Reduction of 3D Ambisonic to 2D using plane-wave decomposition

Thibaut Carpentier

Audio Engineering Society

Convention Paper

Presented at the 152nd Convention
2022 May, Online

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

Reduction of 3D Ambisonic to 2D using plane-wave decomposition

Thibaut Carpentier[1]

[1] *STMS, IRCAM – CNRS – Sorbonne Université – Ministère de la Culture, Paris, France*

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

ABSTRACT

This articles deals with the reduction – or "projection" or "downscaling" – of a 3D-encoded Ambisonic sound field ("full-sphere" or "periphonic") into a 2D representation ("horizontal-only" or "planar" or "pantophonic"). We show that the reduction operation can be equivalently achieved by 1) applying conversion formula of the normalization factors, or 2) performing a plane-wave decomposition of the original sound field and re-encoding the resulting plane waves to 2D Ambisonic. The latter approach provides greater flexibility in adapting the content to horizontal-only reproduction.

1 Introduction

The Ambisonic formalism [1] comes from writing the wave equation in the spherical coordinate system, wherein an acoustic pressure field can be decomposed into a Fourier-Bessel series. The series exhibits radial functions (the spherical Bessel functions), and angular functions called spherical harmonics. The latter form an orthonormal basis for the spherical scalar product (integration over the sphere). In practice the Fourier-Bessel series is truncated to a finite number of components, and this spherical harmonic decomposition leads to the directional encoding equations used for 3D Ambisonic representation.
When dealing with a horizontal-only representation, the cylindrical coordinate system can be used, and the Fourier-Bessel series then involves circular harmonics as angular functions. The circular harmonic functions are related to the spherical ones (through a scaling factor that will be discussed in paragraph 3.1), and the 2D-restricted formalism can be considered as a (weighted) subset of the 3D spherical representation.

These 2D and 3D Ambisonic formalisms somehow coexist, and one or the other can be used depending on the context. For a given Ambisonic order N, 2D systems require much less complexity (number of encoding components and decoding loudspeakers) than 3D ones, but they cannot reproduce elevated sound sources. Often, reproduction setups provide a higher density of loudspeakers in the horizontal plane, and thus 2D Ambisonic can potentially enhance the spatial resolution in the horizontal plane (where human auditory localization accuracy is optimal). As a consequence, hybrid representations or mixed-order schemes have been proposed [1, 2, 3, 4] in order to combine the benefits of 2D and 3D playback systems.

It is sometimes desirable to reduce 3D-encoded material to 2D; for instance, this can be useful to reduce the required bandwidth, or to decode a 3D

Ambisonic stream (such as captured by spherical microphone arrays) to planar loudspeaker arrays. This topic has been marginally addressed in a number of previous works [5, 6, 1]. In this paper, we gather and present the information in a comprehensive way, and we discuss some practical considerations. We show that the reduction operation can be equivalently achieved by applying conversion formula of the normalization factors, or by performing a plane-wave decomposition and re-encoding the resulting plane waves to 2D Ambisonic. A somewhat similar approach was previously presented in [7, 8] in order to drive linear or circular arrays of loudspeakers from a 3D sound field representation.

2 Notations and preliminary definitions

In this section, we introduce notation used in the rest of the paper (see e.g. [9, 1] for further details).

2.1 Spherical harmonics

Let's note $\vartheta \equiv (\theta, \phi)$ the angular direction in the spherical coordinate system with $x = \cos\phi \sin\theta$, $y = \sin\phi \sin\theta$, and $z = \cos\theta$,
$\mathbb{I}_N$ the set of Ambisonic indices i.e. $\{(n,m) \in \mathbb{N} \times \mathbb{Z} \,:\, 0 \leq |m| \leq n \leq N\}$, with N being the chosen maximum order,
and χ_N the number of Ambisonic components given by $\chi_N^{2D} = (2N+1)$ and $\chi_N^{3D} = (N+1)^2$ for the 2D and 3D case respectively.

