HAL
open science

Réduction horizontale d'un flux Ambisonic 3D

Thibaut Carpentier

To cite this version:

Thibaut Carpentier. Réduction horizontale d'un flux Ambisonic 3D. 16ème Congrès Français d'Acoustique, CFA2022, Société Française d'Acoustique; Laboratoire de Mécanique et d'Acoustique, Apr 2022, Marseille, France. hal-03848022

HAL Id: hal-03848022

https://hal.science/hal-03848022v1

Submitted on 10 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

16ème Congrès Français d'Acoustique
11-15 Avril 2022, Marseille

Réduction horizontale d'un flux Ambisonic 3D

T. Carpentier[a]

[a]STMS Lab – IRCAM, SU, CNRS, Ministère de la Culture – 1 place Igor Stravinsky, 75004 Paris, France

Cet article s'intéresse à la réduction dans le plan horizontal d'un champ sonore 3D encodé en Ambisonic. Pour réaliser cette projection en 2D, une approche naïve (mais fréquemment observée en pratique) consiste à ignorer les harmoniques tessérales et zonales pour ne conserver que les composantes sectorielles du flux Ambisonic. Toutefois cet élagage brutal de la représentation harmonique engendre une déformation spatiale du champ sonore (on peut notamment observer des lobes secondaires indésirables sur le diagramme de directivité équivalent). Une méthode mathématiquement plus rigoureuse consiste à appliquer des facteurs de conversion de normalisation, qui permettent de transcoder un flux SN3D en SN2D (par exemple). Dans cet article, nous démontrons que ce transcodage est équivalent à la décomposition en ondes planes du champ sonore 3D originel, suivi d'un ré-encodage 2D des ondes planes selon le formalisme des harmoniques cylindriques. Lors du décodage sur un dispositif horizontal de haut-parleurs, cette approche permet de préserver des effets de latéralisation induits par les informations d'élévation, comme cela a déjà été noté par Daniel et al.

1 Introduction

Le formalisme Ambisonic [2] découle de l'écriture de l'équation des ondes en coordonnées sphériques : un champ de pression acoustique peut alors s'exprimer sous la forme d'un développement en série de Fourier-Bessel sphérique. Ce développement fait apparaître des fonctions radiales (les fonctions sphériques de Bessel) et des fonctions à dépendance angulaire appelées harmoniques sphériques. Ces dernières forment une base orthonormale au sens du produit scalaire sphérique (intégration sur la surface de la sphère unité). En pratique, le développement de Fourier-Bessel est tronqué à un ordre N, donc à un nombre fini de composantes, et cette décomposition en harmoniques sphériques conduit aux équations d'encodage directionnel utilisées pour une représentation Ambisonic 3D (parfois dite "périphonique").
Lorsqu'on s'intéresse à une représentation bidimensionnelle (et aussi appelée "pantophonique"), c'est-à-dire restreinte au plan horizontal, le système de coordonnées cylindriques peut être utilisé, et la série de Fourier-Bessel correspondante fait alors apparaître des harmoniques circulaires pour les termes à dépendance angulaire. Les harmoniques circulaires sont liées aux fonctions sphériques (via un facteur de pondération qui sera discuté au paragraphe 2.4), et le formalisme restreint au plan horizontal équivaut à une coupe (pondérée) de la représentation 3D.

Ainsi, les représentations bi– et tri–dimensionnelle coexistent – sans être strictement uniformes – et l'une ou l'autre pourra être utilisée en fonction du contexte. Pour un ordre Ambisonic donné, les systèmes 2D sont moins complexes que leurs homologues 3D : ils nécessitent moins de canaux à l'encodage et moins de haut-parleurs lors du décodage ; en revanche, ils ne peuvent pas naturellement restituer des sources en hauteur. En pratique, les dispositifs de reproduction présentent souvent une densité de haut-parleurs supérieure dans le plan horizontal, et le format Ambisonic 2D peut a priori y offrir une meilleure résolution spatiale (et les capacités de localisation du système auditif sont optimales dans le plan horizontal). Ceci a conduit plusieurs auteurs [2, 19, 11, 7] à proposer des représentations hybrides ou des formats mixtes, de sorte à tirer profit des atouts respectifs des systèmes 2D et 3D.

