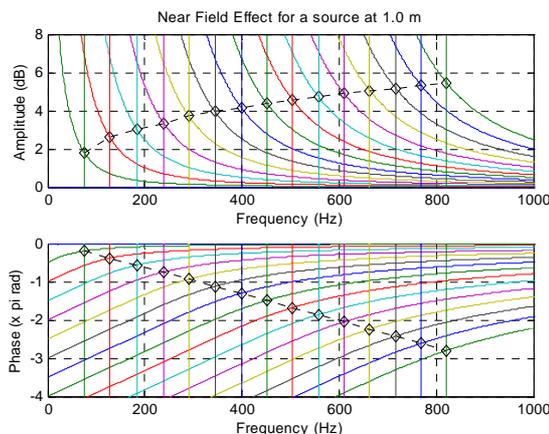


Figure 15 Amplitude and phase of Near Field Functions $F_m(k\rho)$ with $\rho=1\text{m}$ and for orders $m=1$ to 15 (from left to right). Vertical bars mark frequency limits $f_{\text{lim}}^{(m)}$.

For each order m , high-pass filter $H^{(m)}$ has to preserve both amplitude and phase of $F_m(k\rho)$ in the pass-band above $f_{\text{lim}}^{(m)}$; it has to ensure that rejection at lower frequency $f_{\text{lim}}^{(m-1)}$ is sufficient to compensate for the rise of $F_m(k\rho)$ in-between and furthermore to lower it to a negligible level; and finally its slope in the remaining low frequency band must be sufficient regarding the

slope of $F_m(k\rho)$. Filters that fulfill such requirements are at least of order $2m$ when implemented as IIR filters. It's worth highlighting that a simplification occurs in filter combination $H^{(m)}(z).F_m(z)$, since all poles of $F_m(z)$ are zeros of $H^{(m)}(z)$. Figure 16 shows spectra of such combined filters.

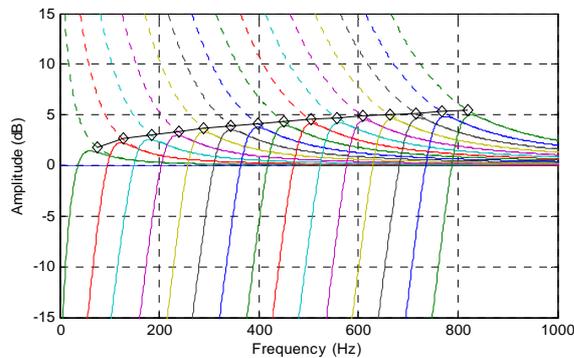


Figure 16 Spectra of high-passed Near Field functions $H^{(m)}(\omega).F_m(\omega)$ for orders $m=1$ to 15 (from left to right) and a source distance $\rho=1\text{m}$. Original Near Field functions $F_m(\omega)$ are shown in dotted line for comparison. Diamonds show for each order m the highest "useful" amplification (at frequency $f_{\text{lim}}^{(m)}$).

Note that no loudspeaker near field compensation (NFC) is involved at this stage, but NFC actually has to be performed, either at the encoding or decoding stage, in order to synthesize the wave front with its expected curvature. It lets possible to directly replace distance coding filters by $H^{(m)}.H_m^{\text{NFC}(\rho/c, R/c)}$ in encoding formula (8), or to apply encoding formula (5) with $H^{(m)}.F_m^{(\rho/c)}$ then apply NFC $1/F_m^{(R/c)}$ (6) in a later stage, *e.g.* the decoding stage. Anyway, final amplification will be lower than shown by Figure 16, especially if the virtual source is beyond the loudspeaker array.

3.4. Distributing bass rendering effort amongst loudspeakers

High-pass filters such as described in previous sections can find another advantageous use for processing an already composed sound field (even without distance coding) just before decoding and rendering it over a loudspeaker array. Indeed, since they restrict the spatial resolution at low frequencies, their effect is to distribute low frequency effort over a larger number of loudspeakers. This can be appreciated when loudspeakers are small and have a poor efficiency for low frequencies.

4. CONCLUSION

A previous study showed the ability of Higher Order Ambisonics to encode arbitrary sound fields, especially including near field effects, as long as a near field compensation (NFC-HOA) is introduced. NFC-HOA so limits the bass-boost due to near field sources or present in microphone array processing, but it cannot prevent from excessive bass-boost/levels when dealing with small source distance or microphone radius. The present paper introduced a variant and still efficient way of spatial coding that is totally safe with respect to signal processing and also electro-acoustic criteria.

