Normalization schemes in Ambisonic: does it matter?

Thibaut Carpentier

Audio Engineering Society

Convention Paper

Presented at the 142nd Convention
2017 May 20–23, Berlin, Germany

This paper was peer-reviewed as a complete manuscript for presentation at this convention.

Normalization schemes in Ambisonic: does it matter?

Thibaut Carpentier[1]

[1] *UMR 9912 STMS IRCAM–CNRS–UPMC, Paris, France*

Correspondence should be addressed to Thibaut Carpentier (thibaut.carpentier@ircam.fr)

ABSTRACT

In the context of Ambisonic processing, various normalizations of the spherical harmonic functions have been proposed in the literature and there is yet no consensus in the community about which one should be preferred (if any). This is a frequent source of confusion for the end users and this may lead to compatibility issues between rendering engines. This paper reviews the different conventions in use, presents an extension of the FuMa scheme to higher orders, and discusses possible pitfalls in the decoding stage.

1 Preliminary definitions

This section collects some definitions and conventions that will be needed later in the article. We expect the reader to be familiar with this material, so the results appear without proof (see e.g. [1] for further details).

1.1 Notations

δ_{ij}	the Kronecker symbol such as $\delta_{ij}=1$ if $i=j$, 0 otherwise
$\boldsymbol{\vartheta}\equiv(\theta,\phi)$	angular direction in the spherical coordinate system with $x=\cos\phi\sin\theta$, $y=\sin\phi\sin\theta$, $z=\cos\theta$
$\|\mathbf{x}\|$	Euclidian norm of $\mathbf{x}$
$\mathbb{S}^2$	unit sphere i.e. $\{\mathbf{x}\in\mathbb{R}^3:\|\mathbf{x}\|=1\}$
K	number of loudspeakers
N	the Ambisonic order
χ_N	number of Ambisonic components i.e. $\chi_N=(2N+1)$ in 2-D, $\chi_N=(N+1)^2$ in 3-D
$\mathbb{I}_N$	the Ambisonic indices i.e. $\{(n,m)\in\mathbb{N}\times\mathbb{Z}:0\leq\|m\|\leq n\leq N\}$
$\binom{n}{k}=\frac{n!}{k!(n-k)!}$	binomial coefficient with $n\geq k\geq 0$

1.2 Real-valued spherical harmonics

The Legendre polynomials $P_n(x)$ are solutions to Legendre's differential equation and can be expressed with a Rodrigues representation $\forall x\in\mathbb{R}\cap[-1,1]$, $\forall n\in\mathbb{N}$

$$P_n(x)=\frac{1}{2^n n!}\frac{d^n}{dx^n}\left(x^2-1\right)^n. \tag{1}$$

The associated Legendre functions $P_n^m(x)$ are defined $\forall x\in\mathbb{R}\cap[-1,1]$, $\left\{(n,m)\in\mathbb{N}^2:m\leq n\right\}$ by

$$P_n^m(x)=(-1)^m\left(1-x^2\right)^{\frac{m}{2}}\frac{d^m}{dx^m}P_n(x). \tag{2}$$

Note that the Condon and Shortley phase factor $(-1)^m$ is sometimes omitted.

Spherical harmonics, which are the solutions of Laplace's equation in spherical coordinates, are an infinite set of harmonic functions defined over the surface of a sphere. A real-valued set of spherical harmonics, $\mathrm{Y}_n^m(\boldsymbol{\vartheta})$, can be defined $\forall(n,m)\in\mathbb{I}_N$ by

$$\mathrm{Y}_n^m(\boldsymbol{\vartheta})=\mathscr{A}_n^{|m|}P_n^{|m|}(\cos\theta)\begin{cases}\cos(m\phi), & \text{for } m\geq 0\\ \sin(|m|\phi), & \text{for } m<0\end{cases} \tag{3}$$

where $\mathscr{A}_n^{|m|}$ is a scalar normalization constant.

2 Normalization schemes

One source of confusion in the context of Ambisonics is that not all authors use the same normalization constant. In the remainder of this paper, we will notate $(\mathscr{A}_n^m)^{\mathscr{N}}$ the normalization factor with respect to the convention $\mathscr{N}$. Daniel [2] provided a comprehensive review of the most frequently used normalization conventions for Ambisonic applications. This section recapitulates the definitions of these schemes.

The fully-normalized (or 4π-normalized) scheme, noted N3D, is given for $\left\{(n,m) \in \mathbb{N}^2 : m \leq n\right\}$ by

$$(\mathscr{A}_n^m)^{\text{N3D}} = \begin{cases} \sqrt{2n+1}, & \text{if } m = 0 \\ (-1)^m \sqrt{2\,(2n+1)\,\frac{(n-m)!}{(n+m)!}}, & \text{if } m \neq 0 \end{cases} \tag{4}$$

N3D is the chosen normalization in various Ambisonic standards (e.g. [3, 4]), however their adoption by the community is yet limited and other schemes are widely used in practice.

When considering horizontal-only harmonics, the fully-normalized factors are expressed $\forall n \in \mathbb{N}$ by

$$(\mathscr{A}_n^n)^{\text{N2D}} = \begin{cases} 1, & \text{if } n = 0 \\ (-1)^n \sqrt{\frac{2^{2n+1}\,n!^2}{(2n)!^2}}, & \text{if } n \neq 0 \end{cases} \tag{5}$$

Alternatively, the Schmidt semi-normalized (SN) convention can be employed, applying the normalization factors

$$(\mathscr{A}_n^m)^{\text{SN3D}} = \begin{cases} 1, & \text{if } m = 0 \\ (-1)^m \sqrt{2\,\frac{(n-m)!}{(n+m)!}}, & \text{if } m \neq 0 \end{cases} \tag{6}$$

for $\left\{(n,m) \in \mathbb{N}^2 : m \leq n\right\}$.

