

Académie de Nantes

ÉCOLE DOCTORALE DE L'UNIVERSITÉ DU MAINE

LE MANS, FRANCE

THÈSE DE DOCTORAT

Spécialité : ACOUSTIQUE

présentée par

Sébastien MOREAU

pour obtenir le grade de Docteur de l'Université du Maine

**Étude et réalisation d'outils avancés d'encodage spatial
pour la technique de spatialisation sonore
Higher Order Ambisonics :
microphone 3D et contrôle de distance**

Version provisoire

Soutenance prévue le 11 juillet 2006

devant le jury composé de

Michel BRUNEAU	Professeur à l'Université du Maine, LAUM, Le Mans	Examineur
Jérôme DANIEL	Docteur Ingénieur, France Télécom R&D, Lannion	Examineur
Claude DEPOLLIER	Professeur à l'Université du Maine, LAUM, Le Mans	Directeur de thèse
Angelo FARINA	Professeur à l'Université de Parme, Parme (Italie)	Rapporteur
Philippe HERZOG	HDR chargé de recherche CNRS, LMA, Marseille	Rapporteur
Rozenn NICOL	Docteur Ingénieur, France Télécom R&D, Lannion	Invitée
Olivier WARUSFEL	Docteur chargé de recherche, responsable de l'équipe d'acoustique des salles, IRCAM, Paris	Examineur

Remerciements

J'aimerais tout d'abord exprimer ma gratitude à Jean-Pierre Petit, Yannick Mahieux et toutes les équipes de recherche du laboratoire TECH/SSTP de France Télécom R&D à Lannion (22) qui m'ont accueilli et épaulé pendant trois ans. Je remercie tout particulièrement Jérôme Daniel : ces travaux de thèse sont le fruit de notre collaboration étroite pendant ces années. J'ai également beaucoup apprécié les conseils et le soutien de Rozenn Nicol, Marc Emerit, Claude Marro, David Virette, Manuel Briand, Jean-Marie Pernaux, Pascal Anselmo. Merci enfin à Adil Chraa et Virginie Durin qui ont apporté leur contribution pendant leurs périodes de stage.

Je remercie vivement Angelo Farina et Philippe Herzog pour avoir accepté la charge d'être rapporteurs et ainsi témoigner de leur intérêt pour les travaux de thèse relatés dans ce document. Je souhaite également exprimer ma reconnaissance à mon directeur de thèse Claude Depollier, ainsi qu'à Michel Bruneau, Rozenn Nicol et Olivier Warusfel pour avoir accepté de faire partie du jury.

Un grand merci également à Marcel Clément pour son aide indispensable à la confection du prototype microphonique, ainsi qu'à Stéphanie Bertet et Olivier Warusfel pour la réalisation des mesures à l'Ircam (Paris), dans les meilleures conditions.

Je remercie chaleureusement mes parents, mes frères pour leurs encouragements incessants. Je remercie enfin Virginie, du fond du cœur, pour sa compréhension, sa présence rassurante et son soutien sans faille.

Table des matières

CHAPITRE 1 – INTRODUCTION	7
Cadre de la thèse	7
Contexte de développement des différents médias	8
Enjeux actuels de la spatialisation sonore	8
Représentation du champ sonore : quel format ?	9
Principaux objectifs et apports de la thèse – Organisation du document	11
RÉFÉRENCES BIBLIOGRAPHIQUES DU CHAPITRE 1	13
CHAPITRE 2 – CONCEPT DE DÉVELOPPEMENT DU CHAMP ACOUSTIQUE EN HARMONIQUES SPHÉRIQUES	15
2.1 PRÉLIMINAIRES MATHÉMATIQUES : L'ESPACE $L^2(S)$ ET LES HARMONIQUES SPHÉRIQUES	16
2.1.1 Espace $L^2(S)$	16
2.1.2 Harmoniques sphériques	18
2.2 DÉVELOPPEMENT EN SÉRIE D'HARMONIQUES SPHÉRIQUES ET APPROXIMATION D'UNE FONCTION DE $L^2(S)$	23
2.2.1 Développement en harmoniques sphériques d'une fonction de $L^2(S)$	23
2.2.2 Approximation d'une fonction de $L^2(S)$	24
2.3 APPLICATION AU CHAMP DE PRESSION ACOUSTIQUE	25
2.3.1 Développement du champ acoustique en harmoniques sphériques	25
2.3.2 Approximation du champ de pression acoustique	31
RÉFÉRENCES BIBLIOGRAPHIQUES DU CHAPITRE 2	39
CHAPITRE 3 – HIGHER ORDER AMBISONICS	41
3.1 REPRÉSENTATION DE SCÈNE SONORE AVEC HOA	42
3.1.1 Format de représentation associé : les signaux HOA	42
3.1.2 Conventions et notations	45
3.2 ENCODAGE SPATIAL ET MANIPULATION DE SCÈNE SONORE	47
3.2.1 Encodage spatial de source virtuelle et de scène sonore naturelle	47
3.2.2 Manipulations de scène sonore HOA	53
3.3 DÉCODAGE SPATIAL : APPROCHE BASIQUE	55
3.3.1 Décodage basique pour un dispositif régulier de haut-parleurs	56
3.3.2 Décodage binaural pour une restitution au casque d'écoute	58

3.4 OPTIMISATION DU DÉCODAGE SPATIAL SUIVANT LE CONTEXTE D'ÉCOUTE	63
3.4.1 <i>Prédiction objective de la localisation auditive d'après caractérisation locale de la propagation acoustique par les vecteurs vitesse V et énergie E</i>	63
3.4.2 <i>Optimisation du décodage pour un dispositif régulier de haut-parleurs</i>	65
RÉFÉRENCES BIBLIOGRAPHIQUES DU CHAPITRE 3.....	71
CHAPITRE 4 – ENCODAGE SPATIAL DE SOURCE VIRTUELLE EN CHAMP PROCHE.....	75
4.1 CARACTÉRISATION PHYSIQUE ET PERCEPTIVE DU CHAMP PROCHE	76
4.1.1 <i>Caractérisation physique du champ proche</i>	76
4.1.2 <i>Perception auditive du champ proche</i>	77
4.2 PROBLÈME DE LA REPRODUCTION DU CHAMP PROCHE D'UNE SOURCE PONCTUELLE HOMOGENÈNE	80
4.2.1 <i>Définition théorique et analyse des filtres de champ proche</i>	80
4.2.2 <i>Prise en compte du champ proche des haut-parleurs de restitution</i>	82
4.3 PROPOSITION D'UNE MÉTHODE DE CONCEPTION DES FILTRES DE CHAMP PROCHE.....	84
4.3.1 <i>Concept de bande fréquentielle utile</i>	84
4.3.2 <i>Application à la conception de filtres de champ proche</i>	87
RÉFÉRENCES BIBLIOGRAPHIQUES DU CHAPITRE 4.....	93
CHAPITRE 5 – ENREGISTREMENT HOA DE SCÈNE SONORE NATURELLE AVEC UN DISPOSITIF SPHÉRIQUE DE MICROPHONES	95
5.1 INTRODUCTION À L'ENREGISTREMENT HOA	96
5.1.1 <i>Microphones différentiels</i>	96
5.1.2 <i>Liens entre les microphones différentiels et les signaux HOA</i>	98
5.2 PRINCIPES THÉORIQUES DU MICROPHONE HOA SPHÉRIQUE.....	100
5.2.1 <i>Modélisation mathématique d'un microphone HOA sphérique continu</i>	100
5.2.2 <i>Discrétisation du microphone HOA sphérique continu</i>	106
5.3 RÉALISATION PRATIQUE D'UN MICROPHONE HOA SPHÉRIQUE	109
5.3.1 <i>Échantillonnage de la surface d'une sphère</i>	109
5.3.2 <i>Régularisation des signaux HOA estimés</i>	116
5.4 RÉALISATION, MESURE ET VALIDATION OBJECTIVE D'UN PROTOTYPE D'ORDRE 4	119
5.4.1 <i>Construction du prototype</i>	119
5.4.2 <i>Confrontation du prototype au modèle théorique de la sphère rigide</i>	124
5.4.3 <i>Reconstruction des directivités associées aux harmoniques sphériques</i>	128
RÉFÉRENCES BIBLIOGRAPHIQUES DU CHAPITRE 5.....	133

CHAPITRE 6 – CONCLUSIONS ET PERSPECTIVES D’AVENIR	137
BIBLIOGRAPHIE.....	149
ANNEXES – PUBLICATIONS	149
AES 114th Convention, Amsterdam, Pays-Bas, 22-25 mars 2003	151
Congrès commun CFA/DAGA, Strasbourg, France, 22-25 mars 2004	169
AES 116th Convention, Berlin, Allemagne, 8-11 mai 2004.....	171
CORESA 2004, Lille, France, 25-26 mai 2004.....	185
AES 120th Convention, Paris, France, 20-23 mai 2006.....	189

Chapitre 1

Introduction

Cadre de la thèse

Les travaux de thèse rapportés dans le présent document appartiennent au domaine de la *spatialisation sonore* qui désigne un ensemble de techniques dont le but est la reconstruction spatiale des sources acoustiques. Mener à bien cette reconstruction nécessite de connaître les informations spatiales et temporelles du champ sonore, de pouvoir les stocker et les transmettre. La Figure 1.1 représente les principes de base d'une technique de spatialisation sonore : l'*encodage spatial* d'une scène sonore dans un *format de représentation*, puis le *décodage spatial* de la scène formatée permettant sa reproduction. Une étape intermédiaire de manipulation du champ sonore avant décodage peut être ajoutée, éventuellement. Le format de représentation, constitué en pratique d'un ensemble de signaux audio temporels auxquels peuvent être associés des paramètres de spatialisation, contient les informations spatiales et temporelles du champ sonore nécessaires à sa reconstruction. L'*encodage spatial* désigne le processus qui permet d'acquérir, de traduire une scène sonore dans un format spécifique de représentation. Le *décodage spatial* concerne quant à lui l'adaptation des données formatées au système de restitution. Si le format est constitué de signaux destinés à alimenter directement les haut-parleurs d'un système de diffusion particulier, comme dans le cas de la stéréophonie à 2 ou 5.1 canaux par exemple, l'étape de décodage n'existe pas. Les principales techniques de spatialisation sonore sont la *stéréophonie* [C78, P97], le *binaural* [Mo92], *Wave Field Synthesis* (WFS) [BVV93, N99, DNM03], et *Higher Order Ambisonics* (HOA) [DNM03, B95, D00].



Figure 1.1 – Principes de base de la spatialisation sonore.

Contexte de développement des différents médias

Qu'il représente un élément de son, d'image, de logiciel, de graphique ou de texte, un élément binaire numérique garde toujours la même forme. De cette façon, la numérisation de l'ensemble des médias permet un rapprochement progressif des secteurs de l'audiovisuel, de l'informatique, des télécommunications et de l'édition : c'est la *convergence des médias*. Les nouvelles normes de diffusion numérique (DAB, DVB, DRM¹, etc.) ouvrent ainsi la voie à beaucoup de nouvelles formes d'applications multimédias proposant des contenus mélangeant texte, son et vidéo, et pour lesquels la spatialisation sonore, notamment, est un véritable enjeu.

Parallèlement, nous assistons à l'émergence d'un nouveau système de diffusion de contenu multimédia : *Internet*. Internet compte en effet de plus en plus d'adeptes à travers le monde et ses capacités techniques s'améliorent de jour en jour. Plusieurs technologies sont déjà disponibles pour accéder aux contenus multimédias comme le *streaming* permettant la transmission d'un contenu audiovisuel en continu (temps réel), ou le téléchargement simple de logiciels, de données variées. Ces technologies, parallèlement au développement de la puissance de traitement et de la capacité de mémorisation des ordinateurs, font dès à présent d'Internet et à plus forte raison dans le futur, un système de diffusion multimédia alternatif aux systèmes classiques de radiodiffusion (voie terrestre, satellite et câble) ou aux supports de stockage numériques tels que le CD, le DVD, etc. La portée mondiale d'Internet est un véritable atout : un utilisateur internaute peut avoir accès à une quantité considérable de contenus multimédias disponibles dans le monde entier. Le second atout majeur est l'*interactivité* offerte par la navigation sur le Web : tout usager d'Internet est ainsi susceptible de recevoir des informations et d'en envoyer à quiconque. Il ne fait aucun doute qu'Internet sera de plus en plus utilisé pour proposer un choix d'applications améliorées et innovantes, notamment audiovisuelles, de grande qualité.

Enjeux actuels de la spatialisation sonore

Dans le contexte actuel de développement des différents médias, la spatialisation sonore doit relever un certain nombre de défis techniques :

- *Améliorer la qualité de restitution de l'espace sonore*. La numérisation du son a permis d'améliorer la dynamique du signal, d'obtenir une bande passante optimale et de garantir cette

¹ Respectivement *Digital Audio Broadcasting*, *Digital Video Broadcasting*, et *Digital Radio Mondiale*.

qualité sonore à chaque écoute. Désormais, l'auditeur s'intéresse à la qualité de rendu spatial du son qu'il souhaite le plus réaliste possible, notamment dans les applications interactives telles que les jeux vidéos ou les applications de réalité virtuelle.

- *Adapter la reproduction sonore à l'environnement d'utilisation.* La convergence des médias et les nouvelles normes de diffusion numérique permettent aux utilisateurs d'accéder à des contenus multimédias dans des situations de plus en plus nombreuses et variées : utilisation solitaire et interactive devant l'ordinateur ou au contraire plurielle et passive devant la télévision, en situation de mobilité, etc. Simultanément, les équipements sonores des particuliers se diversifient, avec notamment une percée des systèmes multivoies 5.1 sur le marché grand public. D'un point de vue technique, cela nécessite de pouvoir reproduire correctement une scène sonore à partir de différents systèmes, parfois imposés par le contexte d'utilisation : casque d'écoute, dispositif stéréophonique conventionnel à deux haut-parleurs, multicanaux (5.1, 7.1, etc.), systèmes embarqués dans les automobiles, les trains, les avions, etc.

- *S'intégrer dans les nouvelles applications interactives.* La spatialisation sonore concerne de plus en plus de nouveaux domaines d'application où l'interactivité est de règle : la téléphonie fixe ou portable (conversation multi-locuteurs spatialisée dans le but d'améliorer l'intelligibilité, le confort d'écoute, etc.), la navigation Web 3D, les appareils multimédias portatifs, la réalité virtuelle, les jeux vidéos, etc. Pour s'intégrer facilement à ces nouvelles applications, la spatialisation sonore doit pouvoir offrir une relative souplesse dans la représentation des informations acoustiques, notamment en termes de modification de la position des sources sonores dans l'espace.

Représentation du champ sonore : quel format ?

La multiplication et la variété des applications sonores spatialisées et des systèmes de diffusion associés posent la question du format de représentation du champ acoustique : comment représenter les informations spatiales et temporelles d'une scène sonore de façon à s'intégrer facilement aux nouvelles applications sonores interactives et s'adapter aux diverses situations d'écoute ? Outre la précision de la représentation du champ acoustique et sa possibilité de se faire en trois dimensions, un « bon » format répond dans l'idéal à un certain nombre de critères qui sont principalement :

- la possibilité d'*adapter la reproduction de la scène sonore* à de multiples systèmes de diffusion, souvent imposés par le contexte d'écoute ;
- la possibilité de *manipuler* la scène sonore encodée. Ce critère devient de plus en plus important dans les applications où l'utilisateur interagit avec l'environnement virtuel, comme dans les jeux vidéo ou les applications de réalité virtuelle par exemple ;
- enfin, la *taille de la représentation*, c'est-à-dire le nombre de signaux audio et éventuellement la quantité de paramètres de spatialisation, doit bien évidemment rester raisonnable par rapport aux capacités de mémorisation et à la puissance de calcul des appareils multimédias numériques actuels.

Le Tableau 1.1 confronte les principaux formats de représentation de scène sonore aux trois critères précédemment énoncés. Ces principaux formats sont :

- le *format stéréophonique* que nous définissons comme un ensemble de signaux destinés à alimenter directement des haut-parleurs placés dans une configuration prédéfinie (2.0, 4.0, 5.1, etc.). Les sources sonores sont chacune reconstruites à partir d'une paire (2D) ou d'un triplet (3D) de haut-parleurs [P97] ;
- le *format binaural*, composé des deux *signaux binauraux* que capteraient les tympans d'un auditeur en situation d'écoute réelle, c'est-à-dire comportant les effets de filtrage créés par l'interaction des ondes sonores avec le corps de l'auditeur, notamment la tête, les pavillons des oreilles et le torse [Mo92] ;
- le *format HOA* associé à la technologie *Higher Order Ambisonics*, constitués de signaux qui résultent de la décomposition en harmoniques cylindriques (2D) ou sphériques (3D) d'un champ acoustique en un point [DNM03, B95, D00]. Le nombre de ces signaux dépend de l'ordre M maximal de la décomposition : $(M+1)^2$ en 3D et $2M+1$ en 2D ;
- enfin, le format « *tout paramétrique* » composé par les signaux et les paramètres de position de chaque source sonore. Ce format de représentation qui n'est pas lié à une technique de spatialisation sonore particulière sert uniquement à transmettre ou stocker séparément les éléments constituant une scène acoustique.

Notons que nous ne mentionnons pas dans cette comparaison le format de représentation associé à la technique de spatialisation sonore *Wave Field Synthesis* [BVV93, N99, DNM03]. Cette technique se base en effet sur une description du champ acoustique complètement liée au dispositif de restitution par des propriétés de réciprocité acoustique. Les signaux résultants de l'encodage WFS n'ont pas pour vocation d'être transmis au public et exploités avec les systèmes de diffusion sonore habituels. La technique WFS est cependant particulièrement bien adaptée à la reproduction sonore spatialisée à partir d'une installation fixe (dans une salle de concert, de cinéma par exemple) et destinée à une large audience.

L'observation du Tableau 1.1 montre que le format HOA semble réaliser un bon compromis entre souplesse d'utilisation et quantité d'informations à transmettre ou stocker². En effet, ce format décrit de façon rationnelle une scène sonore 3D qui peut subir quelques manipulations globales (rotation, distorsion de perspective), très utiles dans les applications interactives. D'autre part, son indépendance vis à vis du système de restitution lui permet de s'adapter à diverses situations d'écoute. Enfin, le format HOA est *scalable*, c'est-à-dire que la taille d'une scène sonore encodée peut être réduite en cas de besoin tout en préservant la cohérence des informations

² Bien que le format « tout paramétrique » soit attrayant au prime abord, le nombre des signaux acoustiques nécessaires pour décrire une scène sonore complexe s'élève rapidement. D'autre part, ce format ne peut décrire à l'heure actuelle que des scènes sonores synthétiques, l'enregistrement naturel supposant la mise au point d'une méthode universelle d'extraction de source sonore difficilement imaginable dans un futur proche.

acoustiques. Cette réduction consiste simplement à diminuer l'ordre M de la description en ne conservant que le sous-ensemble des signaux correspondants. Cela se traduit finalement par une baisse de la résolution spatiale de la scène sonore encodée.

Format de représentation	Manipulation de la scène encodée	Dispositif de restitution	Taille de la scène encodée
<i>Stéréophonie</i> (2.0, 4.0, 5.1, etc.)	Impossible, scène sonore figée	Fixe, lié au format	1 signal par haut-parleur
<i>Binaural</i>			2 signaux
<i>HOA</i> (ordre M)	Globale (rotation, distorsion de perspective)	Adaptable, indépendant du format	3D : $(M+1)^2$ 2D : $2M+1$ signaux
<i>Format</i> « tout paramétrique »	Chaque source de façon indépendante		1 signal par source + paramètres de position spatiale

Tableau 1.1 – Comparaison des principaux formats de représentation de scène sonore.

Principaux objectifs et apports de la thèse – Organisation du document

Bien que le format HOA possède dans le contexte actuel un fort potentiel pour la transmission et l'exploitation de contenu audio spatialisé, son utilisation reste encore très marginale. Cela s'explique en partie par le manque d'outils diversifiés de création de contenu sonore. Quelques outils logiciels existent actuellement pour l'encodage et le décodage HOA de source virtuelle aux trois premiers ordres, notamment sous la forme de *plugins VST (Virtual Studio Technology* de Steinberg) qui s'utilisent facilement avec beaucoup de logiciels audio grand public et professionnels [FU98, GI03, MF03]. Cependant, l'enregistrement naturel reste à l'heure actuelle limitée à l'ordre 1 avec un microphone de type *Soundfield* [F79b, F79a] dont les principes remontent aux années 70 [CG77]. La réalisation de microphones HOA d'ordres supérieurs se trouve notamment confrontée à des problèmes d'amplification excessive dans un domaine fréquentiel basse fréquence, équivalents à ceux rencontrés dans la construction de microphones différentiels d'ordres supérieurs à 1 [E00a].

Le principal objectif de ces travaux de thèse est de concevoir et de valider des nouvelles méthodes d'encodage spatial pour la technologie de spatialisation sonore *Higher Order Ambisonics* afin de combler certains manques actuels et favoriser son développement. Plus précisément, les apports de la thèse concernent :

- l'encodage HOA de source virtuelle en champ proche. Alors que la plupart des techniques de spatialisation sonore se base sur l'hypothèse que les sources reconstruites (et les haut-

parleurs) émettent à une distance infinie de l'auditeur (hypothèse d'onde plane), les nouvelles applications sonores spatialisée et interactives appartenant au domaine de la réalité virtuelle ou des jeux vidéos par exemple, permettent à l'utilisateur de se déplacer à proximité des différentes sources acoustiques qui peuplent son environnement virtuel. Cependant, l'encodage HOA d'une source virtuelle proche de l'auditeur implique en théorie des opérations de filtrage souvent impossibles à réaliser en pratique [D00]. Nous proposons au Chapitre 4 une nouvelle méthode de conception des filtres décrivant l'effet de champ proche des sources sonores, qui permettent de reconstruire le champ acoustique au voisinage d'une source émettrice, notamment à l'intérieur d'un dispositif de haut-parleurs, tout en limitant l'énergie globale délivrée par ces haut-parleurs. Ces travaux ont fait l'objet d'une publication [DM04] et d'un brevet d'invention [DM05a].

- *l'enregistrement HOA d'ordre supérieur à 1*. Nous présentons au Chapitre 5 une étude sur la conception des microphones HOA 3D permettant d'enregistrer les ordres supérieurs à 1 de la décomposition en harmoniques sphériques du champ acoustique, et basée sur l'emploi de dispositifs sphériques. Plusieurs publications [MD04, MDB06, MDC04] et un brevet d'invention [DM05b] ont résulté de cette étude.

- *la réalisation d'un microphone HOA 3D d'ordre 4*. L'étude théorique des microphones HOA sphériques a permis la réalisation d'un prototype 3D d'ordre 4, constitué de 32 capsules disposées à la surface d'une sphère rigide. Ce prototype a fait l'objet d'une campagne de mesures et d'une validation objective dont les résultats sont présentés au Chapitre 5 et dans [MDB06].

- *l'élaboration d'une suite logicielle d'outils de création de contenu sonore HOA* sous la forme de *plugins* VST s'intégrant à un large panel de logiciels audio grand public et professionnels [MDC04]. Cette suite logicielle a ensuite été développée par l'équipe son3D de France Télécom R&D (Lannion) et permet notamment de faire fonctionner notre prototype microphonique et d'écouter le résultat sur différents systèmes d'écoute en temps réel.

L'ensemble des publications parues dans le cadre de ces travaux de thèse est annexé au présent document.

Avant d'entrer dans le vif du sujet, nous présentons au deuxième chapitre de ce document le concept de développement en harmoniques sphériques du champ de pression acoustique sur lequel se base la représentation d'une scène sonore avec la technique *Higher Order Ambisonics*. Nous tenterons de comprendre l'intérêt d'ajouter dans le format HOA des informations concernant les composantes harmoniques d'ordres supérieurs à 1 et illustrerons l'impact sur la reconstruction potentielle du champ sonore, notamment en termes d'élargissement de la zone d'écoute. Les cas particuliers d'un champ de pression acoustique engendré par une onde plane et une onde sphérique seront étudiés avec attention. Le troisième chapitre a pour objectif de donner un aperçu complet et relativement détaillé de la technologie HOA qui intègre les récentes évolutions et souligne les limitations actuelles. Nous étudierons tour à tour l'encodage spatial HOA de source virtuelle et réelle, la manipulation de scène sonore et le décodage spatial basique et optimisé pour différents systèmes de reproduction.

Références bibliographiques du Chapitre 1

- [C78] CONDAMINES, R. *Stéréophonie - Cours de relief sonore théorique et appliqué*. Paris: Masson, 1978, 301 p.
- [P97] PULKKI, V. *Virtual Sound Source Positioning Using Vector Base Amplitude Panning*. J. Audio Eng. Soc., 1997, Vol. 45, n° 6, p. 456-466.
- [Mo92] MOLLER, H. *Fundamentals of Binaural Technology*. Applied Acoustics, 1992, Vol. 36, p. 171-218.
- [BVV93] BERKHOUT, A.J., DE VRIES, D., et VOGEL, P. *Acoustic control by wave field synthesis*. J. Acoust. Soc. Am., 1993, Vol. 93, n° 5, p. 2764-2778.
- [N99] NICOL, R. *Restitution sonore spatialisée sur une zone étendu : application à la téléprésence*. Thèse de Doctorat, Le Mans: Université du Mans, 1999.
- [DNM03] DANIEL, J., NICOL, R., et MOREAU, S. *Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging*. 114th AES Convention, preprint 5788, Amsterdam, Pays-Bas, 22-25 mars 2003.
- [B95] BAMFORD, J.S. *An Analysis of Ambisonics Sound Systems of First and Second Order*. M. Sc. Thesis, Waterloo: University of Waterloo, 1995.
- [D00] DANIEL, J. *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. Thèse de Doctorat, Paris: Université de Paris 6, 2000.
- [FU98] FARINA, A. et UGOLOTTI, E. *Software Implementation of B-Format Encoding And Decoding*. AES 104th Convention, preprint 4691, Amsterdam, Pays-Bas, 15-20 mai 1998.
- [GI03] GERZONIC INDUSTRIES. *3D-Audio Software*. 2003. <http://www.gerzonic.net> (dernier accès le 28 mai 2006)
- [MF03] MALHAM, D. et FIELD, A. *VST Ambisonic Tools*. York University Music Technology Group, York, 2003. http://www.york.ac.uk/inst/mustech/3d_audio/vst/welcome.html (dernier accès le 28 mai 2006)
- [F79b] FARRAR, K. *Soundfield Microphone - 2*. Wireless World, 1979, novembre, p. 99-103.
- [F79a] FARRAR, K. *Soundfield Microphone*. Wireless World, 1979, octobre, p. 48-50.
- [CG77] CRAVEN, P. et GERZON, M.A. *Coincident Microphone Simulation Covering Three Dimensional Space and Yielding Various Directionnal Outputs*. U.S. Patent 4,042,779, août 1977.

- [E00a] ELKO, G., *Chap. 10. Superdirectional Microphone Arrays*, in *Acoustic Signal Processing for Telecommunication*. 2000. Boston: Kluwer Academic Publishers. p. 181-238.
- [DM04] DANIEL, J. et MOREAU, S. *Further Study of Sound Field Coding with Higher Order Ambisonics*. AES 116th Convention, preprint 6017, Berlin, Allemagne, 8-11 mai 2004.
- [DM05a] DANIEL, J. et MOREAU, S. *Procédé de traitement de données sonores, en particulier en contexte ambiophonique*. Brevet d'invention FR 2866974, septembre 2005.
- [MD04] MOREAU, S. et DANIEL, J. *Study of Higher Order Ambisonic Microphone*. Congrès commun CFA-DAGA'04, Strasbourg, France, 22-25 mars 2004.
- [MDB06] MOREAU, S., DANIEL, J., et BERTET, S. *3D Sound Field Recording with Higher Order Ambisonics - Objective Measurements and Validation of Spherical Microphone*. AES 120th Convention, preprint 6857, Paris, France, 20-23 mai 2006.
- [MDC04] MOREAU, S., DANIEL, J., et CHRAA, A. *Production, transmission et restitution temps réel d'une scène sonore dans un format audio 3D flexible*. CORESA 2004, Lille, France, 25-26 mai 2004.
- [DM05b] DANIEL, J. et MOREAU, S. *Procédé et dispositif de traitement de données sonores en contexte ambiophonique*. Brevet d'invention FR 2858512, février 2005.

Chapitre 2

Concept de développement du champ acoustique en harmoniques sphériques

Pour représenter le champ sonore en trois dimensions, la technique de spatialisation sonore *Higher Order Ambisonics* (HOA) se base sur le développement en série d'harmoniques sphériques de la pression acoustique³. Ce chapitre a pour but de développer et de démontrer l'idée que *le nombre de composantes harmoniques prises en compte dans ce développement se traduit directement en terme d'étendue de la zone de l'espace où le champ acoustique est correctement connu*, et donc potentiellement reconstituable. Pour cela, nous commencerons par décrire le cadre mathématique dans lequel les harmoniques sphériques sont définies. Nous verrons ensuite que dans certaines conditions, ces harmoniques forment une base sur laquelle peuvent être développées certaines fonctions. Enfin, nous appliquerons ces principes au champ de pression acoustique et étudierons en détails les cas particuliers d'un champ sonore engendré par une onde plane et par une onde sphérique.

³ La description du champ acoustique en deux dimensions (2D) se base quant à elle sur un développement du champ sonore en série d'*harmoniques cylindriques* [B98]. Des principes analogues à ceux développés au cours de ce chapitre peuvent lui être associés. Cependant, cette représentation 2D ne permet ni la description de source sonore en élévation, ni celle de la propagation d'une onde sphérique, pourtant nécessaires quand nous étudierons dans les chapitres suivants l'enregistrement HOA d'une scène sonore naturelle et la synthèse d'une source ponctuelle en champ proche.

2.1 Préliminaires mathématiques : l'espace $L^2(S)$ et les harmoniques sphériques

2.1.1 Espace $L^2(S)$

Système de repérage spatial

Pour repérer les points de l'espace Euclidien à trois dimensions, nous utilisons le système de coordonnées sphériques illustré à la Figure 2.1, constitué de trois paramètres distincts :

- l'azimut θ , angle formé dans le plan (xOy) par rapport à l'axe (Ox) , avec $\theta \in [0, 2\pi[$ exprimé en radians ;
- l'élévation δ , angle vertical à partir du plan (xOy) exprimé en radians, avec $\delta \in [-\pi/2, \pi/2]$;
- le rayon r , distance à l'origine exprimée en mètres, avec $r \in [0, +\infty[$.

Les coordonnées cartésiennes usuelles (x, y, z) et les coordonnées sphériques (r, θ, δ) sont liées par le système d'équations suivant :

$$\begin{cases} x = r \cos \theta \cos \delta \\ y = r \sin \theta \cos \delta \\ z = r \sin \delta \end{cases} \quad (2.1)$$

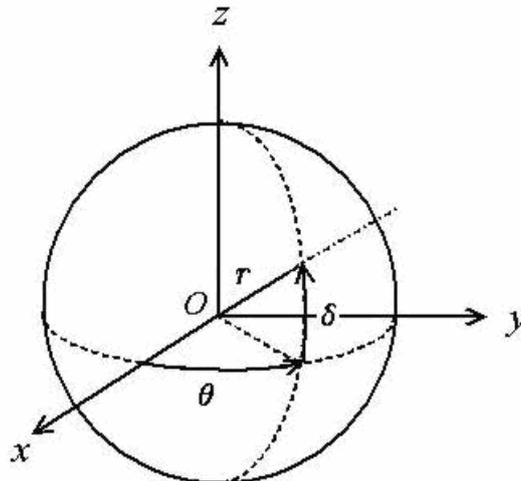


Figure 2.1 – Système de coordonnées sphériques dans lequel tout point de l'espace \mathbb{R}^3 est repéré par son rayon r , son azimut θ , et son élévation δ .

Définition

Nous appelons $L^2(S)$ l'espace vectoriel réel des classes⁴ de fonctions de carré intégrable⁵ sur la sphère $S = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$ [G96, R89].

En repérant les points de l'espace avec le système de coordonnées sphériques défini à la Figure 2.1, les points décrivant la sphère S possèdent tous un rayon unitaire et se différencient les uns des autres par leurs coordonnées angulaires θ et δ . Un vecteur de $L^2(S)$ devient ainsi une fonction de θ et δ .

Produit scalaire et norme dans $L^2(S)$

Dans $L^2(S)$, un produit scalaire se définit en associant à tout couple (f, g) d'éléments de $L^2(S)$ le réel noté $\langle f, g \rangle_S$, nommé *produit scalaire de f et g* et s'exprimant ainsi [G96, R89] :

$$\langle f, g \rangle_S = \frac{1}{4\pi} \int_{\theta=0}^{2\pi} \int_{\delta=-\frac{\pi}{2}}^{+\frac{\pi}{2}} f(\theta, \delta) g(\theta, \delta) \cos \delta d\delta d\theta, \tag{2.2}$$

où $\cos \delta d\delta d\theta / 4\pi$ définit un élément de surface de la sphère S . Ce produit scalaire a les propriétés suivantes :

- $\langle f, f \rangle_S \geq 0$ pour tout f appartenant à $L^2(S)$; $\langle f, f \rangle_S = 0$ si et seulement si $f = 0$;
- $\langle f_1 + f_2, g \rangle_S = \langle f_1, g \rangle_S + \langle f_2, g \rangle_S$ pour tout triplet (f_1, f_2, g) appartenant à $L^2(S)$;
- $\langle af, g \rangle_S = a \langle f, g \rangle_S$ pour tout réel a et tout couple (f, g) appartenant à $L^2(S)$;
- $\langle f, g \rangle_S = \langle g, f \rangle_S$ pour tout couple (f, g) appartenant à $L^2(S)$.

La *norme* $\|f\|_S$ d'une fonction f appartenant à $L^2(S)$ associée à ce produit scalaire est donnée par :

$$\|f\|_S = \sqrt{\langle f, f \rangle_S} = \sqrt{\frac{1}{4\pi} \int_{\theta=0}^{2\pi} \int_{\delta=-\frac{\pi}{2}}^{+\frac{\pi}{2}} |f(\theta, \delta)|^2 \cos \delta d\delta d\theta}. \tag{2.3}$$

L'*énergie* $\|f\|_S^2$ d'une fonction f appartenant à $L^2(S)$ est le carré de sa norme :

$$\|f\|_S^2 = \frac{1}{4\pi} \int_{\theta=0}^{2\pi} \int_{\delta=-\frac{\pi}{2}}^{+\frac{\pi}{2}} |f(\theta, \delta)|^2 \cos \delta d\delta d\theta. \tag{2.4}$$

Orthogonalité, base orthonormée et projection dans $L^2(S)$

L'*orthogonalité* dans $L^2(S)$ se définit de la façon suivante [G96, R89] :

- Deux éléments f et g appartenant à $L^2(S)$ sont dits orthogonaux si leur produit scalaire est nul ;

⁴ *classe* signifie que deux de ces fonctions f et g doivent être identifiées si elles sont *presque partout* égales, c'est-à-dire si $f - g$ est négligeable ou encore de mesure nulle.

⁵ Une fonction f est dite de *carré intégrable* sur S si le carré de son module est sommable sur S . En théorie du signal, de telles fonctions sont appelées des signaux d'énergie finie.

- Soit G une partie de $L^2(S)$, on dit que f est orthogonal à G si f est orthogonal à tous les éléments de G ;
- Soient F et G deux parties de $L^2(S)$, on dit que F est orthogonal à G si tout élément f de F est orthogonal à G .

Une *base orthonormée* de $L^2(S)$ est une suite d'éléments de S orthogonaux deux à deux et de norme unitaire, telle que le seul élément de $L^2(S)$ orthogonal à tous les éléments de la suite est la fonction nulle. Autrement dit, si nous notons $e_0, e_1, \dots, e_m, \dots$ les éléments de la suite, celle-ci forme une base orthonormée si et seulement si :

- $\langle e_p, e_p \rangle_S = 1$ pour tout p appartenant à \mathbb{N} et $\langle e_p, e_q \rangle_S = 0$ pour tout p différent de q ;
- $\langle f, e_p \rangle_S = 0$ pour tout p appartenant à \mathbb{N} si et seulement si $f = 0$.

Soit $e_0, e_1, \dots, e_m, \dots$ une base orthonormée de $L^2(S)$. Soit $\mathbf{H}_M = (e_0, e_1, \dots, e_M)$, l'espace vectoriel des fonctions h de la forme $\alpha_0 e_0 + \alpha_1 e_1 + \dots + \alpha_M e_M$, $M \in \mathbb{N}$. A toute fonction f de $L^2(S)$, nous associons sa *projection orthogonale* f_M sur \mathbf{H}_M définie par :

$$f_M = \langle f, e_0 \rangle_S e_0 + \langle f, e_1 \rangle_S e_1 + \dots + \langle f, e_M \rangle_S e_M . \quad (2.5)$$

La norme $\|f_M - f\|_S$ tend vers zéro lorsque M tend vers l'infini. Autrement dit, f_M converge vers f dans $L^2(S)$ lorsque M tend vers l'infini. L'erreur commise en confondant f_M avec f peut être rendue aussi faible que nécessaire en choisissant M assez grand. Enfin, $\sum_{m=0}^{\infty} \langle f, e_m \rangle_S e_m$ s'appelle la *développement en série de f dans $L^2(S)$ par rapport à la base $e_m, m \in \mathbb{N}$* .

2.1.2 Harmoniques sphériques

Définitions et illustrations

Nous appelons *harmoniques sphériques*[MI68] la famille de fonctions Y_{mn}^σ qui forme une base orthonormée de $L^2(S)$ définie⁶ par [G96, W99] :

$$Y_{mn}^\sigma(\theta, \delta) = \sqrt{(2m+1)\varepsilon_n \frac{(m-n)!}{(m+n)!}} P_{mn}(\sin \delta) \times \begin{cases} \cos n\theta & \text{si } \sigma = 1 \\ \sin n\theta & \text{si } \sigma = -1 \end{cases} \quad (2.6)$$

où θ et δ sont respectivement l'azimut et l'élévation, m et n sont des entiers positifs tels que $n \leq m$, σ prend les valeurs $+1$ ou -1 , et ε_n est égal à 1 si $n = 0$ et égal à 2 si $n > 0$. Nous appelons l'indice m l'*ordre*⁷ de l'harmonique. Le terme en racine carrée permet de normaliser ces fonctions au sens du produit scalaire (2.2) de telle façon que leur norme soit unitaire

⁶ Les harmoniques sphériques sont souvent définies de façon complexe, mais puisque nous nous intéressons à la description de fonctions réelles, le champ acoustique en particulier, nous ne mentionnons que les harmoniques sphériques réelles.

⁷ En règle générale, c'est l'indice n qui est appelé *ordre* de l'harmonique, m en étant le *degré*. Nous dérogeons cependant à cette règle pour rester en conformité avec de précédents travaux concernant la technologie HOA.