Quoiqu'il en soit, il est parfois nécessaire de réduire un matériau encodé en 3D dans une représentation bidimensionnelle ; par exemple, cela peut être utile pour limiter la bande passante requise, ou pour décoder un flux Ambisonic 3D (par exemple capté par un réseau sphérique de microphones) sur un dispositif horizontal de haut-parleurs. Cette problématique a été abordée, à la marge, dans un certain nombre de travaux antérieurs [4, 3, 2]. Dans cet article, nous rassemblons de façon cohérente l'information disponible, et nous examinons quelques aspects pratiques. Nous montrons que l'opération de réduction peut être réalisée, de façon équivalente, en appliquant des formules de conversion de convention, ou en réalisant une décomposition en ondes planes puis en ré-encodant les ondes planes résultantes dans le formalisme 2D. Une approche quelque peu analogue a été précédemment présentée dans [15, 14], dans le but de piloter un réseau linéaire ou circulaire de haut-parleurs alimenté par un champ sonore tridimensionnel.

2 Notations et définitions

2.1 Harmoniques sphériques

Notons $\boldsymbol{\vartheta} \equiv (\theta,\ \phi)$ la direction angulaire dans le repère de coordonnées sphériques, avec $x = \cos\phi\sin\theta$, $y = \sin\phi\sin\theta$, et $z = \cos\theta$. Notons $\mathbb{I}_N$ l'ensemble des indices Ambisonic i.e. $\{(n,m) \in \mathbb{N}\times\mathbb{Z}\ :\ 0 \le |m| \le n \le N\}$, où N désigne l'ordre Ambisonic maximal ; le nombre de composantes Ambisonic χ_N vaut $\chi_N^{2D} = (2N+1)$ et $\chi_N^{3D} = (N+1)^2$ respectivement dans le cas 2D et 3D.
Les harmoniques sphériques à valeurs réelles $\mathrm{Y}_n^m(\boldsymbol{\vartheta})$ sont définies $\forall (n,m) \in \mathbb{I}_N$ par

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

où $\mathcal{A}_n^{|m|}$ est une constante scalaire qui dépend du type de normalisation retenu [2, 1].
$P_n^m(x)$ sont les fonctions de Legendre associées, définies par $\forall x \in \mathbb{R}\cap[-1,1]$, $\left\{(n,m)\in\mathbb{N}^2\ :\ m \le n\right\}$,

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

$P_n(x)$ sont les polynômes de Legendre, qui peuvent s'écrire selon la formule de Rodrigues $\forall x \in \mathbb{R}\cap[-1,1]$, $\forall n \in \mathbb{N}$,

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

2.2 Terminologie

Les composantes $\{(n,m) \in \mathbb{I}_N \ : \ m = 0\}$ sont appelées harmoniques "zonales". Elles ne dépendent pas de la longitude et elles divisent la sphère en zones parallèles à l'équateur.
Les composantes $\{(n,m) \in \mathbb{I}_N \ : \ |m| = n\}$ sont appelées harmoniques "sectorielles". Elles partagent la sphère en secteurs selon les méridiens.
Enfin les composantes $\{(n,m) \in \mathbb{I}_N \ : \ 0 < |m| < n\}$ sont appelées harmoniques "tésserales". Elles forment un motif "en damier" à la surface de la sphère (voir Figure 1).