Similarly, the normalization factors for the 2-D Schmidt semi-normalization writes $\forall n \in \mathbb{N}$

$$(\mathscr{A}_n^n)^{\text{SN2D}} = (-1)^n \sqrt{\frac{2^{2n}\,n!^2}{(2n)!^2}} \tag{7}$$

The popularity of the SN3D scheme has soared in recent years with its integration in several virtual reality environments, following the ambiX format proposal [5].

Yet another convention is the Max-normalization (MaxN), defined such that, $\forall (n,m) \in \mathbb{I}_N$

$$\max_{\boldsymbol{\vartheta}} |\mathrm{Y}_n^m(\boldsymbol{\vartheta})| = 1 \tag{8}$$

It has been shown by Daniel [2] that, in the 2-D case, MaxN is equivalent to SN2D i.e. $\forall n \in \mathbb{N}$

$$(\mathscr{A}_n^n)^{\text{MaxN2D}} = (\mathscr{A}_n^n)^{\text{SN2D}}. \tag{9}$$

It is important to note that the same MaxN definition can be applied to both 2-D and 3-D spherical harmonics. For improved readability, the notation MaxN2D and MaxN3D will be used in the remainder of this article for the 2-D and the 3-D case, respectively.

The MaxN type has been followed in the Furse-Malham (FuMa) scheme [6], sometimes considered a de facto standard for legacy and first order materials. FuMa only differs from MaxN with the inclusion of a $1/\sqrt{2}$ weighting of the sound pressure component, i.e. for $\left\{(n,m) \in \mathbb{N}^2 : m \leq n\right\}$:

$$(\mathscr{A}_n^m)^{\text{FuMa}} = \begin{cases} \frac{1}{\sqrt{2}}, & \text{if } n = 0 \\ (\mathscr{A}_n^m)^{\text{MaxN}}, & \text{if } n > 0 \end{cases}. \tag{10}$$

Such as the MaxN factors, the FuMa normalization can be applied to both 2-D and 3-D spherical harmonics. Furse and Malham [6, 7] have published explicit formulae of their normalization factors up to the order of $N = 3$. The following section provides insights for extension to higher orders.

2.1 Derivation of MaxN factors for $N > 3$

As expressed in equation (8), the MaxN convention is such that each spherical harmonic component reaches a maximum amplitude of 1 on $\mathbb{S}^2$. The derivation of the $(\mathscr{A}_n^m)^{\text{MaxN}}$ factors thus consists in determining the maximum value of $P_n^{|m|}(x)$ with respect to x. To the best of this author's knowledge, there is yet no closed-form nor recursive solution to this problem. Nevertheless, we will derive here the explicit formulae for some of the factors.

First, using well-known properties of the associated Legendre functions [1], it is easy to show that $\forall n \in \mathbb{N}$,

$$\left(\mathscr{A}_n^0\right)^{\text{MaxN}} = 1 \;\text{ and }\; (\mathscr{A}_n^n)^{\text{MaxN}} = (-1)^n \frac{2^n\,n!}{(2n)!} \tag{11}$$

The procedure to derive the other terms (i.e. $0 < m < n$)

consists in computing the derivative of the Legendre function $\mathscr{D}_n^m(x) = \frac{d}{dx}P_n^m(x)$, finding the roots $\{\beta_n^m\}$ such that $\mathscr{D}_n^m(x=\beta_n^m)=0$ and finally evaluate:

$$(\mathscr{A}_n^m)^{\text{MaxN}} = \frac{(-1)^m}{\left|P_n^{|m|}(\beta)\right|} \tag{12}$$

where $\beta = \max\{\beta_n^m\}$ is the location of the maxima.

Using the binomial expansion of the Legendre polynomials:

$$P_n(x) = 2^n \sum_{k=0}^{n} x^k \binom{n}{k} \binom{\frac{n+k-1}{2}}{n} \tag{13}$$

the m^{th} derivative has the following expression:

$$\frac{d^m P_n(x)}{dx^m} = 2^n n! \sum_{k=m}^{n} \frac{x^{k-m}}{(n-k)!\,(k-m)!} \binom{\frac{n+k-1}{2}}{n}. \tag{14}$$

Writing

$$U_n^m(x) = n! \sum_{k=m}^{n} \frac{x^{k-m}}{(n-k)!\,(k-m)!} \binom{\frac{n+k-1}{2}}{n} \tag{15}$$

it can be shown that

$$\frac{dP_n^m(x)}{dx} = 0 \Longleftrightarrow R_n^m(x) = 0 \tag{16}$$

where we introduced the auxiliary polynomial $R_n^m(x)$:

$$R_n^m(x) = mxU_n^m(x) - (1-x^2)U_n^{m+1}(x). \tag{17}$$

The largest root of the $R_n^m(x)$ polynomial is β. Unfortunately, there is no easy way to analytically express the roots of $R_n^m(x)$, and no recurrence formula appear. A closed-form solution can however be expressed for some of the terms. Noting that

$$U_n^n(x) = n! \binom{\frac{2n-1}{2}}{n} \quad \text{and} \quad U_n^{n-1}(x) = xU_n^n(x) \tag{18}$$

we can solve equation (17) for $m = n-1$, leading to

$$x = \frac{1}{\sqrt{n}} = \beta_n^{n-1}. \tag{19}$$

Consequently we obtain $\forall n \geq 2$

$$\left(\mathscr{A}_n^{n-1}\right)^{\text{MaxN}} = \frac{(-1)^{n-1}}{\left|P_n^{n-1}\left(\beta_n^{n-1}\right)\right|} = \frac{(-1)^{n-1}\,2^n\,n!\,\sqrt{n}}{(1-\frac{1}{n})^{\frac{n-1}{2}}(2n)!} \tag{20}$$