The real-valued spherical harmonics $\mathrm{Y}_n^m(\vartheta)$ are defined $\forall (n,m) \in \mathbb{I}_N$ by

$$\mathrm{Y}_n^m(\vartheta) = \mathscr{A}_n^{|m|} P_n^{|m|}(\cos\theta) \begin{cases} \cos(m\phi), & \text{for } m \geq 0 \\ \sin(|m|\phi), & \text{for } m < 0 \end{cases} \tag{1}$$

where $\mathscr{A}_n^{|m|}$ is a scalar constant, dependent on the chosen normalization scheme [1, 10].

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

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

$P_n(x)$ are the Legendre polynomials, which can be expressed $\forall x \in \mathbb{R} \cap [-1,1]$, $\forall n \in \mathbb{N}$,

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} \left(x^2-1\right)^n. \tag{3}$$

2.2 Terminology

The set of $\{(n,m) \in \mathbb{I}_N \,:\, m = 0\}$ components are called "zonal" harmonics. They do not depend on the longitude and divide the sphere into zones parallel to the equator.
The set of $\{(n,m) \in \mathbb{I}_N \,:\, |m| = n\}$ components are called "sectoral" harmonics. They divide the sphere into sectors following the meridians.
The set of $\{(n,m) \in \mathbb{I}_N \,:\, 0 < |m| < n\}$ components are referred to as "tesseral" harmonics. They divide the sphere into a "chequered" pattern, as can be observed in Figure 1.

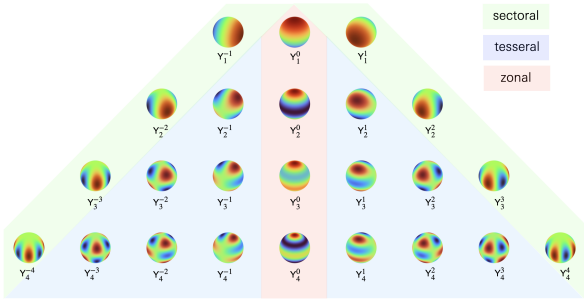

Fig. 1: Zonal (light red), sectoral (light green), and tesseral (light blue) spherical harmonics for $1 \leq n \leq 4$.

2.3 Normalization scheme

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

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

When considering a 2D (horizontal-only) representation, only the sectoral harmonics are used, and the fully-normalized factors (N2D) are expressed $\forall n \in \mathbb{N}$ by

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

With the N2D normalization, the $\mathbf{Y}_N^{\mathrm{N2D}}(\vartheta)$ functions form an orthonormal basis of circular harmonics for the circular scalar product (evaluated on the unit circle).

3 Reducing 3D representation to 2D

This section examines different approaches to reduce a 3D-encoded Ambisonic sound field to 2D.

3.1 Using conversion factors

The conversion factor from normalization $\mathscr{N}_1$ to $\mathscr{N}_2$ is notated $(\alpha_n^m)^{\mathscr{N}_1\to\mathscr{N}_2}$, and defined such as $\forall(n,m)\in\mathbb{I}_N$

$$(\mathscr{A}_n^m)^{\mathscr{N}_1}\,(\alpha_n^m)^{\mathscr{N}_1\to\mathscr{N}_2} = (\mathscr{A}_n^m)^{\mathscr{N}_2}\,. \tag{6}$$

Conversion factors satisfy the reciprocity property i.e.

$$(\alpha_n^m)^{\mathscr{N}_1\to\mathscr{N}_2} = \frac{1}{(\alpha_n^m)^{\mathscr{N}_2\to\mathscr{N}_1}}\,. \tag{7}$$

Daniel [1] provides a review of conversion factors, from which it is easily derived that

$$\forall n\in\mathbb{N},\ \left(\alpha_n^{\pm n}\right)^{\mathrm{N3D}\to\mathrm{N2D}} = \frac{2^n\,n!}{\sqrt{(2n+1)!}}\,. \tag{8}$$

A numerically stable calculation of Eq. 8 is proposed in Annex 1.