(cf. Équation (2.9)). Enfin, les fonctions $P_{mn}(x)$ sont les *polynômes de Legendre associés*, définis pour tout x appartenant à $[-1, 1]$ par :

$$P_{mn}(x) = (1 - x^2)^{\frac{n}{2}} \frac{d^n}{dx^n} P_m(x) \quad (2.7)$$

avec $m \geq 0$, et où $P_m(x)$ est le polynôme de Legendre de première espèce d'ordre m , qui peut être calculé numériquement grâce à une relation de récurrence :

$$\begin{cases} P_0(x) = 1 \\ P_1(x) = x \\ (m+1)P_{m+1}(x) = (2m+1)xP_m(x) - mP_{m-1}(x), \quad m > 1 \end{cases} \quad (2.8)$$

A titre d'exemple, les expressions des harmoniques sphériques d'ordres 0 à 3 sont données explicitement au Tableau 2.1, et illustrées pour ces mêmes ordres sous de formes différentes à la Figure 2.2 et à la Figure 2.3. La première de ces figures représente par une échelle de couleur les variations des harmoniques sphériques à la surface d'une sphère de rayon unitaire ($r=1$) en fonction de l'azimut et de l'élévation. Observons que lorsque θ et δ parcourent la sphère, les harmoniques sphériques Y_{mn}^σ s'annulent selon m cercles (tracés en noir) : n cercles verticaux (azimut θ constant), et $m-n$ cercles horizontaux (élévation δ constante). Entre les cercles d'annulation, les fonctions sont alternativement positives et négatives. La seconde figure représente des surfaces sphériques d'équation $r(\theta, \delta) = Y_{mn}^\sigma(\theta, \delta)$, c'est-à-dire déformées suivant les variations des harmoniques sphériques, laissant apparaître des lobes positifs et négatifs. Notons enfin qu'un ordre m donné est constitué de $2m+1$ harmoniques sphériques distinctes, les fonctions nulles Y_{m0}^{-1} étant ignorées.

Quelques propriétés

Puisqu'elles définissent une base orthonormée de $L^2(S)$, les harmoniques sphériques sont orthogonales deux à deux, ce qui s'exprime mathématiquement :

$$\begin{aligned} \langle Y_{mn}^\sigma, Y_{m'n'}^{\sigma'} \rangle_S &= \frac{1}{4\pi} \int_{\theta=0}^{2\pi} \int_{\delta=-\frac{\pi}{2}}^{+\frac{\pi}{2}} Y_{mn}^\sigma(\theta, \delta) Y_{m'n'}^{\sigma'}(\theta, \delta) \cos \delta \, d\delta \, d\theta \\ &= \delta_{mm'} \delta_{nn'} \delta_{\sigma\sigma'} \end{aligned} \quad (2.9)$$

où m, m', n et n' appartiennent à \mathbb{N} avec $n \leq m$, σ et σ' valent $+1$ ou -1 , et δ_{ij} est le symbole de Kronecker défini par :

$$\begin{cases} \delta_{ii} = 1, \forall i \\ \delta_{ij} = 0, \forall i, j, i \neq j \end{cases}$$

De plus, la norme de chaque harmonique sphérique est unitaire, soit $\|Y_{mn}^\sigma(\theta, \delta)\|_S = 1$ pour tout m et n appartenant à \mathbb{N} tels que $n \leq m$, et $\sigma = \pm 1$.

	Normalisation		Fonctions de Legendre associées	
	$\sqrt{\varepsilon_0 \frac{(2m+1)(m-n)!}{(m+n)!}}$		$P_{mn}(\sin \delta)$	
Ordre $m = 0$ (1 composante)	$Y_{00}^1(\theta, \delta) =$	1	1	
Ordre $m = 1$ (3 composantes)	$Y_{11}^1(\theta, \delta) =$	$\sqrt{3}$	$\cos \delta$	$\times \cos \theta$
	$Y_{11}^{-1}(\theta, \delta) =$	$\sqrt{3}$	$\cos \delta$	$\times \sin \theta$
	$Y_{10}^1(\theta, \delta) =$	$\sqrt{3}$	$\sin \delta$	
Ordre $m = 2$ (5 composantes)	$Y_{22}^1(\theta, \delta) =$	$\sqrt{\frac{5}{12}}$	$3 \cos^2 \delta$	$\times \cos(2\theta)$
	$Y_{22}^{-1}(\theta, \delta) =$	$\sqrt{\frac{5}{12}}$	$3 \cos^2 \delta$	$\times \sin(2\theta)$
	$Y_{21}^1(\theta, \delta) =$	$\sqrt{\frac{5}{3}}$	$3 \cos \delta \sin \delta$	$\times \cos \theta$
	$Y_{21}^{-1}(\theta, \delta) =$	$\sqrt{\frac{5}{3}}$	$3 \cos \delta \sin \delta$	$\times \sin \theta$
	$Y_{20}^1(\theta, \delta) =$	$\sqrt{5}$	$\left(\frac{3 \sin^2 \delta - 1}{2} \right)$	
Ordre $m = 3$ (7 composantes)	$Y_{33}^1(\theta, \delta) =$	$\sqrt{\frac{7}{360}}$	$15 \cos^3 \delta$	$\times \cos(3\theta)$
	$Y_{33}^{-1}(\theta, \delta) =$	$\sqrt{\frac{7}{360}}$	$15 \cos^3 \delta$	$\times \sin(3\theta)$
	$Y_{32}^1(\theta, \delta) =$	$\sqrt{\frac{7}{60}}$	$15 \cos^2 \delta \sin \delta$	$\times \cos(2\theta)$
	$Y_{32}^{-1}(\theta, \delta) =$	$\sqrt{\frac{7}{60}}$	$15 \cos^2 \delta \sin \delta$	$\times \sin(2\theta)$
	$Y_{31}^1(\theta, \delta) =$	$\sqrt{\frac{7}{6}}$	$\left(\frac{3 \cos \delta \cdot (5 \sin^2 \delta - 1)}{2} \right)$	$\times \cos \theta$
	$Y_{31}^{-1}(\theta, \delta) =$	$\sqrt{\frac{7}{6}}$	$\left(\frac{3 \cos \delta \cdot (5 \sin^2 \delta - 1)}{2} \right)$	$\times \sin \theta$
	$Y_{30}^1(\theta, \delta) =$	$\sqrt{7}$	$\sin \delta \cdot \frac{(5 \sin^2 \delta - 3)}{2}$	

Tableau 2.1 – Expressions analytiques des harmoniques sphériques pour les ordres 0 à 3.

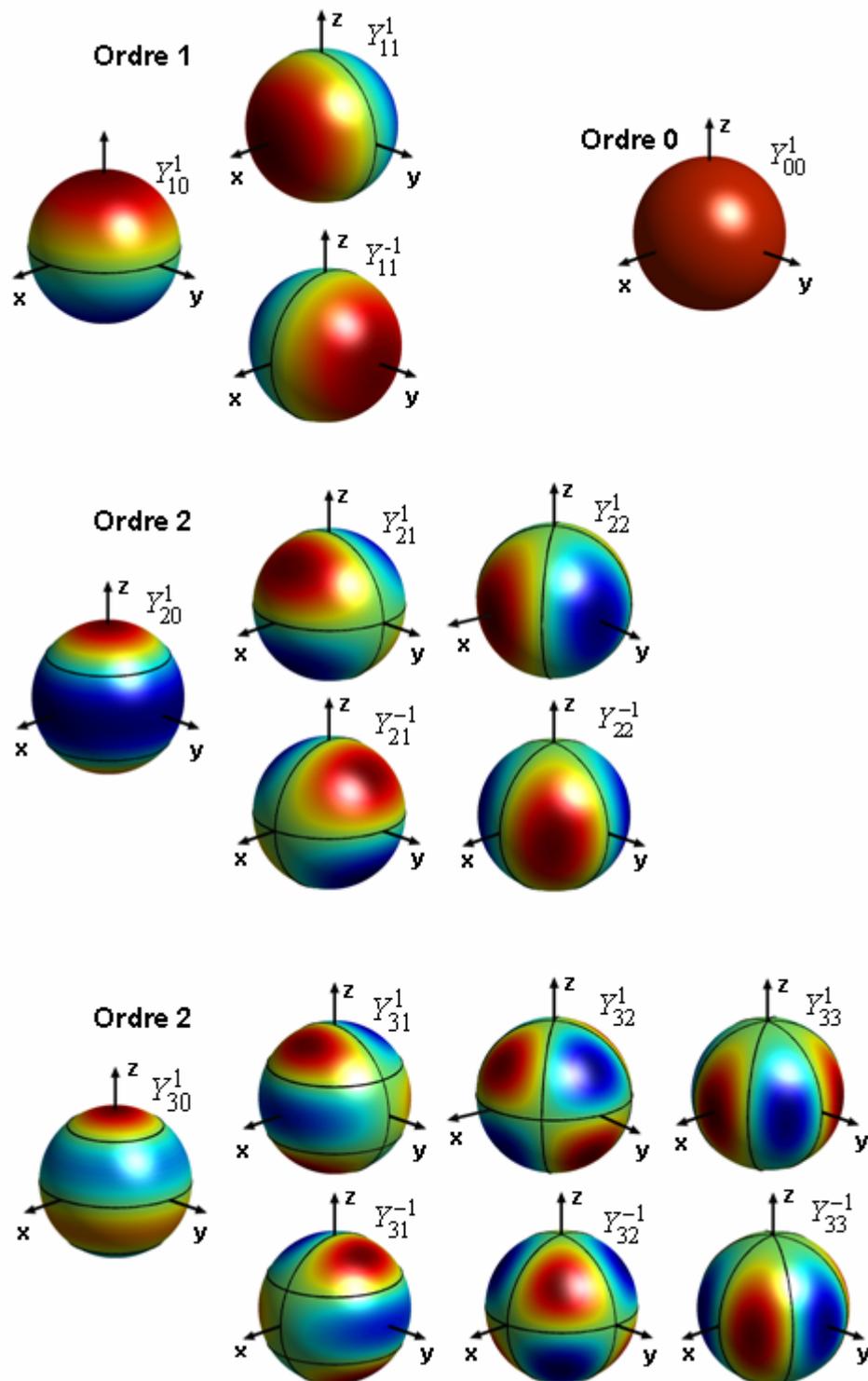


Figure 2.2 – Illustrations des harmoniques sphériques à la surface d'une sphère de rayon unitaire pour les ordres $m = 0$ à 3. Les couleurs du rouge au bleu en passant par le vert correspondent à une échelle respective de valeurs positives à négatives en passant par zéro. Les cercles d'annulation des harmoniques sphériques sont représentés en noir.

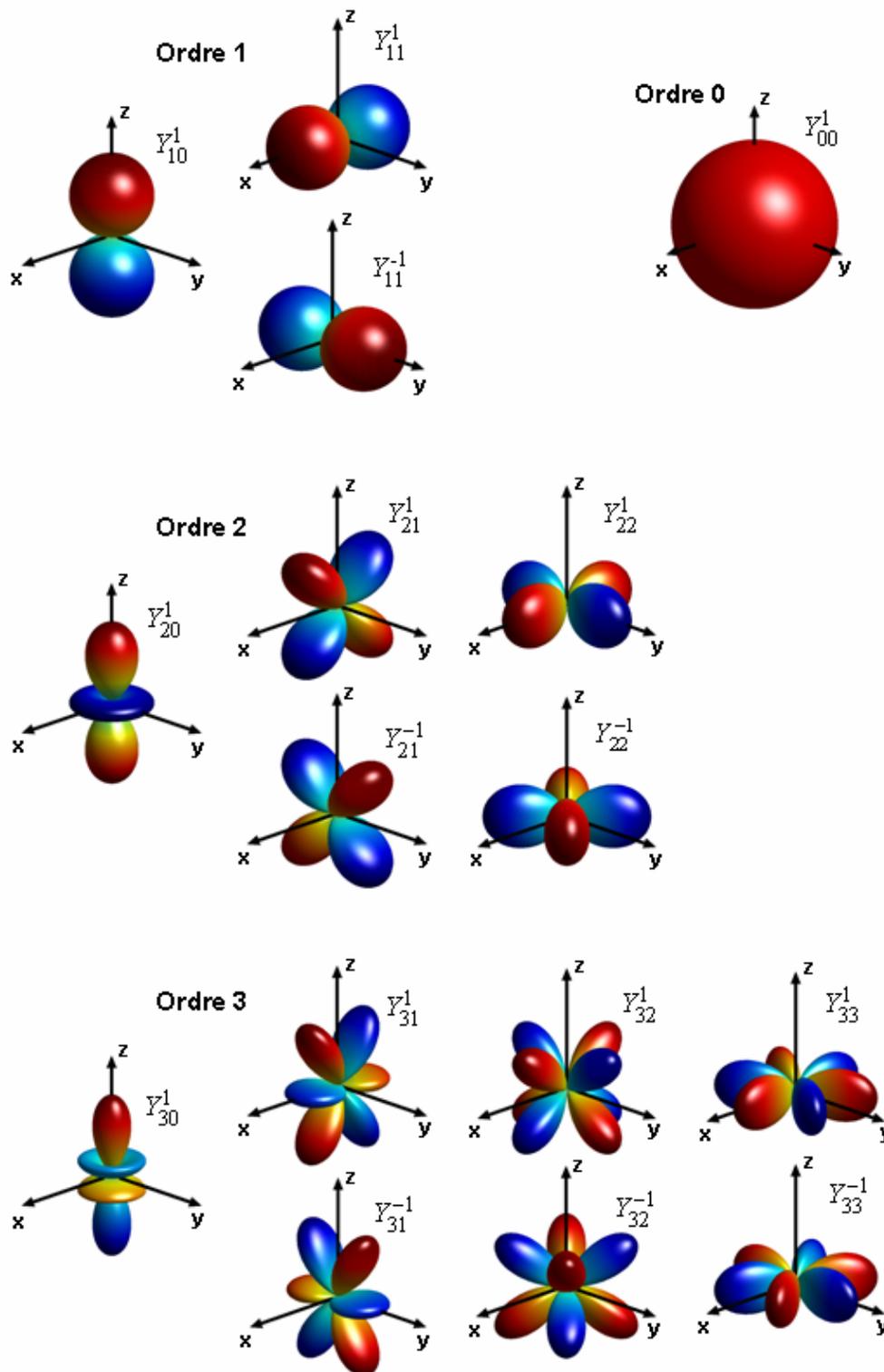


Figure 2.3 – Illustration des harmoniques sphériques pour les ordres $m=0$ à 3 . Les couleurs rouge et bleue indiquent respectivement des valeurs positives et négatives.

Enfin, mentionnons dès à présent un théorème important concernant les harmoniques sphériques et que nous utiliserons par la suite, le *théorème d'addition*. En notant (θ_1, δ_1) et (θ_2, δ_2) deux points distincts de la sphère S de rayon unitaire, et γ l'angle formé par ces deux points, nous avons la relation trigonométrique suivante :

$$\cos \gamma = \cos \theta_1 \cos \theta_2 + \sin \theta_1 \sin \theta_2 \cos(\delta_1 - \delta_2). \quad (2.10)$$

Le théorème d'addition [A85] établit alors que pour un ordre m entier naturel :

$$\sum_{n=0}^m \sum_{\sigma=\pm 1} Y_{mn}^\sigma(\theta_1, \delta_1) Y_{mn}^\sigma(\theta_2, \delta_2) = (2m+1) P_m(\cos \gamma). \quad (2.11)$$

2.2 Développement en série d'harmoniques sphériques et approximation d'une fonction de $L^2(S)$

2.2.1 Développement en harmoniques sphériques d'une fonction de $L^2(S)$

Définitions

Puisque les harmoniques sphériques forment une base orthonormée de $L^2(S)$, toute fonction f appartenant à $L^2(S)$ et donc définie sur la sphère S , peut s'écrire sous la forme d'un développement en série par rapport à cette base, que nous appellerons simplement *développement de f en harmoniques sphériques* :

$$f(\theta, \delta) = \sum_{m=0}^{\infty} \sum_{n=0}^m \sum_{\sigma=\pm 1} f_{mn}^\sigma Y_{mn}^\sigma(\theta, \delta), \quad (2.12)$$

avec

$$\begin{aligned} f_{mn}^\sigma &= \langle f, Y_{mn}^\sigma \rangle_S \\ &= \frac{1}{4\pi} \int_{\theta=0}^{2\pi} \int_{\delta=-\frac{\pi}{2}}^{+\frac{\pi}{2}} f(\theta, \delta) Y_{mn}^\sigma(\theta, \delta) \cos \delta \, d\delta \, d\theta. \end{aligned} \quad (2.13)$$

Cette dernière équation (2.13) exprime la projection de la fonction f sur la base des harmoniques sphériques. L'ensemble des coefficients f_{mn}^σ détermine entièrement la fonction f et constitue la *transformée de Fourier sphérique* de f de façon analogue à la transformée de Fourier des fonctions unidimensionnelles. Par analogie également avec cette dernière, une fonction f pour laquelle les coefficients f_{mn}^σ sont nuls pour tout $m \geq M$, M appartenant à \mathbb{N} , est dite à bande limitée de largeur M .

Spectre d'énergie d'une fonction de $L_2(S)$

Il peut être utile de représenter l'énergie d'une fonction f appartenant à $L^2(S)$ portée par l'ensemble des $2m+1$ coefficients f_{mn}^σ appartenant à un même ordre m . Nous définissons pour cela le *spectre d'énergie de Fourier sphérique* d'une telle fonction f comme la suite infinie de

nombre S_m qui donne l'amplitude de chacun des termes d'ordre m de son développement en harmoniques sphériques :

$$S_m = \frac{1}{2m+1} \sum_{n=0}^m \sum_{\sigma=\pm 1} |f_{mn}^\sigma|^2,$$

où m et n appartiennent à \mathbb{N} , et $\sigma = \pm 1$. Un coefficient S_m élevé indique que la contribution des harmoniques sphériques d'ordre m dans la fonction f est élevée.

2.2.2 Approximation d'une fonction de $L_2(S)$

Troncature du développement en harmoniques sphériques

Le développement en harmonique sphérique d'une fonction f appartenant à $L^2(S)$ converge absolument et uniformément dans $L^2(S)$. Par conséquent, cette même fonction peut être approximée par sa projection f_M sur une partie de la base des harmoniques sphériques, limitée aux ordres inférieurs ou égaux à un entier naturel M :

$$f_M(\theta, \delta) = \sum_{m=0}^M \sum_{n=0}^m \sum_{\sigma=\pm 1} f_{mn}^\sigma Y_{mn}^\sigma(\theta, \delta), \quad M \in \mathbb{N}. \quad (2.14)$$

Ceci revient à *tronquer* le développement en harmonique sphériques (2.12) de la fonction f à l'ordre M . La fonction f_M est ainsi la *meilleure* approximation en moyenne quadratique de la fonction f par une combinaison linéaire des harmoniques sphériques d'ordre inférieur ou égal à M . Autrement dit, lorsque M tend vers l'infini, f_M converge vers f en moyenne quadratique, ou encore $\|f - f_M\|_S^2$ tend vers zéro. L'approximation f_M de la fonction f peut être rendue aussi bonne que voulue en choisissant un ordre M de troncature assez élevé.

Erreur d'approximation

La précision de cette approximation f_M peut être mesurée en considérant par exemple la différence d'énergie entre les fonctions f et f_M :

$$\bar{e}_M = \frac{\|f - f_M\|_S^2}{\|f\|_S^2} = \frac{\int_{\theta=0}^{2\pi} \int_{\delta=-\frac{\pi}{2}}^{+\frac{\pi}{2}} |f(\theta, \delta) - f_M(\theta, \delta)|^2 \cos \delta \, d\delta \, d\theta}{\int_{\theta=0}^{2\pi} \int_{\delta=-\frac{\pi}{2}}^{+\frac{\pi}{2}} |f(\theta, \delta)|^2 \cos \delta \, d\delta \, d\theta}. \quad (2.15)$$

\bar{e}_M s'appelle l'*erreur quadratique moyenne normalisée* d'approximation de la fonction f par la fonction f_M , associée à un ordre M de troncature. La normalisation par $\|f\|_S^2$ permet d'obtenir une erreur \bar{e}_M sans dimension comprise entre 0 et 1, interprétable comme un pourcentage en la multipliant par 100.

Comme l'erreur \bar{e}_M est une erreur moyenne sur l'ensemble de la surface de la sphère, il peut s'avérer utile de définir en complément un indicateur de la dispersion des erreurs d'approximation

autour de cette valeur moyenne \bar{e}_M . Nous pouvons par exemple calculer la variance associée à l'erreur moyenne \bar{e}_M :

$$\sigma_M^2 = \|e_M - \bar{e}_M\|_S^2 = \frac{1}{4\pi} \int_{\theta=0}^{2\pi} \int_{\delta=-\frac{\pi}{2}}^{+\frac{\pi}{2}} |e_M(\theta, \delta) - \bar{e}_M|^2 \cos \delta d\delta d\theta \quad (2.16)$$

L'étude de la valeur maximale de l'erreur d'approximation lorsque θ et δ parcourent l'ensemble de la sphère est une alternative possible à celle de la variance [B95, D01, P00]. Elle ne donne pas forcément d'indication représentative de la dispersion globale de l'erreur mais permet d'en connaître la limite supérieure.

2.3 Application au champ de pression acoustique

2.3.1 Développement du champ acoustique en harmoniques sphériques

Équation d'onde et équation d'Helmholtz

Sous les hypothèses classiques d'acoustique linéaire, en supposant que l'air est un fluide parfait, la pression acoustique obéit à une *équation différentielle de propagation d'ondes*, appelée équation de d'Alembert [B98, F94]. Dans une région de l'espace *ne comprenant pas de source sonore*, cette équation s'écrit :

$$\Delta p(r, \theta, \delta, t) - \frac{1}{c} \frac{\partial^2 p(r, \theta, \delta, t)}{\partial t^2} = 0 \quad (2.17)$$

où (r, θ, δ) sont les coordonnées du point d'observation exprimées dans un système de coordonnées sphériques conforme à la Figure 2.1, t est le temps, c la vitesse de propagation du son dans l'air valant approximativement 340 m/s, et Δp désigne le Laplacien de la pression acoustique p .

Dans le cas particulier des sons purs de fréquence angulaire $\omega = 2\pi f$ exprimée en radians par seconde, la pression acoustique peut s'écrire $p(r, \theta, \delta, t) = p(r, \theta, \delta) e^{i\omega t}$. En introduisant cette expression dans l'équation d'onde (2.17), nous aboutissons à l'*équation d'Helmholtz* [F94] homogène qui ne dépend plus du temps :

$$(\Delta + k^2) p(r, \theta, \delta) = 0, \quad (2.18)$$

où apparaît le nombre d'onde $k = \omega/c$. Les solutions de l'équation d'Helmholtz représentent la partie spatiale des solutions de l'équation d'onde. Même si les sons auxquels nous avons affaire ne sont pas purement sinusoïdaux, il est souvent utile de résoudre des problèmes d'acoustique dans ce cadre là puisqu'un signal temporel peut être représenté sous la forme d'une intégrale de Fourier, c'est-à-dire une combinaison linéaire de fonctions qui dépendent sinusoïdalement du temps.

Série de Fourier-Bessel

Nous faisons l'hypothèse que les sources sonores se situent à l'extérieur d'une sphère de rayon R centrée à l'origine de notre repère. A l'intérieur de cette sphère ($r < R$), le champ acoustique obéit à l'équation d'Helmholtz homogène. La résolution de cette équation par la méthode dite de *séparation des variables* [B98] dans le système de coordonnées sphériques défini à la Figure 2.1 aboutit à une solution générale appelée *série de Fourier-Bessel* :

$$p(kr, \theta, \delta) = \sum_{m=0}^{+\infty} i^m j_m(kr) \sum_{n=0}^m \sum_{\sigma=\pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \delta), \quad (2.19)$$

Les fonctions angulaires $Y_{mn}^{\sigma}(\theta, \delta)$ sont les harmoniques sphériques définies à la Section 2.1.2, auxquelles sont associés les coefficients B_{mn}^{σ} que nous appelons *coefficients de Fourier sphérique*. Les fonctions à dépendance radiale $j_m(kr)$ sont les *fonctions de Bessel sphériques*, que nous pouvons définir pour tout x appartenant à \mathbb{N} par :

$$\begin{cases} j_0(0) = 1 \\ j_m(x) = (-1)^m x^m \left(\frac{1}{x} \frac{d}{dx} \right)^m \frac{\sin x}{x} \end{cases} \quad (2.20)$$

Les premières expressions analytiques des fonctions de Bessel sphériques sont données par :

$$\begin{cases} j_0(x) = \frac{\sin x}{x} \\ j_1(x) = \frac{\sin x}{x^2} - \frac{\cos x}{x} \\ j_2(x) = \left(\frac{3}{x^3} - \frac{1}{x} \right) \sin x - \frac{3}{x^2} \cos x \\ j_3(x) = \frac{\cos x}{x} - \frac{6 \sin x}{x^2} - \frac{15 \cos x}{x^3} + \frac{15 \sin x}{x^4} \end{cases} \quad (2.21)$$

Les illustrations graphiques correspondantes sont regroupées à la Figure 2.4. Enfin, chaque coefficient B_{mn}^{σ} résulte de la projection orthogonale de la pression acoustique p sur l'harmonique sphérique identiquement indiquée :

$$\begin{aligned} i^m j_m(kr) B_{mn}^{\sigma} &= \langle p, Y_{mn}^{\sigma} \rangle_S \\ &= \frac{1}{4\pi} \int_{\theta=0}^{2\pi} \int_{\delta=-\frac{\pi}{2}}^{+\frac{\pi}{2}} p(kr, \theta, \delta) Y_{mn}^{\sigma}(\theta, \delta) \cos \delta d\delta d\theta. \end{aligned} \quad (2.22)$$

Pour les valeurs non nulles de $j_m(kr)$, les coefficients B_{mn}^{σ} peuvent se déduire de la connaissance de la pression acoustique grâce à cette dernière équation.

La série de Fourier-Bessel peut être vue comme un développement en harmoniques sphériques du champ de pression acoustique sur une infinité de surfaces sphérique de rayon kr , c'est-à-dire dans un volume sphérique. Ce développement est entièrement déterminé par les coefficients B_{mn}^{σ} . Les fonctions de Bessel sphériques $j_m(kr)$ décrivent la contribution des différents modes harmoniques en fonction de kr , c'est-à-dire en fonction du rayon d'observation et de la fréquence.

Nous observons notamment sur la Figure 2.4 que ces modes apparaissent progressivement lorsque kr augmente et que leur amplitude maximale décroît dans le même temps pour traduire une énergie globale constante. D'autres approches sont possibles pour interpréter la série de Fourier-Bessel. Par exemple, si nous identifions les harmoniques sphériques à des directivités microphoniques qui scrutent le champ acoustique, plus ces harmoniques sont nombreuses et d'ordre élevé, meilleure est la résolution angulaire, et par conséquent la connaissance du champ acoustique autour du point de référence. Enfin, les harmoniques sphériques sont également interprétables comme les dérivées spatiales successives du champ acoustique au point de référence (cf. Chapitre 5, Section 5.1). Plus ces dérivées sont connues en nombre, plus il est possible d'extrapoler le champ acoustique sur une zone large qui s'étend autour du point de référence, comme dans le cas d'une série de Taylor. Le lien très étroit entre la série de Fourier-Bessel et la série de Taylor a notamment été démontré par Dickins et Kennedy [DK99].

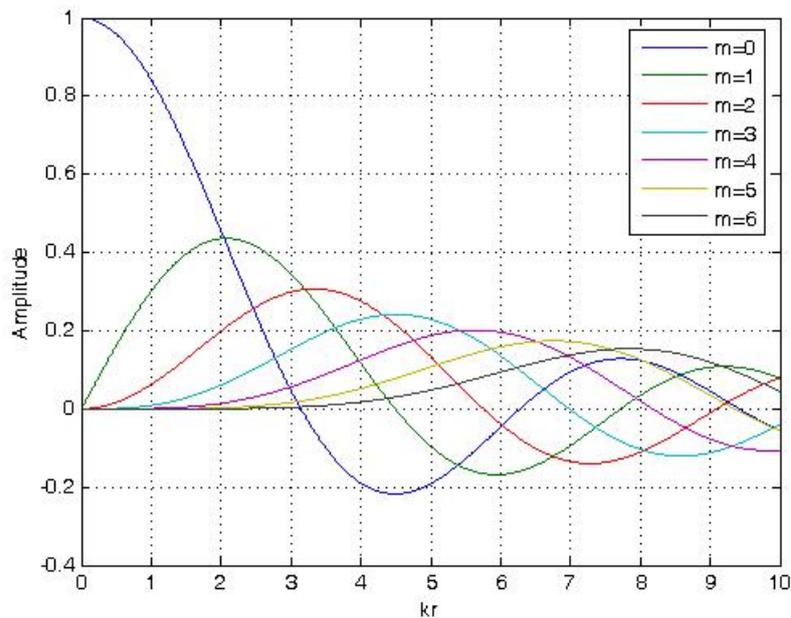


Figure 2.4 – Illustration des fonctions de Bessel sphériques en fonction de kr pour les ordres $m = 0$ à 6 .

Cas particuliers d'une onde plane

Considérons à présent le cas particulier d'une onde plane qui provient d'une direction (θ_p, δ_p) de l'espace. Cette onde transporte un signal de fréquence ω qui se propage avec une amplitude S constante. La pression acoustique résultante s'écrit alors :

$$p(kr, \theta, \delta) = S e^{ikr \cos \gamma}, \quad (2.23)$$

ou encore [B98]

$$p(kr, \theta, \delta) = S \sum_{m=0}^{+\infty} (2m+1) i^m j_m(kr) P_m(\cos \gamma), \quad (2.24)$$

où γ désigne l'angle entre la direction (θ, δ) du point d'observation et la direction (θ_p, δ_p) de provenance de l'onde plane. En faisant appel au théorème d'addition des fonctions de Legendre (cf. Équation (2.11)), nous pouvons écrire :

$$p(r, \theta, \delta) = S \sum_{m=0}^{+\infty} i^m j_m(kr) \sum_{n=0}^m \sum_{\sigma=\pm 1} Y_{mn}^\sigma(\theta_p, \delta_p) Y_{mn}^\sigma(\theta, \delta). \quad (2.25)$$

En identifiant l'équation (2.25) avec la série de Fourier-Bessel (2.19), nous obtenons finalement l'expression analytique des coefficients de la transformée de Fourier sphérique d'un champ de pression acoustique engendré par une onde plane incidente :

$$B_{mn}^\sigma = S Y_{mn}^\sigma(\theta_p, \delta_p), \quad m \in \mathbb{N}, \quad 0 < n < m, \quad \sigma = \pm 1. \quad (2.26)$$

Les coefficients B_{mn}^σ sont ainsi définis par la valeur des harmoniques sphériques dans la direction de provenance de l'onde plane, pondérée par l'amplitude S du signal transporté.

Cas particulier d'une onde sphérique

Le second cas particulier que nous étudions ici est celui d'un champ acoustique engendré par une onde sphérique émise par une source ponctuelle placée à une distance finie de l'origine, au point de coordonnées $\mathbf{r}_s = (r_s, \theta_s, \delta_s)$. Le champ de pression acoustique résultant peut alors s'exprimer :

$$p(r, \theta, \delta) = S \frac{r_s}{e^{-ikr_s}} \frac{e^{-ik(\mathbf{r}-\mathbf{r}_s)}}{|\mathbf{r}-\mathbf{r}_s|}, \quad (2.27)$$

où S est l'amplitude du signal transporté, mesuré à l'origine, et $\mathbf{r} = (r, \theta, \delta)$ décrit la position du point d'observation. Le terme r_s / e^{-ikr_s} permet d'obtenir une phase nulle de référence et une amplitude S à l'origine. Remarquons par ailleurs que l'amplitude de la pression acoustique décroît de façon inversement proportionnelle à la distance entre le point d'observation et la source émettrice. En développant l'équation (2.27), nous aboutissons finalement à l'expression [MI68] :

$$p(r, \theta, \delta) = S \sum_{m=0}^{\infty} i^m j_m(kr) F_m(kr_s) Y_{mn}^\sigma(\theta_s, \delta_s) Y_{mn}^\sigma(\theta, \delta), \quad (2.28)$$

avec

$$F_m(kr_s) = i^{-(m+1)} \frac{h_m^-(kr_s)}{h_0^-(kr_s)} \tag{2.29}$$

où les fonctions $h_m^-(kr_s) = j_m(kr_s) - i n_m(kr_s)$ sont les fonctions de Hankel sphériques divergentes d'ordre m , les fonctions $n_m(kr_s)$ étant les fonctions de Neumann. L'identification terme à terme de l'équation (2.28) avec l'expression générale de la série de Fourier-Bessel (2.19) permet d'obtenir l'expression des coefficients B_{mn}^σ associés à une onde sphérique émise par une source ponctuelle :

$$B_{mn}^\sigma = S Y_{mn}^\sigma(\theta_s, \delta_s) F_m(kr_s) \tag{2.30}$$

Ces coefficients de Fourier sphérique diffèrent de ceux qui définissent une onde plane par un facteur $F_m(kr_s)$ qui dépend de kr_s , c'est-à-dire à la fois de la distance r_s de la source émettrice et implicitement de la fréquence f .

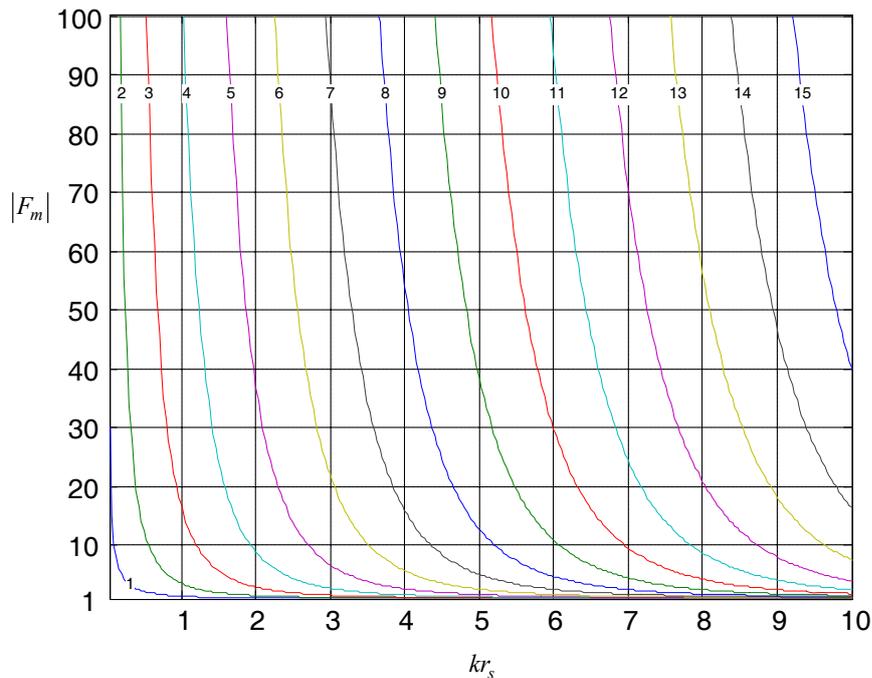


Figure 2.5 – Illustration du module des fonctions $F_m(kr_s)$ pour les ordres 1 à 15 en fonction de la distance r_s à la source et pour les fréquences 100 Hz (a), 200 Hz (b), 500 Hz (c), et 1000 Hz (d). La distance de référence à laquelle l'amplitude S est mesurée est fixée à $d = 1$.

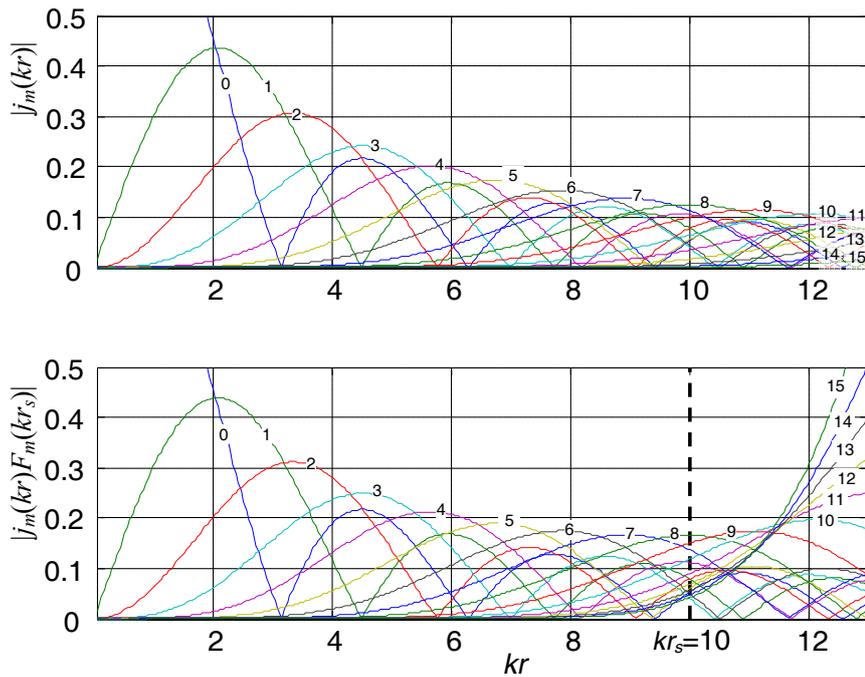


Figure 2.6 – Illustrations des fonctions $|j_m(kr)|$ (haut) et $|j_m(kr)F_m(kr_s)|$ (bas) en fonction de kr et kr_s pour les ordres 0 à 15.

La Figure 2.5 illustre le module des fonctions $F_m(kr_s)$ en fonction de kr_s pour les ordres 1 à 15. Nous constatons que pour un ordre m particulier, le facteur $F_m(kr_s)$ correspond à une amplification qui tend vers l'infini lorsque kr_s tend vers zéro, c'est-à-dire lorsque la fréquence ou la distance à la source tend vers 0. De plus, pour une même valeur de kr_s cette amplification est d'autant plus importante que l'ordre m est élevé. Enfin, lorsque kr_s tend vers l'infini, les fonctions complexes $F_m(kr_s)$ tendent vers la valeur réelle 1. Cela signifie que l'onde sphérique peut être approximativement assimilée à une onde plane à une distance de la source émettrice et une fréquence suffisamment élevées.