Similarly, for the case $m = n-2$, we have:

$$\begin{aligned} U_n^{n-2}(x) &= \tfrac{x}{2!}U_n^{n-1}(x) + \tfrac{n!}{2!}\left(\tfrac{2n-3}{n}\right) \\ &= \tfrac{x^2}{2!}U_n^n(x) + \tfrac{n!}{2!}\left(\tfrac{2n-3}{n}\right) \end{aligned} \tag{21}$$

which, after some simplifications, leads to:

$$x = \beta_n^{n-2} = \sqrt{\frac{5n-4}{n(2n-1)}} \tag{22}$$

and thus $\forall n > 2$,

$$\left(\mathscr{A}_n^{n-2}\right)^{\text{MaxN}} = \frac{(-1)^n}{\left|P_n^{n-2}\left(\beta_n^{n-2}\right)\right|}. \tag{23}$$

The other terms are rather difficult to express. We have thus used a symbolic mathematical computation program (Maple®) to derive $R_n^m(x)$, $\beta_n^m(x)$ and $(\mathscr{A}_n^m)^{\text{MaxN}}$, and the results (up to the order $N=6$) are presented in Annex 1.

However, the estimation of $(\mathscr{A}_n^m)^{\text{MaxN}}$ with this approach yields poor numerical precision for higher orders. A numerically more robust computation first calculates the N3D factors using a set of stable recurrence relations [8] and then applies the corresponding conversion coefficients from N3D to MaxN3D, which will be discussed in the following section.

2.2 Conversion between normalizations

It is clear from the previous definitions that various normalization types only differ from an order-dependent scaling factor. The conversion factor from norm $\mathscr{N}_1$ to $\mathscr{N}_2$ is notated $(\alpha_n^m)^{\mathscr{N}_1\to\mathscr{N}_2}$ and defined such as:

$$\forall (n,m) \in \mathbb{I}_N,\ (\mathscr{A}_n^m)^{\mathscr{N}_1}\,(\alpha_n^m)^{\mathscr{N}_1\to\mathscr{N}_2} = (\mathscr{A}_n^m)^{\mathscr{N}_2} \tag{24}$$

Conversion factors satisfy the reciprocity and transitivity properties i.e.

$$(\alpha_n^m)^{\mathscr{N}_1\to\mathscr{N}_2} = 1/(\alpha_n^m)^{\mathscr{N}_2\to\mathscr{N}_1} \tag{25}$$

$$(\alpha_n^m)^{\mathscr{N}_1\to\mathscr{N}_3} = (\alpha_n^m)^{\mathscr{N}_1\to\mathscr{N}_2}\cdot(\alpha_n^m)^{\mathscr{N}_2\to\mathscr{N}_3} \tag{26}$$

Daniel [2] provided a review of conversion factors, and some of them are recalled below:

$$\forall (n,m) \in \mathbb{I}_N,\quad (\alpha_n^m)^{\text{SN3D}\to\text{N3D}} = \sqrt{2n+1} \tag{27}$$

$$\forall n \in \mathbb{N},\ \left(\alpha_n^{\pm n}\right)^{\text{SN2D}\to\text{N2D}} = \begin{cases} \sqrt{2} & \text{if } n \neq 0 \\ 1 & \text{if } n = 0 \end{cases} \tag{28}$$

In the 2-D case, MaxN and SN2D are equivalent, while FuMa and N2D only differ from a constant scaling:

$$\forall n \in \mathbb{N}, \quad \left(\alpha_n^{\pm n}\right)^{\text{MaxN2D}\to\text{SN2D}} = 1 \tag{29}$$

$$\forall n \in \mathbb{N}, \quad \left(\alpha_n^{\pm n}\right)^{\text{FuMa2D}\to\text{N2D}} = \sqrt{2} \tag{30}$$

Furthermore, using equations (4), (8) and (24) it was possible to numerically evaluate the conversion factors from N3D to MaxN and results are given in Annex 1 up to the order $N = 16$. The computed factors are numerically well balanced and they allow to implement the MaxN and FuMa schemes (the conversion between MaxN and FuMa being trivial) for higher orders, indeed extending data previously published by Furse and Malham for $N \leq 3$.

3 Properties

3.1 Orthonormality

Considering the Hilbert space $L^2\left(\mathbb{S}^2\right)$ under the inner product

$$\left\langle f, g \right\rangle = \frac{1}{4\pi} \int_{\mathbb{S}^2} f\left(\boldsymbol{\vartheta}\right) g^*\left(\boldsymbol{\vartheta}\right) d\boldsymbol{\vartheta}, \tag{31}$$

it is known that real-valued spherical harmonics for the N3D scheme form an orthonormal basis i.e.

$$\left\langle \mathrm{Y}_n^m, \mathrm{Y}_{n'}^{m'} \right\rangle_{\text{N3D}} = \delta_{nn'}\,\delta_{mm'}. \tag{32}$$

With the Schmidt semi-normalization, on the other hand, the spherical harmonics are orthogonal [2]:

$$\left\langle \mathrm{Y}_n^m, \mathrm{Y}_{n'}^{m'} \right\rangle_{\text{SN3D}} = \frac{1}{2n+1}\,\delta_{nn'}\,\delta_{mm'}. \tag{33}$$

More generally, it is clear from equations (31), (32) and (24) that, for any convention $\mathscr{N}$, we have:

$$\left\langle \mathrm{Y}_n^m, \mathrm{Y}_{n'}^{m'} \right\rangle_{\mathscr{N}} = \left(\left(\alpha_n^m\right)^{\text{N3D}\to\mathscr{N}} \right)^2 \delta_{nn'}\,\delta_{mm'}. \tag{34}$$

As a consequence of this, there is no closed-form expression for the scalar product under the FuMa3D and MaxN3D schemes, yet we can state that these conventions do not satisfy the orthonormality property (they are only orthogonal).