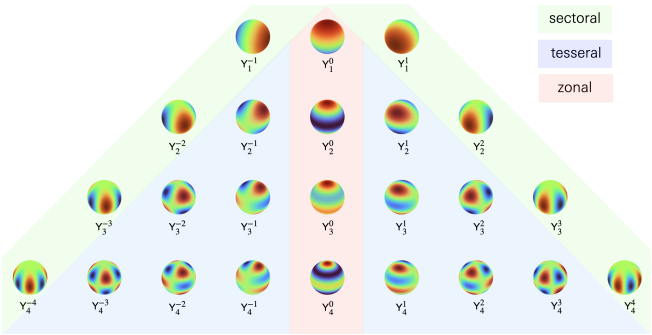

FIGURE 1 – Harmoniques sphériques zonales (rouge), sectorielles (vert), et tésserales (bleu) pour $1 \leq n \leq 4$.

2.3 Convention d'encodage

La convention N3D, ou normalisation-3D, est définie par $\{(n,m) \in \mathbb{N}^2 \ : \ m \leq n\}$

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

Dans le cas d'une représentation uniquement horizontale, seules les harmoniques sectorielles sont utilisées, et la convention N2D s'exprime $\forall n \in \mathbb{N}$ par

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

Avec la convention N2D, les fonctions $\mathbf{Y}_N^{\text{N2D}}(\boldsymbol{\vartheta})$ forment une base orthonormée d'harmoniques circulaires, pour le produit scalaire circulaire (évalué sur le cercle unité).

2.4 Facteurs de conversion

Les facteurs de conversion d'une convention $\mathcal{N}_1$ vers $\mathcal{N}_2$ sont notés $(\alpha_n^m)^{\mathcal{N}_1 \to \mathcal{N}_2}$, et définis tel que $\forall (n,m) \in \mathbb{I}_N$

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

Les facteurs de conversion satisfont la relation de réciprocité

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

Daniel [2] a établi une liste des principaux facteurs de conversion ; il en découle aisément que

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

Nous présentons en Annexe 1 un procédé numériquement stable pour le calcul de l'équation 8.

Le facteur de conversion N3D $\to$ N2D de l'équation 8 a également été obtenu par Thomas et al. [18], mais via une approche différente : en considérant que les transformées de Fourier spatiales cylindriques et sphériques doivent être égales pour une onde plane qui se propage dans le plan horizontal, et en constatant que $\forall\,(\theta,\ \phi)\,,\ \forall (n,m) \in \mathbb{I}_N$

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

Thomas et al. ont abouti à la solution suivante [1] $\forall n \in \mathbb{N}^*$

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

qui est équivalente à la formule de Daniel.

3 Réduction horizontale d'une représentation 3D

3.1 Par décomposition en ondes planes

L'encodage d'une onde plane avec une direction d'incidence $\boldsymbol{\vartheta_0} \equiv (\theta_0,\ \phi_0)$, et transportant un signal $s_0(t)$, en signaux Ambisonic 3D d'ordre N s'écrit

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

où $\mathbf{Y}_N^{\text{N3D}}(\boldsymbol{\vartheta_0})$ est un vecteur à χ_N^{3D} éléments :

$$\mathbf{Y}_N^{\text{N3D}}(\boldsymbol{\vartheta_0}) = \begin{bmatrix} \mathrm{Y}_0^0(\boldsymbol{\vartheta_0}) \\ \mathrm{Y}_1^{-1}(\boldsymbol{\vartheta_0}) \\ \mathrm{Y}_1^0(\boldsymbol{\vartheta_0}) \\ \mathrm{Y}_1^1(\boldsymbol{\vartheta_0}) \\ \vdots \\ \mathrm{Y}_n^m(\boldsymbol{\vartheta_0}) \\ \vdots \end{bmatrix}\ . \tag{12}$$

Le champ sonore $\mathbf{s}(t)$ peut également être analysé selon une décomposition en ondes planes [16]. Les contributions des ondes planes, c'est-à-dire la fonction de densité d'amplitude, peuvent être estimées par formation de voies (*beamforming*). Dans le domaine des harmoniques sphériques, la technique de formation de voies s'opère simplement par une pondération des composantes harmoniques : le signal capté par un microphone virtuel pointant dans la direction $\boldsymbol{\Omega}$ s'écrit (en supposant un diagramme de directivité à symétrie axiale) [17]