Si on considère maintenant la série de Fourier Bessel dans sa globalité, la contribution des différentes composantes harmoniques en fonction du rayon (et de la fréquence) est finalement décrite par le produit des fonctions de Bessel sphérique $j_m(kr)$ et les fonctions de champ proche $F_m(kr_s)$. L'amplitude des fonctions de Bessel sphériques $j_m(kr)$ expriment une décroissance des niveaux maximums des modes harmoniques au fur et à mesure qu'ils apparaissent tandis que celle des fonctions $F_m(kr_s)$ traduisent une amplification très prononcée de ces mêmes modes lorsque le point d'observation s'approche de la source émettrice. Sur la Figure 2.6 sont représentés en fonction de kr et kr_s les modules des fonctions $j_m(kr)$ et $j_m(kr)F_m(kr_s)$. Nous observons que l'amplitude maximale de ces dernières fonctions baisse dans un premier temps lorsque m augmente, puis s'inverse autour de la valeur de kr_s , c'est-à-dire de la position radiale de la source émettrice, pour traduire une forte amplification. La ligne verticale en pointillée indique la position de la source. À partir de cette position la série de Fourier-

Bessel n'est plus définie puisqu'elle est valable uniquement dans une zone de l'espace qui exclut toute source sonore. Portons une attention particulière aux fonctions $j_m(kr)F_m(kr_s)$ lorsque kr s'approche de kr_s . Outre une amplification de l'amplitude des modes dominants par rapport au cas d'une onde plane, nous remarquons que l'ensemble des modes (même si seulement quelques uns sont représentés) intervient, et ce d'autant plus que kr est proche de kr_s . Il en résulte une pression acoustique qui tend vers l'infini lorsque kr tend vers kr_s , c'est-à-dire lorsque le point d'observation tend à se rapprocher de la source émettrice. Nous reviendrons plus en détails sur ces considérations au Chapitre 4 lorsqu'il s'agira de modéliser et reproduire avec la technologie HOA le champ proche d'une source virtuelle.

2.3.2 Approximation du champ de pression acoustique

Troncature de la série de Fourier-Bessel

De la même façon qu'il est possible d'approximer une fonction de $L^2(S)$ en tronquant son développement en harmoniques sphériques (cf. Section 2.2.2), une approximation p_M d'un champ de pression acoustique peut être obtenue en effectuant une troncature de la série de Fourier-Bessel à un ordre M fini ($M \in \mathbb{N}$) :

$$p_M(kr, \theta, \delta) = \sum_{m=0}^M i^m j_m(kr) \sum_{n=0}^m \sum_{\sigma=\pm 1} B_{mn}^\sigma Y_{mn}^\sigma(\theta, \delta). \quad (2.31)$$

La troncature nous permet de connaître en partie le champ acoustique à partir d'un nombre limité de coefficients de Fourier sphérique B_{mn}^σ . Chaque ordre comportant $2m+1$ coefficients, le nombre K total de coefficients correspondant à une description du champ sonore tronquée à l'ordre M se calcul par la formule :

$$K = \sum_{m=0}^M (2m+1) = (M+1)^2 \quad (2.32)$$

Comme allons le constater par la suite, l'ordre de troncature se révèle être un indicateur de la taille de la zone de l'espace dans laquelle nous connaissons assez bien le champ acoustique. Pour connaître avec précision des variations de pression acoustique de plus en plus petites (correspondant à des fréquences de plus en plus hautes) sur la surface d'une sphère de rayon de plus en plus large, nous devons connaître les coefficients B_{mn}^σ à des ordres de plus en plus élevés.

Analyse de l'erreur d'approximation d'une onde plane

En se référant à l'équation (2.26), une approximation d'une onde plane s'obtient en remplaçant les coefficients B_{mn}^σ dans l'équation (2.31) par $S Y_{mn}^\sigma(\theta_p, \delta_p)$, S étant le signal transporté par cette onde plane. Nous nous intéressons plus particulièrement à l'approximation d'une onde plane dans le plan horizontal d'élévation nulle ($\delta = 0$). Dans ce plan, l'erreur moyenne de troncature peut être calculée en considérant la différence d'énergie entre l'onde plane idéale et l'onde approximée, intégrée sur un périmètre circulaire de rayon kr :

$$\bar{e}_M = \frac{\int_0^{2\pi} |p(kr, \theta, \delta = 0) - p_M(kr, \theta, \delta = 0)|^2 d\theta}{\int_0^{2\pi} |p(kr, \theta, \delta = 0)|^2 d\theta}. \quad (2.33)$$

Cette erreur moyenne peut être complétée par une indication de la dispersion de l'erreur, par exemple en lui associant un calcul de la déviation standard ou de l'erreur maximale. Nous choisissons de calculer l'erreur maximale normalisée, définie sur un périmètre de rayon kr par :

$$e_M^{\max} = \max_{\theta} \frac{|p(kr, \theta, \delta = 0) - p_M(kr, \theta, \delta = 0)|^2}{|p(kr, \theta, \delta = 0)|^2}. \quad (2.34)$$

La Figure 2.7 et la Figure 2.8 représentent respectivement l'erreur moyenne \bar{e}_M d'approximation dans le plan horizontal d'une onde plane pour les ordres de troncature allant de 1 à 15, et l'erreur maximale associée. Il s'avère que cette erreur est indépendante de l'incidence azimutale de l'onde plane (son élévation est supposée nulle). Plusieurs remarques peuvent être faites en observant la Figure 2.7. Nous remarquons tout d'abord que chaque courbe est strictement croissante, excepté dans un intervalle d'erreur trop importante que nous ignorerons. Cela signifie que l'erreur associée à chaque ordre de troncature croît avec kr , c'est-à-dire lorsque la distance à l'origine ou la fréquence augmente. Pour conserver une valeur d'erreur constante, par exemple 7 % représenté sur la Figure 2.7 par la ligne horizontale en pointillés, nous observons d'autre part que lorsque kr augmente, l'ordre M de troncature doit également augmenter.

Il peut être utile de disposer d'un critère permettant de déduire de la valeur de kr l'ordre de troncature requis pour obtenir une qualité d'approximation jugée suffisante. L'observation de la Figure 2.7 montre par exemple qu'une valeur d'erreur de 7 % (-11,5 dB) est approximativement obtenue pour un ordre de troncature M lorsque $kr = M$. Par conséquent, étant donné une fréquence limite supérieure et un rayon d'approximation maximal, une règle simple permet de choisir raisonnablement l'ordre de troncature nécessaire :

$$M = \lceil kr \rceil, \quad (2.35)$$

où $\lceil kr \rceil$ désigne l'entier le plus proche de kr . Ce même critère a déjà été proposé par Ward et Abhayapala [WA01a, WA01b] pour l'approximation d'une onde plane dans un volume sphérique, la valeur d'erreur associée, légèrement plus faible que celle obtenue le plan horizontal, valant 4 % (-14 dB). Par exemple, pour une approximation correcte du champ acoustique dans une zone circulaire de rayon 8,5 cm du plan horizontal (approximativement le rayon d'une tête humaine) et une fréquence maximale de $f = 2000$ Hz, un ordre $M=3$ est nécessaire selon ce critère. Selon ce critère toujours, la technique de spatialisation sonore *Ambisonics* (cf. Chapitre 3) qui se base sur une représentation du champ acoustique sous la forme d'une série de Fourier-Bessel tronquée à l'ordre $M=1$ ne suffit donc pas pour reproduire un champ acoustique au niveau des oreilles d'un auditeur dans un domaine moyenne et haute fréquence.

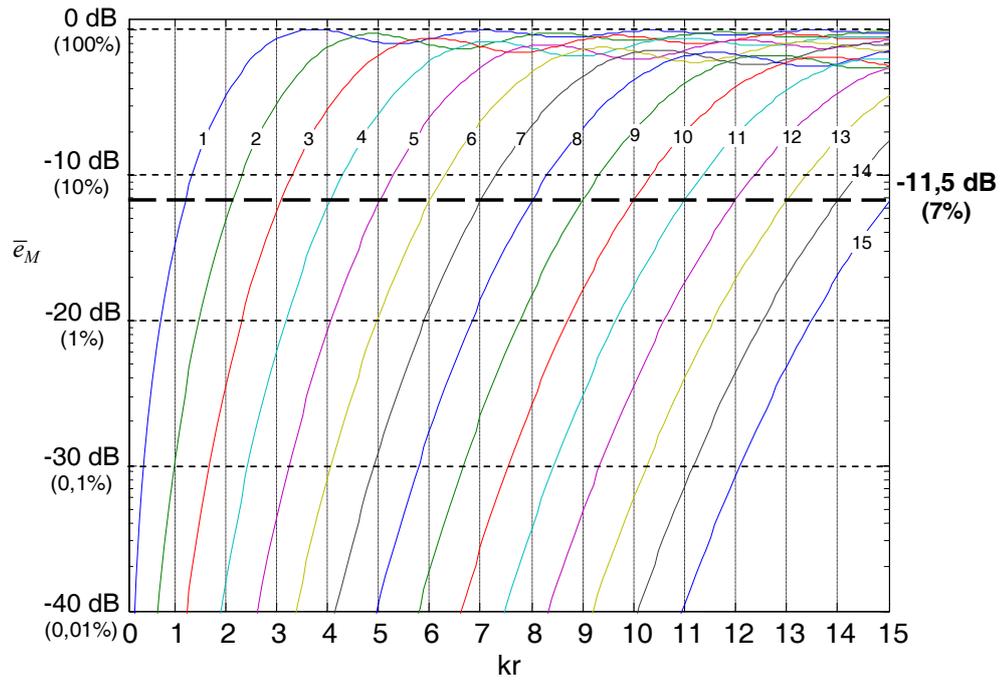


Figure 2.7 – Erreur moyenne \bar{e}_M d'approximation d'une onde plane émise et observée dans le plan horizontal, pour les ordres de troncatures allant de 1 à 15.

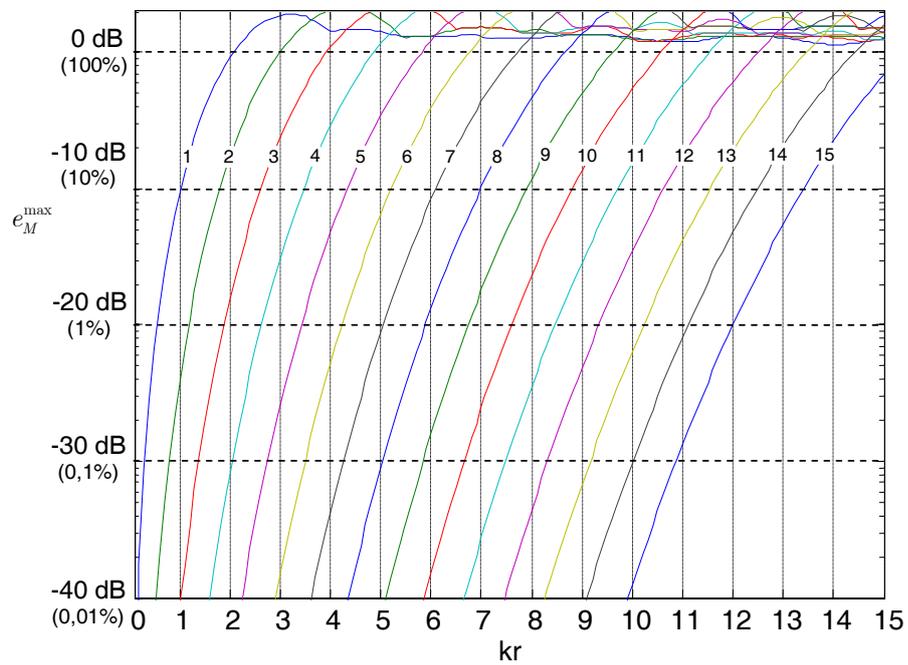


Figure 2.8 – Erreur maximale e_M^{\max} d'approximation d'une onde plane émise et observée dans le plan horizontal, pour les ordres de troncatures allant de 1 à 15.

Pour illustrer de façon plus concrète la qualité d'approximation du champ acoustique engendré par une onde plane en fonction de l'ordre de troncature, nous représentons à la Figure 2.9 la pression sonore p_M tronquée aux ordres M allant de 1 à 5 sur une surface circulaire d'1 mètre de diamètre dans le plan horizontal et pour différentes fréquences. Les courbes étoilées délimitent les zones de l'espace pour lesquelles l'erreur, calculée selon la formule $|p(kr, \theta, \delta = 0) - p_M(kr, \theta, \delta = 0)|^2 / |p(kr, \theta, \delta = 0)|^2$, reste inférieure à 7 %, soit $-11,5$ dB. Conformément aux remarques précédentes, cette zone croît non seulement avec l'ordre de troncature, mais également lorsque la fréquence de l'onde diminue.

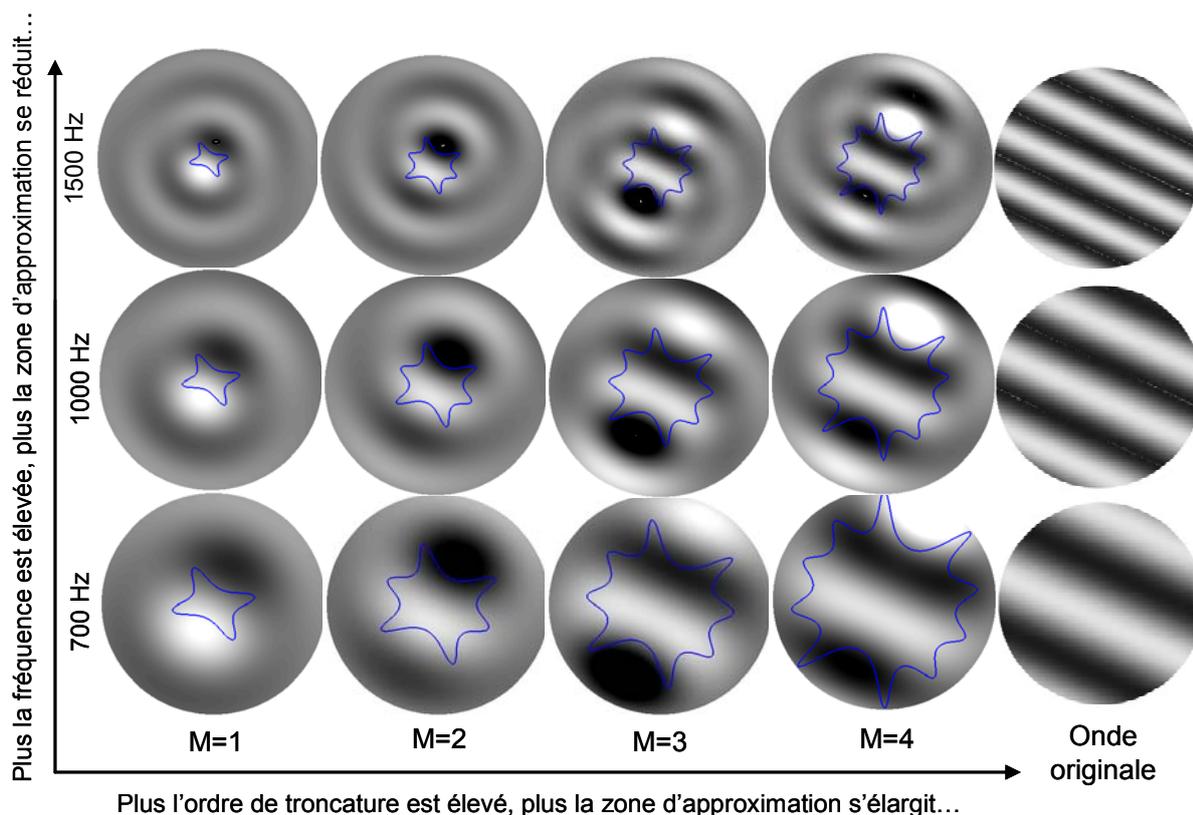


Figure 2.9 - Illustration dans le plan horizontal de l'approximation d'une onde plane en fonction de l'ordre de troncature et de la fréquence. Le diamètre de la zone est de 1 m. La valeur de la pression acoustique est représentée par les niveaux de gris. Les courbes étoilées indiquent la limite des zones d'approximation où l'erreur est inférieure à 7% ($-11,5$ dB).

Analyse de l'erreur d'approximation d'une onde sphérique

Nous étudions à présent l'erreur d'approximation du champ acoustique engendré par une onde sphérique résultant de la troncature de la série de Fourier-Bessel. Cette étude peut être menée à partir de l'expression analytique des coefficients B_{mn}^σ associés à une onde sphérique, donnés à l'équation (2.30), et que nous insérons dans l'expression (2.31). Nous choisissons d'analyser non

pas l'erreur moyenne de troncature sur une zone circulaire, mais l'erreur absolue dans une direction particulière, celle de la source émettrice :

$$e_M = \frac{|p(kr, \theta_s, \delta_s) - p_M(kr, \theta_s, \delta_s)|^2}{|p(kr, \theta_s, \delta_s)|^2}, \tag{2.36}$$

où (θ_s, δ_s) décrit la position angulaire de la source émettrice. Ce choix d'erreur est motivé par le fait que pour un ordre de troncature fini, comparativement au cas d'une onde plane, l'erreur d'approximation se dégrade brutalement et très fortement lorsque l'on s'approche de la source. L'erreur e_M est représentée sur la Figure 2.11 pour les ordres 1 à 20. À titre de comparaison, l'erreur de troncature associée à une onde plane dans la direction de provenance (θ_s, δ_s) est représentée à la Figure 2.10. Notons que cette dernière figure est très proche de l'erreur maximale représentée à la Figure 2.8, en cohérence avec les illustrations de la Figure 2.9 où l'erreur constante indiquée par la courbe étoilée correspond à un rayon minimum dans la direction de provenance de la source. Dans le cas particulier d'une onde sphérique, contrairement à celui d'une onde plane, la diminution de l'erreur d'approximation n'est pas régulière lorsque kr augmente. Elle tend en effet vers 1 à tous les ordres lorsque kr tend vers kr_s , et les différentes courbes associées à ces différents ordres de troncature se tassent jusqu'à ne plus traduire d'amélioration significative de l'approximation. Dans le même temps, l'amplitude du champ de pression acoustique tend vers l'infini. Ces deux observations sont totalement cohérentes avec les commentaires de la Figure 2.6 qui soulignaient la contribution théorique de l'infinité des modes harmoniques à l'emplacement de la source, entraînant un niveau de pression sonore infini et une erreur d'approximation maximale.

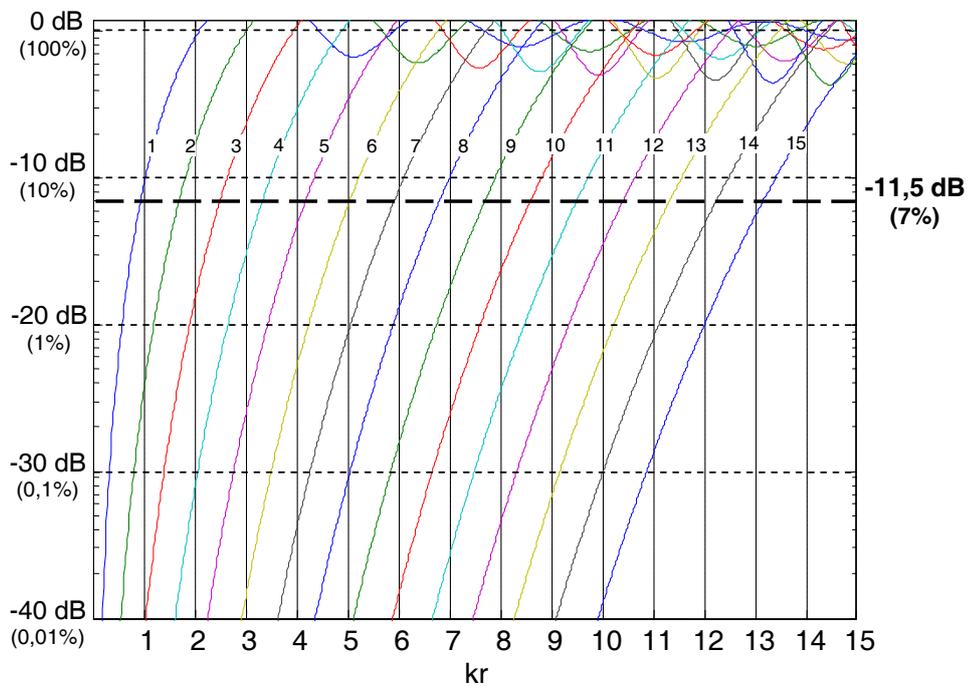


Figure 2.10 – Erreur e_M d'approximation d'une onde plane évaluée dans sa direction de provenance et pour les ordres de troncature allant de 1 à 15.

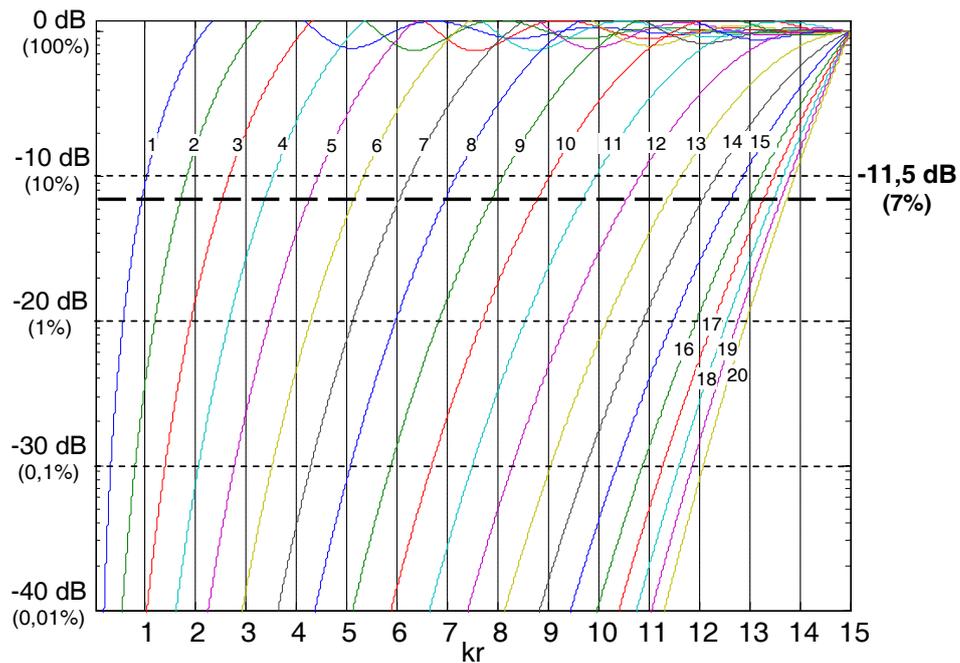


Figure 2.11 – Erreur e_M d'approximation d'une onde sphérique évaluée dans sa direction de provenance et pour les ordres de troncature allant de 1 à 15.

Comparativement à la Figure 2.9, la Figure 2.12 représente sur une surface circulaire d'1 mètre de diamètre dans le plan horizontal et pour différentes fréquences, la pression sonore p_M engendrée par la propagation d'une onde sphérique et tronquée aux ordres M allant de 1 à 5. La source émettrice, représentée par un point de couleur verte, est placée à un rayon de 1 m. Les courbes étoilées délimitent les zones de l'espace pour lesquelles l'erreur, calculée selon la formule $|p(kr, \theta, \delta = 0) - p_M(kr, \theta, \delta = 0)|^2 / |p(kr, \theta, \delta = 0)|^2$, reste inférieure à 7 %, soit -11,5 dB. Ces illustrations confirment que les ordres les plus élevés contribuent moins efficacement à l'extension de la zone d'approximation délimitée par la courbe étoilée au fur et mesure que cette zone s'approche de la position de la source.

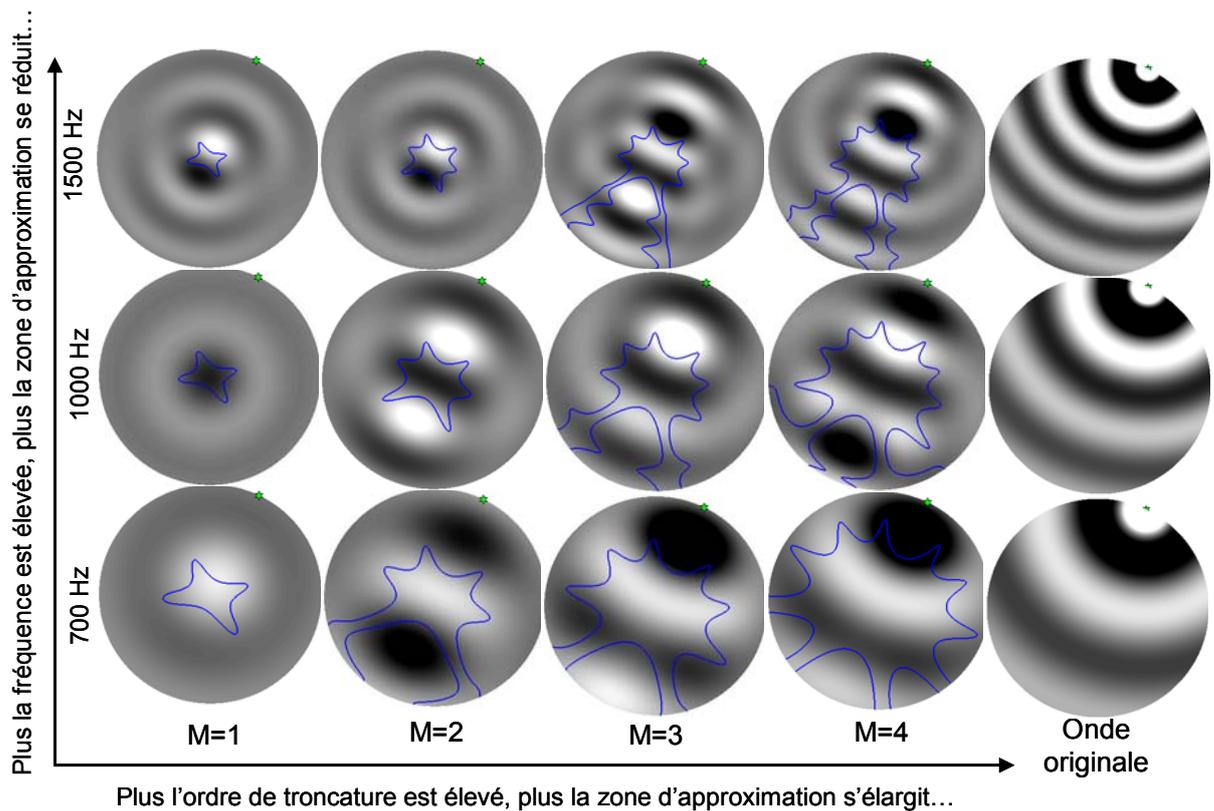


Figure 2.12 – Illustration dans le plan horizontal de l'approximation du champ de pression acoustique engendré par une onde sphérique, en fonction de l'ordre de troncature et de la fréquence. Le diamètre de la zone est de 1 m. La valeur de la pression acoustique est représentée par les niveaux de gris. Les courbes étoilées indiquent la limite des zones d'approximation où l'erreur est inférieure à 7% (-11,5 dB).

Références bibliographiques du Chapitre 2

- [CG77] CRAVEN, P. et GERZON, M.A. *Coincident Microphone Simulation Covering Three Dimensional Space and Yielding Various Directional Outputs*. U.S. Patent 4,042,779, août 1977.
- [B98] BRUNEAU, M. *Manuel d'acoustique fondamentale*. Paris: Hermès, 1998, 576 p.
- [G96] GROEMER, H. *Geometric Applications of Fourier Series and Spherical Harmonics*. Cambridge: Cambridge University Press, 1996, 343 p.
- [R89] REINHARD, H. *Cours de mathématique du signal*. Paris: Dunod Université, 1989p.
- [MI68] MORSE, P.M. et INGARD, K.U. *Theoretical Acoustics*. New York: Mc Graw-Hill, 1968, 927 p.
- [W99] WILLIAMS, E.G. *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. London, UK: Academic Press, 1999, 306 p.
- [A85] ARFKEN, G. *Mathematical Methods for Physicists, 3rd ed.* San Diego, CA: Academic Press, 1985, 985 p.
- [B95] BAMFORD, J.S. *An Analysis of Ambisonics Sound Systems of First and Second Order*. M. Sc. Thesis, Waterloo: University of Waterloo, 1995.
- [D01] DANIEL, J. *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. Paris: Paris 6, 2001.
- [P00] POLETTI, M.A. *A Unified Theory of Horizontal Holographic Sound Systems*. J. Audio Eng. Soc., 2000, Vol. 48, n° 12, p. 1155-1182.
- [F94] FILIPPI, P.J.T. *Acoustique générale*. Paris: Les éd. de physique, Société française d'acoustique, 1994, 382 p.
- [DK99] DICKINS, G. et KENNEDY, R.A. *Towards Optimal Sound Field Representation*. 106th AES Convention, Munich, 1999.
- [WA01a] WARD, D.B. et ABHAYAPALA, T.D. *Performance Bounds on Sound Field Reproduction Using a Loudspeaker Array*. IEEE WASPAA 2001, New York, octobre 2001.
- [WA01b] WARD, D.B. et ABHAYAPALA, T.D. *Reproduction of a Plane-Wave Sound Field Using an Array of Loudspeakers*. IEEE Transaction on Speech and Audio Processing, 2001, Vol. 9, n° 6, p. 697-707.

Chapitre 3

Higher Order Ambisonics

Higher Order Ambisonics (HOA) est une évolution de la technique de spatialisation sonore *Ambisonics* initiée notamment par Gerzon [G85], Felgett [F74] et Craven [CG77] dans les années 1970, et qui a pour but la reproduction du champ acoustique au voisinage d'un point de l'espace, celui où se situe l'auditeur⁸. Pour cela, la pression acoustique et le vecteur de vitesse particulaire sont mesurés ou simulés en un point, puis reconstruits avec des haut-parleurs. Ces paramètres peuvent également être interprétés comme les premiers coefficients (ordre 0 et ordre 1) de la décomposition du champ acoustique en composantes spatiales : les harmoniques cylindriques en deux dimensions (2D), ou les harmoniques sphériques en trois dimensions (3D). L'extension d'*Ambisonics* à des ordres supérieurs de la décomposition appelée *Higher Order Ambisonics* (HOA), d'abord en 2D par Bamford [B95] puis en 3D par Daniel [D01], permet d'améliorer la reproduction du champ sonore, notamment en élargissant la zone d'écoute (cf. Chapitre 2).

La Figure 3.1 représente les principales possibilités offertes par la technologie HOA en termes d'*encodage spatial*, de *manipulations* et de *décodage spatial* d'une scène sonore. L'encodage spatial peut être envisagé sous deux aspects distincts : la simulation des caractéristiques spatiales du champ acoustique à partir de signaux audio monophoniques (synthèse de source virtuelle), ou l'enregistrement naturel par un dispositif microphonique. Puisque les signaux HOA sont une représentation physique d'une scène sonore indépendante du dispositif de restitution, le décodage spatial peut s'adapter à différents systèmes de diffusion (circulaires, sphériques, stéréophoniques 2.0, 5.1, casque d'écoute, etc.). Dans le présent chapitre, nous décrivons en détails chacune de ces possibilités.

⁸ Signalons d'autre part que d'autres études menées en parallèle et indépendamment de celles sur *Ambisonics* ont proposé des solutions de restitution sonore comparables [CS71].

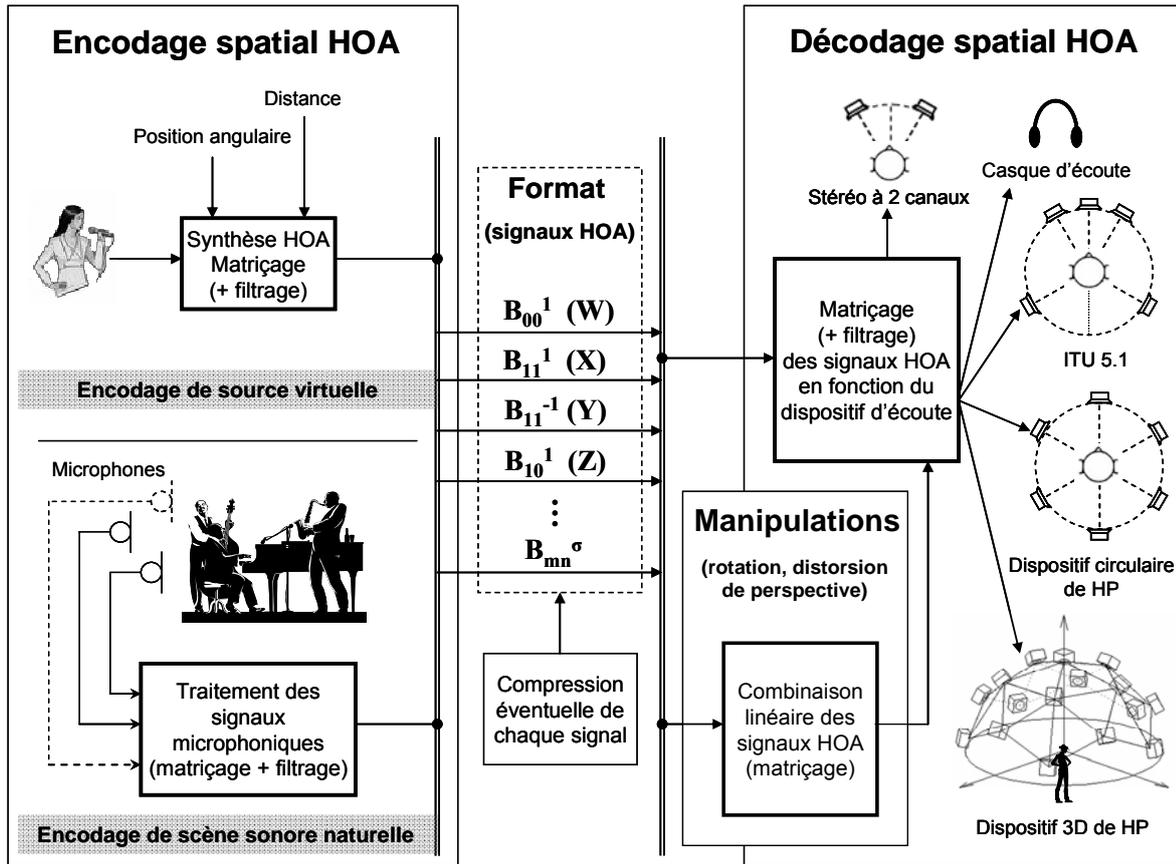


Figure 3.1 – Aperçu global de la technologie HOA : de l'encodage au décodage spatial.

3.1 Représentation de scène sonore avec HOA

3.1.1 Format de représentation associé : les signaux HOA

Définition du format de représentation du champ acoustique : les *signaux HOA*

La représentation d'une scène sonore 3D avec la technologie HOA se base sur l'expression du champ de pression acoustique sous la forme d'une série de Fourier-Bessel, c'est-à-dire un développement en harmoniques sphériques (cf. Chapitre 1). De cette façon, les coefficients de Fourier sphérique B_{mn}^{σ} associés aux harmoniques sphériques identiquement indicées ($m \geq 0$, $0 \leq n \leq m$ et $\sigma = \pm 1$) suffisent à décrire entièrement une scène sonore quelconque et constituent, dans la technologie HOA, le format de représentation de scène sonore. Implicitement exprimés dans le domaine fréquentiel, les coefficients B_{mn}^{σ} peuvent naturellement être transposés dans le domaine temporel au moyen d'une transformée de Fourier inverse :

$$b_{mn}^{\sigma}(t) = \int_{-\infty}^{+\infty} B_{mn}^{\sigma}(\omega) e^{i\omega t} dt. \quad (3.1)$$

Les fonctions $b_{mn}^{\sigma}(t)$ sont des signaux audio au même titre que les signaux monophoniques, stéréophoniques ou encore binauraux ; nous les appelons simplement *signaux HOA*.

En pratique, nous limitons à un ordre maximal M fini les signaux HOA associés à une représentation de scène sonore, ce qui revient à tronquer la série de Fourier-Bessel (cf. Chapitre 2). Chaque groupe de signaux HOA appartenant à un même ordre m étant au nombre de $2m+1$, le nombre total K_M^{3D} de signaux HOA associés à une représentation HOA d'ordre M s'obtient simplement par le calcul suivant :

$$K_M^{(3D)} = \sum_{m=0}^M 2m+1 = (M+1)^2. \quad (3.2)$$

Comme nous l'avons expliqué et illustré au Chapitre 2, l'augmentation de l'ordre maximal M des signaux HOA se traduit par une amélioration de la précision de la description sonore, soit à la fois une amélioration de la précision d'approximation en haute fréquence et un élargissement de la zone de représentation du champ acoustique autour du point de référence, qui deviendra lors de la reproduction le point d'écoute privilégié de l'auditeur.

Cas particulier d'une représentation bidimensionnelle (2D)

Actuellement, la restitution du champ acoustique se fait encore majoritairement par un dispositif de restitution en deux dimensions, c'est-à-dire avec des haut-parleurs uniquement situés dans le plan horizontal contenant les oreilles de l'auditeur. En supposant que les sources sonores encodées émettent dans ce même plan horizontal, les points de l'espace peuvent être décrits dans ce plan par un système de coordonnées polaires (r, θ) , r étant le rayon et θ l'azimut. La représentation du champ acoustique est ainsi simplifiée et le nombre de signaux HOA requis par ordre fortement réduit, comme nous pouvons le constater au Tableau 3.1. Plus précisément, à la base des harmoniques sphériques se substitue alors celle des *harmoniques cylindriques* [B98], notées Y_m^σ et définies de la façon suivante :

$$\begin{cases} Y_0^1(\theta) = 1 \\ Y_m^\sigma(\theta) = \sqrt{2} \cos(m\theta) \quad \text{si } \sigma = 1 \\ Y_m^\sigma(\theta) = \sqrt{2} \sin(m\theta) \quad \text{si } \sigma = -1 \end{cases} \quad (3.3)$$

avec $m \geq 1$. De façon analogue aux coefficients de normalisation des harmoniques sphériques, le facteur $\sqrt{2}$ qui s'applique aux harmoniques cylindriques d'ordre supérieur ou égal à 1 permet de normaliser ces fonctions au sens du produit scalaire défini sur le cercle C de rayon unitaire par :

$$\langle Y_m^\sigma, Y_{m'}^{\sigma'} \rangle_C = \frac{1}{2\pi} \int_{\theta=0}^{2\pi} Y_m^\sigma(\theta) Y_{m'}^{\sigma'}(\theta) d\theta = \delta_{mm'} \delta_{\sigma\sigma'}. \quad (3.4)$$

Le nombre de signaux HOA nécessaires à une description 2D d'ordre M du champ sonore se réduit ainsi à :

$$K_M^{(2D)} = 2M + 1. \quad (3.5)$$

La conversion d'une description 3D vers une description 2D du champ acoustique se fait en ne conservant que les signaux HOA dont l'indice n est égal à l'ordre m et par l'application d'un facteur de conversion :

$$\begin{cases} Y_m^\sigma(\theta) = Y_{00}^1(\theta, \delta = 0) = 1 \\ Y_m^\sigma(\theta) = \sqrt{\frac{2^{2m} m!^2}{(2m+1)!}} Y_{mm}^\sigma(\theta, \delta = 0), \quad \sigma = \pm 1 \end{cases} \quad (3.6)$$

Les premiers facteurs de conversion sont donnés en fonction de l'ordre m dans le Tableau 3.2.

Les systèmes de reproduction sonore HOA en deux dimensions ont fait l'objet de nombreuses études [CS71, B95, P96, P00]. Cependant, si l'emploi des harmoniques cylindriques est plus aisé et se justifie amplement pour la simulation et la restitution d'une onde plane dans le plan horizontal, elles ne sont en revanche pas adaptées à l'enregistrement de scène sonore naturelle ou encore à la description de la propagation d'une onde sphérique en champ proche.