In the case of 2-D horizontal-only Ambisonics, the (circular) scalar product is evaluated on a regular sampling of the unit circle $\mathbb{S}^1$ and the results are similar to the 3-D case i.e.

$$\left\langle \mathrm{Y}_n^m, \mathrm{Y}_{n'}^{m'} \right\rangle_{\text{N2D}} = \delta_{nn'}\,\delta_{mm'} \tag{35}$$

and for any other normalization scheme $\mathscr{N}$:

$$\left\langle \mathrm{Y}_n^m, \mathrm{Y}_{n'}^{m'} \right\rangle_{\mathscr{N}} = \left(\left(\alpha_n^m\right)^{\text{N2D}\to\mathscr{N}} \right)^2 \delta_{nn'}\,\delta_{mm'}. \tag{36}$$

In particular this leads to:

$$\begin{aligned} \left\langle \mathrm{Y}_n^m, \mathrm{Y}_{n'}^{m'} \right\rangle_{\text{SN2D}} &= \tfrac{1}{2}\,\delta_{nn'}\,\delta_{mm'} \\ &= \left\langle \mathrm{Y}_n^m, \mathrm{Y}_{n'}^{m'} \right\rangle_{\text{MaxN2D}} \\ &= \left\langle \mathrm{Y}_n^m, \mathrm{Y}_{n'}^{m'} \right\rangle_{\text{FuMa2D}} \end{aligned} \tag{37}$$

3.2 Range and practical considerations

From the equations given in section 2, it is evident that the different conventions yield very different dynamic ranges for the spherical harmonic functions. To illustrate that, the maximum absolute amplitude of the harmonic functions is plotted in Figure 1 for various conventions and for $N \leq 5$.

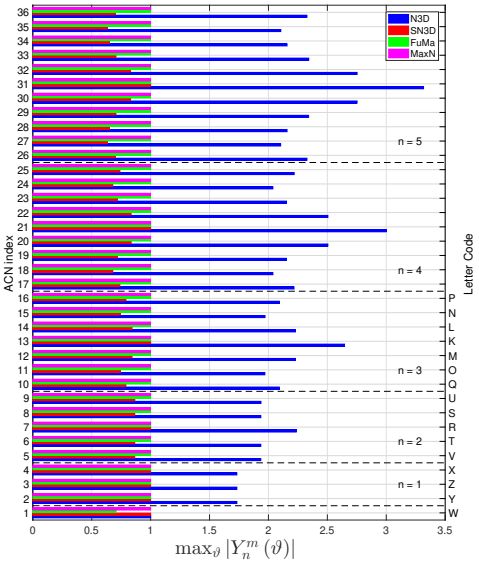

Fig. 1: Maximum amplitude of the spherical harmonics for various normalizations. On the y-axis, the Ambisonic components are sorted according to the Ambisonic Channel Number (ACN) scheme where: $\forall (n,m) \in \mathbb{I}_N,\ \text{ACN}(n,m) = n^2 + n + m + 1$.

By definition, MaxN and FuMa guarantee amplitude normalization i.e. they ensure that the crest level of the harmonics remains in the range [-1;1] when encoding a plane wave with amplitude lower than 1. Similarly, amplitude of components encoded with the SN3D or SN2D flavor will never exceed 1. Such amplitude normalizations might be favorable when storing Ambisonic stream to PCM audio files, as they guarantee that saturation above 0 dBFS will not occur. On the other hand, the N3D and N2D conventions provide energy normalization: when encoding an ideally diffuse sound field, all components exhibit the same RMS level, as already noted by Daniel [2].

When implementing fixed-point digital processing, one might prefer one convention over the other, for numerical accuracy. However, most Ambisonic processor nowadays use floating-point arithmetic, for which there is a priori no significant discrepancies across normalization conventions.
Regarding computational aspects, we have already mentioned that the full- and semi-normalized conventions have generic closed-form expressions and can be efficiently evaluated through simple and stable recursion algorithms. By contrast, there is no straightforward formulation of MaxN and FuMa for higher orders.

Besides these numerical considerations, there is at this point no significant benefit or drawback of one convention versus another. The next section will investigate the possible impact of the normalization scheme at the decoding stage.

4 Normalization schemes and Ambisonic decoding

This section builds on results previously established by Daniel et al. and further emphasizes the role of the chosen normalization $\mathscr{N}$ in the Ambisonic processing chain.

4.1 Sampling Ambisonic Decoder

A plane wave with incidence $\boldsymbol{\vartheta}$ is encoded in the Ambisonic domain by applying a vector $\boldsymbol{c}$ of real gains to the pressure signal S:

$$\boldsymbol{c}^{\mathscr{N}}(\boldsymbol{\vartheta}) = \begin{bmatrix} \mathrm{Y}_0^0(\boldsymbol{\vartheta}) \\ \mathrm{Y}_1^{-1}(\boldsymbol{\vartheta}) \\ \mathrm{Y}_1^0(\boldsymbol{\vartheta}) \\ \mathrm{Y}_1^1(\boldsymbol{\vartheta}) \\ \vdots \\ \mathrm{Y}_n^m(\boldsymbol{\vartheta}) \\ \vdots \end{bmatrix} \tag{38}$$

In practice, the encoding is limited to a finite order N, involving χ_N components in vector $\boldsymbol{c}^{\mathscr{N}}$. As explained in previous sections, the harmonic functions Y_n^m depend on the chosen normalization, although the superscript $\mathscr{N}$ has been omitted here for notational simplicity. Also, note that the components in equation (38) are presented with the ACN ordering, but other layout can be employed as well.