The N3D $\to$ N2D scaling factor in Eq. 8 was also obtained in [11], although from a different approach: equating the cylindrical and spherical spatial Fourier transforms for a plane wave propagating in the horizontal plane, and noting that $\forall(\theta,\,\phi)$, $\forall(n,m)\in\mathbb{I}_N$

$$\begin{aligned}\left(\mathrm{Y}_n^m\right)^{\mathrm{N3D}}(\theta,\,\phi) = {}& \left(\mathrm{Y}_n^m\right)^{\mathrm{N3D}}(\theta,\,0)\,\cos(|m|\phi)\\ &+\left(\mathrm{Y}_n^{-m}\right)^{\mathrm{N3D}}(\theta,\,0)\,\sin(|m|\phi)\,,\end{aligned} \tag{9}$$

Thomas et al. have derived the following solution[1] $\forall n\in\mathbb{N}^*$

$$\left(\alpha_n^{\pm n}\right)^{\mathrm{N3D}\to\mathrm{N2D}} = \frac{\sqrt{2}}{\left(\mathrm{Y}_n^{|n|}\right)^{\mathrm{N3D}}\left(\frac{\pi}{2},\,0\right)} \tag{10}$$

which is equivalent to Daniel's formula.

[1]The result in [11] was obtained for complex-valued spherical harmonics, which we have translated here to real-valued harmonics for consistency with the convention adopted in this article.

3.2 Using plane-wave decomposition

The encoding of a plane wave source carrying a signal $s_0(t)$, with incident direction $\vartheta_0\equiv(\theta_0,\,\phi_0)$, to 3D N^{th}-order Ambisonic signals writes

$$\mathbf{s}(t) = s_0(t)\;\mathbf{Y}_N^{\mathrm{N3D}}(\vartheta_0)\,, \tag{11}$$

where $\mathbf{Y}_N^{\mathrm{N3D}}(\vartheta_0)$ is a vector of χ_N^{3D} components :

$$\mathbf{Y}_N^{\mathrm{N3D}}(\vartheta_0) = \left[\mathrm{Y}_0^0(\vartheta_0),\ \ldots\ \mathrm{Y}_n^m(\vartheta_0),\ \ldots\ \mathrm{Y}_N^N(\vartheta_0)\right]^\top. \tag{12}$$

The sound field $\mathbf{s}(t)$ can also be analyzed by plane-wave decomposition [12]. The plane-wave components, or plane-wave amplitude density function, can be computed by a beamforming system. Beamforming in the spherical harmonics domain reduces to a weight-and-sum combination of the harmonic components: the signal captured by a virtual microphone (with axis-symmetric directivity pattern) steering in direction Ω writes [13]

$$x(t,\Omega) = \mathbf{w}^\top(\Omega)\;\mathbf{s}(t)\,, \tag{13}$$

where $\mathbf{w}(\Omega)$ is a χ_N^{3D} vector of beamforming weights:

$$\left[\mathbf{w}(\Omega)\right]_{nm} = w_n\;\mathrm{Y}_n^m(\Omega)\,. \tag{14}$$

For plane-wave analysis, the hyper-cardioid beamforming weights (a.k.a maximum directivity beamformer or spatial Dirac function) are used:

$$\forall n\in[0,N],\ \ w_n = \frac{1}{(N+1)^2}\,. \tag{15}$$

In order to spatially decompose the sound field, we steer virtual microphones (beams) in K directions Ω_k. These K directions should be chosen appropriately to sample the unit sphere while preserving the discrete orthonormality of the spherical harmonics [14]. For instance, hyper-interpolation grids [15], Fliege grids [16], or spherical t-designs [17] might be used. The required number of directions K depends on the chosen grid properties, and in any case it must satisfy $K\geq(N+1)^2$ for aliasing-free sampling.
The k^{th} hyper-cardioid beam signal writes

$$\begin{aligned}x(t,\Omega_k) &= \mathbf{w}^\top(\Omega_k)\;\mathbf{s}(t)\\ &= \frac{1}{(N+1)^2}\;\left(\mathbf{Y}_N^{\mathrm{N3D}}(\Omega_k)\right)^\top\mathbf{s}(t)\,.\end{aligned} \tag{16}$$

Now, in order to reduce the sound field to 2D, each of the K beamformed signals is re-encoded to 2D

(horizontal-only) Ambisonic, in the corresponding direction Ω_k :

$$\mathbf{y}(t,\Omega_k) = x(t,\Omega_k)\ \mathbf{Y}_N^{\mathrm{N2D}}(\Omega_k)\,. \tag{17}$$