	Ordre	Nombre de signaux HOA	Nombre total cumulé
3D	0	1	1
	1	3	4
	2	5	9
	3	7	16
	m	$2m+1$	$(m+1)^2$
2D	0	1	1
	1	2	3
	2	2	5
	3	2	7
	m>0	2	$2m+1$

Tableau 3.1 – Nombre de signaux HOA décrivant le champ sonore en fonction de la dimension et de l'ordre.

m	1	2	3	4	5	6
$\sqrt{\frac{2^{2m} m!^2}{(2m+1)!}}$	1	$\sqrt{\frac{2}{3}}$	$\sqrt{\frac{8}{15}}$	$\sqrt{\frac{16}{35}}$	$\sqrt{\frac{128}{315}}$	$\sqrt{\frac{256}{693}}$

Tableau 3.2 – Premiers facteurs de conversion des harmoniques sphériques (3D) vers les harmoniques cylindriques (2D).

3.1.2 Conventions et notations

La propriété d'orthonormalité des bases harmoniques sphériques Y_{mn}^σ en 3D et cylindriques Y_m^σ en 2D s'avère utile pour la manipulation des signaux HOA, notamment l'étape de décodage spatial. Cependant, ce choix de normalisation reste quelque peu arbitraire et différentes conventions ont déjà été introduites suivant les auteurs. L'existence de ces différentes conventions pose des problèmes de compatibilité lors de la manipulation des signaux HOA car elles instaurent des relations différentes d'amplitude entre les composantes HOA. Le Tableau 3.3 résume les principales conventions en vigueur ainsi que leurs formules de conversion qui assurent la compatibilité avec les définitions des étapes d'encodage, de manipulation et de décodage que nous donnerons par la suite. Historiquement, la première convention de normalisation des harmoniques sphériques a été donnée par Gerzon pour les ordres 0 et 1 [G73]. Celle-ci ne possède pas de propriétés mathématiques particulières, si ce n'est que sa restriction 2D est conforme aux harmoniques cylindriques Y_0^1 , Y_1^1 , et Y_1^{-1} . Pour limiter l'amplitude des signaux HOA à celle du signal monophonique encodé (cas d'encodage de source virtuelle), Malham et Furse [M95, M05, F99] ont introduit une normalisation valable jusqu'à l'ordre 3 et extensible aux ordres supérieurs [D00], qui permet d'obtenir des harmoniques sphériques et cylindriques bornées entre -1 et +1. Toutefois, dans un souci de compatibilité avec les conventions introduites par Gerzon, c'est-à-dire pour conserver les mêmes rapports d'amplitude entre les différentes composantes, la composante omnidirectionnelle (ordre 0) est pondérée par le facteur $1/\sqrt{2}$. Signalons enfin que nos conventions de référence Y_{mn}^σ pour la représentation 3D et Y_m^σ pour la représentation 2D sur lesquelles nous nous basons dans la suite du présent document sont respectivement conformes aux conventions $Y_{mn}^{\sigma(N3D)}$ et $Y_{mm}^{\sigma(N2D)}$ introduites par Daniel [D01].

Auteurs	m	Conventions	Conv. vers Y_{mn}^σ (3D)	Conv. vers Y_m^σ (2D)
Gerzon [G73]	0	$W = 1$	$Y_{00}^1 = W$	$Y_0^1 = W$
	1	$X = \sqrt{2} \cos \theta \cos \delta$	$Y_{11}^1 = \sqrt{\frac{3}{2}} X$	$Y_1^1 = X$
		$Y = \sqrt{2} \sin \theta \cos \delta$	$Y_{11}^{-1} = \sqrt{\frac{3}{2}} Y$	$Y_1^{-1} = Y$
		$Z = \sqrt{2} \sin \delta$	$Y_{10}^1 = \sqrt{\frac{3}{2}} Z$	_____
Furse et Malham [F99] [M95] [M05]	0	$Y_{00}^{1(\text{FuMa})} = 1/\sqrt{2}$	$Y_{00}^1 = \sqrt{2} Y_{00}^{1(\text{FuMa})}$	$Y_0^1 = \sqrt{2} Y_{00}^{1(\text{FuMa})}$
	1	$Y_{11}^{1(\text{FuMa})} = \cos \theta \cos \delta$	$Y_{11}^1 = \sqrt{3} Y_{11}^{1(\text{FuMa})}$	$Y_1^1 = \sqrt{2} Y_{11}^{1(\text{FuMa})}$
		$Y_{11}^{-1(\text{FuMa})} = \sin \theta \cos \delta$	$Y_{11}^{-1} = \sqrt{3} Y_{11}^{-1(\text{FuMa})}$	$Y_1^{-1} = \sqrt{2} Y_{11}^{-1(\text{FuMa})}$
		$Y_{10}^{1(\text{FuMa})} = \sin \delta$	$Y_{10}^1 = \sqrt{3} Y_{10}^{1(\text{FuMa})}$	_____
	2	$Y_{22}^{1(\text{FuMa})} = \cos 2\theta \cos^2 \delta$	$Y_{22}^1 = \sqrt{\frac{15}{4}} Y_{22}^{1(\text{FuMa})}$	$Y_2^1 = \sqrt{2} Y_{22}^{1(\text{FuMa})}$
		$Y_{22}^{-1(\text{FuMa})} = \sin 2\theta \cos^2 \delta$	$Y_{22}^{-1} = \sqrt{\frac{15}{4}} Y_{22}^{-1(\text{FuMa})}$	$Y_2^{-1} = \sqrt{2} Y_{22}^{-1(\text{FuMa})}$
		$Y_{21}^{1(\text{FuMa})} = \cos \theta \cos 2\delta$	$Y_{21}^1 = \sqrt{\frac{15}{4}} Y_{21}^{1(\text{FuMa})}$	_____
		$Y_{21}^{-1(\text{FuMa})} = \sin \theta \cos 2\delta$	$Y_{21}^{-1} = \sqrt{\frac{15}{4}} Y_{21}^{-1(\text{FuMa})}$	_____
		$Y_{20}^{1(\text{FuMa})} = (3 \sin^2 \delta - 1)/2$	$Y_{20}^1 = \sqrt{5} Y_{20}^{1(\text{FuMa})}$	_____
	3	$Y_{33}^{1(\text{FuMa})} = \cos 3\theta \cos^3 \delta$	$Y_{33}^1 = \sqrt{\frac{105}{24}} Y_{33}^{1(\text{FuMa})}$	$Y_3^1 = \sqrt{2} Y_{33}^{1(\text{FuMa})}$
$Y_{33}^{-1(\text{FuMa})} = \sin 3\theta \cos^3 \delta$		$Y_{33}^{-1} = \sqrt{\frac{105}{24}} Y_{33}^{-1(\text{FuMa})}$	$Y_3^{-1} = \sqrt{2} Y_{33}^{-1(\text{FuMa})}$	
$Y_{32}^{1(\text{FuMa})} = \cos 2\theta \sin \delta \cos^2 \delta$		$Y_{32}^1 = \sqrt{\frac{105}{4}} Y_{32}^{1(\text{FuMa})}$	_____	
$Y_{32}^{-1(\text{FuMa})} = \sin 2\theta \sin \delta \cos^2 \delta$		$Y_{32}^{-1} = \sqrt{\frac{105}{4}} Y_{32}^{-1(\text{FuMa})}$	_____	
$Y_{31}^{1(\text{FuMa})} = \frac{8 \cos \theta \cos \delta (5 \sin^2 \delta - 1)}{11}$		$Y_{31}^1 = \frac{11\sqrt{21}}{16\sqrt{2}} Y_{31}^{1(\text{FuMa})}$	_____	
$Y_{31}^{-1(\text{FuMa})} = \frac{8 \sin \theta \cos \delta (5 \sin^2 \delta - 1)}{11}$		$Y_{31}^{-1} = \frac{11\sqrt{21}}{16\sqrt{2}} Y_{31}^{-1(\text{FuMa})}$	_____	
	$Y_{20}^{1(\text{FuMa})} = \frac{\sin \delta (5 \sin^2 \delta - 3)}{2}$	$Y_{30}^1 = \sqrt{7} Y_{30}^{1(\text{FuMa})}$	_____	
Bamford [B95]	0	$Y_0^{1(\text{Bamf})} = 1$	_____	$Y_0^1 = Y_0^{1(\text{Bamf})}$
	$m \geq 1$	$Y_m^{1(\text{Bamf})} = \cos m\theta$ $Y_m^{-1(\text{Bamf})} = \sin m\theta$	_____	$Y_1^1 = \sqrt{2} Y_1^{1(\text{Bamf})}$ $Y_1^{-1} = \sqrt{2} Y_1^{-1(\text{Bamf})}$

Tableau 3.3 – Principales conventions de normalisation des harmoniques sphériques (3D) et cylindriques (2D) suivant les auteurs, et formules de conversion dans nos conventions de référence. Notons que la réduction d'une restitution 3D à une restitution 2D sous-entend $\delta = 0$.

3.2 Encodage spatial et manipulation de scène sonore

3.2.1 Encodage spatial de source virtuelle et de scène sonore naturelle

Équations d'encodage d'une onde plane

L'hypothèse d'onde plane est tout à fait classique et très répandue : les propriétés théoriques de propagation unidimensionnelle et sans perte d'énergie de l'onde sonore sont simples à modéliser et à mettre en œuvre. Examinons le cas d'un champ acoustique 3D engendré par une onde plane d'incidence $\mathbf{u}_p = (\theta_p, \delta_p)$, vecteur unitaire pointant en direction de provenance de l'onde. L'expression analytique de l'onde plane sous la forme d'une exponentielle, son développement en harmoniques sphériques, et enfin son identification avec la série de Fourier-Bessel (cf. Chapitre 2, Section 2.3.1) aboutit aux *équations d'encodage* HOA d'une onde plane, c'est-à-dire aux équations qui définissent les signaux HOA :

$$B_{mn}^\sigma = S(\omega) \cdot Y_{mn}^\sigma(\theta_p, \delta_p), \quad (3.7)$$

avec $m \in \mathbb{N}$, $0 \leq n \leq m$, $\sigma = \pm 1$, les fonctions Y_{mn}^σ étant les harmoniques sphériques. L'encodage de cette onde plane à l'ordre M peut s'écrire sous une forme matricielle plus générale :

$$\mathbf{b}_M^{3D} = \mathbf{g}_M^{3D} \cdot S(\omega) \quad (3.8)$$

avec

$$\mathbf{b}_M^{(3D)} = \left(B_{00}^1 \ B_{11}^1 B_{11}^{-1} B_{10}^0 \ \cdots \ \underbrace{B_{mn}^1 B_{mn}^{-1} \cdots B_{mn}^1 B_{mn}^{-1} \cdots B_{m0}^1}_{2m+1} \ \cdots \ \underbrace{B_{MM}^1 B_{MM}^{-1} \cdots B_{M0}^1}_{2m+1} \right)^t \quad (3.9)$$

et

$$\mathbf{g}_M^{3D} = \left(Y_{00}^1 \ Y_{11}^1 Y_{11}^{-1} Y_{10}^0 \ \cdots \ \underbrace{Y_{mn}^1 Y_{mn}^{-1} \cdots Y_{mn}^1 Y_{mn}^{-1} \cdots Y_{m0}^1}_{2m+1} \ \cdots \ \underbrace{Y_{MM}^1 Y_{MM}^{-1} \cdots Y_{M0}^1}_{2M+1} \right)^t. \quad (3.10)$$

Le vecteur $\mathbf{b}_M^{(3D)}$ contient les signaux HOA décrivant une scène sonore 3D tronquée à l'ordre M , le vecteur \mathbf{g}_M^{3D} les gains d'encodage associés. Les signaux HOA s'obtiennent donc en appliquant sur le signal S transporté, des gains réels déterminés en fonction de l'incidence de l'onde par les harmoniques sphériques qui deviennent ainsi les *fonctions d'encodage* HOA d'une onde plane. Ces fonctions d'encodage HOA peuvent évidemment être appliquées sur un signal sonore $s(t)$ exprimé dans le domaine temporel, lié à $S(\omega)$ par une transformée de Fourier inverse. Nous obtenons dans ce cas des signaux HOA également exprimés dans le domaine temporel.

En deux dimensions, l'obtention des équations d'encodage d'une onde plane est réalisée de façon analogue au cas 3D. En effet, une onde plane d'angle d'incidence θ_p dans le plan horizontal peut s'écrire mathématiquement sous la forme d'un développement en harmoniques cylindriques :

$$\begin{aligned}
p(r, \theta_p, \omega) &= S(\omega) e^{ikr \cos \theta_p} \\
&= S(\omega) \left(J_0(kr) + 2 \sum_{m=1}^{\infty} i^m J_m(kr) \cos m\theta \right)
\end{aligned} \tag{3.11}$$

où les fonctions $J_m(kr)$ sont les fonctions de Bessel [B98]. L'identification avec l'expression générale du développement en harmoniques cylindriques de la pression acoustique aboutit aux équations d'encodage 2D d'une onde plane :

$$B_m^\sigma = S \cdot Y_m^\sigma(\theta_p, \delta_p), \quad m \in \mathbb{N} \text{ et } \sigma = \pm 1, \tag{3.12}$$

soit sous une forme matricielle :

$$\mathbf{b}_M^{2D} = \mathbf{g}_M^{2D} \cdot S \tag{3.13}$$

avec

$$\mathbf{b}_M^{(2D)} = \left(B_{00}^1 \ B_{11}^1 B_{11}^{-1} \ \cdots \ B_{mm}^1 B_{mm}^{-1} \ \cdots \ B_{MM}^1 B_{MM}^{-1} \right)^t \tag{3.14}$$

et

$$\mathbf{g}_M^{2D} = \left(Y_0^1 \ Y_1^1 Y_1^{-1} \ \cdots \ Y_m^1 Y_m^{-1} \ \cdots \ Y_M^1 Y_M^{-1} \right)^t. \tag{3.15}$$

La structure d'un encodeur HOA 2D et 3D est représentée à la Figure 3.2 Tableau 3.2. Si l'on souhaite simuler un champ sonore engendré par de multiples ondes planes d'incidences diverses, les signaux HOA sont calculés en effectuant la somme de leurs différentes contributions, conformément au *principe de superposition*.

Dans le cas où les signaux HOA sont destinés à être reproduits en deux dimensions, il est toutefois possible d'ajouter des informations de hauteur des sources virtuelles encodées, de façon à reproduire un effet d'éloignement du plan horizontal à une indétermination haut-bas près [D00]. Des informations de hauteur des sources virtuelles peuvent d'ailleurs être automatiquement présentes dans les signaux HOA 2D lorsque ces derniers résultent de la conversion d'une scène sonore HOA 3D complexe. Proposé à l'origine pour l'encodage de source virtuelle à l'ordre 1 [JLP99, PBJ98], ce procédé a rapidement été étendu aux ordres supérieurs [D00] sans faire toutefois l'objet de tests approfondis. Les nouvelles composantes harmoniques sur lesquelles se basent ces principes constituent à un facteur près un sous-ensemble des harmoniques sphériques :

$$\begin{cases} \tilde{Y}_0^1(\theta) = Y_0^1(\theta) \\ \tilde{Y}_m^\sigma(\theta) = Y_m^\sigma(\theta) \cos^m \delta, \quad \sigma = \pm 1 \end{cases} \tag{3.16}$$

Les fonctions $\tilde{Y}_m^\sigma(\theta)$ peuvent se substituer aux harmoniques cylindriques $Y_m^\sigma(\theta)$ dans l'équation (3.15). Cependant, l'emploi de telles fonctions rend l'énergie contenue dans les signaux HOA dépendante de l'élévation des sources virtuelles, de telle façon qu'elle s'affaiblit lorsque ces sources s'éloignent du plan horizontal. Pour remédier à ce problème audible à la restitution, il est possible de reporter dans la composante omnidirectionnelle l'énergie des composantes verticales non prises en compte [D00, JLP99, PBJ98] :

$$B_0^1 = S(\omega) \sqrt{1 + 2 \sum_{m=1}^M (1 - \cos^2 \delta) \delta^m}. \quad (3.17)$$

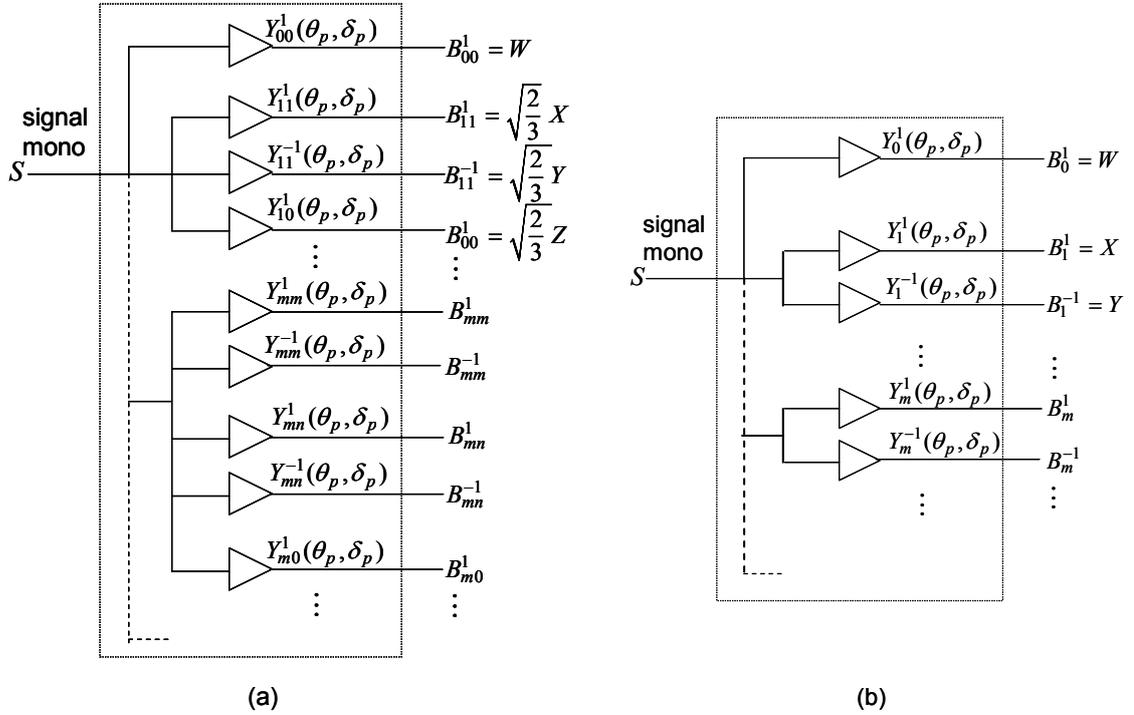


Figure 3.2 – Structure d'un encodeur HOA. (a) : encodeur HOA 3D, (b) : encodeur HOA 2D.

Équation d'encodage d'une onde sphérique - filtres de champ proche

Comme nous l'avons souligné au Chapitre 2, Section 2.3.1, l'écriture d'une source ponctuelle émettant une onde sphérique sous la forme d'un développement en harmonique sphérique permet, par identification avec la série de Fourier-Bessel, d'aboutir à une expression analytique des signaux HOA correspondants :

$$B_{mn}^\sigma = S(\omega) Y_{mn}^\sigma(\theta_s, \delta_s) F_m^{\tau_s}(kr_s), \quad (3.18)$$

avec

$$F_m^{\tau_s}(kr_s) = i^{-(m+1)} \frac{h_m^-(kr_s)}{h_0^-(kr_s)} = \sum_{n=0}^m \frac{(m+n)!}{(m-n)!n!} \left(\frac{-ic}{2\omega r_s} \right)^n, \quad (3.19)$$

où $\tau_s = r_s / c$, c est la vitesse de propagation du son dans l'air, $(r_s, \theta_s, \delta_s)$ sont les coordonnées repérant la position de la source, et les fonctions $h_m^-(kr_s)$ sont les fonctions de Hankel sphériques divergentes d'ordre m . Comparativement aux équations d'encodage d'une onde plane, des filtres $F_m^{\tau_s}(kr_s)$ apparaissent (cf). Ces filtres caractérisent le champ acoustique proche de la source et seront appelés par la suite *filtres de champ proche*. La mise en œuvre de ces filtres pour les ordres

supérieurs à 1 pose problème car leur amplitude diverge en basse fréquence d'autant plus que la source est proche de l'auditeur et que l'ordre est élevé (cf. Figure 2.5, Chapitre 1). Ce phénomène de *bass-boost* est déjà évoqué par Gerzon lorsqu'il considère le filtre $F_1^{r_s}(kr_s)$ du premier ordre pour l'encodage au *B-Format* (HOA d'ordre 1) d'une source virtuelle placée à distance finie de l'auditeur [G92c]. De façon générale, les filtres $F_m^{r_s}(kr_s)$ possèdent une pente d'amplification infinie en basse fréquence de $-m \times 6$ dB par octave.

L'encodage HOA de source virtuelle en champ proche sous-entend d'annuler à la restitution celui des haut-parleurs à l'aide des filtres inverses $1/F_m^{r_{hp}}(kr_{hp})$, r_{hp} désignant la distance des haut-parleurs, pour éviter qu'ils perturbent la reconstruction du champ proche. Cette compensation était déjà recommandée par Gerzon pour la reproduction d'une scène sonore encodée au *B-Format*. Daniel [D03] a récemment proposer de combiner dès l'étape d'encodage spatial les filtres de champ proche associés à la source virtuelle et ceux de compensation du champ proche des haut-parleurs de reproduction. Cette méthode permet de substituer aux filtres $F_m^{r_s}(kr_s)$ dans l'équation (3.18) des filtres d'amplitude finie. Cependant, l'amplitude des filtres résultant reste souvent excessive et par conséquent impossible à mettre en oeuvre, notamment lorsque la source virtuelle est intérieure au dispositif de haut-parleurs. Nous donnerons plus de détails sur l'encodage de source virtuelle à distance finie de l'auditeur au Chapitre 4, dans lequel nous proposerons une nouvelle méthode de conception des filtres de champ proche.

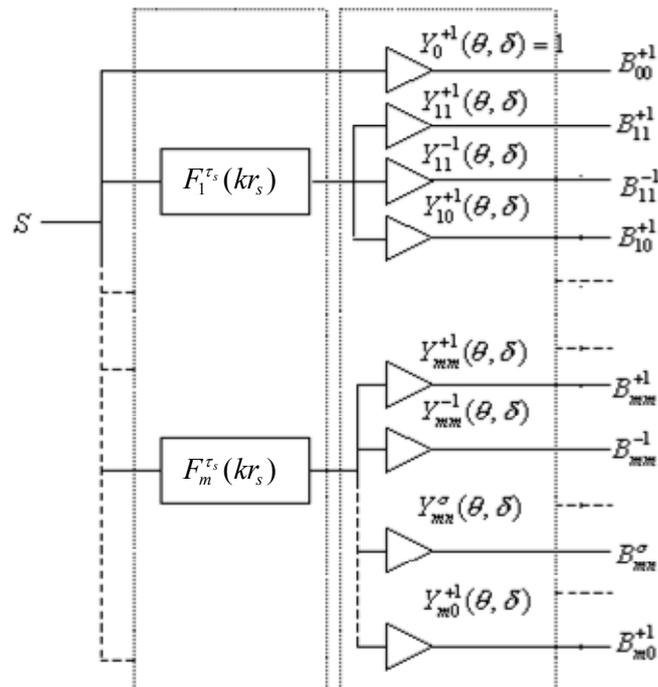


Figure 2.3 – Schéma d'encodage HOA d'une onde sphérique émise par une source ponctuelle homogène de rayon r_s

Enregistrement de champ sonore naturel à l'ordre 1

Pour enregistrer une scène sonore réelle au format HOA, il faut être capable d'estimer chacune des composantes du développement en harmoniques sphériques du champ acoustique en un point de l'espace. Craven et Gerzon [CG77] ont breveté dès 1975 un procédé d'enregistrement du premier ordre constitué par une combinaison tétraédrique de microphones à directivité hypercardioïde quasi coïncident. Ce procédé permet d'estimer au centre du dispositif tétraédrique, la pression (W) et les trois composantes du vecteur de vitesse particulaire (X , Y , Z). Le microphone *SoundField* [F79a, F79b] de la société Calrec est une application commerciale directe du brevet de Craven et Gerzon. Comme le montre la Figure 3.3, les quatre microphones hypercardioïdes sont disposés à la surface d'un tétraèdre régulier et captent des signaux LF (*Left Front* ou avant gauche), RF (*Right Front* ou avant droit), LB (*Left Back* ou arrière gauche) et RB (*Right Back* ou arrière droit). Les signaux HOA s'obtiennent par combinaison linéaire de ces signaux, c'est-à-dire par un simple matricage (cf. Figure 3.4) :

$$\begin{cases} W = LF + LB + RF + RB \\ X = LF - LB + RF - RB \\ Y = LF + LB - RF - RB \\ Z = LF - LB - RF - RB \end{cases} \quad (3.20)$$

Craven et Gerzon [CG77] introduisent d'autre part des filtres égaliseurs qui doivent être appliqués aux signaux (W , X , Y , Z), notamment pour compenser la non-coïncidence des capsules microphoniques inévitable en pratique. Deux filtres différents interviennent suivant l'ordre d'appartenance des signaux, F_W étant appliqué à W et F_X à X , Y et Z :

$$\begin{cases} F_W = \frac{1 + i\omega\tau - \frac{1}{3}\omega^2\tau^2}{1 + \frac{1}{3}i\omega\tau} \\ F_X = \sqrt{6} \left(\frac{1 + \frac{1}{3}i\omega\tau - \frac{1}{3}\omega^2\tau^2}{1 + \frac{1}{3}i\omega\tau} \right) \end{cases} \quad (3.21)$$

où $\tau = R/c$, R étant la distance réelle entre le centre du dispositif tétraédrique et les capsules microphoniques et c la vitesse de propagation du son dans l'air. Notons que le facteur $\sqrt{6}$ permet de rendre les relations (3.20) conformes avec les conventions de normalisation de Gerzon (cf. Tableau 3.3).

Jusqu'à très récemment, le microphone *Soundfield* basé sur le brevet de Craven et Gerzon était le seul outil existant permettant d'enregistrer une scène sonore au format HOA limité à l'ordre 1. Ces dernières années, des solutions nouvelles ont été proposées pour étendre la captation aux ordres supérieurs. Nous présentons nos travaux dans ce domaine au Chapitre 5. Ces travaux ont notamment abouti à la réalisation d'un microphone 3D HOA d'ordre 4.

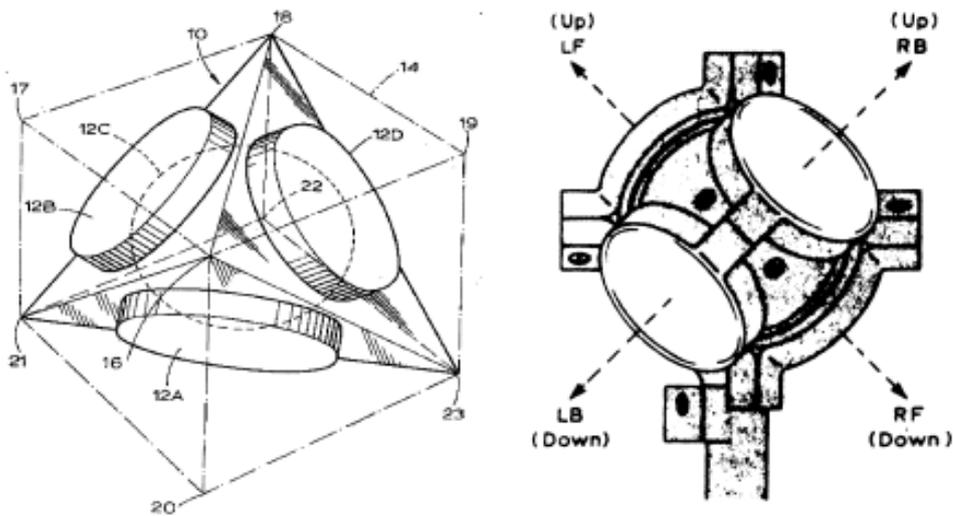


Figure 3.3 – Géométrie du Microphone Soundfield : quatre capsules microphoniques hypercardioides montées sur les faces d'un tétraèdre régulier (à gauche d'après Craven et Gerzon [CG77], à droite d'après Farrar [F79a, F79b]).

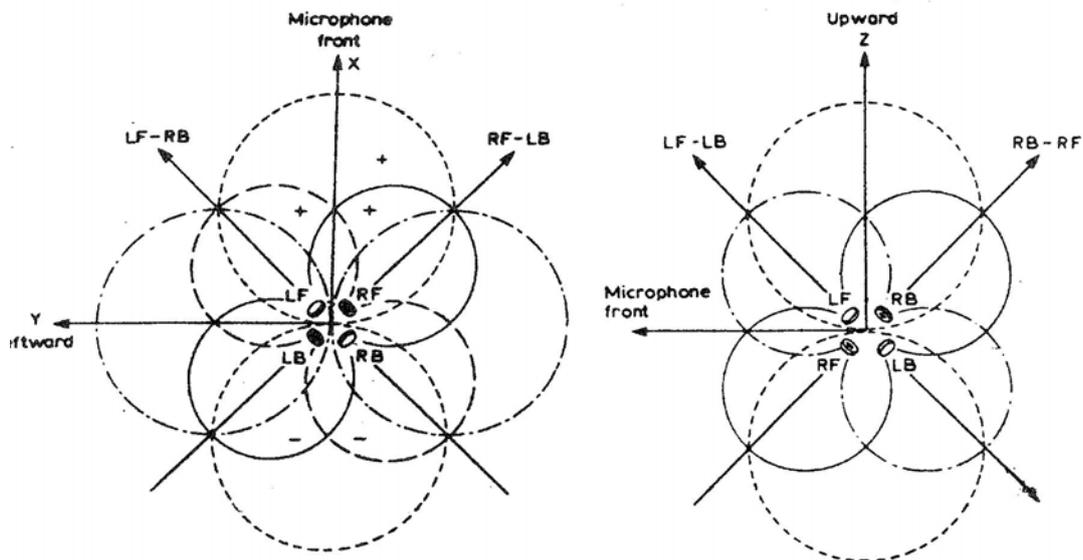


Figure 3.4 – Obtention des signaux (W , X , Y , Z) à partir des signaux microphoniques (LF , LB , RF , RB) captés par le microphone SoundField [F79a] par projection sur le plan horizontal (gauche) et vertical (droit).

3.2.2 Manipulations de scène sonore HOA

Une scène sonore encodée au format HOA peut subir un certain nombre de manipulations comme des *rotations* ou des *distorsions de perspective* dont un cas particulier est la *focalisation*. Ces manipulations sont réalisées à partir d'opérations linéaires simples effectuées sur les signaux HOA, que nous pouvons écrire sous une forme matricielle :

$$\mathbf{b}' = \mathbf{R} \cdot \mathbf{b}, \quad (3.22)$$

où les vecteurs colonne \mathbf{b} et \mathbf{b}' contiennent respectivement les signaux HOA d'origine et les signaux HOA résultant de leur transformation par la matrice carrée \mathbf{R} , cette dernière matrice étant définie en fonction du type de manipulation choisie.

Rotation

De part ses propriétés, une fonction harmonique sphérique d'ordre m se transforme par rotation en une fonction harmonique sphérique de même ordre. La matrice \mathbf{R} qui décrit la rotation possède par conséquent une structure diagonale par blocs, chaque bloc \mathbf{R}_m associé à l'ordre m étant de dimensions $(2m+1) \times (2m+1)$:

$$\mathbf{R} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & \times & \times & \times & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & \times & \mathbf{R}_1 & \times & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & \times & \times & \times & 0 & 0 & 0 & 0 & 0 & \dots \\ \hline 0 & 0 & 0 & 0 & \times & \times & \times & \times & \times & \dots \\ 0 & 0 & 0 & 0 & \times & \times & \times & \times & \times & \dots \\ 0 & 0 & 0 & 0 & \times & \times & \mathbf{R}_2 & \times & \times & \dots \\ 0 & 0 & 0 & 0 & \times & \times & \times & \times & \times & \dots \\ 0 & 0 & 0 & 0 & \times & \times & \times & \times & \times & \dots \\ \hline \vdots & \ddots \end{pmatrix}$$

Une matrice \mathbf{R}_m s'appliquant à un groupe de signaux HOA d'ordre m quelconque peut s'écrire sous la forme d'un produit de trois rotations appliquées successivement autour des axes (Ox) , (Oy) et (Oz) du repère cartésien et d'angles respectifs γ , ϕ , et θ :

$$\mathbf{R}_m = \mathbf{R}_m^x(\gamma) \cdot \mathbf{R}_m^y(\phi) \cdot \mathbf{R}_m^z(\theta). \quad (3.23)$$

En particulier, la matrice $\mathbf{R}_m^z(\theta)$ qui décrit les rotations dans le plan horizontal, c'est-à-dire autour de l'axe (Oz) , possède une expression analytique simple et générique. Elle transforme en effet pour une rotation d'angle θ chaque couple de signaux HOA (B_{mn}^1, B_{mn}^{-1}) en un couple modifié $(B_{mn}^{\prime 1}, B_{mn}^{\prime -1})$ par la relation suivante :

$$\begin{pmatrix} B_{mn}^{\prime 1} \\ B_{mn}^{\prime -1} \end{pmatrix} = \begin{pmatrix} \cos(n\theta) & -\sin(n\theta) \\ \sin(n\theta) & \cos(n\theta) \end{pmatrix} \begin{pmatrix} B_{mn}^1 \\ B_{mn}^{-1} \end{pmatrix} \quad (3.24)$$

Par exemple pour l'ordre $m = 2$, la matrice $\mathbf{R}_2^z(\theta)$ caractérisant une rotation d'angle θ dans le plan azimutal est définie par :

$$\mathbf{R}_2^z(\theta) = \begin{pmatrix} \cos 2\theta & -\sin 2\theta & 0 & 0 & 0 \\ \sin 2\theta & \cos 2\theta & 0 & 0 & 0 \\ 0 & 0 & \cos \theta & -\sin \theta & 0 \\ 0 & 0 & \sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (3.25)$$

La rotation dans le plan horizontal peut être avantageusement utilisée en association avec un *head-tracker* pour compenser les mouvements de tête en écoute binaurale avec casque d'écoute. Dans ce cas, notons cependant que c'est la matrice de rotation *inverse* $(\mathbf{R}_m^z(\theta))^{-1} = \mathbf{R}_m^z(-\theta)$ qui doit être appliquée sur les signaux HOA.

Distorsion de la perspective

Un autre type de manipulation applicable sur les signaux HOA consiste en une *distorsion de la perspective*, c'est-à-dire un resserrement ou un élargissement de la scène frontale. Cette manipulation proposée initialement par Gerzon [GB92] sous le nom de *forward-backward dominance* pour être appliquée sur les composantes HOA d'ordre 0 et 1 (*B-Format*), est réalisée au moyen d'une forme particulière de transformation de Lorentz décrite par la matrice :

$$\mathbf{T}_\lambda = \begin{pmatrix} \frac{\lambda + \lambda^{-1}}{2} & \frac{\lambda - \lambda^{-1}}{2\sqrt{3}} & 0 & 0 \\ \frac{\sqrt{3}(\lambda - \lambda^{-1})}{2} & \frac{\lambda + \lambda^{-1}}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (3.26)$$

le paramètre λ de contrôle étant réel et positif. Remarquons que cette transformation laisse inchangées les composantes B_{11}^{-1} et B_{11}^0 . D'autre part, la matrice \mathbf{T}_λ a été adaptée pour être en conformité avec nos conventions de normalisation des signaux HOA. Une matrice de transformation équivalente à celle de l'équation (3.26) est utilisée par Malham [M90] et obtenue en effectuant le changement de paramètre $\mu = (\lambda - \lambda^{-1}) / (\lambda + \lambda^{-1})$, μ étant compris entre -1 et 1, et en multipliant toutes les composantes par $2 / (\lambda + \lambda^{-1}) = \sqrt{1 - \mu^2}$. Nous obtenons finalement la matrice de transformation :

$$\mathbf{T}_\mu = \begin{pmatrix} 1 & \mu/\sqrt{3} & 0 & 0 \\ \sqrt{3}\mu & 1 & 0 & 0 \\ 0 & 0 & \sqrt{1-\mu^2} & 0 \\ 0 & 0 & 0 & \sqrt{1-\mu^2} \end{pmatrix}. \quad (3.27)$$

La distorsion de perspective change la position originale $(\theta, \delta = 0)$ d'une source sonore vers une nouvelle position $(\theta', \delta = 0)$ donnée par l'équation :

$$\cos \theta' = \frac{\mu + \cos \theta}{1 + \mu \cos \theta}. \quad (3.28)$$

Par conséquent, si $\lambda > 1$ ou $\mu > 0$ toutes les directions se déplacent vers l'avant, et si $\lambda < 1$ ou $\mu < 0$, toutes les directions se déplacent vers l'arrière. Le cas particulier $\lambda = 1$ ou $\mu = 0$ laisse inchangés les signaux HOA, ce qui se traduit par une matrice \mathbf{T}_λ ou \mathbf{T}_μ unitaire. D'autre part, l'amplitude de la source sonore est amplifiée par un facteur $1 + \mu \cos \theta$ pour la méthode de Malham (3.27) et un facteur $(\lambda + \lambda^{-1})/2 + (\lambda - \lambda^{-1}) \cos \theta/2$ pour la méthode de Gerzon (3.26). Enfin, cette transformation est réversible et son inverse s'écrit :

$$(\mathbf{T}_\mu)^{-1} = \frac{1}{\sqrt{1-\mu^2}} \mathbf{T}_{-\mu}. \quad (3.29)$$

L'extension de l'effet de distorsion de perspective aux ordres supérieurs à 1 ne peut pas se baser sur l'équation (3.28) décrivant la distorsion angulaire sous peine de produire des signaux HOA dont les caractéristiques d'onde plane associées ne sont pas préservées.

La *focalisation*, parfois appelée *focus*, est un cas particulier de distorsion de perspective obtenu en fixant $\mu = 1$. Les signaux HOA sont combinés de telle façon à obtenir l'effet d'une prise de son directive effectuée à partir d'un microphone cardioïde. Contrairement au cas général de distorsion de perspective, le principe de la focalisation peut être étendu aux ordres supérieurs à 1 de façon à imiter une directivité microphonique de plus en plus sélective [D01].

3.3 Décodage spatial : approche basique

Bien que nous ayons insisté sur l'indépendance du format HOA par rapport au système de diffusion utilisé pour la restitution, nous ne pouvons évidemment pas espérer reconstruire correctement le champ acoustique en plaçant n'importe où les haut-parleurs. Un certain nombre de règles doivent être respectées pour exploiter au mieux les données objectives du champ sonore contenues en puissance dans les signaux HOA. Nous présentons ici les solutions de décodage spatial dites *basiques*, car basées sur aucun a priori ni aucune procédure d'optimisation quelconque. Nous supposons que les ondes sonores émises par les haut-parleurs peuvent être considérées comme planes dans la zone d'écoute. Nous étudions également l'adaptation du décodage HOA au casque d'écoute par simulation binaurale des haut-parleurs.