At the decoding stage, the so-called "re-encoding principle" [9] can be employed: considering K loudspeakers (supposed to be far enough from the sweet area so that the plane waves assumption holds) with directions $\{\boldsymbol{\varphi}_1, ..., \boldsymbol{\varphi}_K\}$, the re-encoding matrix has χ_N rows and K columns and it writes:

$$\boldsymbol{C}^{\mathscr{N}} = \Big[\boldsymbol{c}^{\mathscr{N}}(\boldsymbol{\varphi}_1)\ \boldsymbol{c}^{\mathscr{N}}(\boldsymbol{\varphi}_2)\ ...\ \boldsymbol{c}^{\mathscr{N}}(\boldsymbol{\varphi}_K)\Big]. \tag{39}$$

Provided that there are enough loudspeakers (i.e. $K \geq \chi_N$), the decoding matrix $\boldsymbol{D}^{\mathscr{N}}$ is given by the pseudo-inverse of $\boldsymbol{C}^{\mathscr{N}}$:

$$\boldsymbol{D}^{\mathscr{N}} = \mathrm{pinv}(\boldsymbol{C}^{\mathscr{N}}) = (\boldsymbol{C}^{\mathscr{N}})^T \left(\boldsymbol{C}^{\mathscr{N}}(\boldsymbol{C}^{\mathscr{N}})^T\right)^{-1}. \tag{40}$$

Daniel et al. [10][2] have shown that, for regular loudspeaker layouts and under the full-normalization ($\mathscr{N} = \mathrm{N3D}$ or $\mathscr{N} = \mathrm{N2D}$ in the 3-D and 2-D case respectively), the previous expression simplifies to:

$$\boldsymbol{D}^{\mathrm{N2D}} = \frac{1}{K}(\boldsymbol{C}^{\mathrm{N2D}})^T \tag{41}$$

which is commonly refers to as the "sampling Ambisonic decoder" (SAD). However, this formula does not hold for other normalization schemes (namely, the formula is only valid for conventions that fullfil the orthonormal property).
Introducing the shorthand notation $\forall n \in \mathbb{N}$

$$\alpha_n^{\mathscr{N}} = \begin{cases} (\alpha_n^n)^{\mathscr{N}\to\mathrm{N3D}} & \text{in the 3-D case} \\ (\alpha_n^n)^{\mathscr{N}\to\mathrm{N2D}} & \text{in the 2-D case} \end{cases} \tag{42}$$

and the $(\chi_N \times \chi_N)$ diagonal matrix $\boldsymbol{\alpha}^{\mathscr{N}}$:

$$\boldsymbol{\alpha}^{\mathscr{N}} = \mathrm{Diag}\Big\{ \alpha_0^{\mathscr{N}} \; \alpha_1^{\mathscr{N}} \; \alpha_1^{\mathscr{N}} \; \alpha_2^{\mathscr{N}} \; \alpha_2^{\mathscr{N}} \; \cdots \; \alpha_N^{\mathscr{N}} \Big\}, \quad (43)$$

it is easy to show that for any convention $\mathscr{N}$,

$$\boldsymbol{C}\boldsymbol{C}^T = K \cdot (\boldsymbol{\alpha}^{\mathscr{N}})^{-2}. \quad (44)$$

This ultimately leads to the generic expression of the sampling decoder matrix:

$$\boldsymbol{D}^{\mathscr{N}} = \frac{1}{K} (\boldsymbol{C}^{\mathscr{N}})^T \, (\boldsymbol{\alpha}^{\mathscr{N}})^2 \quad (45)$$

and equivalently:

$$\boldsymbol{D}^{\mathscr{N}} = \boldsymbol{D}^{\mathrm{N2D}} \, \boldsymbol{\alpha}^{\mathscr{N}}. \quad (46)$$

This is obviously compatible with the expression given in Eq (41) as $\boldsymbol{\alpha}^{\mathscr{N}}$ reduces to identity for the full-normalization.

The decoder can further be optimized by applying a set of correction gains $\{g_0, g_1, ..., g_N\}$ (discussed in the next section) to the encoded signals; the decoding matrix thus becomes:

$$\boldsymbol{D}^{\mathscr{N}} = \boldsymbol{D}^{\mathrm{N2D}} \, \boldsymbol{\alpha}^{\mathscr{N}} \, \mathrm{Diag}\Big\{ g_0 \;\; g_1 \;\; g_1 \;\; \cdots \; g_N \Big\}. \quad (47)$$

Finally, Daniel et al. have derived an equivalent panning function (see also [11, 12]) by combining the encoding equation (38) and the decoding matrix (47): the k^{th} loudspeaker is fed with a signal $S_k = S\,G_k$ such as $\forall k \in [1; K]$

$$K\,G_k = g_0 + 2 \sum_{n=1}^{N} g_n \cos(n\,\gamma_k) \quad \text{in 2-D} \quad (48)$$

$$K\,G_k = \sum_{n=0}^{N} (2n+1)\, g_n \, P_n(\cos\gamma_k) \quad \text{in 3-D} \quad (49)$$

where γ_k denotes the angle between the source and the k^{th} loudspeaker direction.

What is crucial to note here is that the decoding matrix $\boldsymbol{D}$ in equation (45) is dependent on the chosen normalization scheme $\mathscr{N}$. In all his seminal publications, Daniel is employing – in a more or less explicit way – the full-normalization N2D/N3D and therefore he writes the decoding matrix as in equation (41). The author believes that this has unfortunately raised some misusage or incorrect implementation of the Ambisonic decoder e.g. in software tools. Typically the decoding matrix $\boldsymbol{D}$ might have been implemented with equation (41), omitting the term $\boldsymbol{\alpha}^{\mathscr{N}}$ as required in (45). When an orthogonal (but not orthonormal) convention $\mathscr{N}$ is involved, this generates some artifacts in the rendered sound field.

Just for the sake of illustration, we have plotted in Figures 2 and 3 the impact of omitting the $\boldsymbol{\alpha}^{\mathscr{N}}$ term in the decoding matrix $\boldsymbol{D}$. In these figures, it can be observed that neglecting the $\boldsymbol{\alpha}^{\mathscr{N}}$ term causes a distortion of the equivalent panning and directivity functions compared to the proper decoder. The rendering artifacts typically consist in a higher angular spread (and therefore a larger perceived source width) as well as an overall energy loss.

