

Conception et expérimentation d'un système de prise de son et de mixage basé sur la technique de spatialisation sonore Higher Order Ambisonics (HOA)

Sébastien Moreau, Perrine Ganjean

Salle Pleyel
252, rue du Faubourg Saint Honoré
75008 Paris
E-mail : smoreau@sallepleyel.fr
pganjean@sallepleyel.fr

Résumé

Dans le cadre de notre projet portant sur la prise de son et le mixage HOA (Higher Order Ambisonics), un système de captation a été conçu sur la base d'un dispositif circulaire de microphones. Ce système permet en théorie d'obtenir des signaux HOA en appliquant successivement sur les signaux captés des opérations de matricage (combinaison linéaire) et de filtrage. Sa réalisation pratique a pris la forme d'une structure modulable dont les principaux paramètres (taille, nombre et modèle des microphones notamment) se modifient aisément en fonction du contexte musical et des préférences de l'ingénieur du son. Divers enregistrements ont été réalisés et ont permis de valider globalement l'étude théorique. L'analyse du rendu des timbres instrumentaux et de l'espace sonore a révélé l'impact des différents paramètres du dispositif sur la perception auditive et, pour certains, l'intérêt de les choisir en s'écartant des recommandations théoriques. Enfin, la technique HOA a été intégrée dans une station multipiste audionumérique de manière à pratiquer le mixage avec une ergonomie comparable à celle du mixage stéréophonique traditionnel.

Abstract

As part of our project on sound recording and mixing with Higher Order Ambisonics (HOA), a pick-up system has been designed on the basis of a circular microphone array. This system allows in theory to obtain HOA signals by successively matrixing and filtering the picked up signals. Its practical implementation took the form of a flexible structure whose main parameters (especially size, number and type of microphones) are modified according to the musical context and the sound engineer's preferences. Various recordings were made and have validated the theoretical study. The analysis of the rendering of instrumental timbres and sound space revealed the impact of the different array parameters on the auditory perception and the interest to choose some of these parameters by deviating from the theoretical recommendations. Finally, the HOA technology has been integrated into a digital audio workstation to practise the HOA mixing with ergonomics comparable to traditional stereo mixing.



Higher Order Ambisonics (HOA) [1,2] est une évolution de la technique de spatialisation sonore *Ambisonics* apparue dans les années 1970 sous l'influence notable de Michael Gerzon. Cette dernière permet de reproduire un champ sonore 3D grâce à la connaissance de ses caractéristiques directionnelles en un point : les composantes omnidirectionnelle (W) et bidirectionnelles en 3D (X, Y, Z) constituant le Format B (*B-Format*) [3]. Cependant, sa faible résolution spatiale limite la reconstruction du champ sonore à une petite zone de l'espace, notamment en haute fréquence. *Higher Order Ambisonics* permet l'amélioration de cette résolution (et donc d'étendre le *sweet spot*¹ lors de la restitution) en y ajoutant des composantes supplémentaires issues de la décomposition du champ acoustique sur une base de fonctions propres. Ces fonctions propres, associées à l'équation d'onde en coordonnées sphériques (3D),

combinent les harmoniques sphériques et des fonctions de Bessel sphériques. Si l'on restreint l'espace à deux dimensions, les fonctions propres en question deviennent celles de l'équation d'onde en coordonnées polaires et associent dans ce cas les harmoniques cylindriques et des fonctions de Bessel de première espèce. Ces aspects théoriques peuvent paraître complexes, mais ils fondent la légitimité de la technique HOA.

La première approche pratique de la prise de sons basée sur la théorie HOA proposée par Craven et Gerzon a abouti à la mise au point du microphone Soundfield sous la forme d'une combinaison tétraédrique de microphones à directivité hypercardioïde quasi coïncidente, associé à des outils de matricage pour l'encodage et le décodage des signaux [5].

¹ - Le *sweet spot* est un terme utilisé par les ingénieurs du son pour décrire le point focal

Aujourd'hui, dans son principe le plus général, un système de prise de son HOA prend la forme d'un réseau circulaire ou sphérique de microphones dont le but est d'échantillonner l'onde acoustique. Les signaux microphoniques sont destinés à être projetés sur la base des fonctions propres mentionnées précédemment, ce qui définit l'opération d'encodage dans le domaine HOA (plus synthétiquement appelé «encodage HOA»). Il en résulte un ensemble de signaux HOA qui constituent une nouvelle représentation de la scène sonore dans un domaine dual des coordonnées d'espace. Cette représentation possède un certain nombre d'avantages. Il s'agit d'abord d'une représentation **hiérarchique** dans laquelle les signaux sont regroupés par ordre ($m=1$ à M), le niveau de précision spatiale augmentant avec l'ordre. Ainsi, les signaux des premiers ordres se suffisent pour représenter l'intégralité de la scène sonore, mais l'information spatiale associée est pauvre. À l'extrême, l'ordre 0 correspond à une prise de son monophonique et est en soi une représentation complète et parfaitement exploitable de la scène sonore, mais qui ne contient quasiment pas d'information spatiale (en dehors de l'information de distance liée au rapport entre l'énergie du son direct et l'énergie du champ réverbéré). La prise en compte des ordres supérieurs ne permet alors que d'augmenter la résolution spatiale de la scène. Le second avantage de la représentation HOA est qu'elle définit un format **générique** indépendant à la fois du format de prise de son (i.e. les signaux issus des microphones) et du format de restitution (i.e. les signaux alimentant les haut-parleurs). En effet, les signaux HOA doivent passer par une étape préalable de matricage (décodage HOA) avant diffusion sur un dispositif de haut-parleurs. Au passage, on notera que, contrairement à d'autres techniques de spatialisation, la technique HOA n'est pas associée à un dispositif de restitution spécifique : n'importe quelle configuration de haut-parleurs peut être utilisée (même si toutes ne sont pas aussi performantes pour restituer au mieux l'ensemble des caractéristiques spatiales du champ sonore en question). C'est justement le rôle de l'opération de décodage d'adapter les signaux HOA au système d'écoute. Enfin, dernier avantage de la représentation HOA : dans le domaine dual, il est plus facile de **manipuler l'espace sonore** (rotation, zoom).