3.3.1 Décodage basique pour un dispositif régulier de haut-parleurs

Problème du décodage spatial et solutions générales basiques

Le problème du décodage basique s'attache à recréer à partir d'un ensemble de haut-parleurs les composantes spatiales du développement en harmoniques sphériques du champ sonore dans une zone de l'espace où se situe l'auditoire : c'est le *principe de réencodage* [D00, JLP99, WA01b]. En trois dimensions, ce principe implique la relation :

$$\mathbf{C} \cdot \mathbf{s} = \mathbf{b} \quad (3.30)$$

avec

$$\mathbf{C} = \begin{bmatrix} Y_{00}^1(\theta_1, \delta_1) & Y_{00}^1(\theta_2, \delta_2) & \cdots & Y_{00}^1(\theta_L, \delta_L) \\ \vdots & \vdots & \vdots & \vdots \\ Y_{11}^1(\theta_1, \delta_1) & Y_{11}^1(\theta_2, \delta_2) & \cdots & Y_{11}^1(\theta_L, \delta_L) \\ \vdots & \vdots & \vdots & \vdots \\ Y_{M0}^1(\theta_1, \delta_1) & Y_{M0}^1(\theta_2, \delta_2) & \cdots & Y_{M0}^1(\theta_L, \delta_L) \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_L \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} B_{00}^1 \\ \vdots \\ B_{mn}^\sigma \\ \vdots \\ B_{M0}^1 \end{bmatrix},$$

où L est le nombre de haut-parleurs utilisés, la matrice \mathbf{C} contient les vecteurs harmoniques sphériques associés à chaque direction de haut-parleur et organisés suivant ses lignes, le vecteur \mathbf{S} contient les signaux émis par les L haut-parleurs, et enfin le vecteur \mathbf{b} contient les signaux HOA. En deux dimensions, les harmoniques cylindriques se substituent aux harmoniques sphériques dans la matrice \mathbf{C} . Le problème du décodage basique se résume alors à un système de $(M+1)^2$ en 3D ou $2M+1$ en 2D équations linéaires à L inconnus qu'il suffit de résoudre.

En général, une solution exacte au système (3.30) peut être trouvée si le nombre de haut-parleurs est supérieur ou égal à celui des signaux HOA, c'est-à-dire si $L \geq K_M^{3D} = (M+1)^2$ en 3D ou $L \geq K_M^{2D} = 2M+1$ en 2D. L'obtention d'une solution exacte dépend des propriétés de la matrice \mathbf{C} , c'est-à-dire de la position des haut-parleurs. Puisque la matrice \mathbf{C} n'est pas obligatoirement carrée et inversible, une solution générale au problème (3.30) consiste à calculer l'inverse généralisée de la matrice \mathbf{C} , notée \mathbf{D} , encore appelée *pseudo-inverse* de Moore-Penrose [GVL96] :

$$\mathbf{s} = \mathbf{D} \mathbf{b} \quad (3.31)$$

avec

$$\mathbf{D} = \mathbf{C}^t \cdot (\mathbf{C} \cdot \mathbf{C}^t)^{-1}. \quad (3.32)$$

Nous appellerons désormais la matrice \mathbf{D} *matrice de décodage* puisque c'est par elle que sont *décodés* les signaux HOA.

Cas idéal : dispositifs circulaires (2D) ou sphériques (3D) de haut-parleurs

En réalité, la solution (3.31) résout la plupart du temps de façon approximative le système (3.30) et sa précision dépend de la matrice \mathbf{C} , c'est-à-dire du système de restitution utilisé.

Elle peut être rendue exacte si les haut-parleurs sont assez nombreux et disposés *régulièrement* sur un périmètre circulaire (2D) ou à la surface d'une sphère (3D) [CS71, P96, WA01b]. Plus précisément, la régularité du dispositif de restitution signifie que l'ensemble des directions θ_l en 2D et (θ_l, δ_l) en 3D avec $l=1,2,\dots,L$ des haut-parleurs préservent la propriété d'orthonormalité des harmoniques cylindriques et des harmoniques sphériques respectivement, ce qui se traduit mathématiquement par la condition :

$$\mathbf{C} \cdot \mathbf{C}^t = \frac{1}{L} \cdot \mathbf{I}_L \quad (3.33)$$

où \mathbf{I}_L est la matrice identité de dimensions $L \times L$. La matrice de décodage s'exprime alors simplement :

$$\mathbf{D} = \frac{1}{L} \mathbf{C}^t. \quad (3.34)$$

Les équations de décodage du système (3.31) correspondent ainsi à la projection des signaux HOA sur les vecteurs harmoniques associés à la direction de chaque haut-parleur. Ces équations sont dites *basiques* pour les distinguer des solutions *optimisées* que nous présenterons par la suite.

Pour que l'équation (3.31) soit valable en 2D, il suffit simplement de disposer de façon régulière sur cercle un nombre de haut-parleurs au minimum égal à celui des signaux HOA $L \geq K_M^{2D} = 2M + 1$ [CS71, P96]. Nous illustrons à la Figure 3.4 la reconstruction dans le plan horizontal au moyen d'un tel dispositif circulaire de haut-parleurs d'une onde plane tronquée à différents ordres. Pour l'étude plus complexe des dispositifs 3D de haut-parleurs préservant la propriété (3.33) d'orthonormalité des harmoniques sphériques nous nous référons à celle menée en détails au Chapitre 5 dans le contexte de la captation sonore, réciproque de celui de la diffusion. Signalons simplement qu'au-delà de l'ordre 2, l'équation (3.33) ne peut pas être préservée exactement, à moins d'utiliser un nombre déraisonnable de haut-parleurs. Dans ce cas, la matrice pseudo-inverse de décodage (3.32) semble par conséquent mieux adaptée.

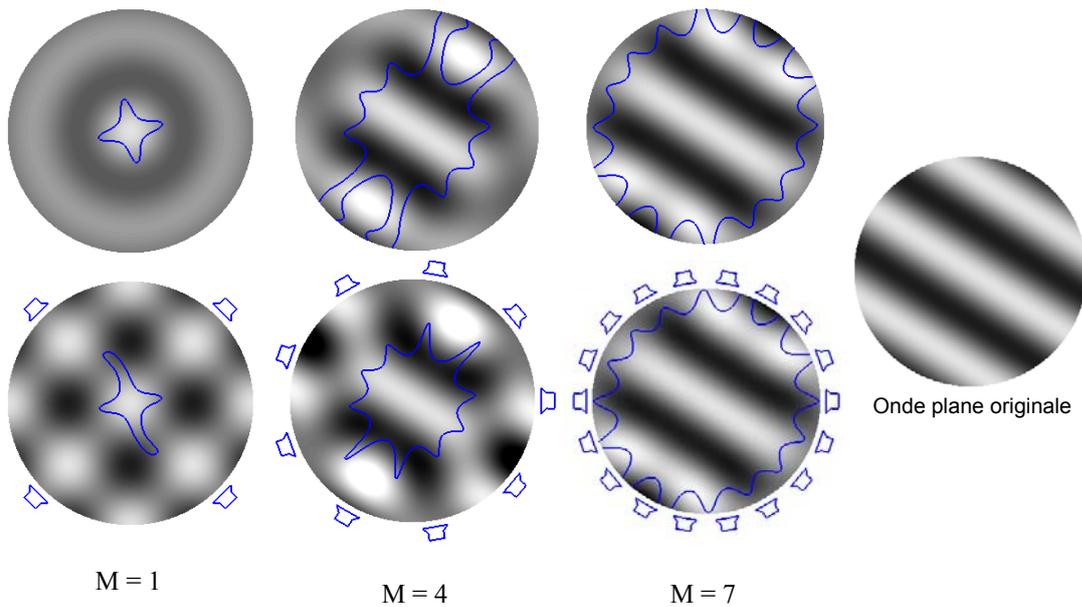


Tableau 3.4 – Reconstruction d’une onde plane (1000 Hz) tronquée aux ordres $M=1$, 4 et 7 dans le plan horizontal (1 m de diamètre) avec un dispositif circulaire de haut-parleurs⁹ (bas). La représentation tronquée théorique¹⁰ associée (haut) sert de référence. Les courbes étoilées délimitent la zone de l’espace dans laquelle le champ acoustique est approximé/reproduit avec une erreur inférieure à 7 % (même définition mathématique de l’erreur qu’au Chapitre 2).

3.3.2 Décodage binaural pour une restitution au casque d’écoute

Pour adapter la restitution d’une scène sonore HOA au casque d’écoute qui possède la particularité de diffuser un signal sonore à chaque oreille de l’auditeur de façon indépendante, il est nécessaire d’ajouter après l’étape de décodage spatial que nous venons d’étudier une nouvelle étape de *synthèse binaurale* des haut-parleurs [Mo92]. Nous devons pour cela connaître les fonctions de transfert entre chacun de ces haut-parleurs et chaque tympan.

⁹ Notons que sur les illustrations, les haut-parleurs dessinés indiquent leurs positions angulaires mais pas leur distance par rapport à la zone d’écoute.

¹⁰ Signalons qu’ici c’est la représentation tronquée du développement en *harmoniques cylindriques* qui est indiquée contrairement aux illustrations du Chapitre 2 dans lesquelles nous prenions en compte le développement en *harmoniques sphériques*. Dans le cas particuliers d’une onde plane émise dans le plan horizontal, ces deux développements donnent toutefois des résultats comparables.

Head-Relative Transfer Functions (HRTF)

Pour déterminer la position d'une source sonore dans l'espace, le système auditif utilise un certain nombre d'*indices*, présents dans les signaux captés par les tympan, et qui résultent de l'interaction du son avec le corps de l'auditeur, notamment le torse, la tête et les pavillons. Les principaux indices sont les différences de temps, d'intensité entre les signaux captés par les deux tympan, et des indices spectraux [B97, C90]. Pour une incidence particulière de source sonore, ces indices sont indépendants du signal émis par la source et invariants dans le temps. Ils sont par conséquent entièrement déterminés par les réponses impulsionnelles (dans le domaine temporel) ou, ce qui revient au même, les fonctions de transferts (dans le domaine fréquentiel) caractérisant la propagation du son entre la source émettrice et chaque tympan. La plupart du temps, on parle plutôt des fonctions de transfert appelées HRTF (*Head-Related Transfer Functions*, ou en français, *fonctions de transfert relatives à la tête*). Autrement dit, tous les indices permettant au système de reconnaître la position angulaire (azimut et élévation) d'une source sonore sont contenus dans la paire d'HRTF correspondant. Quelques bases de HRTF mesurées pour un grand nombre d'incidences de source et sur de nombreux sujets sont disponibles sur Internet [A01, GM95, W02]. La Figure 3.5 illustre par exemple une paire d' HRTF mesurée par le CIPIC (*Center for Image Processing and Integrated Computing*) [A01] sur un mannequin KEMAR correspondant à une source placée à 75° dans le plan horizontal.

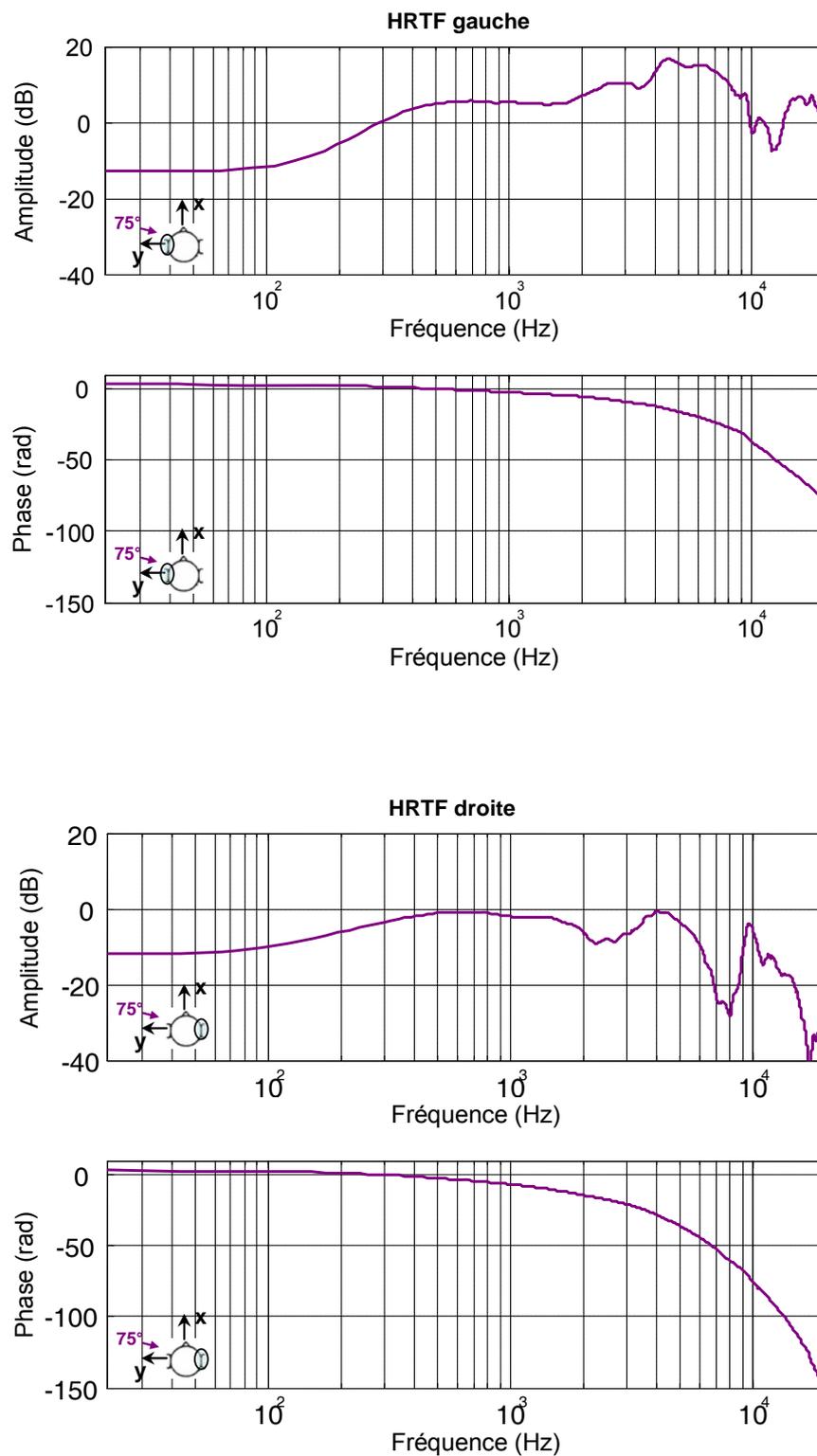


Figure 3.5 – Illustration du spectre d'amplitude et de la phase d'une paire d'HRTF mesurée sur le mannequin KEMAR. (D'après la base HRTF du CIPIC [A01])

Filtrage binaural des signaux HOA

La simulation binaurale des haut-parleurs consiste à filtrer les signaux \mathbf{s} issus des équations de décodage (3.31) par les HRTF correspondant à la position des haut-parleurs et à sommer indépendamment les contributions associées aux oreilles gauche et droite [Mo92]. Les signaux binauraux résultants sont ensuite reproduits au casque d'écoute.

Les opérations de matricage et de filtrage binaural peuvent être regroupées sous la forme d'un ensemble de filtres s'appliquant directement sur les signaux HOA. Nous pouvons en effet exprimer les signaux binauraux S_G et S_D respectivement gauche et droit de la façon suivante :

$$\begin{cases} S_G = \sum_{l=1}^L H_G(\theta_l, \delta_l) \cdot S_l = \mathbf{h}_G^t \cdot \mathbf{s} \\ S_D = \sum_{l=1}^L H_D(\theta_l, \delta_l) \cdot S_l = \mathbf{h}_D^t \cdot \mathbf{s} \end{cases} \quad (3.35)$$

où $H_G(\theta_l, \delta_l)$ et $H_D(\theta_l, \delta_l)$ désignent la paire d'HRTF caractérisant la propagation du son entre le haut-parleur l et les oreilles gauche et droite respectivement de l'auditeur, \mathbf{s} est le vecteur défini à l'équation (3.30) contenant les signaux émis par les haut-parleurs, \mathbf{h}_G^t et \mathbf{h}_D^t les vecteurs constitués par les HRTF gauche et droit associés aux L haut-parleurs :

$$\begin{cases} \mathbf{h}_G = (H_G(\theta_1, \delta_1) \ H_G(\theta_2, \delta_2) \ \cdots \ H_G(\theta_L, \delta_L))^t \\ \mathbf{h}_D = (H_D(\theta_1, \delta_1) \ H_D(\theta_2, \delta_2) \ \cdots \ H_D(\theta_L, \delta_L))^t \end{cases} \quad (3.36)$$

En substituant au vecteur \mathbf{s} dans l'équation (3.35) son expression (3.31), nous obtenons finalement :

$$\begin{cases} S_G = \mathbf{g}^t \cdot \mathbf{b} \\ S_D = \mathbf{d}^t \cdot \mathbf{b} \end{cases} \quad (3.37)$$

avec

$$\begin{cases} \mathbf{g}^t = \mathbf{h}_G^t \cdot \mathbf{D} \\ \mathbf{d}^t = \mathbf{h}_D^t \cdot \mathbf{D} \end{cases} \quad (3.38)$$

Le vecteur \mathbf{b} contient les signaux HOA conformément à l'équation (3.30), les vecteurs \mathbf{g} et \mathbf{d} sont constitués par les fonctions de transfert des signaux HOA directement vers les signaux binauraux S_G et S_L . Une mise en œuvre avantageuse du filtrage binaural peut être obtenue grâce à l'hypothèse de symétrie des HRTF et de celle du dispositif de restitution [D00].

La Figure 3.6 illustre la reconstruction des HRTF représentées à la Figure 3.5 à partir des signaux HOA et en fonction de l'ordre de troncature. Nous observons très clairement que la reconstruction s'améliore en haute fréquence au fur et à mesure que l'ordre augmente. Autrement dit, les courbes d'amplitude et de phase résultant de la troncature du champ acoustique « décrochent » des courbes associées aux HRTF originales de plus en haut dans le spectre. Cela est

directement lié aux propriétés largement discutées au Chapitre 2 d'extension de la zone de correcte approximation du champ acoustique en fonction de l'ordre de troncature et de la fréquence.

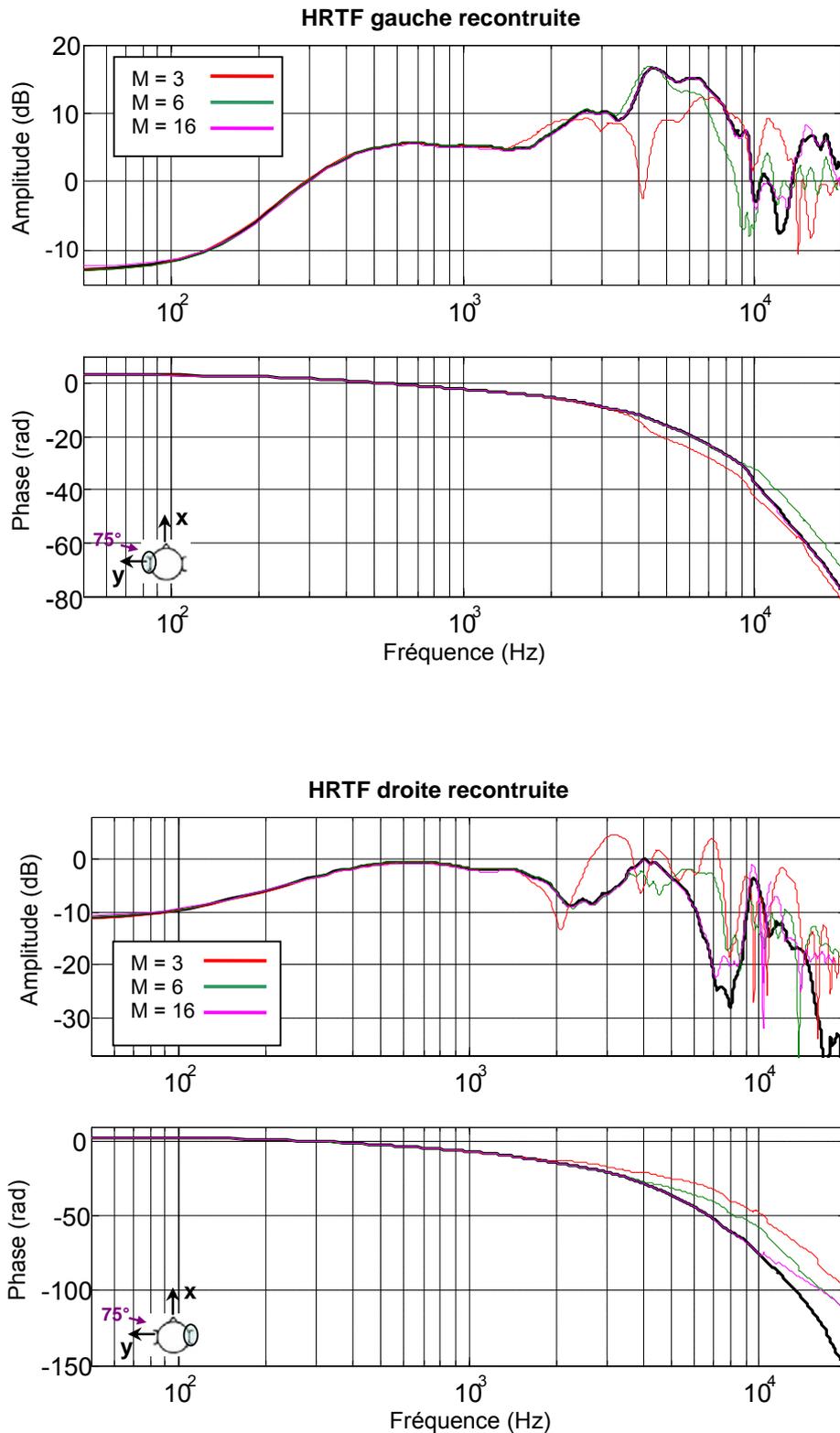


Figure 3.6 – Reconstruction d'une paire de HRTF en fonction de l'ordre des signaux HOA. Cas particuliers pour les ordres 3, 6, et 16.

3.4 Optimisation du décodage spatial suivant le contexte d'écoute

La recherche de décodeurs HOA optimisés est utile notamment lorsque la zone où la reconstruction acoustique satisfaisante est trop petite par rapport à la zone d'écoute réelle. Cette situation arrive par exemple lorsqu'un auditeur s'éloigne de la position d'écoute idéale ou que l'ordre de l'approximation et donc des signaux HOA, est trop faible. Il est parfois possible d'améliorer l'effet *subjectif* de localisation des sources sonores en modifiant le décodage de telle façon que des critères *objectifs* associés soient optimisés. Deux critères importants introduits par Gerzon pour optimiser le décodage du *B-Format*, c'est-à-dire de signaux HOA d'ordre 1, sont le *vecteur vitesse* et le *vecteur énergie*. Leur utilisation peut être avantageusement étendue aux systèmes d'ordres supérieurs.

Un deuxième aspect de l'optimisation du décodage HOA que nous n'abordons pas ici concerne la recherche de formules de décodage optimales associées à des dispositifs irréguliers de haut-parleurs, imposés notamment dans certaines situations d'écoute : dispositif 5.1 pour le *home cinema*, stéréophonie à deux canaux encore majoritaire, disposition contraignante des haut-parleurs dans une automobile, etc. Un tel problème d'optimisation implique la résolution d'un système d'équations non linéaires défini en fonction d'un certain nombre de critères, basés notamment sur les vecteurs vitesse et énergie. Quelques études ont été menées sur le sujet, notamment pour la stéréophonie standard à deux canaux [G85, G77] et le dispositif multicanal 5.1 [GB92, G92b, W03, C03].

3.4.1 Prédiction objective de la localisation auditive d'après caractérisation locale de la propagation acoustique par les vecteurs vitesse \mathbf{V} et énergie \mathbf{E}

Caractérisation locale de la propagation acoustique par le vecteur vitesse \mathbf{V}

Une première quantité physique sur laquelle se base l'optimisation du décodage spatial HOA est le vecteur vitesse \mathbf{V} qui dépend du rapport de la vitesse particulaire sur la pression acoustique :

$$\mathbf{V} = -\frac{1}{\rho c} \frac{\mathbf{v}(r, \theta, \delta)}{p(r, \theta, \delta)} = -\frac{i}{k} \frac{\vec{\nabla} p(r, \theta, \delta)}{p(r, \theta, \delta)}. \quad (3.39)$$

Nous pouvons séparer les parties réelles et imaginaires :

$$\begin{cases} \vec{\Omega}(r, \theta, \delta) = \Re(\mathbf{V}(r, \theta, \delta)) = r_\nu \mathbf{u}_\nu \\ \vec{\Phi}(r, \theta, \delta) = \Im(\mathbf{V}(r, \theta, \delta)) = r_\phi \mathbf{u}_\phi \end{cases} \quad (3.40)$$

où $|\mathbf{u}_\nu| = |\mathbf{u}_\phi| = 1$. La partie réelle $\vec{\Omega}(\mathbf{r})$ reflète les phénomènes de propagation, tandis que la partie imaginaire $\vec{\Phi}(\mathbf{r})$ décrit les phénomènes stationnaires du champ acoustique [DRP99].

Dans le cas d'un champ de pression acoustique engendré par une unique onde plane d'incidence $\mathbf{u}_p = (\theta_p, \delta_p)$, le vecteur vitesse se calcule aisément :

$$\mathbf{V} = \mathbf{u}_p. \quad (3.41)$$

La vecteur \mathbf{V} est ainsi réel, unitaire ($r_V = 1$) et exactement orienté suivant la direction de provenance de l'onde plane ($\mathbf{u}_V = \mathbf{u}$). L'équation (3.39) s'applique également à un champ sonore engendré par une superposition d'ondes planes, émises par exemple par des haut-parleurs suffisamment éloignés et placés à équidistance du point d'observation, et transportant un signal sonore identique à une amplitude et une opposition de phase près. Au centre d'un tel dispositif de haut-parleurs, le vecteur vitesse s'exprime ainsi :

$$\mathbf{V} = \frac{\sum_{l=1}^L g_l \mathbf{u}_l}{\sum_{l=1}^L g_l} = r_V \mathbf{u}_V, \quad (3.42)$$

où r_V et \mathbf{u}_V correspondent respectivement à la norme et à la direction du vecteur vitesse \mathbf{V} , le vecteur unitaire \mathbf{u}_l désigne la direction du $l^{\text{ème}}$ haut-parleur auquel est associé le gain g_l . Ces gains sont réels si les ondes planes sont en phase ou en opposition de phase. Dans ce cas, le vecteur \mathbf{V} est également réel et le vecteur unitaire \mathbf{u}_l indique une direction généralement différente de chaque vecteur \mathbf{u}_l puisque plusieurs haut-parleurs émettent à la fois. Cela entraîne d'autre part que la norme r_V du vecteur vitesse est différente de la valeur 1. Il résulte finalement d'une superposition d'onde plane émise en phase ou en opposition de phase à une distance égale du point d'observation, un vecteur vitesse \mathbf{V} qui pointe dans la direction apparente de provenance d'une onde plane virtuelle dont la vitesse apparente de propagation c_{loc} est donnée par l'équation [DRP98] :

$$c_{loc} = \frac{c}{r_V}. \quad (3.43)$$

La définition du vecteur vitesse donnée à l'équation (3.42) rejoint celle donnée par Gerzon dans le cadre de la recherche de décodeur optimaux associés au *B-Format* [G92d].

Caractérisation locale de la propagation acoustique par le vecteur énergie \mathbf{E}

Pour compléter la caractérisation locale de la propagation acoustique, Gerzon [G92d] a introduit une seconde quantité vectorielle, le *vecteur énergie* \mathbf{E} , défini par la relation :

$$\mathbf{E} = \frac{\sum_{l=1}^L |g_l|^2 \mathbf{u}_l}{\sum_{l=1}^L |g_l|^2} = r_E \mathbf{u}_E, \quad (3.44)$$

où les notations de l'équation (3.42) sont conservés, r_E et \mathbf{u}_E correspondent respectivement à la norme et à la direction du vecteur \mathbf{E} . Bien que ne correspondant pas à une grandeur physique particulière, le vecteur énergie est homogène à une intensité acoustique et peut s'interpréter comme une indication de la *direction moyenne de provenance de l'énergie* donnée par \mathbf{u}_E , ainsi

que du *taux de concentration de l'énergie* dans cette direction donné par la norme r_E [DRP98, DNM03]. Cette dernière norme r_E , positive par définition, reste toujours strictement inférieure à 1 sauf dans le cas particulier d'une onde plane émise par un haut-parleur unique, r_E valant exactement 1 et le vecteur \mathbf{u}_E étant orienté suivant la direction de provenance du haut-parleur émetteur.

Liens avec la localisation auditive

Au centre d'un dispositif concentrique de haut-parleurs, l'étude des vecteurs vitesse et énergie définis précédemment permet de prédire l'effet subjectif de localisation auditive d'une source virtuelle produite par la superposition des ondes émises, supposées planes et synchrones. Cette assertion se fonde sur un ensemble d'hypothèses liant la caractérisation objective du champ acoustique par les vecteurs \mathbf{V} et \mathbf{E} aux mécanismes de la perception auditive de l'espace.

La première hypothèse formulée par Makita [M62] établit que la direction perçue par un auditeur d'un son de fréquence inférieure à 700 Hz, c'est-à-dire lorsque le corps ne fait pas obstacle à la propagation acoustique, se confond avec la normale au front d'onde évaluée à sa position. Par conséquent, la direction du vecteur vitesse calculée à la position de l'auditeur s'avère être un bon indicateur de la direction apparente de la source virtuelle. La norme r_V peut être quant à elle associée à la stabilité ou la robustesse de l'effet subjectif de localisation, ou plus largement au confort d'écoute [G92d]. Comme nous l'avons indiqué précédemment, cette norme r_V est en fait liée à la vitesse apparente de propagation de l'onde virtuelle par la relation $c_{loc} = c/r_V$, où c désigne la vitesse de propagation du son dans l'air. L'altération de la vitesse de propagation de l'onde induit une variation des indices de localisation utilisés par le système auditif, qui explique notamment l'instabilité de la perception. Plus précisément, la source virtuelle tend à se rapprocher du plan médian de l'auditeur lorsque $r_V < 1$ et à s'en éloigner lorsque $r_V > 1$. Une cohérence parfaite entre la direction indiquée par le vecteur \mathbf{u}_V et les indices de localisation auditive est obtenue pour $r_V = 1$.

La seconde hypothèse concerne le domaine des moyennes et hautes fréquences (au-dessus de 700 Hz), celui où la propagation du son est perturbée par la présence de l'auditeur, notamment sa tête. Dans ce domaine fréquentiel, Gerzon [G92d] postule que la direction perçue par l'auditeur se confond avec celle du vecteur énergie \mathbf{E} , c'est-à-dire avec la direction moyenne de provenance de l'énergie. De façon analogue à la norme r_V , le taux de concentration d'énergie dans la direction \mathbf{u}_E donné r_E est un indicateur de la précision de localisation et du confort d'écoute.

3.4.2 Optimisation du décodage pour un dispositif régulier de haut-parleurs

Pour garantir la meilleure restitution sonore d'un point de vue perceptif, les considérations précédentes sur les vecteurs vitesse et énergie imposent quelques règles à respecter : $\mathbf{u}_V = \mathbf{u}_E = \mathbf{u}_p$ où \mathbf{u}_p désigne la direction de provenance de l'onde plane virtuelle, $r_V = 1$ et r_E est

maximal. Gerzon aboutit ainsi à un certain nombre de « théorèmes » [G73] relatifs au décodage du *B Format* et qui donnent des recommandations sur le placement des haut-parleurs et les formules associées. Ces recommandations, notamment la distribution régulière des haut-parleurs sur un cercle ou une sphère, sont en tout point conformes aux nôtres, données pendant l'étude portant sur le décodage basique. La prise en compte dans la technologie HOA des ordres supérieurs à 1 dans la description du champ acoustique permet une extension de la zone d'écoute et améliore la reconstruction physique et donc perceptive du champ acoustique pour un auditeur placé au centre du système de restitution. La maximisation de la norme r_E reste toutefois utile si l'auditeur se trouve en position excentrée. Si le décalage par rapport au centre du dispositif est trop important, il peut être préférable d'envisager un décodage dit *in-phase*, autrement appelé *controlled-opposites*, proposé par Malham [M92] et garantissant une émission sonore en phase de tous les haut-parleurs. Ces deux procédures d'optimisation se réalisent en modifiant les formules du décodage basique selon la méthode générale que nous présentons maintenant.

Principe du décodage modifié

Tout en maintenant la direction des vecteurs vitesse et énergie dans la direction de la source virtuelle, le décodage basique (avec $L > K$) peut être modifié en appliquant sur les signaux HOA avant la matrice de décodage des gains g_m propres à chaque ordre m [D00] :

$$\mathbf{D}_{\text{mod}} = \mathbf{D} \cdot \mathbf{\Gamma}_M \quad (3.45)$$

où \mathbf{D} est la matrice de décodage, et $\mathbf{\Gamma}_M$ désigne une matrice diagonale, définie par :

$$\mathbf{\Gamma}_M = \begin{cases} \text{diag} \left[(g_0 \ g_1 g_1 \ g_2 g_2 \ \cdots \ g_M g_M)^t \right] & (2D) \\ \text{diag} (g_0 \ g_1 g_1 g_1 \ \cdots \ \underbrace{g_m \ \cdots \ g_m}_{2m+1} \ \cdots \ \underbrace{g_M \ \cdots \ g_M}_{2M+1})^t & (3D) \end{cases} \quad (3.46)$$

Ces gains g_m sont déterminés suivant un certain nombre de critères d'optimisation, comme par exemple les critères $\max r_E$ et *in-phase* que nous décrirons par la suite. Ces critères doivent être associés à une contrainte de conservation de l'amplitude ou de l'énergie totale émise. Les gains peuvent alors s'écrire sous la forme d'un produit $g_m = g_0 g_m'$ où la valeur de g_0 dépend du choix entre la préservation de l'amplitude du champ de pression reconstruit au point de référence et celle de l'énergie totale restituée :

$$\begin{cases} g_0 = 1 & \text{(préservation de l'amplitude)} \\ g_0 = \sqrt{\frac{L}{E}} & \text{(conservation de l'énergie totale)} \end{cases} \quad (3.47)$$

avec

$$E = \begin{cases} g'_0 + 2 \sum_{m=1}^M g_m'^2 & (2D) \\ \sum_{m=0}^M (2m+1) g_m'^2 & (3D) \end{cases} \quad (3.48)$$

Solutions de décodage $\max r_E$ et *in-phase*

La maximisation de la norme r_E du vecteur énergie pour un décodage HOA d'ordre M quelconque est réalisé dans [D00, DRP98]. Les coefficients g_m' résultants et le paramètre g_0 garantissant la conservation de l'énergie sont donnés par :

$$(2D) \quad \begin{cases} g'_m = \cos \frac{m\pi}{2M+2}, & g_0 = \sqrt{\frac{L}{M+1}} \\ r_E = \cos \frac{\pi}{2M+2} (= g'_1) \end{cases} \quad (3.49)$$

et

$$(3D) \quad \begin{cases} g'_m = P_m(r_E), & g_0 = \sqrt{\frac{L}{E}} \\ r_E (= g'_1) \text{ est la plus grande racine de } P_{M+1} \end{cases} \quad (3.50)$$

où L est le nombre de haut-parleurs du dispositif de restitution. Nous remarquons que la norme r_E converge vers 1 lorsque M tend vers l'infini, et ce d'autant plus rapidement en 2D qu'en 3D.

Les solutions de décodage *in-phase* introduites initialement par Malham [M92] pour le décodage du *B-Format* (ordre 1), ont été étendues à l'ordre 2 par Furse [F99], Monro [M00] puis aux ordres supérieurs par Daniel [D00]. Le principe consiste à diminuer les gains des haut-parleurs au fur et à mesure qu'ils s'éloignent de la position de la source virtuelle, jusqu'à les annuler complètement dans la position diamétralement opposée. Ce critère d'optimisation est utile lorsqu'un auditeur est très excentré pour minimiser les effets néfastes des haut-parleurs qui émettent en opposition de phase mais induit en contrepartie une dégradation importante du champ acoustique reconstruit au voisinage de la position d'écoute idéale, c'est-à-dire au centre du dispositif de restitution. Les gains g_m' et le paramètre g_0 vérifiant ce critère sont les suivants :

$$(2D) \quad \begin{cases} g'_m = \frac{M!}{(M+m)!(M-m)!}, & g_0 = \sqrt{\frac{L}{E}} \\ r_E = \frac{2M}{2M+1} \end{cases} \quad (3.51)$$

et

$$(3D) \quad \begin{cases} g'_m = \frac{M!(M+1)!}{(M+m+1)!(M-m)!}, & g_0 = \sqrt{\frac{L(2M+1)}{M+1}} \\ r_E = \frac{M}{M+1} \end{cases} \quad (3.52)$$

Comparaison et combinaison des différentes stratégies de décodage

Le Tableau 3.5 donne explicitement les paramètres et caractéristiques des solutions de décodage basique, $\max r_E$ et *in-phase* pour les premiers ordres en 2D et 3D. Nous y voyons tout d'abord que la norme r_V du vecteur vitesse est unitaire pour un décodage basique d'ordre quelconque, ce qui signifie une excellente reconstruction perceptive de l'espace en basse fréquence. D'autre part, la norme r_E du vecteur énergie converge vers 1 lorsque l'ordre M tend vers l'infini. Par conséquent, la qualité perceptive de l'espace sonore reconstruit s'améliore lorsque l'ordre augmente, ce qui est conforme avec l'idée que la zone de correcte approximation du champ acoustique s'élargit (cf. Chapitre 2). Remarquons par ailleurs que la précision de localisation prédite en haute fréquence par la valeur de r_E est toujours meilleure à ordre égal en 2D qu'en 3D, ce qui traduit une zone d'approximation plus large en 2D qu'en 3D. Nous constatons également que la maximisation de la norme r_E dans l'étape de décodage se fait aux dépens de la norme r_V dont la valeur passe en dessous de 1, de telle façon que $r_V = r_E$. Dans ce cas, les critères objectifs de prédiction de la qualité perceptive sont également d'autant meilleurs que l'ordre est élevé, avec toujours une longueur d'avance à ordre égal pour le cas 2D. Enfin, le décodage *in-phase* est caractérisé par une forte dégradation du critère r_V , la valeur de r_E correspondant à celle du décodage basique. Cela signifie que les caractéristiques de propagation d'énergie restent relativement correctes mais que les propriétés de propagation de phase ne sont plus préservées. Ces dernières s'améliorent toutefois petit à petit lorsque l'ordre augmente. Il est enfin possible de combiner ces différentes stratégies de décodage en les appliquant sur des bandes fréquentielles distinctes dans le but de trouver le meilleur compromis possible pour la restitution sonore dans une zone large qui s'étend jusqu'aux haut-parleurs.