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

1. Le résultat établi dans [18] a été obtenu pour une base d'harmoniques sphériques complexes ; nous l'avons ici traduit pour des harmoniques réelles, par souci de cohérence avec la convention adoptée dans cet article.

où $\boldsymbol{w}(\boldsymbol{\Omega})$ est un vecteur à χ_N^{3D} éléments qui contient les poids de beamforming :

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

Pour opérer une décomposition en ondes planes, on utilise une formation de voies à directivité hypercardioïde (qui maximise l'indice de directivité) telle que

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

Afin de décomposer spatialement le champ sonore, des microphones virtuels (ou faisceaux) sont orientés vers K directions $\boldsymbol{\Omega_k}$. Ces K directions doivent être choisies adéquatement, de sorte à échantillonner la sphère unité tout en préservant la propriété d'orthonormalité (discrète) des harmoniques sphériques [21]. Par exemple, les grilles d'hyper-interpolation [20], de Fliege [8], ou des t-designs sphériques [10] peuvent être utilisés. Le nombre requis de directions K dépend des caractéristiques de la grille choisie, et dans tous les cas il est tel que $K \geq (N+1)^2$ pour éviter l'*aliasing* d'échantillonnage.
Le signal capté par le $k^{\text{ième}}$ faisceau hypercardioïde s'écrit

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

Ensuite, pour réduire le champ sonore en 2D, chacun des K signaux issus de la formation de voie est ré-encodé en Ambisonic 2D, suivant la direction $\boldsymbol{\Omega_k}$:

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

Le flux Ambisonic 2D, noté $\mathbf{y}(t)$, est alors la combinaison linéaire des K signaux ré-encodés

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

En introduisant la notation abrégée $\boldsymbol{\alpha}_N \equiv (\boldsymbol{\alpha}_N)^{\text{N3D}\to\text{N2D}}$, les facteurs de conversion s'écrivent, par définition,

$$\mathbf{Y}_N^{\text{N2D}}(\boldsymbol{\Omega}) = \boldsymbol{\alpha}_N \cdot \mathbf{Y}_N^{\text{N3D}}(\boldsymbol{\Omega}) \, , \tag{19}$$

et en insérant Eq. 19 dans Eq. 18, on obtient

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

Puisque l'ensemble des directions $\{\boldsymbol{\Omega_k}\}$ satisfait la propriété d'orthonormalité discrète des harmoniques sphériques (pour la convention N3D), la sommation sur K se simplifie :

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

Hormis la constante $\frac{K}{(N+1)^2}$, cette dernière équation correspond à l'encodage d'une onde plane transportant un signal $s_0(t)$, avec une direction d'incidence $\boldsymbol{\vartheta_0} \equiv (\theta_0, \; \phi_0)$, en Ambisonic 2D d'ordre N.
Autrement dit, décomposer spatialement un champ sonore 3D en K faisceaux (de directivité maximale) puis ré-encoder ces faisceaux en 2D, est strictement équivalent à l'encodage natif 2D de l'onde originelle.
Nous obtenons donc deux approches équivalentes pour "réduire" un flux Ambisonic 3D en 2D.

Notons que cette équivalence – démontrée dans le cas d'une unique onde plane – reste valide pour un champ sonore 3D constitué d'un nombre quelconque J de signaux sources

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

en vertu du principe de superposition.