The 2D Ambisonic stream $\mathbf{y}(t)$ is the linear superposition of the K re-encoded signals

$$\mathbf{y}(t) = \sum_{k=1}^{K} \mathbf{y}(t,\Omega_k) = \sum_{k=1}^{K} x(t,\Omega_k)\ \mathbf{Y}_N^{\mathrm{N2D}}(\Omega_k)\,. \tag{18}$$

Using the shorthand notation $\alpha_N \equiv (\alpha_N)^{\mathrm{N3D}\to\mathrm{N2D}}$, the definition of the conversion factors writes

$$\mathbf{Y}_N^{\mathrm{N2D}}(\Omega) = \alpha_N \cdot \mathbf{Y}_N^{\mathrm{N3D}}(\Omega)\,, \tag{19}$$

and introducing Eq. 19 in Eq. 18 gives

$$\begin{aligned}\mathbf{y}(t) &= \sum_{k=1}^{K} x(t,\Omega_k)\ \alpha_N \cdot \mathbf{Y}_N^{\mathrm{N3D}}(\Omega_k) \\ &= \frac{1}{(N+1)^2} \sum_{k=1}^{K} \left(\mathbf{Y}_N^{\mathrm{N3D}}(\Omega_k)\right)^{\top} \mathbf{s}(t)\, \alpha_N \cdot \mathbf{Y}_N^{\mathrm{N3D}}(\Omega_k)\,.\end{aligned} \tag{20}$$

As the set of directions $\{\Omega_k\}$ fulfills the discrete orthonormality property of the spherical harmonics (with the N3D scheme), the summation over K simplifies to

$$\begin{aligned}\mathbf{y}(t) &= \frac{1}{(N+1)^2}\ \mathbf{s}(t)\cdot\alpha_N\ K \\ &= \frac{K}{(N+1)^2}\ s_0(t)\ \mathbf{Y}_N^{\mathrm{N3D}}(\vartheta_0)\cdot\alpha_N \\ &= \frac{K}{(N+1)^2}\ s_0(t)\ \mathbf{Y}_N^{\mathrm{N2D}}(\vartheta_0)\,.\end{aligned} \tag{21}$$

Besides a constant scaling factor $\frac{K}{(N+1)^2}$, this exactly corresponds to the encoding of a plane wave source carrying a signal $s_0(t)$, with incident direction $\vartheta_0 \equiv (\theta_0,\ \phi_0)$, in 2D N^{th}-order Ambisonic.
In other words: spatially decomposing a 3D sound field in K beams (with maximum directivity), and re-encoding the beams to 2D, is strictly equivalent to a native 2D encoding of the original sound field.
This gives us two equivalent approaches to "reduce" a 3D Ambisonic stream to 2D.

Note that, the (equivalence) results presented above for a single plane wave, holds true for a 3D sound field made of an arbitrary number J of source signals

$$\mathbf{s}(t) = \sum_{j=1}^{J} s_j(t)\ \mathbf{Y}_N^{\mathrm{N3D}}(\vartheta_j)\,, \tag{22}$$

by virtue of superposition principle.

3.3 Discarding tesseral and zonal harmonics

Another procedure for reducing 3D Ambisonic to 2D consists in simply discarding the tesseral and zonal harmonics from the 3D representation, keeping only its sectoral components. It is clear from the previous sections (3.1 and 3.2) that such "naive" truncation is not rigorous in the sense of mathematics. And yet, this approach is not so uncommon in the practical field. Just for the sake of illustration, we have plotted in Figure 2 the equivalent directivity function of a 3D sound field reduced to 2D with the proper conversion factor $(\alpha_N)^{\mathrm{N3D}\to\mathrm{N2D}}$, and without (naively discarding the tesseral and zonal components). It can be observed that the shape of the main lobe is mostly preserved, but deformation of the side lobes is noticeable.
Indeed, when neglecting the conversion factor $(\alpha_N)^{\mathrm{N3D}\to\mathrm{N2D}}$, we misestimate the contribution of the elevation angle θ in the 2D representation.