Type de décodage	Ordre	r_V	r_E	E	g'_m	
Basique	2D	1	1	0.667	3	(1)
		2	1	0.800	5	(1,1)
		3	1	0.857	7	(1,1,1)
		4	1	0.889	9	(1,1,1,1)
		M	1	$\frac{M}{M+1}$	$2M+1$	$(g'_m = 1)$
	3D	1	1	0.5	4	(1)
		2	1	0.667	9	(1,1)
		3	1	0.750	16	(1,1,1)
		4	1	0.800	25	(1,1,1,1)
		M	1	$\frac{M}{M+1}$	$(M+1)^2$	$(g'_m = 1)$
max r_E	2D	1	0.707	0.707	2	(0.707)
		2	0.866	0.866	3	(0.866,0.5)
		3	0.924	0.924	4	(0.924,0.707,0.383)
		4	0.951	0.951	5	(0.951,0.809,0.588,0.309)
		M	$r_V = r_E$	$\cos \frac{\pi}{2M+2}$	$M+1$	$\left(g'_m = \cos \frac{m\pi}{2M+2} \right)$
	3D	1	0.577	0.577	2	(0.577)
		2	0.775	0.775	3.6	(0.775,0.4)
		3	0.861	0.861	5.75	(0.861,0.612,0.305)
		4	0.906	0.906	8.441	(0.906,0.732,0.501,0.246)
		M	$r_V = r_E$	–	–	$(g'_m = P_m(r_E))$
In-phase	2D	1	0.500	0.667	1.5	(0.5)
		2	0.667	0.800	1.944	(0.667,0.167)
		3	0.750	0.857	2.310	(0.750,0.3,0.05)
		4	0.8000	0.889	2.627	(0.8,0.4,0.114,0.014)
		M	$M/M+1$	$\frac{2M}{2M+1}$	–	$\left(g'_m = \frac{M^i}{(M+m)!(M-m)!} \right)$
	3D	1	0.333	0.500	1.333	(0.333)
		2	0.500	0.667	1.8	(0.5,0.1)
		3	0.600	0.750	2.286	(0.6,0.2,0.029)
		4	0.667	0.800	2.778	(0.667,0.286,0.071,0.008)
		M	$\frac{M}{M+2}$	$\frac{M}{M+1}$	$\frac{(M+1)^2}{2M+1}$	$\left(g'_m = \frac{M!(M+1)!}{(M+m+1)!(M-m)!} \right)$

Tableau 3.5 – Paramètres et caractéristiques des principales solutions de décodage 2D et 3D pour les configurations régulières de haut-parleurs. (D'après [D00])

Références bibliographiques du Chapitre 3

- [G85] GERZON, M.A. *Ambisonics in Multichannel Broadcasting and Video*. J. Audio Eng. Soc., 1985, Vol. 33, n° 11, p. 859-871.
- [F74] FELLGETT, P.B. *Ambisonic reproduction of directionality in surround-sound systems*. Nature, 1974, Vol. 252, p. 534-538.
- [CG77] CRAVEN, P. et GERZON, M.A. *Coincident Microphone Simulation Covering Three Dimensional Space and Yielding Various Directional Outputs*. U.S. Patent 4,042,779, août 1977.
- [CS71] COOPER, D.H. et SHIGA, T. *Discrete-Matrix Multichannel Stereo*. J. Audio Eng. Soc., 1971, Vol. 20, n° 5, p. 346-360.
- [B95] BAMFORD, J.S. *An Analysis of Ambisonics Sound Systems of First and Second Order*. M. Sc. Thesis, Waterloo: University of Waterloo, 1995.
- [D01] DANIEL, J. *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. Paris: Paris 6, 2001.
- [B98] BRUNEAU, M. *Manuel d'acoustique fondamentale*. Paris: Hermès, 1998, 576 p.
- [P96] POLETTI, M.A. *The Design of Encoding Functions for Stereophonic and Polyphonic Sound Systems*. J. Audio Eng. Soc., 1996, Vol. 44, n° 11, p. 948-963.
- [P00] POLETTI, M.A. *A Unified Theory of Horizontal Holographic Sound Systems*. J. Audio Eng. Soc., 2000, Vol. 48, n° 12, p. 1155-1182.
- [G73] GERZON, M.A. *Periphony: With-Height Sound Reproduction*. J. Audio Eng. Soc., 1973, Vol. 21, n° 1, p. 2-10.
- [M95] MALHAM, D. *Basic Ambisonics* 1995. http://www.york.ac.uk/inst/mustech/3d_audio/ambsn1ct.htm (dernier accès le 23 mai 2006)
- [M05] MALHAM, D. *Second and Third Order Ambisonics - the Furse-Malham Set*. 2005, http://www.york.ac.uk/inst/mustech/3d_audio/secon2or.html#ref2 (dernier accès le 23 mai 2006)
- [F99] FURSE, R. *3D Audio Links and Information*. 1999. <http://www.muse.demon.co.uk/3daudio.html> (dernier accès le 23 mai 2006)
- [D00] DANIEL, J. *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. Thèse de Doctorat, Paris: Université de Paris 6, 2000.

-
- [M05] MALHAM, D.G. *Second and Third Order Ambisonics - the Furse-Malham Set*. 2002. http://www.york.ac.uk/inst/mustech/3d_audio/secondor.html#ref2 (dernier accès le 23 mai 2006)
- [JLP99] JOT, J.M., LARCHER, V., et PERNAUX, J.M. *A Comparative Study of 3-D Audio Encoding and Rendering Techniques*. AES 16th International Conference on Spatial Sound Reproduction, Rovaniemi, Finlande, 10-12 avril 1999.
- [PBJ98] PERNAUX, J.M., BOUSSARD, P., et JOT, J.M. *Virtual Sound Source Positioning and Mixing in 5.1 Implementation on the Real-Time System Genesis*. DAFX'98, Barcelone, Espagne, 19-21 novembre 1998.
- [G92c] GERZON, M.A. *The Design of Distance Panpots*. AES 92nd Conv., Vienna, 1992.
- [D03] DANIEL, J. *Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format*. AES 23rd International Conference, Helsingor, Danemark, 23-25 mai 2003.
- [F79a] FARRAR, K. *Soundfield Microphone*. Wireless World, 1979, octobre, p. 48-50.
- [F79b] FARRAR, K. *Soundfield Microphone - 2*. Wireless World, 1979, novembre, p. 99-103.
- [GB92] GERZON, M.A. et BARTON, G.J. *Ambisonic Decoders for HDTV*. AES 92nd Convention, preprint n° 3345, Vienne, Autriche, 24-27 mars 1992.
- [M90] MALHAM, D.G. *Ambisonics - A Technique for Low Cost, High Precision, Three Dimensional Sound Diffusion*. International Computer Music Conference, p. 118-120, Glasgow, Scotland, septembre 1990.
- [WA01b] WARD, D.B. et ABHAYAPALA, T.D. *Reproduction of a Plane-Wave Sound Field Using an Array of Loudspeakers*. IEEE Transaction on Speech and Audio Processing, 2001, Vol. 9, n° 6, p. 697-707.
- [GVL96] GOLUB, G. et VAN LOAN, C. *Matrix Computations, 3rd edition*. London: The Johns Hopkins University Press, 1996, 694 p.
- [Mo92] MOLLER, H. *Fundamentals of Binaural Technology*. Applied Acoustics, 1992, Vol. 36, p. 171-218.
- [B97] BLAUERT, J. *Spatial Hearing. The Psychophysics of Human Sound Localization*. revised ed, Cambridge, Massachusetts: The MIT Press, 1997, 494 p.
- [C90] CANÉVET, G., *Audition binaurale et localisation auditive. Aspects physiques et psychoacoustiques*, in *Psychoacoustique et perception auditive, chap. 3*, é. INSERM, Editor. 1990. p. 83-122.
- [A01] ALGAZI, V.R., et al. *The CIPIC HRTF Database*. IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics, New York, USA, 21-24 octobre 2001.

-
- [GM95] GARDNER, B. et MARTIN, K. *HRTF Measurements of a KEMAR*. J. Acoust. Soc. Am., 1995, Vol. 97, p. 3907-3908.
- [W02] WARUSFEL, O. *LISTEN HRTF database*. 2002. <http://recherche.ircam.fr/equipes/salles/listen/> (dernier accès le 25 mai 2006)
- [G77] GERZON, M.A. *The Optimum Choice of Surround Sound Encoding Specification*. AES 56th Convention, preprint 1199, Paris, France, 1977.
- [G92b] GERZON, M.A. *Psychoacoustic Decoders for Multiplespeaker Stereo and Surround-Sound*. AES 93rd Convention, preprint 3406, San Francisco, USA, octobre 1992.
- [W03] WIGGINS, B., et al. *The Design and Optimisation of Surround Sound Decoders Using Heuristic Methods*. Proceedings of UKSim 2003, Conference of the UK Simulation Society, p. 106-114, 2003.
- [C03] CRAVEN, P. *Continuous Surround Panning for 5-Speaker Reproduction*. AES 24th International Conference on Multichannel Audio, Banff, Canada, mai 2003.
- [DRP99] DANIEL, J., RAULT, J.B., et POLACK, J.D. *Acoustic Properties and Perceptives Implications of Stereophonic Phenomena*. AES 16th International Conference on Spatial Sound Reproduction, Rovaniemi, Finlande, 10-12 avril 1999.
- [DRP98] DANIEL, J., RAULT, J.B., et POLACK, J.D. *Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions*. AES 105th Convention, preprint 4795, 26-29 septembre 1998.
- [G92d] GERZON, M.A. *General Metatheory of Auditory Localisation*. AES 92nd Convention, preprint 3306, Vienne, Autriche, 23-29 mars 1992.
- [DNM03] DANIEL, J., NICOL, R., et MOREAU, S. *Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging*. 114th AES Convention, preprint 5788, Amsterdam, Pays-Bas, 22-25 mars 2003.
- [M62] MAKITA, Y. *On the Directional Localization of Sound in the Stereophonic Sound Field*. E.B.U. Review, 1962, Vol. A, n° 73, p. 102-108.
- [M92] MALHAM, D.G. *Experience With Large Area 3-D Ambisonic Sound Systems*. Proc. of the Institute of Acoustics, Vol. 14, p. 209-215, 1992.
- [M00] MONRO, G. *In-phase corrections for Ambisonics*. Proceedings of the International Computer Music Conference (ICMC), p. 292-295, Berlin, Allemagne, 27 août -1er septembre 2000.

Chapitre 4

Encodage spatial HOA de source virtuelle en champ proche

À l'heure où les applications de réalité virtuelle ou par exemple les jeux vidéo permettent à l'utilisateur d'évoluer dans des environnements virtuels toujours plus réalistes, la question de la reproduction des sources sonores en champ proche paraît plus que jamais d'actualité. Plusieurs études récentes ont été consacrées à ce problème, notamment dans le contexte de la restitution binaurale [BR99], de la *Wave Field Synthesis* [V97], d'*Higher Order Ambisonics* [D03] et enfin à partir d'une technologie hybride WFS/HOA [SH01, SH02]. Les difficultés rencontrées pour reproduire le champ acoustique proche des sources concernent principalement la quantité importante d'énergie devant être fournie par le dispositif de restitution. L'étude réalisée par Daniel [D03] apporte des solutions à ce problème, qui reste toutefois limitées dans le cas particulièrement critique où la source virtuelle est intérieure au dispositif de restitution. Nous présentons dans ce chapitre une nouvelle méthode de conception des filtres de champ proche basée sur un critère objectif qui permette de reconstruire une source intérieure à un dispositif de restitution en assurant une focalisation de l'énergie acoustique sur l'emplacement de la source, tout en limitant celle délivrée par les haut-parleurs.

4.1 Caractérisation physique et perceptive du champ proche

4.1.1 Caractérisation physique du champ proche

Qu'est-ce que le champ proche ?

Les fondements théoriques de la technologie HOA présentés au Chapitre 3 se basent pour l'essentiel sur l'hypothèse de *champ lointain*, c'est-à-dire que nous supposons que les sources virtuelles et les haut-parleurs de restitution émettent infiniment loin de l'auditeur, ou en tout cas *suffisamment* loin pour que les ondes sonores qui se propagent puisse être assimilées à des ondes planes. De cette façon, l'amplitude de la pression acoustique reste relativement homogène dans la zone de l'espace considérée. Par opposition au champ lointain, nous appelons *champ proche* le champ acoustique rayonné à proximité d'une source sonore et caractérisé par une forte instabilité, c'est-à-dire de fortes variations de la pression sonore et de la vitesse particulière sur de petits déplacements. Les notions assez vagues de *proximité* et de *petits déplacements* dépendent en réalité de la distance à la source et de la longueur d'onde du signal émis. La zone de champ proche correspond aux points d'observation dont la distance à la source est petite devant la longueur d'onde ($kr \ll 1$). Inversement, lorsque cette distance d'observation est grande devant la longueur d'onde ($kr \gg 1$), la propagation acoustique est conforme à l'hypothèse de champ lointain. À une même distance de la source, un point de l'espace peut ainsi appartenir à la zone de champ proche dans un domaine basse fréquence, et à la zone de champ lointain dans un domaine haute fréquence. En pratique, il est courant d'associer l'hypothèse d'onde plane à la condition $kr > 10$.

Cas particulier d'une source sonore ponctuelle

L'étude du rayonnement immédiat d'une source ponctuelle homogène en champ libre permet de caractériser physiquement le phénomène de champ proche. Pour plus de simplicité, nous supposons que cette source se situe à l'origine de notre repère. Elle engendre ainsi une pression acoustique et une vitesse particulière en tout point de l'espace ($r \neq 0$) qui peuvent respectivement s'écrire :

$$p(r) = \frac{i\rho_0 ck}{4\pi} Q \frac{e^{-ikr}}{r}, \quad (4.1)$$

et

$$\begin{aligned} v(r) &= -\frac{1}{i\rho_0 ck} \frac{\partial p}{\partial r}(r) \\ &= \frac{Q}{4\pi} \frac{1}{r^2} (1 + ikr) e^{-ikr}, \end{aligned} \quad (4.2)$$

où ρ_0 est la masse volumique de l'air, Q le débit volumique instantané (intensité de la source), c la vitesse du son dans l'air, k le nombre d'onde et r la distance à la source strictement positive. Rappelons que la dépendance en temps $e^{i\omega t}$ est implicite dans ces deux expressions. Remarquons de plus que la pression acoustique et la vitesse particulière ne dépendent que du rayon r grâce aux propriétés de rayonnement sphérique de la source sonore.

L'équation (4.1) montre que l'amplitude de la pression acoustique tend vers l'infini lorsque r tend vers zéro, c'est-à-dire lorsque l'on s'approche de la source émettrice, et décroît suivant une loi en $1/r$ lorsqu'on s'en éloigne, c'est-à-dire très rapidement à proximité de la source puis de façon plus monotone. La vitesse particulière dont l'expression est donnée à l'équation (4.2) est plus complexe à caractériser. Nous pouvons cependant étudier son comportement dans la zone de champ proche, c'est-à-dire lorsque kr est très petit devant 1 ($kr \ll 1$). En effet, dans ce cas le facteur d'amplitude entre parenthèses dans l'équation (4.2) est approximativement égal à 1 et l'amplitude de la vitesse particulière décroît selon une loi en $1/r^2$, soit beaucoup plus rapidement que la pression sonore avec laquelle elle est par ailleurs en quadrature de phase. En hypothèse de champ lointain, lorsque kr est négligeable devant 1 ($kr \gg 1$), la vitesse particulière se propage avec une amplitude qui décroît suivant une loi en $1/r$, de façon identique à la pression acoustique avec laquelle elle est cette fois en phase.

4.1.2 Perception auditive du champ proche

Effet de proximité

La localisation des sources sonores dans l'espace est effectuée par le système auditif grâce à un certain nombre d'indices présents dans les signaux acoustiques captés par les tympanes. Les principaux indices sont les différences interaurales de temps (*Interaural Time Difference* ou ITD) et d'intensité (*Interaural Level Difference* ou ILD) pour la localisation auditive suivant l'azimut (dans un plan horizontal), et les indices de modification spectrale pour la localisation suivant l'élévation (dans un plan vertical) [B97, C90]. Ces indices sont habituellement définis pour des ondes acoustiques supposées planes au niveau des oreilles de l'auditeur. Dans ce cas particulier, l'amplitude de la pression acoustique reste relativement homogène à l'emplacement de l'auditeur même lorsqu'il se déplace, et les différences d'intensité constatées entre les signaux captés par les tympanes s'expliquent principalement par l'effet de masquage acoustique induit par la tête.

Si l'auditeur se place à proximité d'une source sonore, en pratique à une distance inférieure à 1 mètre [BDR99], il expose ses tympanes à des phénomènes acoustiques instables. Ces phénomènes qui restent confinés au voisinage de la source sont à l'origine d'écarts importants de niveau de pression acoustique entre les deux oreilles de l'auditeur, et ce d'autant plus que la source est située sur le côté, le cas le plus critique correspondant à un azimut $\theta_s = \pm\pi/2$. En effet, dans ce dernier cas le rapport entre la distance séparant la source et l'oreille la plus proche (dite *ipsilatérale*) et la distance de la source à l'oreille la plus éloignée (dite *contralatérale*) diminue fortement à mesure que l'auditeur s'approche de la source. Comme l'amplitude de la pression sonore est dans cette zone de l'espace inversement proportionnelle à la distance de la source et que l'effet de diffraction du son par la tête est renforcé, il en résulte que les différences interaurales d'intensité entre les signaux captés par les tympanes (ILD) augmentent fortement à toutes les fréquences : c'est ce qu'on

appelle l'*effet de proximité*¹¹ (ou effet de champ proche). La Figure 2.2 montre les changements dans les indices ILD et ITD en fonction de la distance à la source, calculés pour un modèle de tête sphérique et mesurés sur le mannequin Kemar par Brungart et ses collègues [BR99]. Les différences interaurales d'intensité (ILD) augmentent rapidement, particulièrement aux distances inférieures à 50 cm, et d'autant plus que la source s'éloigne de l'axe frontal et que la fréquence est haute. Remarquons enfin que les différences interaurales de temps (ITD) augmentent quant à elles de façon insignifiante lorsque la distance à la source diminue.

Incidence de l'effet de champ proche sur la perception auditive de l'espace

Comparativement à une source lointaine, la proximité d'une source acoustique modifie les indices de la localisation auditive et amplifie notamment fortement l'ILD, y compris dans un domaine moyenne et basse fréquence où cet indice n'est pas utile pour la localisation en hypothèse de champ lointain. Puisque le système auditif est capable de détecter des différences interaurales d'intensité de l'ordre de 0,8 dB [B97], ces modifications entraînent vraisemblablement un changement de la perception. Même si encore peu d'études subjectives ont été menées sur le sujet, les travaux réalisés par Brungart [Bg98] montrent que d'une part cette amplification de l'ILD ne perturbe pas la localisation angulaire des sources sonores, et d'autre part qu'elle s'ajoute comme un indice supplémentaire de la perception auditive de l'espace, caractéristique de l'écoute en champ proche. Il s'avère par conséquent judicieux de chercher à reproduire le champ acoustique à proximité des sources sonores dans les applications où l'utilisateur évolue dans un environnement virtuel, dans le but d'améliorer ses sensations de réalisme et d'immersion sonore.

¹¹ Cet effet de proximité est à rapprocher de celui, bien connu des ingénieurs du son, obtenu en plaçant un microphone directif en champ proche (à proximité d'une voix ou d'un instrument par exemple).

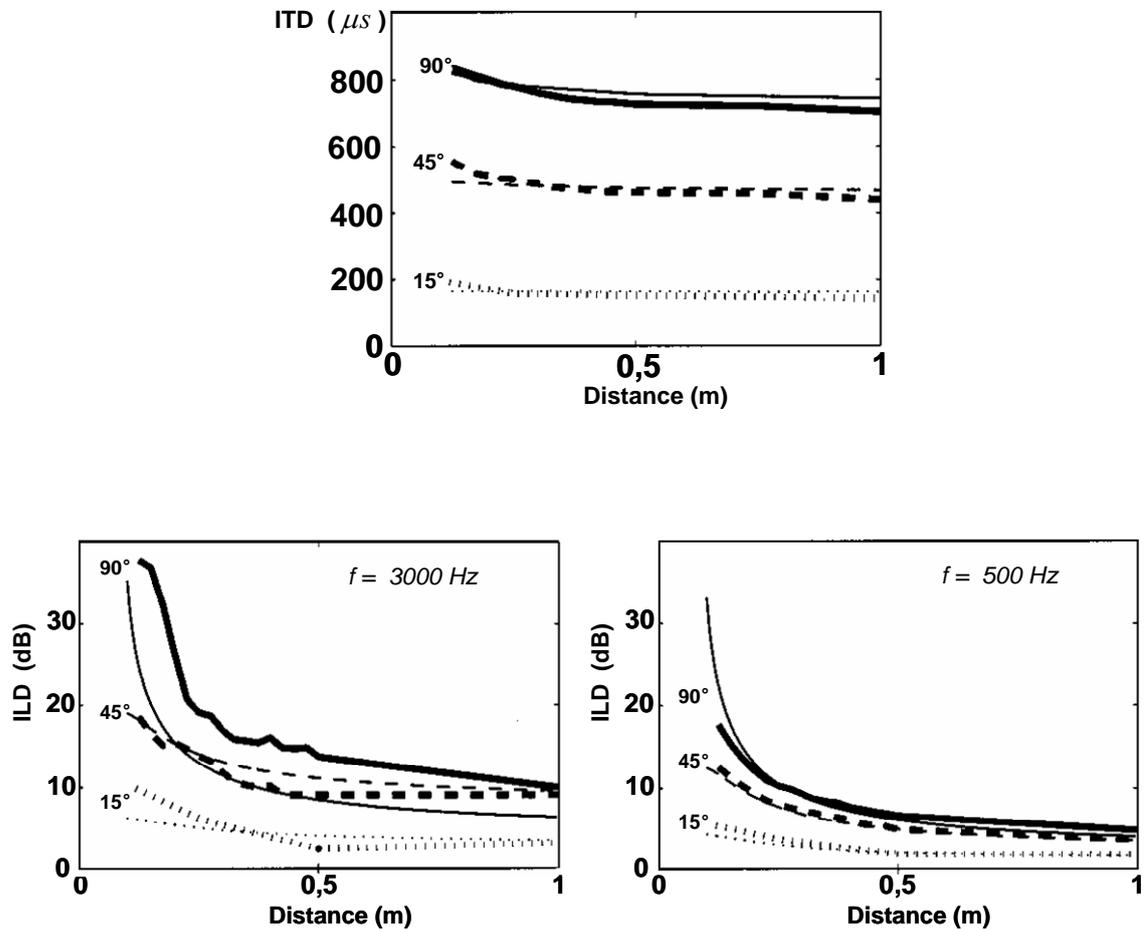


Figure 2.4.1 – Variations des Différences Interaurales d'Intensité (ILD) et des Différences Interaurales de Temps (ITD) en fonction de la distance (de 0,1 à 1 m) de la source pour un modèle de tête humaine sphérique (lignes fines) et pour le mannequin KEMAR (lignes épaisses). Cas particuliers d'un signal de fréquence 3000 Hz et 500 Hz, émis par une source d'azimut successif 15°, 45° et 90° et d'élévation nulle. (D'après [BR99])

4.2 Problème de la reproduction du champ proche d'une source ponctuelle homogène

Pour reproduire une source ponctuelle placée à distance finie de l'auditeur, les propriétés de propagation sphérique de l'onde émise nécessitent d'introduire dans l'encodage HOA des opérations de filtrage impliquant parfois des amplifications infinies, et donc impossibles à implanter telles quelles. La proposition récente de Jérôme Daniel [D03] de combiner dans l'étape d'encodage HOA ces opérations de filtrage avec celles, nécessaires à la restitution, de compensation du champ proche des haut-parleurs, permet désormais d'obtenir des filtres d'amplitude finie. Cependant, ces filtres mettent en jeu des amplifications qui restent excessives dans le cas où la source virtuelle est intérieure au dispositif de restitution.

4.2.1 Définition théorique et analyse des filtres de champ proche

Définition théorique des filtres de champ proche

Rappelons tout d'abord les équations d'encodage HOA d'une onde sphérique virtuelle présentées au Chapitre 3 :

$$B_{mn}^\sigma = S(\omega) \cdot Y_{mn}^\sigma(\theta_s, \delta_s) F_m^{\tau_s}(\omega), \quad (4.3)$$

avec

$$\begin{aligned} F_m^{\tau_s}(\omega) &= \frac{h_m^-(kr_s)}{h_0^-(kr_s)} i^{-(m+1)} \\ &= \sum_{n=0}^m \frac{(m+n)!}{(m-n)!n!} \left(\frac{1}{i2\omega\tau_s} \right)^n \end{aligned} \quad (4.4)$$

où $(r_s, \theta_s, \delta_s)$ sont les coordonnées sphériques de la source virtuelle émettrice, S le signal émis et $F_m^{\tau_s}(\omega)$ les filtres de champ proche caractérisant la propagation acoustique à proximité de la source. Ces filtres sont définis pour une distance r_s de source particulière, correspondant à une unique constante de temps $\tau_s = r_s/c$, c étant la vitesse de propagation du son dans l'air.

Analyse des filtres de champ proche

Les filtres $F_m^{\tau_s}(\omega)$ d'ordre m supérieur ou égal à 1 sont caractérisés par une amplification en basse fréquence de pente $-m \times 6$ dB/octave, déjà évoquée sous le terme de « *bass boost* » par Gerzon dans un article sur la simulation de l'effet de distance [G92c]. Pour l'ordre $m=1$ par exemple, la fonction de transfert du filtre $F_1^{\tau_s}(\omega)$ s'exprime ainsi :

$$F_1^{\tau_s}(\omega) = 1 + \frac{1}{i\omega\tau_s}. \quad (4.5)$$

Le comportement du filtre en basse et haute fréquence peut être évalué en supposant la fréquence ω tour à tour petite et grande. Dans le premier cas, l'hypothèse basse fréquence, l'expression du filtres $F_1^{\tau_s}(\omega)$ donnée à l'équation (4.5) peut être approchée par $F_1^{\tau_s}(\omega) \approx 1/i\omega\tau_s$, caractéristique

d'une amplification et d'un déphasage de -90° figuré par le facteur imaginaire pur i au dénominateur. Dans le second cas qui concerne les hautes fréquences, $F_1^{\tau_s}(\omega)$ est approximativement unitaire, ce qui signifie que le filtre n'entraîne aucune modification de l'amplitude et de la phase du signal auquel il est appliqué. Nous reconnaissons en effet dans l'équation (4.5) l'expression de la fonction de transfert d'un filtre amplificateur basse fréquence de type RC standard caractérisé par sa constante de temps τ_s , une pente de -6 dB par octave, une fréquence d'amplification à $+3$ dB définie par ω_0 et un déphasage qui atteint -90° en basse fréquence. La fréquence d'amplification à $+3$ dB est inversement proportionnelle à la distance r_s de la source. Autrement dit, plus la source est éloignée, plus la fréquence de coupure ω_0 est basse.

La Figure 4.2 illustre les filtres de champ proche pour une distance de source r_s égale à 1 mètre et pour les ordres 1 à 15. Leurs caractéristiques décrites précédemment sont facilement observables, notamment les pentes de -6 dB/octave. Alors que les filtres d'ordre 1 semblent être réalisable en pratique, les choses se compliquent à partir de l'ordre 2. En effet, l'amplification devient de plus en plus problématique en basse fréquence (d'autant plus que la distance à la source est petite) lorsque l'ordre augmente.

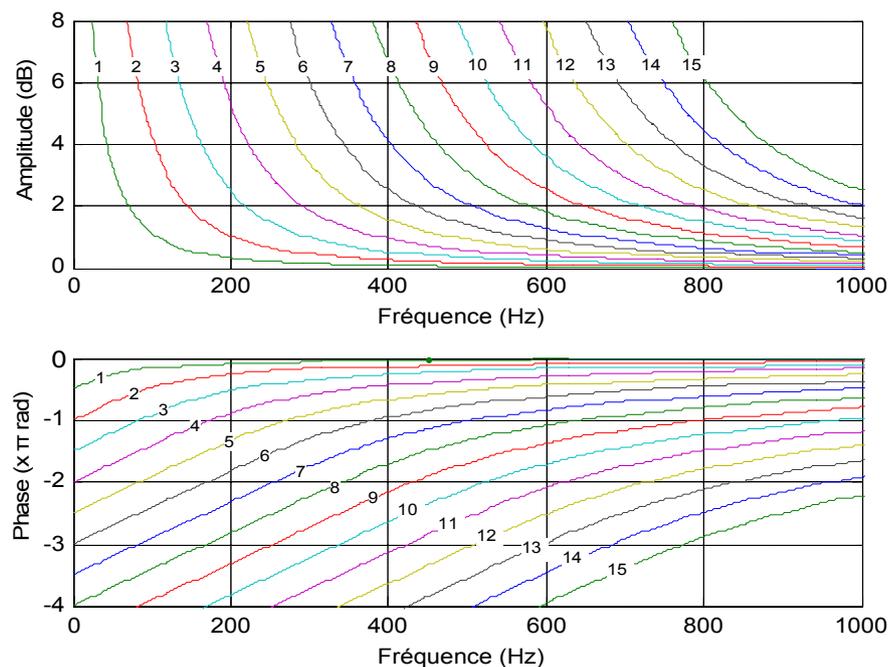


Figure 4.2 – Amplitude (haut) et phase (bas) des filtres de champ proche d'ordres 1 à 15 en fonction de la fréquence, pour une source placée à $r_s=1$ m du point d'observation.

4.2.2 Prise en compte du champ proche des haut-parleurs de restitution

Principe de la pré-compensation du champ proche des haut-parleurs de restitution

Pour reproduire une source en champ proche, il est nécessaire de prendre en compte l'effet de proximité des haut-parleurs. Plus précisément, il s'agit de compenser cet effet en appliquant sur les signaux devant être diffusés lors de l'étape de décodage des filtres inverses aux filtres $F_m^{\tau_{hp}}(\omega)$ définis à l'équation (4.4). Comme l'a récemment suggéré Daniel [D03], cette compensation du champ proche des haut-parleurs peut être associée par avance aux filtres de champ proche des sources encodées, c'est-à-dire dès l'étape d'encodage spatial¹². De cette manière, l'amplification excessive en basse fréquence requise par les filtres de proximité peut être limitée. Les filtres résultants s'écrivent alors comme le rapport entre les filtres de champ proche de la source encodée et ceux des haut-parleurs de restitution :

$$H_m^{\tau_s, \tau_{HP}}(\omega) = \frac{F_m^{\tau_s}(\omega)}{F_m^{\tau_{HP}}(\omega)}, \quad m \geq 0. \quad (4.6)$$

$\tau_s = r_s / c$ et $\tau_{HP} = r_{HP} / c$ sont les constantes de temps respectivement associées aux filtres $F_m^{\tau_s}(\omega)$ et $F_m^{\tau_{HP}}(\omega)$, r_s étant la distance de la source sonore et r_{HP} la distance des haut-parleurs¹³ au point d'écoute central. Contrairement aux filtres de champ proche, les filtres $H_m^{\tau_s, \tau_{HP}}(\omega)$ possèdent en basse fréquence une amplitude maximale finie qui tend vers la valeur $m \times 20 \log_{10}(r_{hp} / r_s)$ dB, positive pour les sources placées à l'intérieur du dispositif de restitution ($r_s < r_{hp}$) et négative pour les sources placées à l'extérieur ($r_s > r_{hp}$). La Figure 2.4.3 représente les amplitudes correspondant aux filtres $H_m^{\tau_s, \tau_{HP}}(\omega)$ d'ordre 1 à 8 pour une source intérieure et une source extérieure au dispositif de reproduction. Avec cette méthode de pré-compensation du champ proche des haut-parleurs, la structure de l'encodeur HOA de source virtuelle en champ proche reste la même que celle illustrée à la Figure 2.3, les filtres $H_m^{\tau_s, \tau_{HP}}(\omega)$ remplaçant simplement les filtres $F_m^{\tau_s}(\omega)$.

La définition des filtres $H_m^{\tau_s, \tau_{HP}}(\omega)$ se base sur le choix d'une distance de référence r_{hp} des haut-parleurs par rapport au point d'écoute privilégié supposé être le centre d'un dispositif de restitution concentrique. Cette contrainte apparente n'en est finalement pas une puisqu'il est possible d'adapter pendant l'étape de décodage spatial les signaux HOA au rayon réel du dispositif. L'adaptation se fait grâce aux formules suivantes :

¹² Signalons également que des travaux sur la reproduction du champ proche d'une source virtuelle avec une technique hybride *Wave Field Synthesis* / HOA ont été réalisés par Sontacchi et Höldrich [SH01, SH02]. Le procédé développé comporte une première étape d'encodage spatial du type *Wave Field Synthesis* dans laquelle les équations d'encodage prennent en compte non seulement la distance des sources virtuelles, mais également celle des haut-parleurs de restitution. Les signaux encodés sont ensuite convertis dans le format HOA pour leur exploitation (transmission, manipulation, décodage spatial).

¹³ Nous supposons un dispositif de restitution concentrique de rayon r_{HP} .

$$B'_{mn}{}^\sigma = \frac{F_m^{\tau_1}(\omega)}{F_m^{\tau_2}(\omega)} B_{mn}^\sigma, \quad m \geq 0, \quad 0 \leq n \leq m \text{ et } \sigma = \pm 1, \quad (4.7)$$

où $\tau_1 = r_1/c$ et $\tau_2 = r_2/c$, r_1 étant la distance d'origine de référence des haut-parleurs et r_2 le rayon réel du dispositif de restitution.

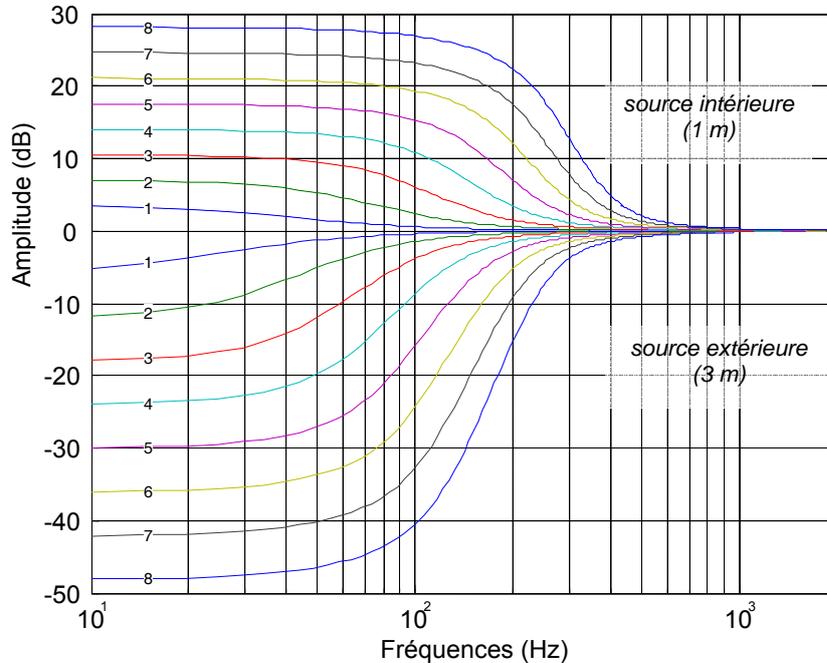


Figure 2.4.3 – Amplitude des filtres $H_m^{\tau_s, \tau_{HP}}(\omega)$ incluant l'effet de proximité des sources sonores encodées ainsi que la compensation de celui des haut-parleurs de restitution. Cas particuliers pour des sources extérieure et intérieure au dispositif de restitution, les haut-parleurs se trouvant à une distance de 1,5 m par rapport à la position d'écoute centrale.

Pour réaliser les filtres $H_m^{\tau_s, \tau_{HP}}(\omega)$, une première méthode consiste à effectuer l'implantation sous la forme de filtres RIF (filtres Réponse Impulsionnelle Finie). Il suffit pour cela de calculer la transformée de Fourier inverse de la réponse fréquentielle idéale des filtres $F_m^{\tau_s}(\omega)$ donnée à l'équation (4.4). Dans ce cas, les filtres doivent être calculés pour chaque valeur de la distance r_s et leur longueur doit être suffisante pour limiter de façon raisonnable le phénomène d'« oscillations de Gibbs ». Le coût de traitement résultant de l'opération de filtrage peut finalement se révéler très important. Il peut être par conséquent plus avantageux d'envisager une implantation des filtres $H_m^{\tau_s, \tau_{HP}}(\omega)$ d'ordre m sous la forme de filtres RII (filtres à Réponse Impulsionnelle Infinie), c'est-à-dire de filtres récursifs, à partir de cellule d'ordre 1 et d'ordre 2 [D03].

Avantages et limites de la méthode

La prise en compte de l'effet de proximité des haut-parleurs de restitution est nécessaire pour une reproduction fidèle du champ acoustique. Elle est également avantageuse si on la combine dès l'étape d'encodage aux filtres de champ proche associés aux sources virtuelles. Il en résulte en effet des filtres stables dont l'amplitude en basse fréquence, contrairement aux filtres de champ proche considérés seuls, converge. L'adaptation possible de la distance des haut-parleurs avant décodage n'en limite pas l'utilisation.

Lorsque les sources virtuelles sont plus éloignées du point de référence que les haut-parleurs ($r_s > r_{hp}$), c'est-à-dire si elles se situent à l'extérieur d'un dispositif de restitution circulaire ou sphérique, la réalisation de tels filtres semble relativement aisée. En effet, la pré-compensation du champ proche des haut-parleurs compense dans ce cas largement la pente d'amplification des filtres $F_m^{r_s}(\omega)$ caractérisant la proximité de la source virtuelle, et les filtres $H_m^{NFC(r_s/c, r_{hp}/c)}(\omega)$ résultants correspondent à une atténuation basse fréquence maximale de $m \times 20 \log_{10}(r_{hp}/r_s)$ dB (cf. Figure 2.3). En revanche, si les sources virtuelles émettent à l'intérieur du dispositif de diffusion, la compensation du champ proche des haut-parleurs ne permet plus de limiter raisonnablement l'amplification en basse fréquence due à l'effet de proximité de la source virtuelle. Il en résulte une amplification d'autant plus importante que la source est proche de l'auditeur et que le rayon du dispositif de restitution est grand. La méthode de pré-compensation du champ proche des haut-parleurs reste dans ce cas insuffisante. Cette constatation nous a amené à proposer une méthode alternative de conception de filtres de champ proche. Cette solution est basée sur le *concept de bande fréquentielle utile*.