Si ces propriétés font d'HOA une technique particulièrement adaptée aux nouvelles technologies interactives de diffusion de contenus multimédias, son développement se heurte actuellement au manque d'outils mis à disposition des ingénieurs du son pour l'enregistrement et la postproduction. Nous présentons dans cet article l'état d'avancement des travaux que nous menons à la salle Pleyel sur la prise de son et le mixage HOA. La première partie expose les fondements théoriques de la technique de spatialisation dans un espace à deux dimensions décrit par un système de coordonnées polaires (r, θ). La deuxième partie s'intéresse à la conception d'un système de prise de son HOA sous la forme d'un dispositif circulaire de microphones. La troisième partie concerne la mise en œuvre de la méthode de prise de son et fait le bilan des premiers enregistrements réalisés. Enfin, l'intégration du dispositif de captation et plus largement de la technique HOA dans une station audionumérique de mixage et de postproduction est décrite et discutée dans la quatrième partie.

Higher Order Ambisonics (HOA)

Représentation du champ sonore : les signaux HOA

Pour représenter une scène sonore en deux dimensions et permettre sa reproduction, la technique de spatialisation sonore HOA se fonde sur le développement en série du champ de pression acoustique sur la base des fonctions propres de l'équation d'onde en coordonnées cylindriques (rayon r , azimut θ). Dans une zone de l'espace centrée à l'origine du repère et ne comprenant pas de sources sonores, le champ de pression acoustique p s'exprime ainsi sous la forme d'une série de Fourier-Bessel [1, 2] :

$$p(kr, \theta) = \sum_{m=0}^{+\infty} i^m J_m(kr) \sum_{\sigma=\pm 1} B_m^\sigma Y_m^\sigma(\theta), \tag{1}$$

où $k = 2\pi f/c$ est le nombre d'onde, c la vitesse du son dans l'air et m l'ordre des composantes. Les fonctions propres associées des fonctions de Bessel de première espèce à dépendance radiale $J_m(kr)$ illustrées à la figure 1 à des harmoniques cylindriques $Y_m^\sigma(\theta)$ représentées pour les huit premiers ordres à la figure 2 et sont définies par :

$$\begin{cases} Y_0^1(\theta) = 1 \\ Y_m^\sigma(\theta) = \sqrt{2} \cos(m\theta) \text{ si } \sigma = 1 \\ Y_m^\sigma(\theta) = \sqrt{2} \sin(m\theta) \text{ si } \sigma = -1 \end{cases} \tag{2}$$

Les coefficients B_m^σ , qui représentent les coefficients de la série de Fourier-Bessel, suffisent à décrire entièrement le champ acoustique et constituent dans la technique HOA le format de représentation du champ sonore. Exprimés ici implicitement dans le domaine fréquentiel, ces coefficients constituent des signaux audio au même titre que des signaux monophoniques, stéréophoniques ou binauraux. Nous les appelons signaux HOA.

La série de Fourier Bessel (1) implique une infinité de signaux HOA B_m^σ . En pratique, la représentation d'une scène sonore doit être limitée à un ordre maximal M fini, soit à un nombre restreint de $2M+1$ signaux HOA. Cette restriction du nombre de signaux correspond à une troncature de l'équation (1) qui ne permet plus de décrire exactement le champ sonore mais seulement de façon approximative. La précision de l'approximation qui influera directement lors de la reproduction sur la taille de la zone d'écoute [2, 4], augmente avec l'ordre M . Le format 2D Ambisonics ou B-Format introduit par Gerzon est un cas particulier de cette représentation HOA qui revient à limiter l'ordre maximal de la décomposition à $M=1$. Seules les trois premières composantes sont ainsi prises en compte : $B_0^1(W), B_1^1(X)$ et $B_1^{-1}(Y)$.

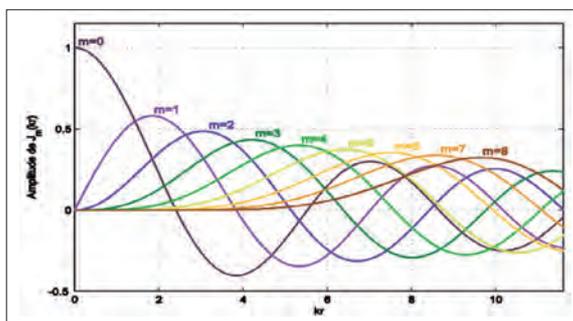


Fig. 1 : Illustration des fonctions de Bessel J_m en fonction de kr pour les ordres $m = 0$ à 8
Illustration of Bessel functions J_m as a function of kr up to order $m = 8$

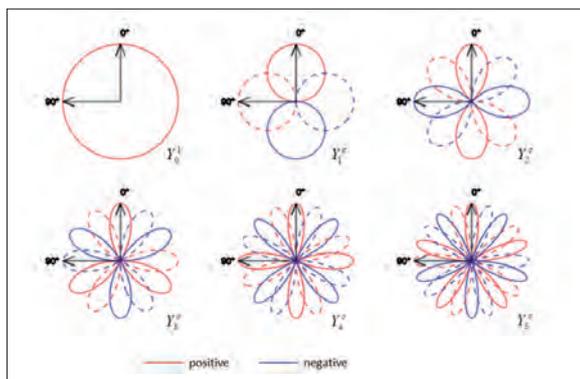


Fig. 2 : Illustration des harmoniques cylindriques pour les ordres $m = 0$ à 5
 Cylindrical harmonics up to order 5

Encodage spatial HOA

Comme en stéréophonie, deux méthodes distinctes permettent d'encoder (créer) des signaux HOA. La première consiste à simuler la position des sources sonores en appliquant sur des signaux audio monophoniques des gains adaptés (panpot^2 d'amplitude). La seconde méthode a pour but la captation directe de signaux HOA naturels à partir d'un dispositif microphonique.