For this purpose, it has been proposed to discard the low frequency part of spatial (ambisonic) components, where problematic bass-boost incidentally arises, as long this part is shown to provide few improvement of sound field reconstruction over a given targeted area. So the main idea of our variant coding strategy is to apply high-pass filters on HOA components, which also performs a frequency dependent truncation of the spatial resolution (or order) of the sound field representation. Frequency limits appear to roughly obey an affine law of the components order m , while being inversely proportional to the target area radius. Special results concern distance coding of virtual sources, especially when these are simulated in the reproduction area and even in the listening area. An additional criterion has been introduced, that is focussing synthetic energy field on the source point, which helps refining high-pass filters specifications and ensures a good compromise between accuracy of wave front reconstruction over the area below the source distance and sensible amplitude field properties beyond.

The design of such high-pass filters and their combination with near field modeling filters is being optimized. For now, their computational cost is higher than NFC filters described in [7]. Therefore, the latter encoding scheme still might be preferred for virtual sources that are known to be outside the loudspeaker array. Note that Near Field Compensation is still required for the correct rendering of wave front curvature. Now, it is not indispensable to introduce it at the encoding stage since high-pass filters themselves solve any bass-boost problem.

Listening tests are planned for subjective evaluation of spatial rendering using circular arrays with 48 loudspeakers and more.

5. REFERENCES

- [1] M.A. Gerzon, *Periphony : With-Height Sound Reproduction*. J. Audio Eng. Soc., 1973. vol. 21(1): p. 2-10.
- [2] J.S. Bamford, *An Analysis of Ambisonics Sound Systems of First and Second Order*. 1995, University of Waterloo: Waterloo, Ont., Canada.
- [3] J. Daniel, J.-B. Rault, and J.-D. Polack. *Ambisonic Encoding of Other Audio Formats for Multiple Listening Conditions*. in *AES 105th Convention*. 1998. San-Francisco, USA.
- [4] D. Malham. *Higher order Ambisonic systems for the spatialisation of sound*. in *ICMC99*. 1999. Beijing.
- [5] R. Nicol and M. Emerit. *3D-Sound Reproduction over an Extensive Listening Area: A Hybrid Method Derived from Holophony and Ambisonic*. in *AES 16th Int. Conference on Spatial Sound Reproduction*. 1999. Rovaniemi, Finland.
- [6] J. Daniel, *Représentation de Champs Acoustiques, Application à la Transmission et à la Reproduction de Scènes Sonores Complexes dans un Contexte Multimédia*. 2000, University of Paris 6: Paris, France. <http://gyronymo.free.fr/>
- [7] J. Daniel. *Spatial Sound Encoding Including Near Field Effect : Introducing Distance Coding Filters and a Viable, New Ambisonic Format*. in *AES 23rd International Conference*. 2003.
- [8] J. Daniel, R. Nicol, and S. Moreau. *Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging*. in *AES 114th Convention*. 2003. Amsterdam.
- [9] A. Laborie, R. Bruno, and S. Montoya. *A New Comprehensive Approach of Surround Sound Recording*. in *AES 114th Convention*. 2003. Amsterdam.
- [10] G.W. Elko, R.A. Kubli, and J.M. Meyer, *Audio system based on at least second-order eigenbeams*. 2003, US patent 0147539A1, MH Acoustics, LLC, a Delaware corporation: US.
- [11] S. Moreau and J. Daniel. *Study of Higher Order Ambisonic Microphone*. in *7ème Congrès Français d'Acoustique (Joint congress CFA-DAGA'04)*. 2004. Strasbourg, France. <http://gyronymo.free.fr/audio3D/publications>
- [12] J. Meyer and G. Elko. *A Highly Scalable Spherical Microphone Array Based on an Orthonormal Decomposition of the Soundfield*. in *IEEE ICASSP-02*. 2002. Orlando, Florida, USA.
- [13] R. Nicol, *Restitution Sonore Spatialisée sur une Zone Etendue: Application à la Téléprésence*. 1999, Université du Maine: Le Mans, France. http://gyronymo.free.fr/audio3D/Guests/RozenNicol_PhD.html