4.3 Proposition d'une méthode de conception des filtres de champ proche

4.3.1 Concept de bande fréquentielle utile

Caractérisation de l'approximation du champ acoustique dans une zone limitée de l'espace

Nous avons constaté aux précédents chapitres que d'une façon générale, lorsque nous considérons une zone de l'espace quelconque sans source, les composantes spatiales participent à la description (et donc la reconstruction) du champ acoustique à partir d'une distance du point de référence d'autant plus grande, proportionnellement à la longueur d'onde, que l'ordre est élevé. L'apparition progressive des différents modes harmoniques au fur et à mesure que le rayon r augmente et la longueur d'onde diminue est en effet décrite dans le développement en harmoniques sphériques du champ acoustique par les fonctions de Bessel sphérique. Ce phénomène est facilement observable sur les simulations numériques du champ de pression acoustique résultant de la troncature de la série de Fourier-Bessel, notamment dans les cas particuliers d'une onde plane (cf. Figure 2.9) et d'une onde sphérique (cf. Figure 2.12), ce dernier cas présentant toutefois un tassement de l'étendue de la reconstruction à proximité de la source.

Adoptons à présent un point de vue différent. Si nous considérons une zone d'écoute limitée à la restitution, caractérisée par un rayon r_{lim} indépendant de la fréquence, nous déduisons que les signaux HOA ne sont significativement utiles à la reconstruction du champ acoustique sur la zone de reconstruction visée qu'*au-delà d'une fréquence $f_{\text{lim}}^{(m)}$ qui croît avec l'ordre m* . De plus, cette fréquence limite est d'autant plus élevée pour chaque ordre que la taille de la zone d'écoute visée est petite. Pour illustrer ce principe, reprenons par exemple la Figure 2.9 qui illustre l'approximation d'une onde plane dans le plan horizontal en fonction de l'ordre de troncature du champ sonore et de la fréquence du signal transporté. Nous représentons ainsi à la Figure 4.4, de façon schématique, les fréquences utiles à la reconstruction du champ acoustique dans une zone circulaire de 30 cm de diamètre et pour une tolérance d'erreur moyenne¹⁴ de reconstruction du champ acoustique de 7 %. La Figure 4.5 est une généralisation du schéma de la Figure 4.4 à toutes les fréquences et aux signaux HOA d'ordres quelconques.

Le concept de bande fréquentielle utile s'applique également au cas particulier de l'approximation d'une onde sphérique dans une zone de l'espace limitée. Nous pouvons cependant ajouter quelques commentaires. D'après l'hypothèse de champ libre de laquelle découle la série de Fourier-Bessel (cf. Chapitre 2 Section 2.3.1), nous savons en effet que dans le cas d'une onde sphérique, contrairement à celui d'une onde plane, l'étendue radiale de la zone d'écoute se limite au rayon r_s de la source émettrice sans jamais pouvoir l'atteindre. Le champ de pression acoustique est en effet théoriquement infini à l'emplacement de la source.

¹⁴ Nous utilisons la définition de l'erreur moyenne \bar{e}_M donnée au Chapitre 2.

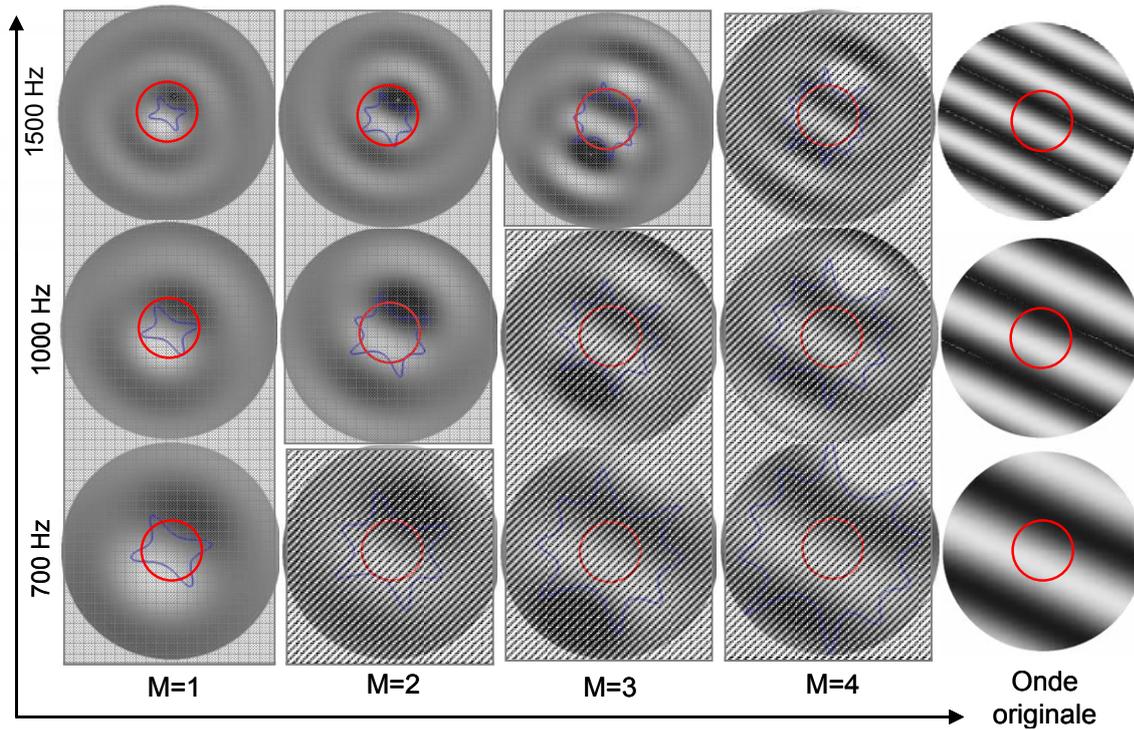


Figure 4.4 – Illustration dans le plan horizontal de l'approximation du champ acoustique en fonction de l'ordre de troncature et de la fréquence. Le cercle rouge de 30 cm de diamètre représente la zone d'écoute visée pour la restitution. Les courbes étoilées délimitent les zones de l'espace où l'approximation est inférieure à 7 % (-11,5 dB). Pour chaque ordre de troncature, les fréquences utiles à la description du champ acoustique à l'intérieur du cercle rouge, définies en fonction d'une tolérance d'erreur moyenne de 7 %, sont représentées en transparence grisée ; les fréquences inutilisées pour cette reconstruction, car concernant une zone en dehors du cercle, sont rayées.

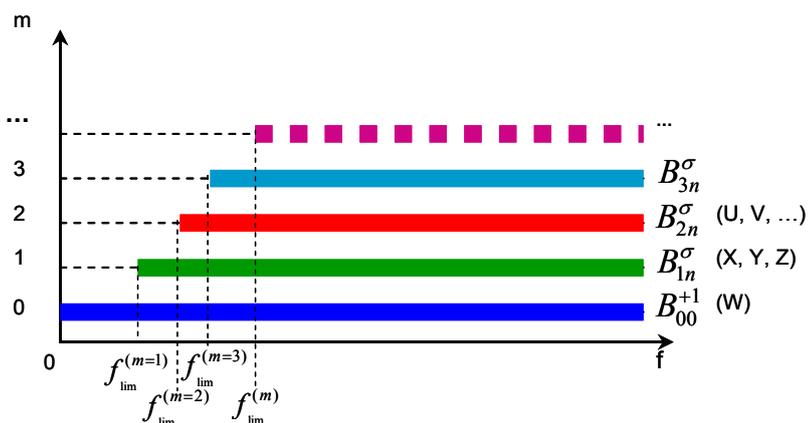


Figure 4.5 – Représentation schématique des bandes fréquentielles utiles à la description du champ acoustique dans une zone limitée de l'espace. La fréquence limite à partir de laquelle les signaux HOA sont supposés être utiles croît avec l'ordre m auquel ils appartiennent.

4.3.2 Application à la conception de filtres de champ proche

Dans le cas particulier de la reconstruction d'une source ponctuelle homogène, le concept de bande fréquentielle utile expliqué précédemment nous amène à conclure que si l'on vise une reconstruction du champ acoustique dans une zone limitée de l'espace caractérisée par un rayon r_{lim} , le respect de l'encodage théorique des composantes spatiales HOA dans un domaine basse fréquence où les problèmes d'amplification excessive surviennent n'est pas réellement utile. Par conséquent, il est possible de réduire cette amplification dès l'étape d'encodage spatial, par exemple en leur appliquant sur les signaux HOA des filtres passe-haut, ce qui revient à effectuer une troncature du champ acoustique en fonction de la fréquence.

Choix des fréquences de coupures : critère de focalisation d'énergie

Pour définir les fréquences limites $f_{\text{lim}}^{(m)}$ en dessous desquelles nous jugeons que les signaux HOA d'ordre m ne sont plus utiles à la reconstruction du champ acoustique dans la zone de restitution visée, nous introduisons un critère dit de *focalisation d'énergie*. Ce critère permet de garantir, lors du cas critique de la reproduction d'une source virtuelle intérieure au dispositif de haut-parleurs, que l'énergie restituée est focalisée sur l'emplacement de cette source virtuelle. Tout en assurant une reproduction cohérente, cette contrainte permet dans le même temps de limiter l'amplitude de la pression acoustique au-delà du rayon de la source virtuelle.

La Figure 4.6 illustre pour quatre fréquences distinctes l'amplitude absolue de la pression acoustique engendrée par une source ponctuelle homogène de rayon 1 m, et approximée par la troncature de la série Fourier-Bessel à différents ordres. Plus précisément, cette amplitude de la pression est représentée en fonction du rayon et dans la direction de la source, c'est-à-dire sur le demi axe passant par elle. Comme nous pouvons le remarquer, cette amplitude absolue croît jusqu'à atteindre un maximum d'autant plus haut que l'ordre est élevé, puis s'atténue rapidement. Pour chaque ordre de troncature, ce maximum coïncide avec l'emplacement de la source virtuelle à une fréquence d'autant plus haute que cet ordre est élevé. Par exemple, les fréquences 92 Hz, 212 Hz, 388 Hz et 604,5 Hz correspondent à un maximum de pression acoustique pour les ordres respectifs 1, 3, 6 et 10 de troncature comme nous pouvons le constater sur la Figure 4.6. Aux fréquences inférieures à celles-ci, les pics d'amplitude se déplacent au-delà de la position théorique de la source virtuelle. Ainsi, pour ces fréquences dites *limites* et aux ordres de troncature associés, la représentation du champ acoustique est optimale au sens de notre critère de focalisation d'énergie sur le lieu de la source virtuelle. Ces fréquences limites définissent ainsi la limite inférieure des bandes fréquentielles utiles pour les ordres concernés (cf. Figure 4.5). De façon plus générale, les fréquences $f_{\text{lim}}^{(m)}$ associées à chaque ordre m de troncature et à une distance de source virtuelle r_s se définissent selon notre critère simplement d'après les valeurs de $kr_{\text{lim}}^{(m)}$, indépendantes de cette distance et correspondant aux valeurs maximales de la pression acoustique tronquée à l'ordre m :

$$f_{\text{lim}}^{(m)} = \frac{kr_{\text{lim}}^{(m)} c}{2\pi r_s}. \quad (4.8)$$

Nous avons implicitement supposé jusqu'à présent la reconstruction d'une onde sphérique d'après l'expression adaptée de la série de Fourier en 3D. Cependant, dans le cas où la restitution se fait en deux dimensions à partir d'un dispositif circulaire de haut-parleurs, l'effet de proximité d'une source ponctuelle homogène peut être décrit approximativement en introduisant les fonctions de champ proche dans le développement en harmoniques cylindriques du champ acoustique donné à l'équation [DM04] :

$$p_M = S \left(J_0(kr) + 2 \sum_{m=1}^M i^m J_m(kr) F_m^{\tau_s}(kr_s) \cos m\gamma \right), \quad (4.9)$$

où les fonctions $J_m(kr)$ sont les fonctions de Bessel, les fonctions $F_m^{\tau_s}(kr_s)$ sont les fonctions de champ proche, et la constante de temps τ_s vaut r_s/c , c étant la vitesse de propagation du son dans l'air. Les valeurs $kr_{\text{lim}}^{(m)}$ sont alors légèrement différentes de celles obtenues dans le cas 3D, car elles sont définies d'après l'expression en fonction de l'ordre. Le Tableau 4.1 donne pour les cas 2D et 3D les valeurs de kr_{lim} associées aux dix premiers ordres de troncature, ainsi que les fréquences limites $f_{\text{lim}}^{(m)}$ correspondant à un rayon de source virtuelle $r_s = 1$ m.

Pour illustrer l'intérêt du critère de focalisation d'énergie, nous considérons la reconstruction d'une source virtuelle à l'intérieur d'un dispositif de restitution circulaire (2D) émettant un signal de fréquence 506 Hz correspondant, selon le critère de focalisation d'énergie et pour le cas 2D, approximativement à la fréquence limite inférieure de la bande fréquentielle utile associée aux signaux HOA d'ordre 9 (cf. Tableau 4.1). La Figure 4.7 représente dans le cas 2D le champ de pression acoustique engendrée par une source de rayon $r_s = 1$ m, tronqué aux ordres 8, 9, 10, 11, 12 et 15 conformément à l'équation (4.9), et sa reconstruction avec un dispositif de haut-parleurs circulaire de rayon 1,5 m. Cette figure montre effectivement que l'énergie est focalisée sur la source pour l'ordre 9 à cette fréquence 506 Hz, et que la reconstruction du front d'onde est dans ce cas correcte dans la zone de rayon inférieur à r_s . Au dessus de l'ordre 9, nous observons que l'amplitude du champ de pression acoustique s'amplifie rapidement dans la zone intermédiaire délimitée par le rayon de la source et celui des haut-parleurs, sans amélioration conséquente de la reconstruction acoustique dans la zone d'écoute. Le critère de focalisation d'énergie semble par conséquent être un bon compromis entre la qualité de l'effet de champ proche reproduit et l'énergie délivrée par les haut-parleurs pour y arriver.

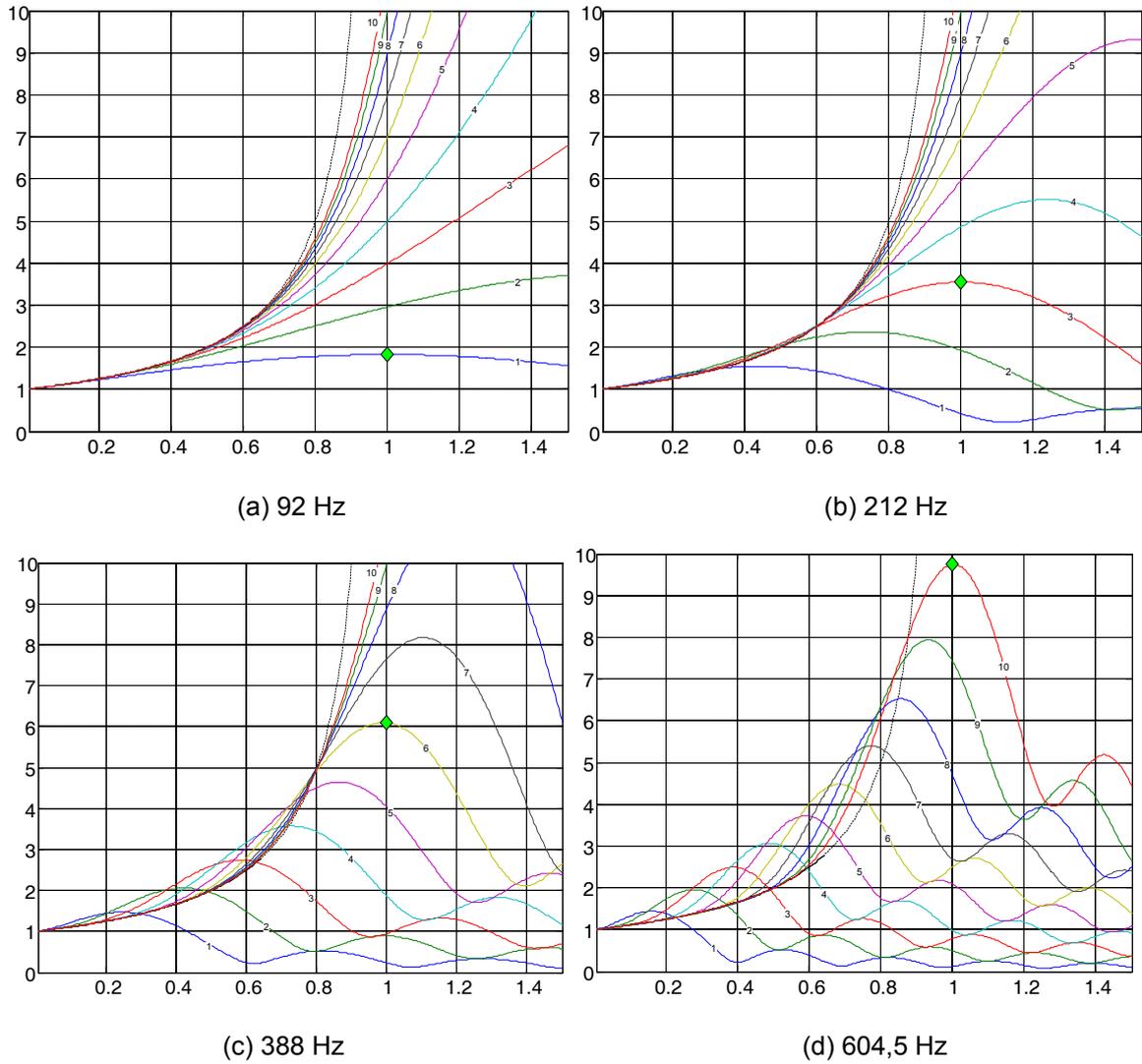


Figure 4.6 – Reconstruction de la pression acoustique tronquée aux ordres 1 à 10 engendrée par une source ponctuelle de rayon 1 m, et pour différentes fréquences. La courbe en pointillées représente l'amplitude de l'onde sphérique idéale, suivant une loi en $r_s / |\mathbf{r} - \mathbf{r}_s|$.

M	1	2	3	4	5	6	7	8	9	10
$kr_{\text{lim}}(3D)$	1.699	2.836	3.919	4.981	6.031	7.074	8.110	9.132	10.147	11.171
$f_{\text{lim}}(3D)$	91.937	153.464	212.068	269.535	326.354	382.793	438.854	494.157	549.081	604.493
$kr_{\text{lim}}(2D)$	1.405	2.355	3.417	4.429	5.393	6.404	7.409	8.357	9.347	10.345
$f_{\text{lim}}(2D)$	76.028	127.435	184.903	239.665	291.830	346.538	400.921	452.220	505.791	559.796

Tableau 4.1 – Fréquences limites associées aux amplitudes maximales du champ acoustique tronqué aux ordres 1 à 10 en 3D (développement en harmoniques sphériques) et 2D (développement en harmoniques cylindriques), pour une source virtuelle située à un rayon $r_s = 1$ m.

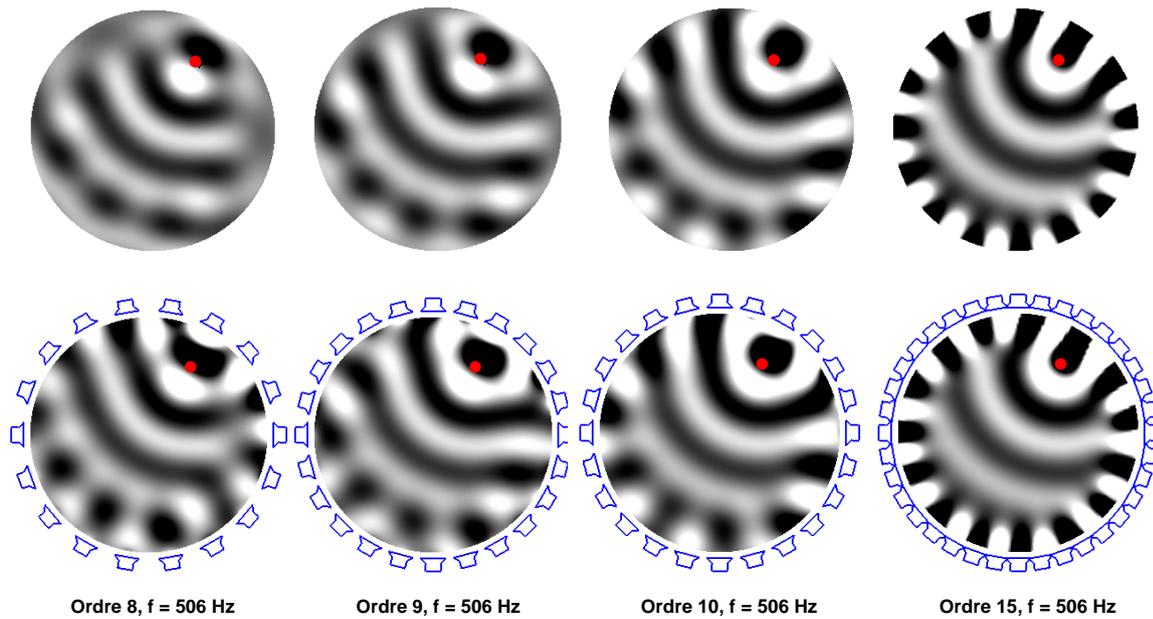


Figure 4.7 – Illustration 2D de la troncature aux ordres 8, 9, 10 et 15 du champ de pression acoustique engendrée par une source virtuelle de rayon $r_s = 1$ m et émettant un signal de fréquence 506 Hz (haut), et de sa reconstruction avec un dispositif de haut-parleurs circulaire de rayon 1,5 m (bas). Le rayon de la zone d’illustration vaut 1,5 m, le point rouge désigne la position de la source virtuelle. Notons que la fréquence 506 Hz correspond approximativement, selon le critère de focalisation d’énergie, à la fréquence limite de la bande fréquentielle utile associée aux signaux HOA d’ordre 9 (cf. Tableau 4.1).

Spécifications des filtres passe-haut

Comme le suggère la Figure 4.5, pour conserver la partie *utile* des signaux HOA en fonction de l’ordre d’appartenance, une solution consiste à leur appliquer des filtres passe-haut $H^{(m)}$, ce qui revient à tronquer la série de Fourier-Bessel en fonction de la fréquence. Pour être optimaux, ces filtres $H^{(m)}$ doivent répondre à un certain nombre de spécifications. Premièrement, leur pente d’atténuation doit compenser suffisamment la pente de $-m \times 6$ dB par octave des filtres de champ à partir des fréquences limites $f_{\text{lim}}^{(m)}$ en dessous desquelles les composantes d’ordre m ne sont plus utiles, selon le critère de focalisation d’énergie, à la reconstruction du champ de pression acoustique dans la zone de l’espace considérée. Dans l’idéal, la pente du filtre $H^{(m)}$ associé à l’ordre m doit également assurer une réjection suffisante à la fréquence limites $f_{\text{lim}}^{(m-1)}$ du filtre de champ proche d’ordre $m-1$. Cela implique que l’ordre du filtre $H^{(m)}$ doit être égal au minimum à $2m$ lorsqu’il est implanté sous la forme d’un filtre RII (Réponse Impulsionnelle Infinie). Enfin, pour reconstruire convenablement l’onde sphérique et sa courbure, la phase des filtres $H^{(m)}$ doit impérativement avoir une réponse nulle dans leurs bandes passantes respectives, ou encore une réponse linéaire traduisant un retard constant en fonction de la fréquence. Un exemple de spectre d’amplitude des filtres passe-haut $H^{(m)}$ est représenté à la Figure 4.8 pour les ordres 1 à 15, les losanges verts indiquant les fréquences limites $f_{\text{lim}}^{(m)}$ associées à chaque ordre de troncature (cas 2D). Le spectre d’amplitude résultant de la combinaison des filtres $H^{(m)}$ et des filtres de champ

proche est représenté à la Figure 4.9 pour les ordres 1 à 15. Les losanges verts indiquent encore les fréquences limites $f_{\text{lim}}^{(m)}$, qui correspondent ainsi au niveau d'amplification « suffisant » des signaux HOA d'ordre m associé. Comme nous pouvons le constater, ces niveaux restent relativement modérés (autour de 3 à 4 dB) et augmentent peu avec l'ordre m . Ils semblent vérifier d'autre part une loi affine, à partir de laquelle nous pourrions déduire approximativement les niveaux d'amplification maximaux utiles associés aux ordres supérieurs à 15. Les filtres résultants se substituent ainsi aux filtres de champ proche théoriques F_m^{rs} dans l'étape d'encodage spatial d'une onde sphérique (cf. Figure 2.3). Signalons enfin que la compensation du champ proche des haut-parleurs de restitution n'est pas prise en compte dans cet encodage, et doit par conséquent être réalisée avant le décodage des signaux HOA.

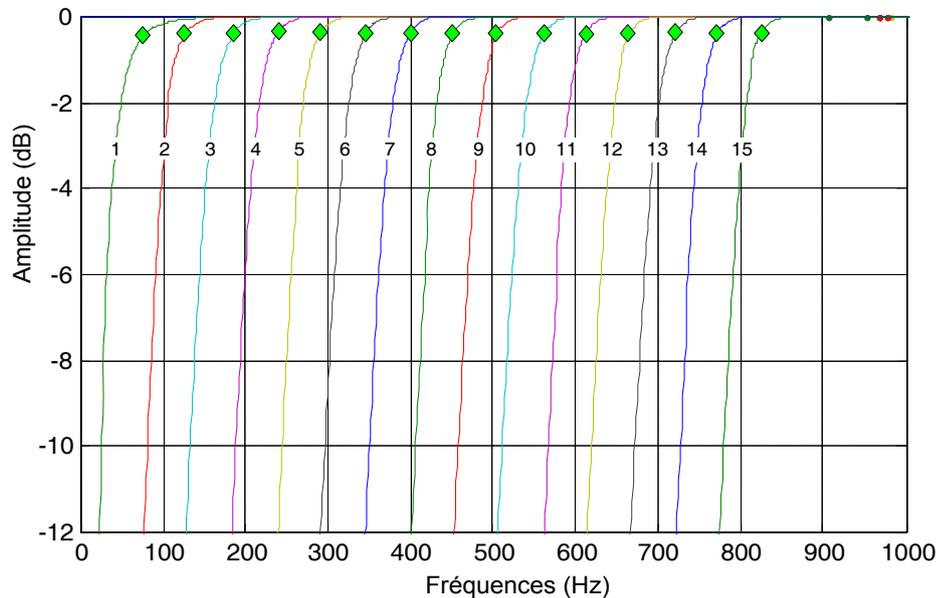


Figure 4.8 – Spectre d'amplitude des filtres passe-haut conçus pour limiter l'amplification excessive des basses fréquences induite par les filtres de champ proche pour les ordres 1 à 15. Les losanges verts indiquent les fréquences limites définies par le critère de focalisation d'énergie en fonction de l'ordre et reportées dans le Tableau 4.1.

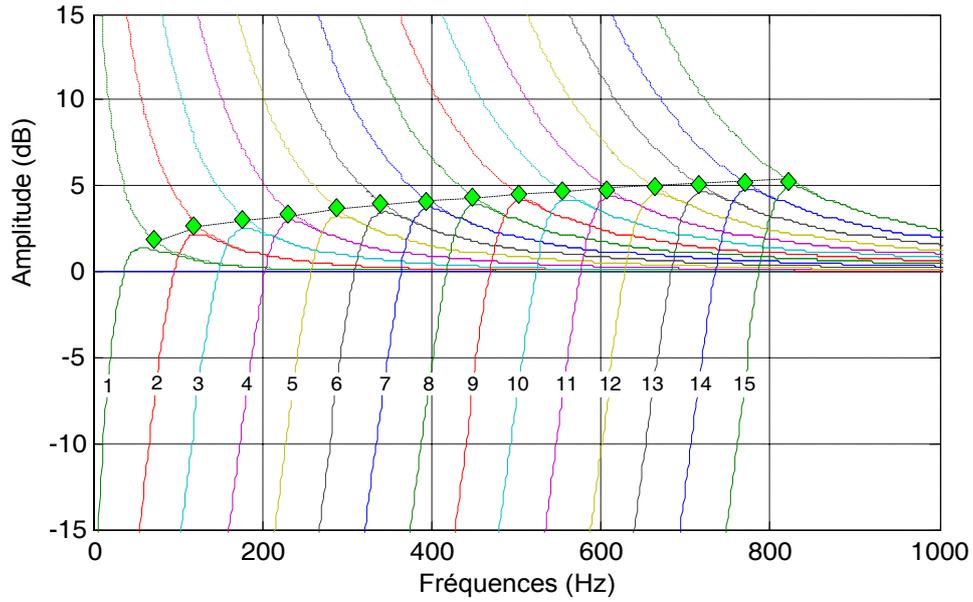


Figure 4.9 – Spectres d'amplitude de la combinaison des filtres de champ proche $F_m^{\tau_s}$ et des filtres passe-haut $H^{(m)}$ pour les ordres 1 à 15 et un rayon de source 1,5 m. Les fonctions de champ proche initiales sont représentées en lignes pointillées à titre de comparaison. Les losanges indiquent les fréquences limites à partir desquelles chaque ordre est supposé être « utile » à la reconstruction acoustique selon le critère de focalisation d'énergie.

Références bibliographiques du Chapitre 4

- [BR99] BRUNGART, D.S. et RABINOWITZ, W.M. *Auditory localization of nearby sources. Head-related transfer functions*. J. Acoust. Soc. Am., 1999, Vol. 106, n° 3, pt. 1, p. 1465-1479.
- [V97] VERHEIJEN, E.N.G. *Sound reproduction by Wave Field Synthesis*. Ph. D. Thesis, Delft: Delft University of Technology, 1997.
- [D03] DANIEL, J. *Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format*. AES 23rd International Conference, Helsingor, Danemark, 23-25 mai 2003.
- [SH01] SONTACCHI, A. et HÖLDRICH, R. *Further Investigations on 3D Sound Fields Using Distance Coding*. DAFX-01, Limerick, Irlande, 6-8 décembre 2001.
- [SH02] SONTACCHI, A. et HÖLDRICH, R. *Distance Coding in 3D Sound Field*. AES 21st Conference, St. Petersburg, Russie, 1er-3 juin 2002.
- [B97] BLAUERT, J. *Spatial Hearing. The Psychophysics of Human Sound Localization*. revised ed, Cambridge, Massachusetts: The MIT Press, 1997, 494 p.
- [C90] CANÉVET, G., *Chap. 3 Audition binaurale et localisation auditive. Aspects physiques et psychoacoustiques*, in *Psychoacoustique et perception auditive*. 1989. Paris: INSERM. p. 83-122.
- [BDR99] BRUNGART, D.S., DURLACH, N.I., et RABINOWITZ, W.M. *Auditory localization of nearby sources. II. Localization of a broadband source*. J. Acoust. Soc. Am., 1999, Vol. 106, n° 4, (Pt. 1), p. 1956-1968.
- [Bg98] BRUNGART, D.S. *Near-Field Auditory Localisation*. Ph. D. Thesis, Cambridge, MA: Massachusetts Institute of Technology, 1998.
- [G92c] GERZON, M.A. *The Design of Distance Panpots*. AES 92nd Conv., Vienna, 1992.
- [DM04] DANIEL, J. et MOREAU, S. *Further Study of Sound Field Coding with Higher Order Ambisonics*. AES 116th Convention, preprint 6017, Berlin, Allemagne, 8-11 mai 2004.

Chapitre 5

Enregistrement HOA de scène sonore naturelle avec un dispositif sphérique de microphones

Jusque très récemment, l'enregistrement d'un champ sonore réel au format HOA restait limité à une résolution spatiale d'ordre 1, suivant des principes théoriques remontant aux années 70 [CG77, F79a, F79b]. Nous nous intéressons dans ce chapitre à l'encodage HOA de champ acoustique naturel d'ordre supérieur à 1 au moyen d'un *dispositif sphérique de microphones*. Parallèlement à nos travaux, de tels dispositifs sphériques ont été proposés pour diverses applications : *beamforming* [M01a, ME02], analyse de champ acoustique [BR04, G01, G05, GRS02, GRS03, GRS04], reproduction sonore spatialisée [BA05a, BA05b]. D'autres stratégies ont au contraire été envisagées, comme par exemple un dispositif microphonique à rayons multiples [L03]. Nous rapportons dans un premier temps une étude théorique menée sur la conception des microphones sphériques destinés à l'enregistrement HOA. Dans un second temps, nous présentons la réalisation et la validation objective d'un prototype 3D d'ordre 4 constitué de 32 capsules réparties sur la surface d'une sphère rigide.

5.1 Introduction à l'enregistrement HOA

5.1.1 Microphones différentiels

Définition et propriétés

Un microphone différentiel d'ordre M est un microphone dont la sensibilité est proportionnelle à une combinaison linéaire des dérivées spatiales de la pression acoustique d'ordre 0 à M inclus en un point de l'espace [E00a]. Un tel microphone enregistre un signal du type :

$$s_M(\omega) = \sum_{m=0}^M c_m \frac{d^m}{dr^m} p(r, \omega) \Big|_{r=r_0} \quad (5.1)$$

où r_0 désigne la position du microphone orienté selon la direction r , et les coefficients c_m sont des facteurs réels qui pondèrent les différentes contributions de chaque ordre au signal $s_M(\omega)$.

Pour étudier les principales propriétés d'un microphone différentiel, nous observons son comportement en hypothèse d'onde plane. Ce cas particulier est essentiel car nous supposons par la suite que vues d'un dispositif microphonique relativement petit devant la longueur d'onde, les ondes sonores sont rapidement assimilables à des ondes planes au fur et à mesure que l'on s'en éloigne. Rappelons tout d'abord que dans une zone de l'espace ne comprenant pas de source sonore, une onde plane dont la direction d'incidence est définie par le vecteur de norme unitaire $\mathbf{u}_s = (\theta_s, \delta_s)$ engendre un champ de pression acoustique que l'on peut exprimer ainsi (dépendance en temps $e^{-i\omega t}$) :

$$p(k, r, \gamma) = p_0 e^{ikr \cos \gamma}, \quad (5.2)$$

γ étant l'angle entre la position du microphone et l'incidence de l'onde plane, et p_0 la pression acoustique mesurée à l'origine. En insérant cette expression dans l'équation (5.1), nous obtenons finalement :

$$s_M(r, \omega) = \sum_{m=0}^M c_m p_0 (ik \cos \gamma)^m \quad (5.3)$$

s_M dépend à la fois de l'angle γ et de la fréquence. La dépendance angulaire est entièrement contenue dans la fonction $\sum_{m=0}^M c_m (\cos \gamma)^m$ qui oscille autour de zéro et qui définit la *directivité* du microphone. Ce microphone enregistre alors, en fonction de sa directivité (de son orientation et de la direction d'incidence de l'onde plane), un signal dont l'amplitude passe par des maxima et des minima en s'annulant régulièrement. Le microphone différentiel le plus élémentaire, d'ordre 0, enregistre un signal proportionnel à la pression acoustique. Comme la pression en un point ne dépend pas de la direction de provenance de l'onde, il capte de façon identique dans toutes les directions de l'espace ; sa directivité est dite *omnidirectionnelle*. Un microphone différentiel d'ordre 1 mesure à la fois la pression et sa dérivée première. En fonction de la valeur des coefficients c_0 et c_1 qui leur sont respectivement associés, la directivité du microphone varie. Le même raisonnement s'applique aux microphones différentiels d'ordres 2 et supérieurs et d'une façon générale, plus l'ordre est élevé, plus la directivité du microphone peut être étroite dans sa

direction d'orientation. Enfin, en ce qui concerne la dépendance en fréquence du signal s_M dans l'expression (5.3), nous remarquons qu'un microphone différentiel d'ordre M se comporte à la manière d'un filtre passe-haut dont la pente d'atténuation est de $6 \times m$ dB par octave.

Construction d'un microphone différentiel

En pratique, un microphone différentiel mesure des différences finies de pression acoustique entre différents points très rapprochés de l'espace. Il n'enregistre par conséquent qu'une approximation du différentiel de pression. En effet, la dérivée $m^{\text{ième}}$, notée $f^{(m)}(x)$, d'une fonction $f(x)$ m fois dérivable par rapport à sa variable x est définie par :

$$f^{(m)}(x) = \lim_{h \rightarrow 0} \frac{f^{(m-1)}(x-h) - f^{(m-1)}(x+h)}{2h}, \quad m > 0. \quad (5.4)$$

La dérivée d'ordre m de la pression acoustique p en un point peut donc être approchée par la différence finie de sa dérivée d'ordre $m - 1$ entre deux points de l'espace séparés par une distance finie d suffisamment petite comparée à la longueur d'onde :

$$p^{(n)}(r, k) \simeq \frac{p^{(n-1)}(r + d/2, k) - p^{(n-1)}(r - d/2, k)}{d} \quad (5.5)$$

La réalisation d'un microphone différentiel d'ordre 0 ne pose pas de problème particulier puisqu'il mesure simplement la pression sonore au point de référence. En ce qui concerne la captation d'un différentiel de pression d'ordre 1, nous pouvons considérer la différence de pression captée par deux microphones de pression (ordre 0) très proches, séparés par une distance d suffisamment petite et orientés selon r :

$$\left. \frac{dp}{dr} \right|_{r=0} \simeq \frac{p(k, r - d/2) - p(k, r + d/2)}{d}. \quad (5.6)$$

En réalité, ce type de microphone est souvent réalisé plus simplement à partir d'un capteur de vitesse particulaire, proportionnel à la dérivée première de la pression sonore et constitué d'une membrane ouverte sur les deux faces. La dérivée seconde de la pression acoustique en un point s'obtient quant à elle à partir de deux microphones différentiels d'ordre 1 orientés de façon opposée :

$$\left. \frac{d^2 p}{dr^2} \right|_{r=0} \simeq \frac{p'(k, r - d/2) - p'(k, r + d/2)}{d} \quad (5.7)$$

Enfin, de façon générale, la dérivée $m^{\text{ième}}$ de la pression acoustique en un point peut être estimée à partir d'une combinaison de microphones différentiels d'ordre inférieur et dont l'ordre cumulé est égal à m . Signalons cependant que les microphones différentiels d'ordre supérieur à trois restent actuellement impossibles à réaliser à cause de l'amplification extrême qu'ils nécessitent en basse fréquence et qui augmente de plus les erreurs de calibration, de bruit des microphones, etc. [E00a]

5.1.2 Liens entre les microphones différentiels et les signaux HOA

Représentation 3D du champ sonore par la mesure des différentiels successifs de pression

Dans une zone de l'espace 3D ne comprenant pas de source, la pression acoustique varie en fonction d'un déplacement dans l'espace, c'est-à-dire en fonction des trois variables x , y et z du repère cartésien. Un déplacement induit une variation de la pression acoustique que l'on peut quantifier grâce à ses dérivées successives au point initial (x_0, y_0, z_0) , ou encore grâce à son développement de Taylor [DK99] :

$$p(x, y, z, \omega) = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} p_{l,m,n}(\omega) (x - x_0)^l (y - y_0)^m (z - z_0)^n \quad (5.8)$$

avec

$$p_{l,m,n}(\omega) = \frac{\partial^l}{\partial x^l} \frac{\partial^m}{\partial y^m} \frac{\partial^n}{\partial z^n} p(x, y, z, \omega) \Big|_{x=x_0, y=y_0, z=z_0} . \quad (5.9)$$

Le développement de Taylor est une représentation du champ de pression sonore alternative au développement en série d'harmoniques sphériques (cf. Chapitre 2). Comme pour ce dernier, le champ sonore se réduit à un ensemble de coefficients, ici les dérivées partielles $p_{l,m,n}(\omega)$ successives de la pression acoustique en un point de l'espace, qui traduisent la participation de chacun des polynômes de la série. L'ordre du développement correspond dans ce cas à la somme des puissances l , m et n , respectivement associées à x , y et z . Si l'on tronque la série de Taylor à un ordre fini, nous obtenons une approximation du champ sonore qui se dégrade lorsque l'on s'éloigne du point de référence.