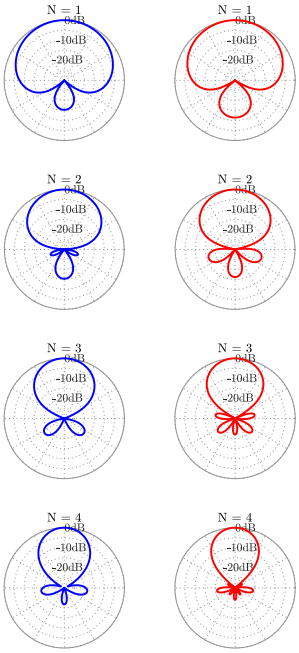

Fig. 2: Equivalent directivity pattern of a virtual source with incident direction $\vartheta_0 \equiv (\theta_0,\ \phi_0) = (45°,\ 0°)$, encoded in 3D Ambisonic, and reduced to 2D. Left side (blue): with proper 2D reduction operation as detailed in Eq. 21. Right side (red): naively discarding the tesseral and zonal harmonics from the 3D representation, and keeping only the sectoral harmonics. Results are displayed for encoding order $1 \leq N \leq 4$.

4 Role of θ information in 2D Ambisonic

Signals encoded in the horizontal-only formalism (2D Ambisonic) actually convey information about the elevation angle θ of the virtual source(s) – although with a top-bottom indeterminacy:

$$\forall(\theta,\,\phi)\,,\;\; \mathbf{Y}_N^{\mathrm{N2D}}(\theta,\,\phi) = \mathbf{Y}_N^{\mathrm{N2D}}(-\theta,\,\phi)\,. \qquad (23)$$

The role of θ information in 2D Ambisonic has been extensively studied by Daniel, notably in [1, 18, 19]: when decoding Ambisonic signals over a horizontal loudspeaker layout, it is not possible to naturally reproduce elevation effects; however, the θ information can still be useful to render lateralization effect.

For a virtual source reproduced by L loudspeakers, with directions $\mathbf{u}_l$, and driven with real gains G_l, the velocity vector [20] in the central listening position is real-valued and writes (assuming superposition of plane waves):

$$\mathbf{V} = \frac{\sum_{l=1}^{L} G_l\,\mathbf{u}_l}{\sum_{l=1}^{L} G_l} = r_v\,\mathbf{u}_v\,, \qquad (24)$$

with r_v denoting its norm.

Daniel et al. [18] have shown that, in the low-frequency domain, the synthesized wave field can be (locally) considered as a plane wave with apparent direction $\mathbf{u}_v$ and propagation speed c/r_v (c being the speed of sound), and the resulting interaural time difference (ITD) for a listener in the central listening position is scaled by a factor r_v (compared to the ITD perceived for a natural plane wave). As r_v implicitly depends on the elevation angle θ of the virtual source, the predicted lateralization is influenced by θ (see Figure 3).

If the listener's head is static, the lateralization effect is characterized only by the projection of $\mathbf{V}$ onto the interaural axis $\mathbf{y}$. Due to the axial symmetry (in the low-frequency domain, the diffraction effect is negligible and so the head behaves like a rigid sphere), this generates a cone of confusion where the angles of incidence of the waves are ambiguous [19]. The opening angle of the cone of confusion is given by $\gamma = \arccos(\mathbf{V}\cdot\mathbf{y})$, and is displayed in Figure 4.

With slight head rotations around the z-axis (yaw rotation of the listener), the variations of the interaural phase differences can be perceived as an "artificial" height effect, and the apparent elevation angle (measured from the horizontal plane) is given by $\hat{\theta} = \arccos(r_v)$.

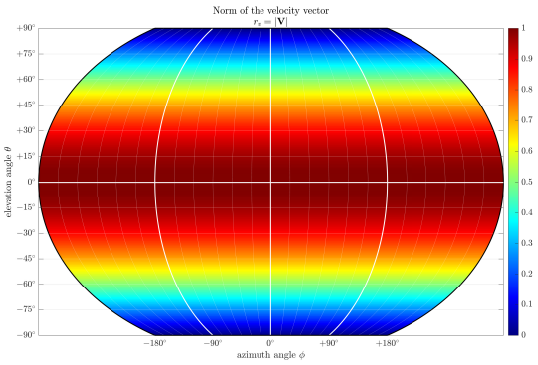

Fig. 3: Norm r_v of the velocity vector, as a function of the virtual source direction $\vartheta \equiv (\theta,\,\phi)$, for a 2D N^{th}-order Ambisonic stream decoded over $2N+2$ loudspeakers regularly arranged on a circle.