3.2 En ignorant les harmoniques tésserales et zonales

Une autre procédure pour réduire un flux Ambisonic 3D en 2D consiste à simplement ignorer les harmoniques tésserales et zonales de la représentation tridimensionnelle, et ne conserver que les composantes sectorielles. D'après la section précédente (3.1), il est clair que cette troncation "naïve" n'est pas mathématiquement rigoureuse. Pourtant, on peut constater que cette approche est assez couramment utilisée en pratique. À des fins d'illustration, nous avons affiché sur la Fig. 2 le diagramme de directivité équivalent d'un champ sonore 3D réduit en 2D, avec application des facteurs de conversion $(\boldsymbol{\alpha}_N)^{\text{N3D}\to\text{N2D}}$, et sans (en supprimant naïvement les harmoniques tésserales et zonales). On peut observer que le lobe principal est globalement préservé ; toutefois des déformations des lobes secondaires sont manifestes.
En effet, en négligeant le facteur de conversion $(\boldsymbol{\alpha}_N)^{\text{N3D}\to\text{N2D}}$, on mésestime la contribution de l'angle d'élévation θ dans la représentation 2D.

4 Rôle de l'information θ dans la représentation Ambisonic 2D

Les signaux encodés dans le formalisme bidimensionnel (2D Ambisonic) véhiculent une information concernant l'angle d'élévation θ des sources virtuelles – toutefois avec une indétermination haut-bas :

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

Le rôle joué par cette information dans la représentation 2D a été minutieusement étudié par Daniel, notamment dans [2, 5, 6] : lorsque les signaux Ambisonic sont décodés sur un système pantophonique, il n'est pas possible de reproduire naturellement des effets de hauteur ; cependant, l'information θ peut être utile pour produire un effet de latéralisation.
Lorsqu'une source virtuelle est reproduite par L haut-parleurs,

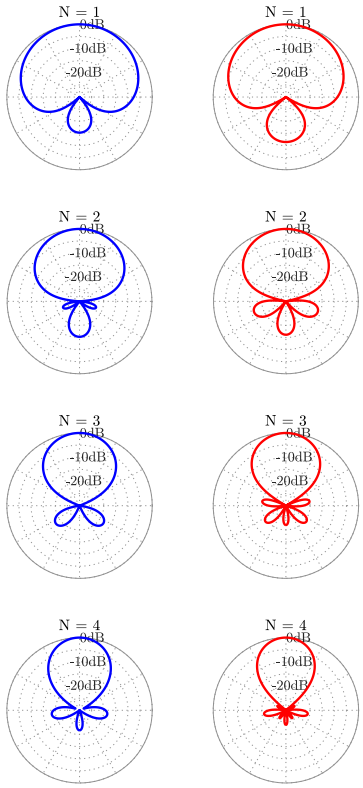

FIGURE 2 – Diagramme de directivité équivalent pour une source virtuelle d'incidence $\boldsymbol{\vartheta_0} \equiv (\theta_0,\ \phi_0) = (45°,\ 0°)$, encodée en Ambisonic 3D, et projetée en 2D. Côté gauche (en bleu) : en appliquant les formules de réduction tel que détaillé dans Eq. 21. Côté droit (en rouge) : en ignorant naïvement les harmoniques tésserales et zonales de la representation 3D, pour ne conserver que les harmoniques sectorielles. Les résultats sont affichés pour un ordre d'encodage maximal $1 \leq N \leq 4$.

de directions $\mathbf{u}_l$, alimentés par des gains réels G_l, le vecteur vélocité [9] est à valeur réelle, et il s'écrit (en supposant un auditeur centré et une superposition d'ondes planes) :

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

où r_v désigne la norme de $\mathbf{V}$.
Daniel et al. [5] ont montré que, en basses fréquences, le champ sonore synthétisé se comporte localement comme une onde plane de direction apparente $\mathbf{u}_v$ et de vitesse de propagation apparente c/r_v (où c désigne la vitesse du son) ; les différences interaurales de temps (ITD) résultantes, pour un auditeur en position centrale, sont plus faibles (avec un facteur multiplicatif r_v) que les ITD perçues dans le cas d'une onde plane naturelle de même incidence. Puisque r_v dépend implicitement de l'angle d'élévation θ de la source virtuelle, l'effet perceptif de latéralité est lui aussi fonction de θ (voir Fig. 3).