La synthèse HOA peut être illustrée avec le cas simple d'une onde plane d'incidence θ S transportant un signal S. La décomposition associée en série de Fourier-Bessel est connue [12]. On en déduit aisément les signaux HOA qui s'obtiennent en pondérant le signal audio par les harmoniques cylindriques [2] :

$$B_m^\sigma(\theta) = S \cdot Y_m^\sigma(\theta_S). \quad (3)$$

Cette équation livre une nouvelle interprétation de l'encodage spatial : on se rend compte qu'il s'apparente à une prise de son par des microphones directifs 2D dont la directivité s'identifie aux harmoniques cylindriques. Les signaux HOA d'ordre 0 et 1 seraient ainsi issus respectivement d'un microphone virtuel omnidirectionnel et de deux microphones virtuels bidirectionnels agencés perpendiculairement (cf. Figure 2), et ainsi de suite.

L'encodage d'une scène sonore naturelle est beaucoup moins évident. En effet, s'il est possible d'enregistrer avec des microphones usuels les signaux HOA jusqu'à l'ordre 1, il n'existe pas de directivités équivalentes aux harmoniques cylindriques d'ordres supérieurs.

D'autre part, l'enregistrement direct des signaux HOA nécessiterait de placer tous les microphones en un même point de référence, ce qui n'est évidemment pas réalisable.

Quelques solutions ont déjà été proposées par le passé, concernant notamment la prise de son en trois dimensions à partir de structures microphoniques sphériques [4-7]. Nous présentons dans la deuxième partie, une méthode d'enregistrement 2D qui permet d'obtenir des signaux HOA naturels d'ordre quelconque à partir d'un dispositif microphonique circulaire.

2- La panoramique sonore est la répartition des sons entre les différentes voies de diffusion du contenu sonore.

Que son interface de gestion soit matérielle ou logicielle, elle tend à se présenter sous la forme standard d'un taper pour les contextes stéréo (deux canaux), c'est-à-dire une roulette (éventuellement crantée, à course finie ou non, etc.) ou un élément d'interface graphique qui permet, pour une piste stéréo donnée, d'augmenter ou diminuer l'intensité (dB) relative du signal entrant entre les deux canaux de sortie (qui sont en fait, dans la plupart des cas, des bus). On parle de pan pot pour panoramic potentiometer, c'est-à-dire potentiomètre panoramique.

Décodage spatial HOA

Contrairement aux signaux stéréophoniques, les signaux HOA sont indépendants du système de diffusion et doivent être adaptés en fonction du contexte d'écoute avec le but de recréer les composantes harmoniques originales de la série de Fourier Bessel (principe de ré-encodage) [2]. Ce processus d'adaptation, ou encore de décodage HOA, consiste à alimenter chaque haut-parleur du dispositif en fonction de sa position angulaire par un signal combinant les signaux HOA de façon linéaire [2] :

$$S_l(\theta) = \sum_{m=0}^M \sum_{\sigma=\pm 1} g_l^{m\sigma} B_m^\sigma, \text{ avec } 0 \leq l \leq L. \quad (4)$$

L est le nombre total de haut-parleurs du dispositif de diffusion et $g_l^{m\sigma}$ les gains pondérant les signaux HOA alimentant le $l^{\text{ème}}$ haut-parleur et déterminés en fonction de sa position angulaire.

Par exemple, en disposant en cercle et de façon régulière $L=2M+2$ haut-parleurs (ce qui garantit une reconstruction du champ acoustique conforme à celui décrit par des signaux HOA d'ordre M [2, 4]), l'application du principe de ré-encodage permet de calculer les gains [2] :

$$g_l^{m\sigma} = \frac{1}{2M+2} \cdot Y_m^\sigma(\theta) \quad (5)$$

Il existe également des procédures d'optimisation du calcul de ces gains $g_l^{m\sigma}$ ayant pour but d'améliorer du point de vue perceptif la restitution sonore dans certains cas particuliers (trop faible résolution des signaux HOA pour la zone d'écoute visée par exemple) [2].

Conception d'un dispositif circulaire de microphones pour l'enregistrement HOA

Fondements théoriques

Puisque la captation directe des signaux HOA caractérisant une scène sonore naturelle n'est pas réalisable, une méthode alternative doit être envisagée. Cette méthode est suggérée par la définition des signaux HOA contenu dans la série de Fourier-Bessel (1) de manière implicite [8]. Les signaux HOA étant les coefficients du développement de la pression acoustique sur la base des fonctions propres, il suffit pour les obtenir, de projeter la pression sur chaque fonction propre :

$$B_m^\sigma = \frac{1}{i^m J_m(kr)} \cdot \frac{1}{2\pi} \int_{\theta=0}^{2\pi} p(kr, \theta) Y_m^\sigma(\theta) d\theta. \quad (J_m(kr) \neq 0) \quad (6)$$

Cette équation s'interprète de la façon suivante : Ayant fixé le rayon à une valeur finie $r=R$, la première étape consiste à enregistrer la pression acoustique avec des microphones omnidirectionnels disposés sur un périmètre circulaire continu, puis ces signaux microphoniques sont combinés linéairement (opération symbolisée par l'intégrale), enfin un filtrage $1/i^m J_m(kr)$ est appliqué.