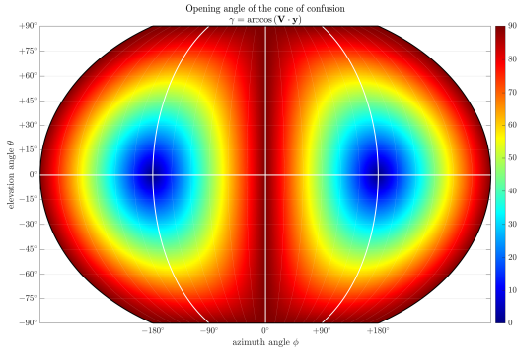

Fig. 4: Opening angle γ of the confusion cone, as a function of the virtual source direction $\vartheta \equiv (\theta,\,\phi)$, for a 2D N^{th}-order Ambisonic stream decoded over $2N+2$ loudspeakers regularly arranged on a circle.

5 Practical considerations

From the previous sections, we stress that, for decoding 3D Ambisonic to a planar loudspeaker array, it is preferable to apply proper reduction formula (as Eq. 8) — in order to preserve the lateralization cues conveyed by the θ information in the circular harmonics formalism — rather than just discarding the tesseral and zonal harmonics.

Still, the practical decoding of 3D material to horizontal-only setups remains an unresolved challenge for many productions, as artistic concern has to be favored over strict mathematical correctness. For instance, the top-bottom indeterminacy of 2D Ambisonic is sometimes problematic as it can make the decoded scene unrealistic or confusing. Often, the lower hemisphere of the 3D scene is unwelcome in 2D

playback. Similarly, sounds originating from the upper region of the sphere can be contentious. As observed in Figure 3, when θ tends towards the zenith angle, the velocity vector r_v tends towards 0 (a similar remark applies to the energy vector r_E). For a sound source located exactly at the zenith (sometimes referred to as "the voice of god"), the 2D Ambisonic stream even reduces to its omnidirectional component (W) as

$$\forall n \in \mathbb{N},\ \forall \phi,\ \mathbf{Y}_n^n(\theta = 0,\, \phi) = 0 \quad \text{if } n \neq 0\,. \quad (25)$$

As a consequence, such elevated sound objects are evenly distributed across all loudspeakers during decoding stage. Even though this can produce kind of a fly-over effect (for moving sources), such behaviour is sometimes undesired (depending on the intention of the content creator).

The two projection approaches presented in paragraphs 3.1 and 3.2 (with conversion formula or with a beamforming system) are theoretically equivalent; however, the plane-wave beamformer method is practically more convenient, as it offers more flexibility in the 2D reduction process. Indeed, the plane-wave amplitude density function could be altered in order to transform the 3D content for 2D playback. Typically, a gain function (or a spectral filter) is applied to the signals steered from the virtual microphones (Eq. 16), depending on their look direction Ω_k — usually, a relatively smooth function of θ_k should be used. That way, it is possible to dim or filter out sounds coming for the lower or upper regions of the sphere. Such strategy is basically similar to the directional loudness modifications proposed in [21].
Another possibility is to modify the directions Ω_k of the virtual beams during 2D re-encoding (Eq. 17), for instance by warping them towards the equator. Again, this is similar, in principle, to the spatial warping effect presented in [21].

Besides localized sounds, the 3D–to–2D projection may also be problematic for diffuse fields. It is known that a perfectly diffuse 3D sound field (uncorrelated plane waves isotropically coming from all directions over the sphere) leads to equal-power Ambisonic components with the N3D normalization scheme [1]. Similarly, a horizontal-only diffuse field (made of uncorrelated plane waves distributed on the circle) generates equal-power N2D Ambisonic signals. However, the 2D reduction of a 3D diffuse field is not horizontally diffuse. This is a side effect of the conversion factors $(\alpha_N)^{\text{N3D}\to\text{N2D}}$. In that case, it might thus be preferable to ignore the α_N weighting, in order to preserve the diffuseness property during 2D reduction. Again, this is a decision that should be left to the content producer.