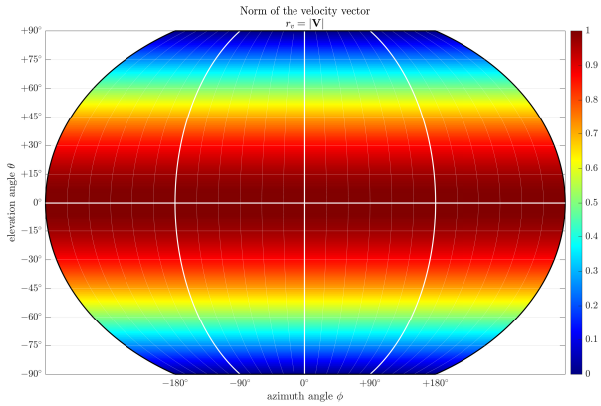

FIGURE 3 – Norme r_v du vecteur vélocité, en fonction de la direction $\boldsymbol{\vartheta} \equiv (\theta,\ \phi)$ de la source virtuelle, pour un flux Ambisonic 2D d'ordre N décodé sur $2N+2$ haut-parleurs régulièrement répartis sur un cercle.

Lorsque la tête de l'auditeur est fixe, l'effet de latéralisation est uniquement déterminé par la projection de $\mathbf{V}$ sur l'axe interaural $\mathbf{y}$. En raison de la symétrie axiale (en basses fréquences, les effets de diffractions sont négligeables et la tête de l'auditeur se comporte comme une sphère rigide), cela se traduit par un cône de confusion où les angles d'incidence des ondes ne peuvent être déterminés de façon univoque [6]. L'angle d'ouverture de ce cône d'ambiguïté est donné par $\gamma = \arccos(\mathbf{V} \cdot \mathbf{y})$, et est représenté sur la Figure 4.
Lorsque l'auditeur réalise de petites rotations autour de l'axe vertical $\mathbf{z}$ (rotations *yaw*), les variations de différences interaurales de phase peuvent être perçue comme un effet de hauteur "artificielle", et l'angle d'élévation perçu (mesuré par rapport au plan horizontal) est donné par $\hat{\theta} = \arccos(r_v)$.

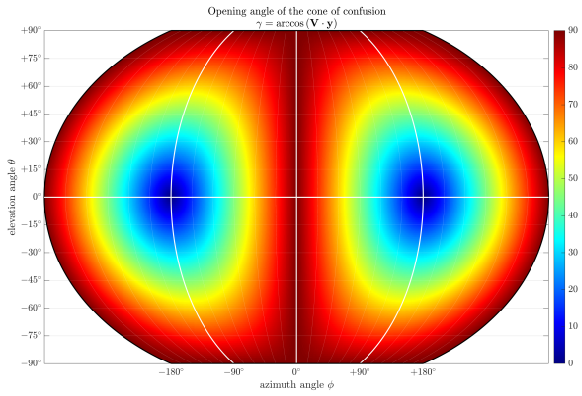

FIGURE 4 – . Angle d'ouverture γ du cône de confusion, en fonction de la direction $\boldsymbol{\vartheta} \equiv (\theta,\ \phi)$ de la source virtuelle, pour un flux Ambisonic 2D d'ordre N décodé sur $2N+2$ haut-parleurs régulièrement répartis sur un cercle.

5 Quelques considérations pratiques

Au vu des sections précédentes, nous soulignons que, lors du décodage d'un flux Ambisonic 3D sur un dispositif horizontal de haut-parleurs, il est préférable d'appliquer la formule de projection (voir Eq. 8), de sorte à préserver les indices de latéralisation véhiculés par l'information θ dans le

formalisme des harmoniques circulaires. Les harmoniques tésserales et zonales ne doivent pas être brutalement supprimées.