Deux difficultés doivent cependant être surmontées pour mettre en œuvre un tel dispositif de prise de son : l'emploi d'un nombre fini et limité de microphones, alors que l'intégrale dans l'équation (6) repose sur un signal continu de pression, et la réalisation des filtres $1/i^m J_m(kr)$, en particulier pour les fréquences où les fonctions de Bessel s'annulent.

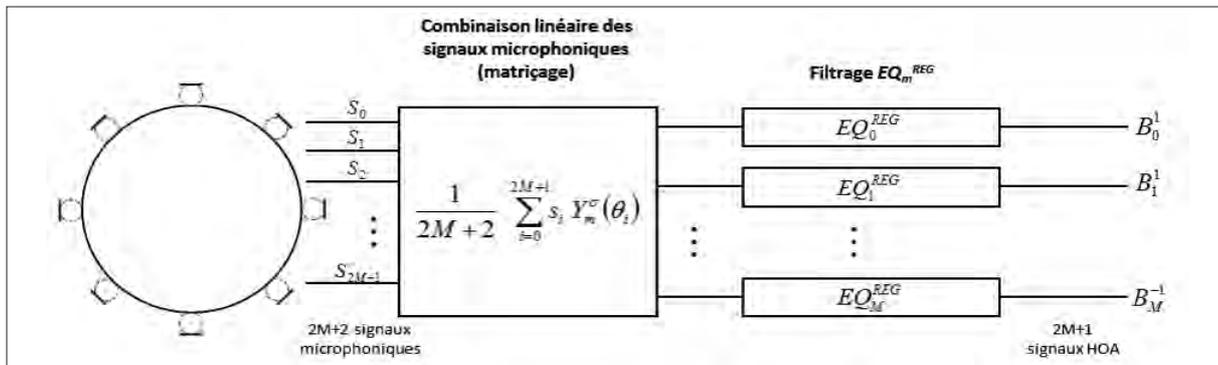


Fig. 3 : Représentation schématique d'un dispositif de prise de son HOA
Schematic view of a HOA microphone

Discrétisation du cercle microphonique continu

L'emploi d'un nombre fini de microphones ne permet pas d'enregistrer la pression acoustique sur un périmètre circulaire continu. Cependant, par analogie au théorème classique d'échantillonnage de Shannon dans le domaine temporel, les signaux HOA peuvent être déduits de façon exacte jusqu'à un ordre fini M en positionnant $2M+2$ microphones aux sommets d'un polygone régulier (échantillonnage régulier de l'azimut), soit aux angles :

$$\theta_i = \frac{\pi i}{2M+2}, \quad i = 0, \dots, 2M+1 \quad (7)$$

Les signaux HOA s'obtiennent alors grâce à la formule suivante, correspondant à la discrétisation de l'équation (6) :

$$B_m^\alpha(\omega) = \frac{1}{i_m J_m(kr)} \cdot \frac{1}{2M+2} \sum_{i=0}^{2M+1} p(kr, \theta_i) \quad (8)$$

La discrétisation du cercle microphonique engendre un double phénomène d'aliasing qui parasite l'acquisition des signaux HOA. Dans le domaine spatial, l'effet résulte du sous-échantillonnage des modes harmoniques supérieurs à M (fréquences spatiales) et tout de même présents dans les signaux captés par les microphones en fonction du rayon du dispositif de prise de son et de la fréquence (cf. Figure 1). Dans le domaine temporel, la fréquence f_{al} au-dessus de laquelle le phénomène de repliement apparaît dépend, selon le critère de Shannon, de la distance Δd entre deux microphones adjacents :

$$f_{al} = \frac{c}{2\Delta d} \quad (9)$$

où c est la vitesse du son dans l'air.

Problèmes liés à la mise en œuvre du filtrage

L'étape de filtrage $1/i^m J_m(kr)$ n'est pas réalisable en l'état puisqu'il est indéterminé lorsque les fonctions de Bessel s'annulent. D'autre part, lorsque les fonctions de Bessel ont une valeur proche de zéro, les filtres se caractérisent par des phénomènes excessifs d'amplification qui dénotent un manque d'information utile dans les signaux captés et rendent le dispositif d'enregistrement particulièrement sensible aux imperfections comme le bruit de fond des microphones et leurs erreurs de positionnement par rapport au modèle théorique.

Pour éviter l'annulation du terme de filtrage, une solution consiste à remplacer les microphones de pression par des microphones cardioïdes (ou plus généralement une

combinaison de microphones de pression et de microphones à gradient de pression), auquel cas le terme de filtrage devient [2, 4] :

$$EQ_m = \frac{1}{i^m [\alpha J_m(kR) - i(1-\alpha)J'_m(kR)]} \quad (10)$$

avec $(i^m [\alpha J_m(kR) - i(1-\alpha)J'_m(kR)] \neq 0)$,

où α caractérise le coefficient de cardioïcité des microphones ($0 \leq \alpha \leq 1$). On constate alors que, pour $\alpha \neq 0$ et $\alpha \neq 1$, EQ_m est défini pour l'ensemble des fréquences audibles et pour tout dispositif de rayon non nul.