6 Conclusion

This contribution gathered theoretical and practical information concerning the reduction of 3D-encoded Ambisonic content to a horizontal-only representation. Even though planar loudspeaker arrays cannot naturally reproduce elevated sound sources, the inclination angle information conveyed in 2D Ambisonic formalism can be useful to render lateralization or "artificial" height effect. When naively discarding tesseral and zonal harmonics from the 3D representation, the inclination angle information is degraded. It is therefore preferable to properly apply projection formula when decoding a periphonic stream to a pantophonic loudspeaker setup. We have demonstrated that the 3D-to-2D projection can be equivalently achieved by applying conversion formula of the normalization factors, or by estimating a plane-wave amplitude density and then re-encoding the resulting plane waves to 2D Ambisonic. Nevertheless, the 3D-to-2D reduction cannot be blindly automated in most productions, as decisions have to be taken depending on the actual content of the sound scene; these decisions are left to the content creator. The plane-wave decomposition approach herein presented offers more flexibility in adapting the content to 2D representation.

References

[1] 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.

[2] Travis, C., "A New Mixed-Order Scheme for Ambisonic Signals," in *Proc. of the 1st Ambisonics Symposium*, Graz, Austria, 2009.

[3] Käsbach, J., Favrot, S., and Buchholz, J., "Evaluation of a mixed-order planar and periphonic Ambisonics playback implementation," in *Proc. of Forum Acusticum*, Aalborg, Denmark, 2011.

[4] Favrot, S., Marschall, M., Käsbach, J., Buchholz, J., and Weller, T., "Mixed-Order Ambisonics Recording and Playback for Improving Horizontal Directionality," in *Proc. of the 131st Convention of the Audio Engineering Society (AES)*, New York, NY, USA, 2011.

[5] 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.

[6] Daniel, J., "Evolving Views on HOA : From Technological To Pragmatic Concerns," in *Proc. of the 1st Ambisonics Symposium*, Graz, Austria, 2009.

[7] Okamoto, T., "Angular spectrum decomposition-based 2.5D higher-order spherical harmonic sound field synthesis with a linear loudspeaker array," in *Proc. of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 180 – 184, New Paltz, NY, USA, 2017.

[8] Okamoto, T., "2.5D higher order ambisonics for a sound field described by angular spectrum coefficients," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 326 – 330, Shanghai, China, 2016.

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

[10] Carpentier, T., "Normalization schemes in Ambisonic: does it matter?" in *Proc 142nd AES Convention*, Berlin, Germany, 2017.

[11] Thomas, M. R. P., Ahrens, J., and Tashev, I., "A Method for Converting Between Cylindrical and Spherical Harmonic Representations of Sound Fields," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4723 – 4727, Florence, Italy, 2014.

[12] Rafaely, B., "Plane-wave decomposition of the sound field on a sphere by spherical convolution," *Journal of the Acoustical Society of America*, 116(4), pp. 2149–2157, 2004.

[13] Rafaely, B., *Fundamentals of Spherical Array Processing – Second edition*, Springer-Verlag, 2019.

[14] Zotter, F., "Sampling Strategies for Acoustic Holography / Holophony on the Sphere," in *Proc. of the Annual German Conference on Acoustics (DAGA)*, Rotterdam, The Netherlands, 2009.

[15] Womersley, R. S. and Sloan, I. H., "How Good can Polynomial Interpolation on the Sphere be?" *Advances in Computational Mathematics*, 14, pp. 195 – 226, 2001.

[16] Fliege, J. and Maier, U., "The distribution of points on the sphere and corresponding cubature formulae," *IMA Journal of Numerical Analysis*, 19(2), pp. 317 – 334, 1999.

[17] Hardin, R. and Sloane, N., "McLaren's Improved Snub Cube and Other New Spherical Designs in Three Dimensions," *Discrete Computational Geometry*, 15(4), pp. 429 – 441, 1996.

[18] 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.

[19] Daniel, J., Rault, J.-B., and Polack, J.-D., "Acoustic properties and perceptive implications of stereophonic phenomena," in *Proc. of the 16th AES International conference on Spatial Sound Reproduction*, Arktikum, Rovaniemi, Finland, 1999.

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

[21] Kronlachner, M. and Zotter, F., "Spatial transformations for the enhancement of Ambisonic recordings," in *Proc. of the 2nd International Conference on Spatial Audio (ICSA)*, Erlangen, Germany, 2014.