Pour autant, le décodage de contenus 3D sur des systèmes de reproduction horizontaux demeure une question ouverte pour nombre de productions, car des considérations artistiques doivent aussi être prises en compte, par-delà la stricte rigueur mathématique. Par exemple, l'indétermination haut-bas du formalisme Ambisonic 2D peut être problématique, car la scène décodée peut paraître irréaliste ou confuse. Le plus souvent, la portion inférieure de la scène 3D est indésirable lors d'une restitution pantophonique. De la même manière, les sons émanant de la partie supérieure de la sphère peuvent être sujets à controverse. Comme on peut l'observer sur la Figure 3, lorsque θ tend vers la position zénithale, le vecteur vélocité r_v tend vers 0 (il en est de même du vecteur énergie r_E). Pour une source sonore située exactement au zénith (ce que l'on désigne parfois comme "the voice of god"), le flux Ambisonic 2D se réduit même à la seule composante omnidirectionnelle (W), c'est-à-dire

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

En conséquence, les objets sonores en hauteur se voient distribués, lors du décodage, de façon homogène sur l'ensemble des haut-parleurs. Bien que cela puisse produire une sorte d'effet de survol (pour des sources en mouvement), ce comportement est parfois indésirable (tout dépend de l'intention du créateur de contenu ou du mixeur).

Les deux techniques de réduction présentées au paragraphe 3.1 (via les formules de conversion ou par une décomposition en ondes planes) sont théoriquement équivalentes ; cependant, la méthode par ondes planes issues de formation de voies se révèle en pratique plus adéquate, car elle offre plus de flexibilité dans le processus de réduction. En effet, la fonction de densité d'amplitude des ondes planes peut être modifiée, de sorte à adapter le contenu 3D pour une diffusion 2D. Typiquement, on pourra appliquer un gain (ou un filtrage) aux signaux extraits des microphones virtuels (Eq. 16), en fonction de leur direction d'observation $\boldsymbol{\Omega_k}$ – on choisira de préférence un gain qui est une fonction lente de l'angle θ_k. De la sorte, on peut atténuer ou éliminer les sons provenant de régions inférieures ou supérieures de la sphère. Cette stratégie est en substance similaire aux modifications directionnelles d'intensité proposées dans [12].
Une autre possibilité est de modifier les directions $\boldsymbol{\Omega_k}$ des faisceaux lors de l'étape de ré-encodage 2D (Eq. 17), par exemple en les "compressant" vers l'équateur. Cette déformation spatiale de la scène sonore est, elle aussi, similaire à l'effet de *warping* spatial présenté dans [12].

Mis à part les sons localisés, la projection 3D vers 2D peut également être problématique dans le cas de champs diffus. Il a déjà été montré [2] qu'un champ sonore 3D parfaitement diffus (infinité d'ondes planes décorrélées et provenant de façon isotrope de toutes les directions de la sphère) se traduit par une puissance égale entre toutes les composantes Ambisonic, lorsque la convention N3D est adoptée. De la même façon, un "champ diffus horizontal" (constitué d'ondes planes décorrélées et distribuées sur toute la surface du cercle unité) produit des signaux d'égale puissance pour un flux Ambisonic encodé selon N2D. En revanche, la projection 2D d'un champ diffus 3D n'est pas "horizontalement diffuse". Ceci est une conséquence directe des facteurs de conversion $(\boldsymbol{\alpha}_N)^{\text{N3D}\rightarrow\text{N2D}}$. Dans ce cas de figure, il peut être souhaitable d'ignorer la pondération $\boldsymbol{\alpha}_N$, de sorte à préserver la propriété de diffusivité lors de la réduction 2D. À nouveau, cette décision est du ressort du créateur de contenu.