Cependant, puisque le dénominateur dans l'expression d' EQ_m tend vers zéro avec la fréquence d'autant plus vite que l'ordre est grand, il subsiste dans ces filtres une amplification importante en basse fréquence d'autant plus prononcée que l'ordre est élevé [4, 8]. Or, les zones fréquentielles concernées par ces amplifications excessives ne contribuent pas significativement à la reconstruction de la scène sonore si l'on restreint la zone d'écoute à un disque de rayon fixe et que l'on souhaite obtenir cette reconstruction avec une résolution homogène sur l'ensemble du spectre des fréquences audibles [4, 8]. Ces zones fréquentielles concernent en effet dans ce cas le champ acoustique au-delà de la zone d'écoute. Par conséquent, il n'est pas utile de s'obstiner à extraire en force des données du champ acoustique faiblement présentes dans le signal et corrompues par l'effet dominant des imperfections du système de prise de son (bruit de fond et erreurs de positionnement des microphones notamment). Un processus de régularisation de Tikhonov limitant les phénomènes d'amplification en basse fréquence peut donc être appliqué sur les filtres instables EQ_m . Les filtres résultants s'expriment ainsi :

$$EQ_m^{REG}(kr) = \frac{\overline{EQ_m}(kr)}{|\overline{EQ_m}(kr)|^2 + \lambda^2} \quad (11)$$

où $\overline{EQ_m}$ désigne le conjugué complexe de EQ_m et λ est le paramètre de régularisation compris entre 0 et 1 dont la valeur idéale permet de préserver les informations utiles dans le signal et d'éliminer les informations inconsistantes. Comme illustré à la figure 3, les filtres EQ_m^{REG} se substituent finalement dans notre procédé d'enregistrement aux filtres $1/i^m J_m(kr)$ théoriques.

Pratique de la prise de son HOA

Présentation du dispositif de prise de son et des enregistrements effectués

L'étude théorique rapportée précédemment nous a permis de réaliser un dispositif expérimental de prise de son HOA [9]. Nous avons fait pour cela fabriquer sur mesure une structure en étoile, constituée d'un anneau central percé sur la tranche tous les 5° pour y fixer des tiges en aluminium de différentes longueurs (cf. Figure 4). Ce dispositif est facile à monter, sa taille et sa configuration angulaire sont variables et ses tiges s'adaptent à tous types de pinces microphoniques. Le matricage et le filtrage des signaux captés (cf. Figure 3) ainsi que le décodage des signaux HOA sont réalisés au moyen de *plugins VST³* (Orange Labs) [10], insérés dans le logiciel audio numérique *Reaper*.

Tous les traitements sont exécutés en temps réel pour permettre une écoute instantanée du rendu sonore sur différents systèmes de reproduction.



Fig. 4 : Dispositif de prise de son HOA
Microphone array for HOA recording

Plusieurs enregistrements HOA de différents ensembles musicaux ont été effectués à la salle Pleyel et au Conservatoire national supérieur de musique et de danse de Paris. Le dispositif microphonique HOA a été utilisé soit comme système principal de prise de son de manière analogue au couple stéréophonique (cf. Figure 5), soit comme système d'ambiance lors de concerts en public. Ces enregistrements ont été l'occasion d'expérimenter la prise de son HOA en faisant varier l'ordre (jusqu'à $M=4$), la taille, la directivité des microphones, et le paramètre de régularisation λ des filtres. Plus de détails sur les configurations testées sont données dans les références [8, 9].



Fig. 5 : Séance d'enregistrement HOA d'un quatuor de clarinette à la salle Pleyel

Premiers résultats des écoutes

Cette section synthétise les premiers résultats informels de nos expérimentations sur la prise de son HOA [8], [9].

Directivité des microphones

Les enregistrements ont été réalisés avec des microphones à directivités cardioïde (Neumann KM184), hypocarديوïde (Schoeps MK21), et omnidirectionnelle (DPA 4006). Lorsque cela était possible, les systèmes étaient montés conjointement avec un décalage de 5° sur l'anneau central de la structure pour que la comparaison soit la plus objective possible. Les microphones à directivité cardioïde ont rapidement été écartés malgré le rendu précis de la localisation des sources sonores car ils nous semblaient trop dégrader le timbre. Le tableau 1 donne une lecture parallèle des principales caractéristiques que nous avons associées à l'utilisation des microphones à directivités omnidirectionnelle et hypocarديوïde.

Microphones à directivité hypocarديوïde	Microphones omnidirectifs
Localisation assez précise	Localisation peu précise
Durété des timbres, manque de graves	Timbres naturels, réponse spectrale large
Sources assez proches, salle petite	Sensation d'enveloppement, salle large
Image plus robuste aux mouvements de l'auditeur	Faible robustesse de l'image

Tabl. 1 : Tableau comparatif des caractéristiques du rendu sonore en fonction de la directivité des microphones utilisés
Comparative view of sound characteristics as a function of microphone directivity

Rayon du dispositif HOA

Le choix du rayon du dispositif résulte d'un compromis difficile entre la volonté de limiter les phénomènes d'aliasing⁴ en haute fréquence et celle de disposer d'un maximum de données acoustiques exploitables en basse fréquence (cf. section 2.2 et 2.3). De plus, le placement du système microphonique, notamment lorsqu'il est utilisé comme dispositif principal de prise de son, est susceptible de faire varier de façon significative le rendu des timbres et de l'espace.

4- Dans le traitement du signal l'aliasing ou crénelage fait référence à un effet qui provoque des signaux différents pour devenir indiscernables. Il se réfère également à la distorsion qui se produit lorsque le signal reconstruit à partir d'échantillons est différent du signal continu d'origine.