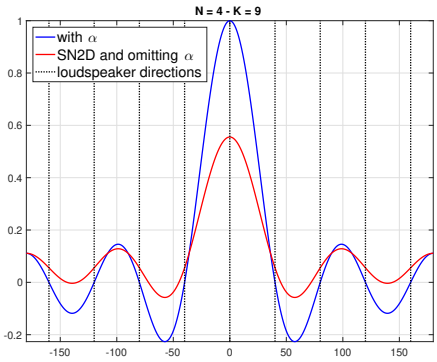

Fig. 2: Equivalent panning function for 2-D Ambisonic rendering with $N = 4$ and using a regular minimum speaker layout ($K = \chi_N$). Blue: with proper decoding matrix as in equation (45). Red: with $\mathscr{N} = \mathrm{SN2D}$ and omitting the term $\boldsymbol{\alpha}^{\mathscr{N}}$ in the decoding matrix.

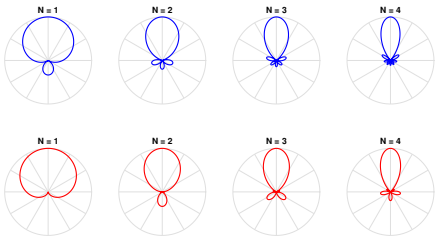

Fig. 3: Equivalent directivity diagram for 2-D Ambisonic rendering with order $N \leq 4$. The directivity patterns are normalized to 1, and the radial scale is linear. Top (blue): with proper decoding matrix as in equation (45). Bottom (red): with $\mathscr{N} = \mathrm{SN2D}$ and omitting the term $\boldsymbol{\alpha}^{\mathscr{N}}$ in the decoding matrix.

As a consequence, the author would recommend to apply the N2D/N3D convention whenever possible. If the encoded materials happen to be with a different convention $\mathscr{N}$, it is advised to convert from $\mathscr{N}$ to N2D/N3D prior to the decoding stage. Such workflow is depicted in Figure 4 and ensures direct compatibility with the formalism published by Daniel.

It is also worth noting that, as long as the proper decoding matrix $\boldsymbol{D}$ (45) is applied, the equivalent panning functions (equations (48) and (49)) are independent of the chosen normalization $\mathscr{N}$. This means that the complete Ambisonic processing chain (encoding+decoding) is independent of the normalization scheme: the generated sound field is the same, regardless of the choice of $\mathscr{N}$.

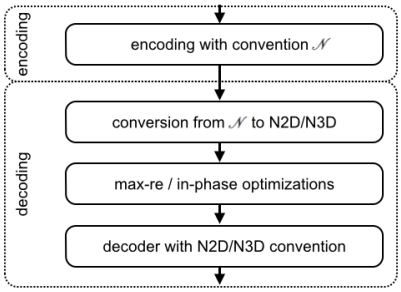

Fig. 4: Recommended workflow where the Ambisonic stream is converted to N2D/N3D prior to optimization and decoding; this ensures straightforward compatibility with the equations provided by Daniel et al.

4.2 Optimized decoding criteria

Daniel et al. [10, 2] have shown that the basic decoder (i.e. $\forall n \in [0;N],\ g_n = 1$) can only achieve proper reconstruction of the sound field in the listening area up to a given frequency. They have therefore proposed alternative decoding strategies to further optimize the perceived spatial rendering, and they have derived a set of corrective gains $\{g_0, g_1, ..., g_N\}$ to be applied to the Ambisonic components. Their work notably resulted in the widely used "max-r_E" and "in-phase" decoding criteria.
In the light of the previous sections, it is legitimate to wonder whether the optimized decoding criteria are dependent on the chosen normalization $\mathscr{N}$. With equations (48) and (49), we have seen that the loudspeaker gains $\{G_k\}$ are independent of $\mathscr{N}$. Consequently, the energy vector $\boldsymbol{E}$ defined as [13]

$$\boldsymbol{E} = \frac{\sum_{k=1}^{K} G_k^2 \boldsymbol{u_k}}{\sum_{k=1}^{K} G_k^2} = r_E \boldsymbol{u_E} \tag{50}$$

transpires to be independent of the normalization scheme. We can thus conclude that the "max-r_E" correction gains, derived by Daniel et al. so as to maximize the norm of $\boldsymbol{E}$, can operate with any convention $\mathscr{N}$. A similar remark applies to the "in-phase" gains, obtained such as $\forall k \in [1;K],\ G_k \geq 0$. The formulae for "max-$r_E$" and "in-phase" decoding are briefly reminded in the Annex 2 for the sake of completeness.

5 Conclusion

Normalization conventions are a frequent source of confusion for Ambisonic users, and when incorrectly used may lead to compatibility issues or rendering artifacts. This paper discussed the various normalization schemes in use in the field of Ambisonic processing, and derived some previously unpublished results. First, we proposed some insights for extending the MaxN and FuMa conventions to higher orders, and we provided numerical table for conversion factors up to the order $N = 16$. Then, we have highlighted both theoretical and practical properties of the different conventions. Finally, we have investigated possible pitfalls related to normalization at the decoding stage. To conclude, we insist that particular attention should be paid to the chosen normalization both in published works and in practical implementation of Ambisonic.

Acknowlegments

The author would like to thank Clément Cornuau and Olivier Warusfel for fruitful discussions and advices.

References

[1] Abramowitz, M. and Stegun, I. A., *Handbook of Mathematical Functions*, General Publishing Company, 1970.

[2] Daniel, J., *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*, Ph.D. thesis, Université de Paris VI, 2001.

[3] Chapman, M., Ritsch, W., Musil, T., Zmölnig, J., Pomberger, H., Zotter, F., and Sontacchi, A., "A Standard for Interchange of Ambisonic Signal Sets," in *Proc. of the 1st Ambisonics Symposium*, Graz, Austria, 2009.