6 Conclusion

Dans cet article, nous avons réuni des connaissances théoriques et pratiques concernant la réduction horizontale d'un contenu 3D encodé en Ambisonic. Bien que la diffusion sur un dispositif pantophonique de haut-parleurs ne permette pas de reproduire naturellement des sources en hauteur, l'information d'élévation véhiculée dans le format Ambisonic 2D peut être utile pour produire un effet de latéralisation ou un effet de hauteur "artificielle". Pour préserver ces propriétés, il est donc préférable d'appliquer les formules de projection lors du décodage d'un flux périphonique sur un système horizontal de haut-parleurs. Nous avons montré que la projection 2D vers 3D peut être réalisée, de façon équivalente, via les formules de conversion de normalisation, ou en réalisant une décomposition en ondes planes suivie d'un ré-encodage 2D des ondes planes résultantes. Néanmoins, la réduction 3D vers 2D ne peut pas s'automatiser de façon aveugle, car dans la plupart des productions, des choix doivent être faits en fonction du contenu de la scène sonore. Ces choix relèvent du mixeur ou du créateur de contenu. Enfin, l'approche par décomposition en ondes planes ici présentée offre une plus grande flexibilité pour adapter le contenu à un rendu bidimensionnel.

Références

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

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

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

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

[5] J. Daniel, J.-B. Rault, and J.-D. Polack. Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions. In

Proc. of the 105th AES Convention, San Francisco, CA, USA, September 1998.

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

[7] S. Favrot, M. Marschall, J. Käsbach, J. Buchholz, and T. Weller. 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, October 2011.

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

[9] M. A. Gerzon. General metatheory of auditory localization. In *Proc. of the 92nd AES Convention*, Vienna, Austria, Mar 1992.

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

[11] J. Käsbach, S. Favrot, and J. Buchholz. Evaluation of a mixed-order planar and periphonic ambisonics playback implementation. In *Proc. of Forum Acusticum*, Aalborg, Denmark, 2011.

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

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

[14] T. Okamoto. 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)*, pages 326 – 330, Shanghai, China, 2016.

[15] T. Okamoto. 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)*, pages 180 – 184, New Paltz, NY, USA, October 2017.

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

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

[18] M. R. P. Thomas, J. Ahrens, and I. Tashev. 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)*, pages 4723 – 4727, Florence, Italy, May 2014.

[19] C. Travis. A new mixed-order scheme for ambisonic signals. In *Proc. of the 1st Ambisonics Symposium*, Graz, Austria, June 2009.

[20] R. S. Womersley and I. H. Sloan. How good can polynomial interpolation on the sphere be ? *Advances in Computational Mathematics*, 14 :195 – 226, 2001.

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

Annexe : Évaluation numérique des facteurs de conversion 3D vers 2D

En raison de la fonction factorielle, l'expression des facteurs de conversion $\left(\alpha_n^{\pm n}\right)^{\text{N3D}\to\text{N2D}}$ dans l'Eq. 8 est inadaptée pour une évaluation numérique avec des valeurs de n relativement grandes. Pour y remédier, prenons le logarithme de l'Eq. 8, de sorte que $\forall n \in \mathbb{N}$

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

où $\Gamma\left(x\right)$ désigne la fonction Gamma et $\ln$ le logarithme népérien. En introduisant la fonction Beta $B\left(x,\ y\right)$ (aussi appelée intégrale d'Euler de première espèce), qui satisfait

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

et en utilisant la formule de multiplication de Legendre

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

il est possible de simplifier les équations 8 et 26 et, après simplifications, nous aboutissons à $\forall n \in \mathbb{N}$

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

Enfin, grâce à l'approximation asymptotique de Stirling pour n tendant vers l'infini, nous établissons que

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

En pratique, la formule de Stirling procure une approximation acceptable même pour des valeurs assez faibles de n. L'erreur relative est inférieure à 2% dès que $n \geq 2$.

Les harmoniques semi-normalisées au sens de Schmidt (notées SN3D et SN2D) sont aussi fréquemment employées dans la communauté Ambisonic [13]. En utilisant Eq. 29, on peut facilement établir l'expression numériquement stable pour le facteur de conversion correspondant :

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

et donc pour de grandes valeurs de n,

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