3- Les plugins VST utilisés pour l'étude ont été développés par la société Orange Labs

Nous avons testé plusieurs tailles de rayon de 26 cm à 93 cm et tenté d'optimiser à chaque fois la position du dispositif de prise de son. Nous avons tout d'abord remarqué que les restitutions de l'acoustique du lieu et des timbres paraissaient d'autant plus naturelles et la sensation d'enveloppement plus agréable que le rayon était grand. À l'inverse, l'acoustique du lieu était moins prégnante et possédait une réverbération moins présente et plus courte lorsque le rayon diminuait. Si les choses sont moins nettes concernant la précision de la localisation des sources sonores, les dispositifs de grands rayons (plus de 60 cm) nous ont parus tout de même plus performants en général. Il semblerait enfin que la taille du dispositif soit faiblement dépendante de celle de l'ensemble musical enregistré puisque nous avons préféré un même rayon (86 cm) pour par exemple un orchestre symphonique et un quatuor de clarinettes.

Filtrage EQ_m^{REG}

Le choix du paramètre de régularisation λ qui permet dans l'idéal de filtrer les données acoustiques erronées dans les signaux audio sans éliminer les informations fiables (cf. Section 2.3) est délicat. En effet, notre parti pris pour la réalisation d'une structure microphonique facilement modulable nous empêche de baser l'optimisation des filtres sur des mesures acoustiques comme cela se ferait dans le cas d'une structure fixe. Nous avons par conséquent appliqué une méthode d'essai-erreur consistant à tester successivement différentes valeurs du paramètre λ jusqu'à aboutir au meilleur compromis sonore sur les rendus des timbres et de l'espace. Il est apparu logiquement que des valeurs trop élevées du paramètre λ (au-dessus de $\lambda=0,07$ dans notre cas) entraînaient une dégradation importante de la localisation des sources sonores et un déséquilibre spectral en faveur des basses fréquences.

À l'inverse, des valeurs trop basses de ce paramètre (en dessous de $\lambda=0,05$ dans notre cas) amélioreraient la précision de localisation mais provoquaient un déséquilibre spectral en faveur des fréquences hautes. En outre, le meilleur compromis obtenu ne permettait toujours pas de respecter convenablement les timbres des instruments. Nous avons finalement jugé préférable de supprimer complètement l'étape de filtrage (cf. Figure 3) pour que les timbres des instruments retrouvent leur naturel à la restitution. Cette solution induit en contrepartie une imprécision dans la localisation des sources sonores. Mais cette localisation reste globalement cohérente et pourra être améliorée en ajoutant des microphones d'appoints placés à proximité des instruments et replacés virtuellement dans l'espace lors de l'étape de mixage (cf. Section 4).

Confrontation entre la théorie et la pratique

Les premiers enregistrements que nous avons réalisés ont validé de manière générale la théorie de la prise de son HOA avec un dispositif microphonique circulaire. Ils nous ont également permis d'acquérir un peu plus d'expérience de terrain et d'apprécier l'importance pour le rendu sonore de chaque paramètre intervenant dans la conception du système. Nous en tirons quelques enseignements qui devront être confirmés et complétés par la suite. Premièrement, il apparaît clairement que l'emploi de microphones à directivité omnidirectionnelle doit être privilégié contrairement à ce que nous suggère la théorie.

En effet, si la précision de la localisation n'est pas aussi bonne qu'avec des microphones directifs, ils apportent un rendu plus fidèle des timbres instrumentaux et de l'acoustique du lieu qui nous paraissent essentiels.

Ensuite, le rayon du dispositif nous semble devoir être grand pour favoriser le rendu spatial, notamment la sensation d'enveloppement. En outre, les phénomènes d'*aliasing*⁴ spatial qui dégradent en théorie la qualité sonore d'autant plus que le rayon est grand, ne sont en réalité pas très gênants d'un point de vue perceptif.

Enfin, nous ne sommes pas encore parvenus à concevoir des filtres EQ_m^{REG} permettant de restituer convenablement le timbre des instruments. Contre toute attente cependant, la suppression complète de l'étape de filtrage s'est avérée être en pratique une solution satisfaisante. En effet, si la localisation des sources sonores est en contrepartie moins précise, le rendu de l'espace sonore global et des timbres instrumentaux nous a paru tout à fait bon.

Mixage et postproduction HOA

Présentation de l'environnement de travail

Le choix d'un environnement de travail pour le mixage et la postproduction HOA n'est pas aisé. En effet, la majorité des outils actuels (consoles de mixage et logiciels audio-numériques notamment) sont dédiés à la stéréophonie et ne permettent pas de manipuler directement des signaux HOA. Pour obtenir une ergonomie de mixage et de postproduction proche de celle de la stéréophonie, nous avons choisi de travailler avec le logiciel *Reaper*, dont la grande flexibilité autorise une intégration poussée de la technique HOA. Ce logiciel permet en effet de définir pour chaque piste et chaque bus interne le nombre de canaux qu'elle véhicule sans a priori sur le contenu. La technologie HOA en elle-même est intégrée sous la forme d'une suite logicielle de *plugins VST*³, et qui fournit les principaux outils de mixage : traitement liés au dispositif de prise de son HOA, *panpot*² d'amplitude pour le placement d'une source virtuelle dans l'espace, décodage des signaux HOA pour l'adaptation au contexte d'écoute. Ces *plugins* sont entièrement paramétrables grâce à des fichiers de configuration qui permettent de les adapter à la géométrie du dispositif de prise de son ou au système de diffusion par exemple. Ils s'insèrent directement sur les bus multicanaux par lesquels transitent les signaux HOA.

Exemple d'une session de mixage/postproduction

Pour illustrer de façon concrète le mixage HOA, nous présentons une session de travail basique réalisée pour l'enregistrement et la postproduction d'un quatuor de clarinette à la salle Pleyel [9]. Cet enregistrement a été effectué avec un dispositif de prise de son HOA circulaire d'ordre 4 constitué de dix microphones omnidirectionnels, et quatre microphones d'appoint à directivité hypocarديوide placés à proximité des instruments (cf. Figure 5). La figure 6, page suivante, est une capture d'écran de la fenêtre d'édition du logiciel *Reaper* montrant les cinq pistes audio : une piste à 10 canaux contenant les signaux captés par le dispositif microphonique HOA et 4 pistes monophoniques accueillant les signaux des microphones d'appoint. Les méthodes traditionnelles d'enregistrement et de montage multipistes peuvent être ici appliquées sans difficulté, y compris concernant la piste multicanale.