[4] Herre, J., Hilpert, J., Kuntz, A., and Plogsties, J., "MPEG-H 3D Audio—The New Standard for Coding of Immersive Spatial Audio," *IEEE Journal of selected topics in signal processing*, 9(5), pp. 770 – 779, 2015.

[5] Nachbar, C., Zotter, F., Deleflie, E., and Sontacchi, A., "ambiX - A suggested ambisonics format," in *Proc. of the 3rd Ambisonics Symposium*, Lexington, KY, USA, 2011.

[6] Malham, D., *Higher order Ambisonic systems. Space in Music – Music in Space.*, Master's thesis, University of York, 2003.

[7] Malham, D., "Higher Order Ambisonic systems for the spatialisation of sound," in *Proc. of the International Computer Music Conference*, pp. 484 – 487, Beijing, China, 1999.

[8] Holmes, S. A. and Featherstone, W. E., "A unified approach to the Clenshaw summation and the recursive computation of very high degree and order normalised associated Legendre functions," *Journal Of Geodesy*, 76(5), pp. 279 – 299, 2002.

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

[10] Daniel, J., Rault, J.-B., and Polack, J.-D., "Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions," in *Proc. of the 105th AES Convention*, San Francisco, CA, USA, 1998.

[11] Neukom, M. and Schacher, J. C., "Ambisonics Equivalent Panning," in *Proc. of the International Computer Music Conference*, Belfast, Ireland, 2008.

[12] Neukom, M., "Ambisonic Panning," in *Proc. of the 123rd AES Convention*, New York, NY, USA, 2007.

[13] Gerzon, M. A., "General metatheory of auditory localization," in *Proc. of the 92nd AES Convention*, Vienna, Austria, 1992.

Annex 1: Derivation of MaxN factors and conversion between N3D and MaxN

n	m	$R_n^m(x)$	β_n^m
2	1	$3/4x^2 - 3/8$	$1/2\sqrt{2}$
3	1	$\frac{15}{32}x^3 - \frac{11}{32}x$	$1/15\sqrt{165}$
3	2	$\frac{15}{16}x^2 - \frac{5}{16}$	$1/3\sqrt{3}$
4	1	$\frac{35}{192}x^4 - \frac{45}{256}x^2 + \frac{5}{256}$	$1/28\sqrt{378+14\sqrt{393}}$
4	2	$\frac{35}{64}x^3 - \frac{5}{16}x$	$2/7\sqrt{7}$
4	3	$\frac{35}{32}x^2 - \frac{35}{128}$	$1/2$
5	1	$\frac{105}{2048}x^5 - \frac{63}{1024}x^3 + \frac{29}{2048}x$	$\frac{1}{105}\sqrt{6615+210\sqrt{231}}$
5	2	$\frac{105}{512}x^4 - \frac{21}{128}x^2 + \frac{7}{512}$	$1/15\sqrt{90+15\sqrt{21}}$
5	3	$\frac{315}{512}x^3 - \frac{147}{512}x$	$1/15\sqrt{105}$
5	4	$\frac{315}{256}x^2 - \frac{63}{256}$	$1/5\sqrt{5}$
6	1	$\frac{231}{20480}x^6 - \frac{133}{8192}x^4 + \frac{35}{6144}x^2 - \frac{7}{24576}$	≈ -0.95945693
6	2	$\frac{231}{4096}x^5 - \frac{119}{2048}x^3 + \frac{133}{12288}x$	$1/33\sqrt{561+132\sqrt{5}}$
6	3	$\frac{231}{1024}x^4 - \frac{315}{2048}x^2 + \frac{21}{2048}$	$1/22\sqrt{165+11\sqrt{137}}$
6	4	$\frac{693}{1024}x^3 - \frac{273}{1024}x$	$1/33\sqrt{429}$
6	5	$\frac{693}{512}x^2 - \frac{231}{1024}$	$1/6\sqrt{6}$

n	m	$(\mathscr{A}_n^m)^{\text{MaxN}}$
0	0	1
1	0	1
1	1	-1
2	0	1
2	1	$-\frac{2}{3} \approx -0.666666666669$
2	2	$\frac{1}{3} \approx 0.333333333334$
3	0	1
3	1	$-\frac{\sqrt{15}}{8} \approx -0.484122918283$
3	2	$\frac{3}{10\sqrt{3}} \approx 0.173205080757$
3	3	$-\frac{1}{15} \approx -0.0666666666668$
4	0	1
4	1	$-\frac{12544}{5}\frac{1}{\sqrt{406-14\sqrt{393}}\sqrt{378+14\sqrt{393}}(3+\sqrt{393})} \approx -0.378778472729$
4	2	$\frac{14}{135} \approx 0.103703703705$
4	3	$-\frac{16}{945}\sqrt{3} \approx -0.0293257279589$
4	4	$\frac{1}{105} \approx 0.00952380952383$
5	0	1
5	1	$-\frac{525}{2}\frac{1}{\sqrt{4410-210\sqrt{231}}(9+\sqrt{231})} \approx -0.310788294986$
5	2	$-\frac{150}{7}\frac{1}{(-9+\sqrt{21})\sqrt{90+15\sqrt{21}}(1+\sqrt{21})} \approx 0.06896808398$
5	3	$-\frac{5}{1792}\sqrt{2}\sqrt{15} \approx -0.0152824374304$
5	4	$\frac{5}{3024}\sqrt{5} \approx 0.00369720234375$
5	5	$-\frac{1}{945} \approx -0.0010582010582$
6	0	1
6	1	≈ -0.26339046203
6	2	$-\frac{363}{560}\frac{1}{(-4+\sqrt{5})(3+2\sqrt{5})} \approx 0.0491803922547$
6	3	$\frac{170368}{315}\frac{1}{(-29+\sqrt{137})\sqrt{319-11\sqrt{137}}\sqrt{165+11\sqrt{137}}(3+\sqrt{137})} \approx -8.995858E^{-3}$
6	4	$\frac{121}{70000} \approx 0.00172857142857$
6	5	$-\frac{8}{48125}\sqrt{5} \approx 3.7171000145E^{-4}$
6	6	$\frac{1}{10395} \approx 9.6200096229E^{-5}$