[22] 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.

Annex 1: Numerical evaluation of the conversion factors

Due to the factorial function, the expression of the conversion factor $(\alpha_n^{\pm n})^{\mathrm{N3D}\to\mathrm{N2D}}$ in Eq. 8 is not suitable for numerical evaluation with somewhat large values of n. Instead, taking the logarithm of Eq. 8, we have $\forall n \in \mathbb{N}$

$$\begin{aligned}\ln\left(\alpha_n^{\pm n}\right)^{\mathrm{N3D}\to\mathrm{N2D}} &= n\,\ln 2 \\ &+ \ln\Gamma(n+1) \\ &- \frac{1}{2}\,\ln\Gamma(\,2(n+1)\,)\,. \end{aligned} \tag{26}$$

where $\Gamma(x)$ denotes the Gamma function and ln the natural logarithm. Now, introducing the Beta function $B(x,\,y)$ (a.k.a. Euler integral of the first kind), which satisfies

$$B(x,\,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}\,, \tag{27}$$

and using the Legendre duplication formula

$$\Gamma(2\,x) = (2\,\pi)^{-1/2}\;(2)^{2x-1/2}\;\Gamma(x)\;\Gamma\left(x+\frac{1}{2}\right)\,, \tag{28}$$

we can further simplify Eq. 8 and 26 to $\forall n \in \mathbb{N}$

$$\left(\alpha_n^{\pm n}\right)^{\mathrm{N3D}\to\mathrm{N2D}} = \sqrt{\frac{1}{2}\;B\left(n+1,\,\frac{1}{2}\right)}\,. \tag{29}$$

Finally, using Stirling's approximation for large values of n, we establish that

$$\left(\alpha_n^{\pm n}\right)^{\mathrm{N3D}\to\mathrm{N2D}} \sim \left(\frac{\pi}{4\,(n+1)}\right)^{1/4}\,. \tag{30}$$

A numerical evaluation of the $(\alpha_n^{\pm n})^{\mathrm{N3D}\to\mathrm{N2D}}$ conversion factors using Eq. 29 and Eq. 30 is presented in Figure 5. It can be observed that Stirling's formula actually provides a fair approximation even for small values of n, the relative error being less than 2% as soon as $n \geq 2$.

The Schmidt semi-normalization (noted SN3D and SN2D) is also popular in Ambisonic applications [22]. Using Eq. 29, we can easily derive the (numerically stable expression of the) conversion factor for this scheme :

$$\left(\alpha_n^{\pm n}\right)^{\mathrm{SN3D}\to\mathrm{SN2D}} = \begin{cases} 1, & \text{if } n = 0 \\ \sqrt{\frac{2n+1}{4}\;B\left(n+1,\,\frac{1}{2}\right)}, & \text{if } n \neq 0 \end{cases} \tag{31}$$

and, for large values of n,

$$\left(\alpha_n^{\pm n}\right)^{\mathrm{SN3D}\to\mathrm{SN2D}} \sim \sqrt{n+\frac{1}{2}}\;\left(\frac{\pi}{4\,(n+1)}\right)^{1/4}\,. \tag{32}$$

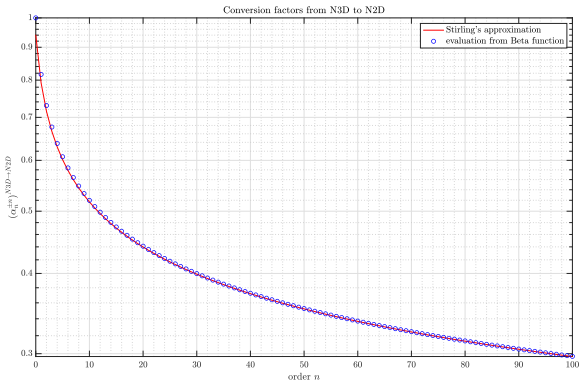

Fig. 5: Numerical evaluation of the conversion factors $(\alpha_n^{\pm n})^{\mathrm{N3D}\to\mathrm{N2D}}$. Using the Beta function Eq. 29 (blue dots), and Stirling's approximation Eq. 30 (red line) for large values of n.