Fig. 6: Capture d'écran de la fenêtre d'édition du logiciel Reaper
Screen view of the Reaper edit window

Le routage des signaux et les différents traitements appliqués pendant le mixage sont représentés schématiquement à la figure 7.

Comme nous pouvons le constater, la méthode de travail est très proche de celle utilisée en stéréophonie traditionnelle. Un plugin VST inséré sur la piste du dispositif HOA de prise de son assure la conversion des signaux captés en signaux HOA. Chaque piste d'appoint comporte les effets audio habituels (délai, égalisation, compression, etc.) et un module VST de panoramique HOA qui convertit le signal monophonique en signaux HOA. Les niveaux d'envoi de chacune des tranches de la console virtuelle dans les bus Master ou Auxiliaire sont commandés par des faders⁵ mono ou multicanaux. Ne disposant pas d'effet audio au format HOA, notre configuration intègre une réverbération externe Lexicon 960L capable de traiter des signaux stéréophoniques 5.0.

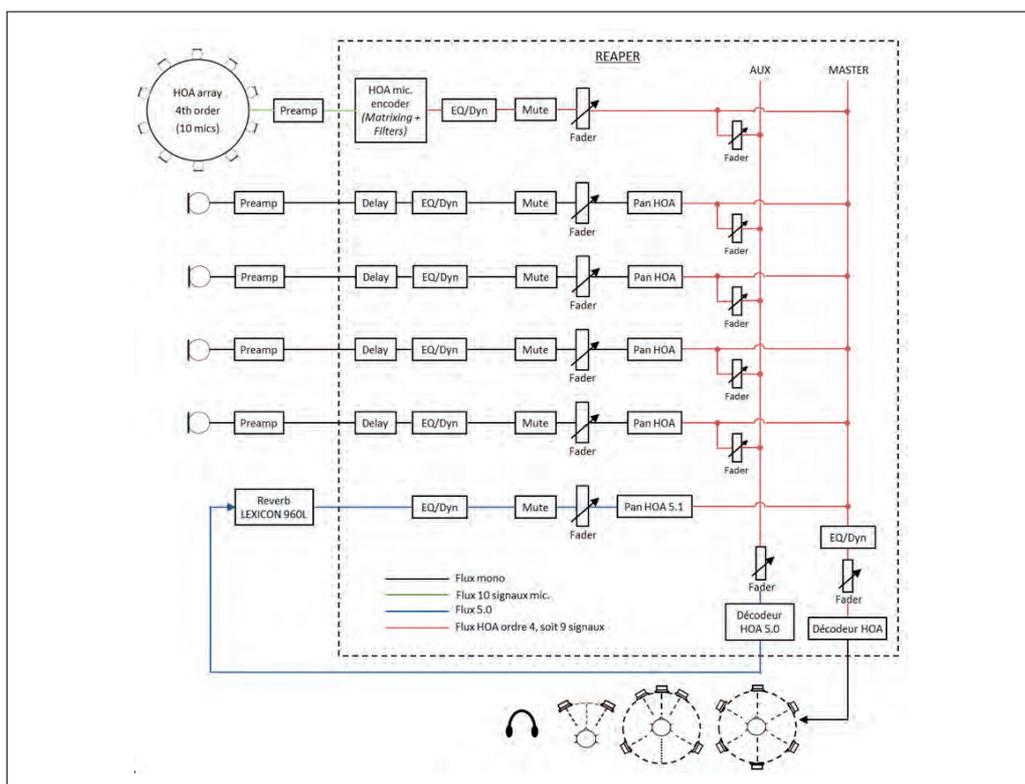
Pour cela, les signaux sortant du bus Auxiliaire dédié à cet effet sont précédemment convertis au format 5.0 par un décodeur HOA. Ceux qui retournent de l'appareil de réverbération subissent le traitement inverse : chaque signal est replacé dans l'espace par l'intermédiaire d'un plugin de panpot² HOA en fonction de la position du haut-parleur qu'il est censé alimenter. Enfin, le décodeur HOA inséré sur la tranche Master permet de modifier en temps réel le choix du système de diffusion pour le monitoring.

Perspectives de développement

Si les méthodes précédemment exposées nous semblent pouvoir favoriser le développement de la pratique de la prise de son et du mixage HOA, il reste encore beaucoup à faire.

Le principal problème réside actuellement dans le manque d'outils mis à la disposition des ingénieurs du son. Malgré la possibilité de travailler avec la suite de plugins VST³ et encore confidentielle, nous avons nous-mêmes été confronté à l'inexistence d'effets audio, notamment de réverbération, traitant directement des signaux HOA. Cela nous a obligé à complexifier la procédure de mixage en utilisant un format stéréophonique multicanal intermédiaire. Pour espérer faire d'HOA une vraie alternative à la stéréophonie multicanale et bénéficier de ses atouts dans le contexte actuel de déploiement des nouvelles technologies (diffusion Internet, par exemple), la priorité nous semble être de développer des outils de création sonore au format HOA, par exemple en intégrant cette technologie dans les logiciels et les consoles audionumériques (sans remettre en cause totalement la philosophie de travail issue de la pratique stéréophonique), et en développant des dispositifs de prise de son laissant si possible aux ingénieurs du son quelques libertés (structure modulable, choix des microphones, etc.).