n	m	$(\alpha_n^m)^{\text{N3D}\to\text{MaxN}}$
0	0	1.0
1	0	0.57735026919
1	1	0.57735026919
2	0	0.4472135955
2	1	0.516397779496
2	2	0.51639777949
3	0	0.377964473009
3	1	0.448210728507
3	2	0.507092552838
3	3	0.478091443734
4	0	0.333333333333
4	1	0.399267567488
4	2	0.463777062004
4	3	0.490713286644
4	4	0.450748935856
5	0	0.301511344578
5	1	0.362922539256
5	2	0.426163708974
5	3	0.462622285519
5	4	0.474835735249
5	5	0.429772247481
6	0	0.277350098113
6	1	0.334763434872
6	2	0.395330166312
6	3	0.43387220974
6	4	0.456632874877
6	5	0.460569772659
6	6	0.412911819275
7	0	0.258198889747
7	1	0.312182337842
7	2	0.36990470086
7	3	0.408361635381
7	4	0.434486446188
7	5	0.449169762022
7	6	0.447920638705
7	7	0.398910715738
8	0	0.242535625036
8	1	0.293577938011
8	2	0.34861576791
8	3	0.386239716959
8	4	0.413398644118
8	5	0.432008709342
8	6	0.441417660804
8	7	0.436684774402
8	8	0.387000239101
9	0	0.229415733871
9	1	0.277917409819
9	2	0.330510080374
9	3	0.367046044173
9	4	0.394301171025
9	5	0.414511980016
9	6	0.428102250922
9	7	0.433833766168
9	8	0.426644132573
9	9	0.37667837259
10	0	0.218217890236
10	1	0.264504480582
10	2	0.314893895632
10	3	0.350280782906
10	4	0.377216758342
10	5	0.398027938843
10	6	0.413528632263
10	7	0.42354770856
10	8	0.42659402767
10	9	0.417608899883
10	10	0.367600451289
11	0	0.208514414057
11	1	0.252851804043
11	2	0.301259060124
11	3	0.335516363369
11	4	0.361944995077
11	5	0.382872526048
11	6	0.399270823659
11	7	0.411375713535
11	8	0.418739251043
11	9	0.419756174729
11	10	0.409422670075
11	11	0.359520333786

n	m	$(\alpha_n^m)^{\text{N3D}\to\text{MaxN}}$
12	0	0.2
12	1	0.242607301514
12	2	0.289227414807
12	3	0.322407723299
12	4	0.348248753276
12	5	0.369043342751
12	6	0.385826370147
12	7	0.399012910511
12	8	0.408561967983
12	9	0.413884596253
12	10	0.413326209776
12	11	0.401958449071
12	12	0.352256547973
13	0	0.19245008973
13	1	0.233509710599
13	2	0.278512499839
13	3	0.310680348618
13	4	0.335907633997
13	5	0.356437663978
13	6	0.373325500088
13	7	0.387070982904
13	8	0.397818994451
13	9	0.405381900006
13	10	0.409095213689
13	11	0.407287608788
13	12	0.395113170465
13	13	0.345671732811
14	0	0.185695338177
14	1	0.225360670157
14	2	0.268893536959
14	3	0.300115708978
14	4	0.324730968317
14	5	0.34492782323
14	6	0.361760766181
14	7	0.375770200865
14	8	0.387194188772
14	9	0.396031846671
14	10	0.40201248546
14	11	0.404430424959
14	12	0.401614768472
14	13	0.388802749537
14	14	0.339659591874
15	0	0.179605302027
15	1	0.218006578435
15	2	0.260197620489
15	3	0.290538725188
15	4	0.314558008581
15	5	0.334388348067
15	6	0.351073265124
15	7	0.365171281752
15	8	0.376969632674
15	9	0.38656369923
15	10	0.393868036686
15	11	0.398562813382
15	12	0.399920317506
15	13	0.396279349876
15	14	0.38295807346
15	15	0.334136303584
16	0	0.174077655956
16	1	0.211326436216
16	2	0.252287346628
16	3	0.281807924599
16	4	0.30525436712
16	5	0.324705182521
16	6	0.341187713078
16	7	0.355266542827
16	8	0.367255282082
16	9	0.377302283238
16	10	0.3854190875
16	11	0.391468604038
16	12	0.395101168504
16	13	0.395578222888
16	14	0.391253263971
16	15	0.377521867141
16	16	0.329034686388

Annex 2: Correction gains for optimized decoding

The results published by Daniel et al. [10, 2] for the "max-r_E" and "in-phase" criteria are briefly reminded below:

max-r_E optimization

In the 2-D case:

$$\left(g_n\right)_{\max r_E} = \cos\left(\frac{n\pi}{2N+2}\right) \quad \text{for} \quad 0 \leq n \leq N \quad (51)$$

In the 3-D case:

$$\left(g_n\right)_{\max r_E} = \begin{cases} 1, & \text{for} \quad n = 0 \\ P_n\left(r_E\right), & \text{for } 1 \leq n \leq N \end{cases} \quad (52)$$

where r_E is the largest root of the $P_{N+1}\left(x\right)$ polynomial.

in-phase optimization

In the 2-D case:

$$\left(g_n\right)_{\text{in-phase}} = \begin{cases} 1, & \text{for } n = 0 \\ \frac{(N!)^2}{(N+n)!\,(N-n)!}, & \text{for } 1 \leq n \leq N \end{cases} \quad (53)$$

In the 3-D case:

$$\left(g_n\right)_{\text{in-phase}} = \begin{cases} 1, & \text{for } n = 0 \\ \frac{(N)!\,(N+1)!}{(N+n+1)!\,(N-n)!}, & \text{for } 1 \leq n \leq N \end{cases} \quad (54)$$