Fig. 7 : Synthétique des traitements audio appliqués lors du mixage HOA d'un quatuor de clarinette



Nos premières expérimentations de prise de son ont montré qu'il existait une différence significative entre les préceptes issus de la spéculation théorique et ceux issus de l'action pratique. Il faudra par conséquent amplifier ces expérimentations pour éprouver en profondeur la technologie HOA, prendre du recul et forger finalement des automatismes comme cela s'est produit pour la stéréophonie. Des solutions devront également être trouvées à des problèmes très concrets comme la gêne occasionnée par le dispositif de prise de son lors d'un concert avec public.

Conclusion

L'étude de la théorie sur laquelle se fonde la technologie HOA nous a permis de concevoir un système de prise de son réaliste sous la forme d'un dispositif circulaire de microphones.

Pour expérimenter un tel système, nous avons fait fabriquer une structure microphonique en étoile avec une résolution angulaire de 5° et s'adaptant à tous types de microphones. Les outils nécessaires à la prise de son et au mixage HOA ont été intégrés dans une station audio-numérique *Reaper* au moyen de *plugins* VST développés par la société Orange Labs.

Les enregistrements réalisés ont permis de valider globalement la théorie tout en relativisant certains aspects. Nous avons choisi de privilégier le rendu spectral au détriment de la précision de localisation des sources sonores. Les meilleurs résultats ont ainsi été obtenus avec un système de prise de son basé sur des microphones omnidirectionnels et sans appliquer l'étape de filtrage, contrairement aux recommandations théoriques. L'effet produit par l'*aliasing*⁴ spatial s'est en outre révélé peu gênant, ce qui a permis l'utilisation de systèmes de prise de son de grande taille favorisant la sensation d'enveloppement sonore. Il a été rendu possible de mixer directement dans le format HOA les signaux issus du dispositif de captation avec des signaux issus de microphones d'appoint et d'appliquer des effets audio (égalisation, réverbération, etc.) avec une ergonomie comparable à celle du mixage stéréophonique traditionnel. Même si davantage d'outils de prise de son et de mixage (effets audio notamment) devront être développés pour faciliter la production musicale, il est envisageable à très court terme de créer et transmettre en temps réel des contenus multimédias au format HOA grâce aux nouvelles technologies de diffusion comme l'Internet. La capacité d'adaptation des signaux HOA au contexte d'utilisation (système de diffusion sonore, débit du réseau de transmission), la possibilité qu'ils pourraient offrir à l'utilisateur d'interagir avec la scène sonore (par exemple la faire tourner, zoomer sur une partie de l'espace), et les faibles ressources informatiques mobilisées pour ces opérations font de la technique HOA un candidat sérieux face à la stéréophonie.

Remerciements

Nous tenons à remercier Virginie Burgun et Olivier Rosset pour leurs travaux de stage menés dans le cadre de notre projet de prise de son et de mixage HOA. Beaucoup d'éléments de cet article sont issus de cette collaboration. Nous sommes également très reconnaissants envers

les musiciens qui ont accepté d'être enregistrés, Orange Labs pour avoir mis à notre disposition leurs *plugins* VST de mixage HOA, et le Conservatoire national de musique et de danse de Paris qui nous a conviés à des séances d'enregistrements particulièrement instructives. Nous remercions enfin Michael Williams pour la fabrication de la structure de prise de son.

Références bibliographiques

- [1] J.S. Bamford, *An Analysis of Ambisonic Sound Systems of First and Second Order*, Master of Science Thesis, University of Waterloo, Waterloo, Ont., Canada, 1995.
- [2] J. Daniel, *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*, Thèse de Doctorat, Université de Paris 6, 2000.
- [3] M.A. Gerzon, *Ambisonics in multichannel broadcasting and video*, *J. Audio Eng. Soc.*, vol. 33, no. 11, pp. 859-871, Nov. 1985.
- [4] S. Moreau, *Étude et réalisation d'outils avancés d'encodage spatial pour la technique de spatialisation sonore Higher Order Ambisonics: microphone 3D et contrôle de distance*, Thèse de Doctorat, Université du Maine, 2006.
- [5] P. Craven et M. A. Gerzon, *Coincident Microphone Simulation Covering Three Dimensional Space and Yielding Various Directional Outputs*, U.S. Patent 4,042,779, 1977.
- [6] J. Meyer et G. Elko, *A Highly Scalable Spherical Microphone Array Based on an Orthonormal Decomposition of the Soundfield*, ICASSP 2002, vol. 2, pp. 1781-1784.
- [7] A. Laborie, R. Bruno, et S. Montoya, *A New Comprehensive Approach of Surround Sound Recording*, AES 114th Convention, Amsterdam, The Netherlands, 2003.
- [8] V. Burgun, *Conception et réalisation d'un système microphonique 2D pour la technologie de spatialisation sonore Higher Order Ambisonics (HOA)*, Mémoire de fin d'études de la Formation Supérieure aux Métiers du Son, CNSMDP, Paris, France, 2011.
- [9] O. Rosset, *Optimisation d'un système microphonique 2D et pratique avancée du mixage pour la technologie de spatialisation sonore Higher Order Ambisonics (HOA)*, Mémoire de fin d'études de la Formation Supérieure aux Métiers du Son, CNSMDP, Paris, France, 2012.
- [10] Suite logicielle "HOA VST Plug-ins - Manuel d'utilisation. Orange Labs, Lannion, France, 2008.
- [11] M. Bruneau, *Introduction aux théories de l'acoustique*, Université du Maine, Le Mans, 1983.
- [12] Philip M. Morse and K. Uno Ingard. *Theoretical Acoustic*, McGraw-Hill, 1968.

⁵- Un fader ou tirette, est un bouton de commande rectiligne réglant le niveau d'un signal audio.