Les signaux $p_{l,m,n}(\omega)$ s'obtiennent naturellement à partir de microphones différentiels. Notons que leur expression fait intervenir des dérivées partielles puisque la pression sonore s'écrit ici en trois dimensions et non suivant une seule direction r comme dans l'équation (5.1). Dans le cas d'une onde plane, l'expression générale des coefficients $p_{l,m,n}(\omega)$ associés au développement de Taylor à l'origine du repère devient :

$$p_{l,m,n}(\theta_s, \delta_s, \omega) = p_0 (ik)^{l+m+n} (\cos \theta_s \cos \delta_s)^l (\sin \theta_s \cos \delta_s)^m (\sin \delta_s)^n . \quad (5.10)$$

Nous retrouvons les propriétés directionnelles et fréquentielles propres aux microphones différentiels que nous avons déjà évoquées, et les mêmes difficultés de mise en œuvre au-delà du troisième ordre. À titre d'exemple, les coefficients $p_{l,m,n}(\omega)$ d'ordres 0 à 2 sont donnés explicitement au Tableau 5.1.

$$\begin{array}{l}
\text{Ordre 0} \\
\text{Ordre 1} \\
\text{Ordre 2}
\end{array}
\left\{
\begin{array}{l}
p_{0,0,0} = p_0 \\
p_{1,0,0} = \frac{\partial p}{\partial x} \Big|_{r=0} = p_0 i k \cos \theta_s \cos \delta_s \\
p_{0,1,0} = \frac{\partial p}{\partial y} \Big|_{r=0} = p_0 i k \sin \theta_s \cos \delta_s \\
p_{0,0,1} = \frac{\partial p}{\partial z} \Big|_{r=0} = p_0 i k \sin \delta_s \\
p_{1,1,0} = \frac{\partial^2 p}{\partial x \partial y} \Big|_{r=0} = -p_0 \frac{k^2}{2} \sin(2\theta) \cos^2 \delta \\
p_{1,0,1} = \frac{\partial^2 p}{\partial x \partial z} \Big|_{r=0} = -p_0 \frac{k^2}{2} \cos \theta \sin(2\delta) \\
p_{0,1,1} = \frac{\partial^2 p}{\partial y \partial z} \Big|_{r=0} = -p_0 \frac{k^2}{2} \sin \theta \sin(2\delta) \\
p_{2,0,0} = \frac{\partial^2 p}{\partial x^2} \Big|_{r=0} = -p_0 k^2 \cos^2 \theta \cos^2 \delta \\
p_{0,2,0} = \frac{\partial^2 p}{\partial y^2} \Big|_{r=0} = -p_0 k^2 \sin^2 \theta \cos^2 \delta \\
p_{0,0,2} = \frac{\partial^2 p}{\partial z^2} \Big|_{r=0} = -p_0 k^2 \sin^2 \delta
\end{array}
\right.$$

Tableau 5.1 – Expression des coefficients caractérisant le développement de Taylor d'une onde plane (ordres 0 à 2).

Lien avec les signaux HOA

Pour décrire le champ acoustique engendré par une onde plane incidente, le Tableau 5.1 révèle que jusqu'à l'ordre 1 inclus les coefficients du développement en série de Taylor $p_{0,0,0}$, $p_{1,0,0}$, $p_{0,1,0}$ et $p_{0,0,1}$ sont respectivement proportionnels aux signaux HOA B_{00}^1 , B_{11}^1 , B_{11}^{-1} et B_{10}^1 (cf. Chapitre 3) et possèdent la même dépendance angulaire. Il est par conséquent tout à fait possible en théorie d'effectuer un enregistrement HOA d'ordre 1 à partir de microphones différentiels : un microphone de pression (ordre 0) et trois microphones à gradient de pression (ordre 1) orientés selon les trois axes du repère cartésien, placés tous les quatre en un même point. Au-delà de l'ordre 1 cependant, la correspondance terme à terme entre les signaux HOA et les coefficients du développement de Taylor n'est plus possible. Pour l'ordre 2 par exemple, ces coefficients sont au nombre de six contre cinq pour les signaux HOA. Seuls les trois coefficients $p_{1,1,0}$, $p_{1,0,1}$ et $p_{0,1,1}$ sont directement proportionnels à trois signaux HOA, respectivement B_{22}^{-1} , B_{21}^1 et B_{21}^{-1} . Nous remarquons d'autre part que $p_{0,0,2}$ est proportionnel à B_{20}^1 si on lui retranche une composante d'ordre 0.

Bien qu'il n'y ait pas entre eux de relation directe au-delà de l'ordre 1, le développement du champ de pression acoustique en série d'harmoniques sphériques et celui en série de Taylor sont étroitement liés. Ce dernier, plus général cependant, ne tient pas compte des propriétés physiques de propagation des ondes de pression et contient de ce fait un certain nombre de redondances. Ces redondances se traduisent par un nombre plus important de coefficients nécessaires à la description du champ sonore, et ceci d'autant plus que l'ordre est élevé (cf. [DK99], Tableau 1). Si l'on considère a priori le développement en série de Taylor comme une solution de l'équation d'onde, ce dernier est équivalent au développement en série d'harmoniques sphériques [DK99].

S'il est possible, comme le suggère le développement de Taylor, d'enregistrer un champ sonore sur une zone étendue de l'espace en mesurant en un point les dérivées successives de la pression sonore en trois dimensions avec des microphones différentiels, ceci nécessite d'acquérir, pour une résolution équivalente, un nombre plus important de signaux que pour un enregistrement HOA. Il est de plus techniquement difficile de rassembler autour d'un même point de l'espace un ensemble important de capteurs très proches les uns des autres. Nous proposons dans la suite de ce chapitre une méthode alternative de conception des microphones HOA. Nous nous basons pour cela sur une interprétation de la série de Fourier-Bessel qui suggère de déduire les signaux HOA de la mesure d'informations du champ acoustique sur une surface sphérique.

5.2 Principes théoriques du microphone HOA sphérique

La captation directe des signaux HOA caractérisant un champ sonore réel ne peut être réalisée en pratique pour des raisons évidentes : inexistence de microphones directifs d'ordre supérieur à 1, impossibilité de placer plusieurs microphones en un même point de l'espace. Une méthode alternative, indirecte, consiste à estimer les signaux HOA à partir d'un ensemble de données acoustiques mesurées par des microphones usuels, le champ de pression acoustique par exemple, et d'un modèle mathématique associé, connu, qui permet de lier ces données aux signaux HOA.

5.2.1 Modélisation mathématique d'un microphone HOA sphérique continu

Dispositif basé sur la transformée de Fourier sphérique du champ acoustique

Dans un volume de l'espace exempt de source sonore, le champ de pression acoustique peut se modéliser sous la forme d'une série de Fourier Bessel (cf. Chapitre 2), c'est-à-dire un développement en série d'harmoniques sphériques en un point qui définit l'origine de notre repère spatial. L'expression analytique de la transformée de Fourier sphérique (Équation (2.22)) suggère une méthode théorique d'estimation des signaux HOA à partir de la mesure du champ de pression acoustique sur une surface sphérique continue. Cette méthode, illustrée schématiquement à la Figure 5.1, consiste à projeter la pression acoustique mesurée par des microphones de pression sur la base des harmoniques sphériques, puis d'égaliser les signaux ainsi obtenus suivant l'ordre au moyen des filtres EQ_m :

$$B_{mn}^{\sigma} = EQ_m(kR) \int_{\theta=0}^{2\pi} \int_{\delta=-\pi/2}^{\pi/2} p(kR, \theta, \delta) Y_{mn}^{\sigma}(\theta, \delta) \cos \delta d\delta d\theta \quad (5.11)$$

avec

$$EQ_m(kR) = \frac{1}{i^m j_m(kR)}, \quad (5.12)$$

où R est le rayon de la surface sphérique. Cette méthode n'est cependant valable que si $j_m(kR) \neq 0$. Or, les fonctions de Bessel sphériques $j_m(kr)$, illustrées pour les premiers ordres à la Figure (2.4), oscillent autour de zéro et s'annulent donc régulièrement. L'équation (5.12) reste par conséquent indéterminée pour un certain nombre de fréquences qui varient avec l'ordre des signaux HOA, et qui correspondent aux zéros des fonctions de Bessel sphériques. D'autre part, lorsque ces fonctions ne sont pas nulles mais de valeur très petite, les filtres EQ_m induisent une amplification extrême, impossible à réaliser en pratique. L'estimation des signaux HOA B_{mn}^{σ} sur une bande fréquentielle large comme la bande des fréquences audibles paraît d'ores et déjà problématique avec ce type de microphone sphérique.

L'énorme effort d'estimation traduit par EQ_m reflète un manque d'information utile dans les signaux captés, au lieu d'enregistrement. Le microphone est alors particulièrement sensible aux imperfections du système de mesure (bruit de fond et réponse non idéale des capteurs, perturbation acoustique engendrée par la structure du dispositif, etc.) qui introduisent des erreurs dans le modèle mathématique et qui, amplifiées à l'extrême, rend inefficace l'estimation des signaux HOA. Les difficultés d'estimation rencontrées s'expliquent en partie par un manque de directivité du microphone sphérique, les différentes composantes harmoniques étant déterminées uniquement à partir des différences de phase entre les points de mesures. En fonction du rayon R du microphone, sa géométrie sphérique induit un certain nombre de fréquences indéterminées dites *singulières*, propres à chaque ordre, et qui correspondent aux zéros des fonctions de Bessel sphériques. Par exemple, le signal B_{00}^1 ne peut être déterminé pour les fréquences $f_a = ac/2R$ où a est un entier naturel et c la vitesse du son dans l'air, soit pour un nombre d'onde multiple de $k = 2R$. Une première solution pour éviter le problème des fréquences indéterminées est d'effectuer la mesure sur plusieurs rayons distincts [G01, L03]. Une seconde méthode, plus facile à mettre en œuvre et particulièrement efficace, consiste à utiliser un microphone sphérique directif.

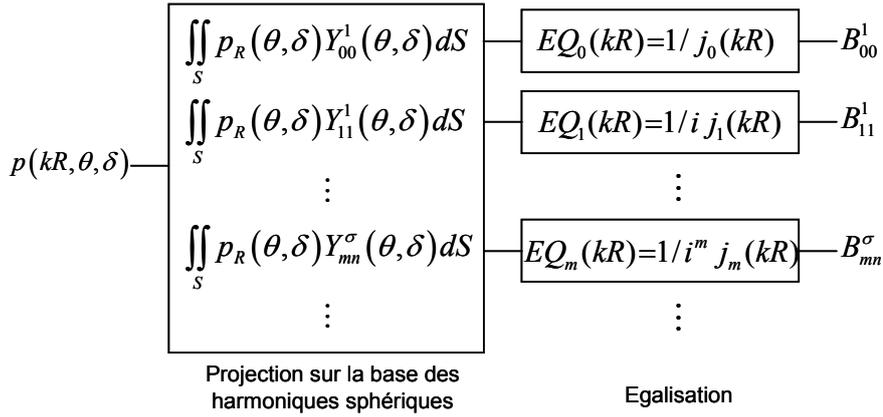


Figure 5.1 – Description schématique du fonctionnement d'un microphone HOA sphérique continu.

Utilisation d'un microphone sphérique directif

L'emploi d'un microphone sphérique directif permet d'améliorer l'estimation des signaux HOA. Un microphone directif standard d'ordre 1 orienté radialement vers l'extérieur, enregistre en un point un signal proportionnel à une combinaison de la pression acoustique et de sa dérivée première :

$$s(kR, \theta, \delta) = \alpha p(kR, \theta, \delta) + (1 - \alpha) \rho c \frac{\partial p(kR, \theta, \delta)}{\partial kR}, \quad (5.13)$$

où ρc caractérise l'impédance de l'air. Typiquement, nous choisissons $0 \leq \alpha \leq 1$ pour normaliser la réponse et avoir une valeur maximale de 1 pour l'incidence ($\theta = 0^\circ, \delta = 0^\circ$). En introduisant dans l'équation (5.13) l'expression du champ de pression acoustique sous la forme d'une série de Bessel (cf. Équation (2.19)), nous obtenons finalement :

$$s(kR, \theta, \delta) = \sum_{m=0}^{\infty} W_m(kR) \sum_{0 \leq n \leq m, \sigma = \pm 1} B_{mn}^\sigma Y_{mn}^\sigma(\theta, \delta) \quad (5.14)$$

avec

$$W_m(kR) = i^m (\alpha j_m(kR) - (1 - \alpha) j_m'(kR)), \quad (5.15)$$

où $j_m'(kr)$ est la dérivée des fonctions de Bessel sphériques $j_m(kr)$ par rapport à la variable kr . Grâce à la propriété d'orthonormalité des harmoniques sphériques, les signaux HOA s'obtiennent ainsi :

$$B_{mn}^\sigma = \frac{1}{W_m(kR)} \int_{\theta=0}^{2\pi} \int_{\delta=-\frac{\pi}{2}}^{\frac{\pi}{2}} p(R, \theta, \delta) Y_{mn}^\sigma(\theta, \delta) \sin \theta d\theta d\delta, \quad (5.16)$$

avec $W_m(kR) \neq 0$. L'équation (5.16) suggère une méthode d'estimation des signaux HOA similaire à celle illustrée à la Figure 5.1, mais avec $EQ_m(kR) = 1/W_m(kR)$. Le spectre d'amplitude des fonctions $W_m(kR)$ est illustré à la Figure 5.2 pour les huit premiers ordres et le cas particulier des microphones à directivité cardioïde, c'est-à-dire avec $\alpha = 1/2$. Nous constatons qu'elle ne

s'annule qu'en $kr = 0$ (et pour les modes $m \geq 1$). Cela signifie que l'équation (5.16), contrairement à l'équation (5.11), est définie sur toute la bande des fréquences audibles (si $r > 0$). Il est de plus possible d'envisager l'emploi dans l'équation (5.14) de coefficients α qui dépendent de la fréquence pour traduire plus précisément le comportement des microphones directifs réels qui tendent à devenir omnidirectifs en basse fréquence.

Une autre approche pour réaliser une mesure directive est l'emploi d'une structure diffractante, par exemple une sphère rigide¹⁵, pour la réalisation du microphone. En effet, il se forme autour d'une telle structure des zones d'ombre et de fort rayonnement acoustique en fonction de la position des sources dans l'espace. Une sphère rigide a déjà été utilisée par Gary W. Elko pour la conception d'un microphone différentiel d'ordre 1 dont l'orientation est ajustable en trois dimensions [ENP97], et proposée dans d'autres études parallèles à nos travaux pour l'enregistrement du champ sonore en 3D [MA03, LD05a] ou la construction de microphone à directivité supérieure à 1 [LD05a]. Les effets de diffraction du son par une sphère rigide sont bien connus et facilement modélisables. Lorsque l'on introduit une sphère rigide de rayon R dans un champ acoustique, la pression sonore résultante peut s'écrire comme la somme du champ de pression en l'absence de la sphère et du champ de pression diffracté par celle-ci [MI68, B98] :

$$p_a(r, \theta, \delta) = \sum_{m=0}^{\infty} \left(j_m(kr) - \frac{j'_m(kR)}{h'_m(kR)} h_m^-(kr) \right) \sum_{n=-m}^m \sum_{\sigma=\pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \delta), \quad (5.17)$$

où h_m^- est la fonction de Hankel divergente, et h_m^-' sa dérivée première par rapport à la variable kr . Un cas particulier, très intéressant car permettant d'obtenir une directivité maximale, concerne l'enregistrement de la pression acoustique à la surface de la sphère rigide ($r = R$). L'expression du champ de pression acoustique devient [MI68] :

$$p(kR, \theta, \delta) = \sum_{m=0}^{\infty} W_m(kR) \sum_{n=-m}^m \sum_{\sigma=\pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \delta) \quad (5.18)$$

avec

$$W_m(kR) = \left(\frac{i^{m-1}}{(kR)^2 h_m^-(kR)} \right). \quad (5.19)$$

Les signaux HOA peuvent une fois de plus être déduits de l'équation (5.18) en projetant la pression mesurée sur la base des harmoniques sphériques, puis en égalisant les signaux résultants :

$$B_{mn}^{\sigma} = EQ_m(kR) \int_{\theta=0}^{2\pi} \int_{\delta=-\pi/2}^{\pi/2} p_R(\theta, \delta) Y_{mn}^{\sigma}(\theta, \delta) \cos \delta d\delta d\theta. \quad (5.20)$$

avec

¹⁵ Une surface est dite *acoustiquement rigide* si son impédance acoustique est supérieure à l'impédance caractéristique du milieu qui l'entoure, l'air en l'occurrence (on dit impédance acoustique infinie).

$$EQ_m(kR) = i^{-m+1} (ka)^2 h_m^{-1}(ka) \quad (5.21)$$

En adaptant l'expression de l'égalisation EQ_m , la Figure 5.1 illustre les principes de bases d'un microphone HOA sphérique continu basé sur l'équation (5.20), c'est-à-dire pouvant estimer les signaux HOA à partir de la mesure du champ de pression acoustique à la surface d'une sphère rigide de rayon R . La fonction $|W_m(kR)|$, représentée à la Figure 5.2, ne s'annule qu'en $ka = 0$ (et pour les ordres $m > 0$). L'équation (5.20) est par conséquent définie sur l'ensemble du spectre des fréquences audibles (si $r > 0$). La mesure du champ de pression à la surface d'une sphère rigide par des microphones omnidirectionnels¹⁶ semble être également une bonne solution pour réaliser un enregistrement HOA. Cette méthode, bien qu'en apparence un peu moins robuste pour l'ordre 1 en basse fréquence que l'utilisation de microphones à directivité cardioïde (cf. Figure 5.2), est plus simple à mettre en œuvre et la perturbation du champ acoustique engendrée par la structure du dispositif mieux contrôlée.

Même si rendre la mesure directive permet d'envisager l'estimation des signaux HOA sur une bande fréquentielle large, l'égalisation spectrale EQ_m qui compense la directivité intrinsèque de chaque capteur et qui inclut, le cas échéant, la perturbation créée par la structure diffractante, reste problématique pour un dispositif de petite taille en basse fréquence puisque son amplitude diverge (cf. Figure 5.3). Ce phénomène qui traduit les difficultés d'estimation rencontrées par le microphone s'explique facilement. Comme nous l'avons précisé au Chapitre 2, à ordre égal la description du champ acoustique concerne une zone de l'espace plus éloignée du point de référence en basse fréquence qu'en haute fréquence. Par conséquent, un microphone sphérique dont le centre correspond précisément à ce point de référence doit fournir un effort d'estimation plus important en basse fréquence, c'est-à-dire pour les fréquences où les informations recherchées sont les plus faiblement présents, qui se traduit par une forte amplification. Un tel microphone est donc très sensible aux erreurs du système d'enregistrement (bruit des capteurs, erreurs de position, etc.), ce qui rend l'estimation des signaux HOA non consistante. Pour garantir la robustesse de l'estimation des signaux HOA, l'amplitude des filtres égaliseurs EQ_m doit être limitée. Comme nous le verrons par la suite, cette limitation, nécessaire techniquement, est de plus tout à fait sensée d'un point de vue physique si l'on vise une zone de restitution dont la taille reste constante en fonction de la fréquence. Nous utiliserons pour cela une méthode de *régularisation* des filtres EQ_m , mais devons auparavant transposer dans le domaine discret le principe du microphone HOA sphérique continu.

¹⁶ Notons qu'il est inutile de placer des microphones directifs à la surface d'une sphère rigide puisque la vitesse particulaire y est théoriquement nulle.

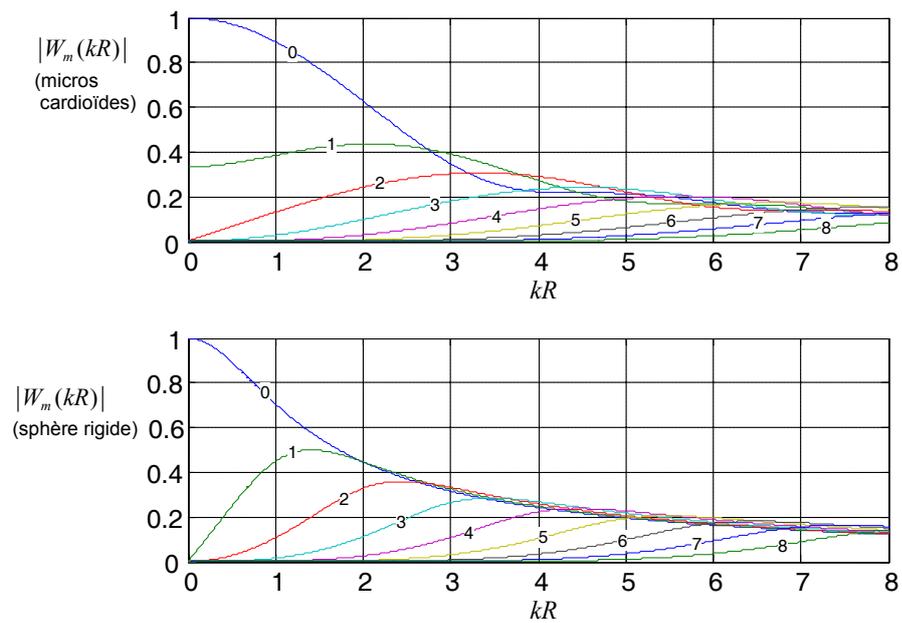


Figure 5.2 – Illustration du spectre d'amplitude des filtres d'égalisation spectrale W_m en fonction de la directivité du dispositif de captation : microphones cardioïde montés sur une sphère acoustiquement transparente (haut) et capteurs de pression à la surface d'une sphère rigide.

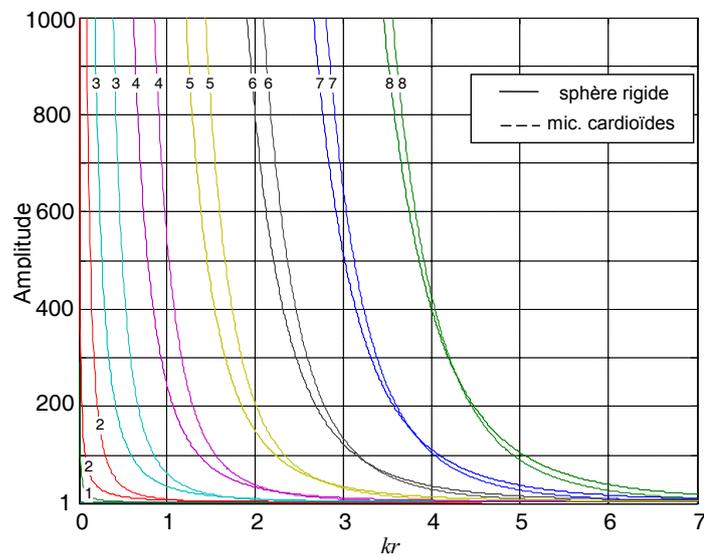


Figure 5.3 – Amplitude des filtres d'égalisation EQ_m en fonction de la directivité du microphone HOA sphérique.

5.2.2 Discrétisation du microphone HOA sphérique continu

Transposition du problème d'estimation des signaux HOA dans le domaine discret

En réalité, la réalisation d'un microphone HOA nécessite l'emploi d'un nombre fini de capteurs. Nous supposons pour cela, à la surface du microphone sphérique, un ensemble de Q points indicés repérés par leurs positions angulaires respectives (θ_q, δ_q) , avec $1 \leq q \leq Q$. Un microphone placé au $q^{\text{ième}}$ point capte un signal $s(kR, \theta_q, \delta_q)$ que nous savons modéliser par un développement en harmoniques sphériques :

$$S_q(kR, \theta_q, \delta_q) = \sum_{m=0}^{+\infty} W_m(kR) \sum_{n=-m}^m \sum_{\sigma=\pm 1} B_{mn}^\sigma Y_{mn}^\sigma(\theta_q, \delta_q), \quad (5.22)$$

avec

$$\begin{cases} W_m(kR) = \alpha j_m(kR) + i(\alpha - 1)j_m'(kR) & (\text{mic. cardioïdes}) \\ W_m(kR) = i^{-m+1}(kR)^2 h_m^{-'}(kR) & (\text{mic. omnidirectifs sur sphère rigide}) \end{cases}$$

Le nombre de capteurs utilisé dépend évidemment du nombre de signaux HOA que l'on souhaite estimer. Pour que le dispositif de captation soit en mesure de distinguer $K = (M + 1)^2$ signaux HOA distincts, l'échantillonnage spatial doit être constitué d'au moins autant de points de mesure, soit $Q \geq K = (M + 1)^2$. L'ensemble des mesures effectuées par les capteurs peut être décrit par un système de Q équations linéaires à K inconnues :

$$\mathbf{T} \cdot \mathbf{b} = \mathbf{s}, \quad (5.23)$$

avec

$$\mathbf{T} = \mathbf{Y} \cdot \text{diag}[W_m(kR)], \quad 0 \leq m \leq M, \quad (5.24)$$

$$\mathbf{Y} = \begin{pmatrix} Y_{00}^1(\theta_1, \delta_1) & Y_{11}^1(\theta_1, \delta_1) & \dots & Y_{M0}^1(\theta_1, \delta_1) \\ Y_{00}^1(\theta_2, \delta_2) & Y_{11}^1(\theta_2, \delta_2) & \dots & Y_{M0}^1(\theta_2, \delta_2) \\ \vdots & \vdots & \ddots & \vdots \\ Y_{00}^1(\theta_Q, \delta_Q) & Y_{11}^1(\theta_Q, \delta_Q) & \dots & Y_{M0}^1(\theta_Q, \delta_Q) \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} B_{00}^1 \\ B_{11}^1 \\ \vdots \\ B_{M0}^1 \end{pmatrix}, \quad \text{et } \mathbf{s} = \begin{pmatrix} S_1 \\ S_2 \\ \vdots \\ S_Q \end{pmatrix}.$$

Ainsi, les colonnes de la matrice \mathbf{T} correspondent aux harmoniques sphériques échantillonnées aux Q points de mesure, pondérées en fonction de l'ordre d'appartenance et de la fréquence considérée par $W_m(kR)$, \mathbf{b} contient les signaux HOA que nous cherchons à déterminer, et \mathbf{s} les échantillons de mesure prélevés par les capteurs aux Q positions angulaires sur la sphère.

Résolution « naïve » du problème d'estimation au sens des moindres carrés et instabilité de la solution

En pratique, les difficultés rencontrées pour choisir de façon optimale l'échantillonnage spatial au sens de la propriété d'orthonormalité des harmoniques sphériques imposent souvent un nombre de capteur plus important que celui des signaux HOA à estimer (cf. Section 2.3.1). Dans ce cas, le système (5.23) est *surdéterminé*, c'est-à-dire qu'il comporte plus d'équations que d'inconnues. En général, un tel système n'admet pas de solution exacte sauf si les colonnes de la matrice \mathbf{T} sont linéairement indépendantes, c'est-à-dire si aucune d'elles n'est une combinaison linéaire des

autres. Dans le cas contraire, la démarche la plus courante consiste à chercher une solution qui minimise le résidu quadratique $\|\mathbf{s} - \mathbf{T} \cdot \mathbf{b}\|_2^2$ ¹⁷ [GVL96]. La solution à ce problème de minimisation, notée \mathbf{b}_{MC} , satisfait un nouveau système d'équations linéaires dites *d'équations normales* [GVL96] :

$$\mathbf{T}^* \cdot \mathbf{T} \cdot \mathbf{b}_{MC} = \mathbf{T}^* \cdot \mathbf{s}, \quad (5.25)$$

où \mathbf{T}^* est la matrice transposée conjuguée complexe de la matrice \mathbf{T} . En supposant la matrice \mathbf{T} non singulière, la matrice $\mathbf{T}^* \mathbf{T}$ est inversible. La solution approchée \mathbf{b}_{MC} s'exprime alors :

$$\mathbf{b}_{MC} = (\mathbf{T}^* \cdot \mathbf{T})^{-1} \cdot \mathbf{T}^* \cdot \mathbf{s}. \quad (5.26)$$

En introduisant dans l'équation (5.26) l'expression factorisée (5.24) de la matrice \mathbf{T} , nous obtenons finalement :

$$\mathbf{b}_{MC} = \text{diag} \left[\frac{1}{W_m(kR)} \right] \mathbf{E} \mathbf{s}, \quad (5.27)$$

avec

$$\mathbf{E} = (\mathbf{Y}^t \mathbf{Y})^{-1} \mathbf{Y}^t. \quad (5.28)$$

La matrice réelle \mathbf{E} s'appelle la *matrice pseudo-inverse de Moore-Penrose* associée à \mathbf{Y} [GVL96]. La solution \mathbf{b}_{MC} d'estimation des signaux HOA jusqu'à un ordre M limité à partir d'un nombre fini Q de capteurs disposés sur une surface sphérique est illustrée schématiquement à la Figure 5.4. Notons que si la base des harmoniques sphériques échantillonnées préserve la propriété d'orthonormalité, nous avons la relation $\mathbf{E} = \mathbf{Y}^t$ et les signaux HOA correspondent exactement aux formulations discrètes des équations (5.16) et (5.20) :

$$B_{mn}^\sigma = \frac{1}{W_m(kR)} \frac{1}{Q} \sum_{q=1}^Q S_1 Y_{mn}^\sigma(\theta_q, \delta_q), \quad (5.29)$$

avec $0 \leq m \leq M$, $0 \leq n \leq m$, $\sigma = \pm 1$.

Le résidu théorique se calcule simplement par la formule $\mathbf{r} = \mathbf{T} \cdot \mathbf{b}_{MC} - \mathbf{s}$. Cependant, un faible résidu ne signifie pas forcément que \mathbf{b}_{MC} réalise une bonne approximation de la solution. En effet, la précision de cette solution dépend en premier lieu de sa sensibilité aux petites erreurs introduites dans le système (5.25), notamment au niveau de la matrice \mathbf{s} constituée par les mesures (bruit de fond des capteurs, erreurs de positionnement, etc.). L'égalisation $1/W_m(kR)$ dans l'équation (5.27) est un facteur d'instabilité de la solution \mathbf{b}_{MC} . En effet, quand cette égalisation correspond à une amplification excessive, les petites erreurs dans les matrices \mathbf{s} et \mathbf{T} sont susceptibles de se retrouver amplifiées à l'extrême dans la solution \mathbf{b}_{MC} . Un critère objectif de stabilité du système d'équations linéaires (5.25) est le nombre de conditionnement de la matrice $\mathbf{T}^* \mathbf{T}$ devant être inversée, défini par [GVL96] :

¹⁷ Rappelons que $\|\cdot\|_2$ désigne la norme euclidienne définie au Chapitre 3.

$$\kappa(\mathbf{T}^*\mathbf{T}) = \|\mathbf{T}^*\mathbf{T}\|_2 \cdot \|(\mathbf{T}^*\mathbf{T})^{-1}\|_2. \quad (5.30)$$

Nous avons par conséquent la relation $\kappa(\mathbf{T}^*\mathbf{T}) \geq 1$, et la matrice $\mathbf{T}^*\mathbf{T}$ est d'autant mieux conditionnée que $\kappa(\mathbf{T}^*\mathbf{T})$ est proche de 1. Le nombre de conditionnement est un indicateur de la précision que l'on peut attendre de la résolution d'un système linéaire et apporte une information sur le risque d'instabilité de ce système par rapport aux variations des données du problème. Il ne permet cependant pas de quantifier précisément cette instabilité. Son calcul peut être effectué à partir de sa décomposition en valeurs singulières [GVL96].

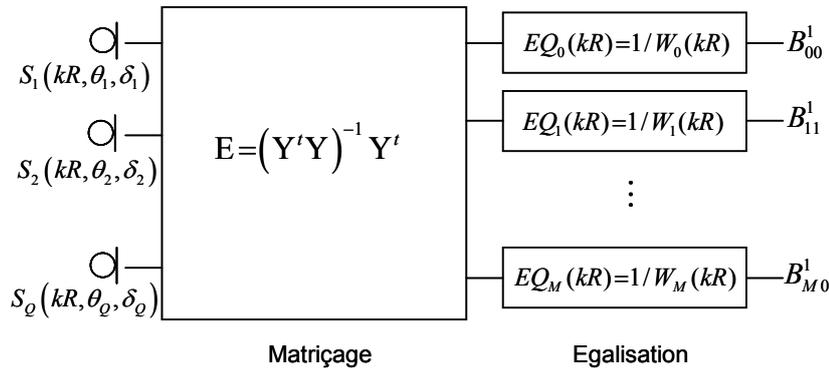


Figure 5.4 – Principe de l'encodage HOA de champ sonore naturel avec un dispositif microphonique sphérique discret.

Aliasing spatial

La discrétisation du microphone sphérique continu entraîne un phénomène d'*aliasing* à la fois dans la dimension temporelle des signaux HOA et dans leur dimension spatiale. La fréquence maximale détectable par le microphone discret dépend de la distance maximale entre les capteurs le constituant. Selon le critère de Shannon, l'intervalle d'échantillonnage Δd , c'est-à-dire la distance entre deux microphones adjacents sur la surface de la sphère, doit être inférieure à la moitié de la plus petite longueur d'onde λ_{\min} présente dans les signaux enregistrés. La fréquence d'*aliasing* f_{al} au-dessus de laquelle l'estimation devient erronée est défini par :

$$f_{\text{al}} = \frac{c}{2 \Delta d} = \frac{c}{2R\gamma}, \quad (5.31)$$

où γ est l'angle maximal entre deux capteurs. Pour augmenter la fréquence f_{al} , il est nécessaire de diminuer la distance entre les microphones en augmentant leur nombre ou en réduisant le rayon du dispositif.

L'*aliasing* spatial intervient également lorsque les fonctions harmoniques sphériques sont sous-échantillonnées. Le nombre Q de microphones utilisés et leurs positions à la surface de la sphère doivent être définis en fonction de l'ordre maximal M que l'on cherche à estimer. Comme nous l'avons déjà signaler, ce nombre doit être au minimum égal à celui des harmoniques sphériques

concernées, soit $Q \geq K = (M+1)^2$. D'autre part, l'échantillonnage des harmoniques sphériques par les Q microphones doit conserver la propriété d'orthonormalité des harmoniques sphériques continues. Dans le cas contraire, même si le nombre des microphones est supérieur à celui des signaux HOA associés aux harmoniques sphériques prises en compte, des dépendances entre les harmoniques discrètes apparaissent, ce qui se traduit par le repliement des unes sur les autres, compromettant la bonne estimation individuelle des signaux HOA (cf. Section 2.3.1). Enfin, une attention particulière doit être portée aux composantes d'ordre supérieur à M qui, si elles sont présentes dans les mesures, perturbent l'estimation des composantes d'ordre inférieur que l'on souhaite estimer. Le niveau de présence de ces modes est déterminé par les fonctions $W_m(kR)$ et dépend par conséquent de la fréquence et du rayon R . Il faut donc être conscient des problèmes d'*aliasing* spatial avant de choisir le rayon du dispositif microphonique.

5.3 Considérations pratiques pour la réalisation d'un microphone HOA sphérique

5.3.1 Échantillonnage de la surface d'une sphère

Échantillonnage indépendant de l'azimut θ et de l'élévation δ

Sur une ligne, soit dans un espace à une dimension, le théorème classique d'échantillonnage de Shannon établit qu'une fonction dont la transformée de Fourier est à support borné peut être exactement reconstruite à partir d'un ensemble fini d'échantillons prélevés uniformément avec une fréquence au moins deux fois supérieure à sa fréquence maximale. Driscoll et Healy [DH94] ont étendu ce théorème aux fonctions définies sur une sphère, et montré que si ces fonctions sont à bande spatiale limitée ($B_{mn}^\sigma = 0$ si $m > M$, cf. Chapitre 2), leurs transformées de Fourier sphérique peuvent être calculées exactement à partir d'une somme finie de valeurs échantillonnées de cette fonction. Pour cela, il suffit d'échantillonner régulièrement et indépendamment l'azimut et l'élévation de la façon suivante :

$$\begin{aligned} \theta_i &= \frac{\pi i}{2M+2}, & i &= 0, \dots, 2M+1 \\ \delta_j &= \frac{2\pi j}{2M+2} & j &= 0, \dots, 2M+1 \end{aligned}$$

La transformée de Fourier sphérique, c'est-à-dire l'ensemble des signaux HOA, s'obtient alors par la formule suivante ([DH94], Théorème 3) :

$$B_{mn}^\sigma = \frac{1}{W_m(kR)} \left(\frac{\sqrt{2\pi}}{2M+2} \right) \sum_{i=0}^{2M+1} \sum_{j=0}^{2M+1} \alpha_i s_R(\theta_i, \delta_j) Y_{mn}^\sigma(\theta_i, \delta_j). \quad (5.32)$$

Les coefficients de pondération α_i compensent la différence entre la densité des échantillons au voisinage des pôles et celle dans la zone équatoriale (même rôle que le facteur $\cos\theta$ qui apparaît lorsque l'on intègre une fonction sur la sphère). Leur expression est donnée dans [DH94]. L'équation (5.32) nous permet de calculer exactement les signaux HOA jusqu'à l'ordre M à partir

de mesures discrètes s_R . Cette méthode nécessite $4(M+1)^2$ capteurs répartis sur la surface sphérique du dispositif de captation. Ce nombre très important est peu optimal car la distribution des points de mesure est beaucoup plus dense au voisinage des pôles que dans la zone équatoriale.

Les polynômes de Legendre étant orthogonaux entre eux (cf. Chapitre 2), il est toutefois possible de diviser par deux le nombre de capteurs requis pour estimer exactement les signaux HOA jusqu'à l'ordre M en associant à l'intégration sur l'élévation δ une *quadrature de Gauss-Legendre* ([A85], Appendix 2). L'élévation δ est alors échantillonné non plus régulièrement, mais suivant les positions données par les racines des $M+1$ ordres du polynôme de Legendre, c'est-à-dire telles que $P_{M+1}(\sin \delta_j) = 0$, $j = 0, \dots, M$. La formule de quadrature (5.32) devient alors :

$$B_{mn}^\sigma = \frac{1}{W_m(kR)} \sum_{i=0}^{M+1} \sum_{j=0}^{2M+1} \alpha_i s_R(\theta_i, \delta_j) Y_{mn}^\sigma(\theta_i, \delta_j), \quad (5.33)$$

les coefficients α_i devant être calculés de façon appropriée [A85]. Ce choix garantit la propriété d'orthonormalité pour les harmoniques sphériques d'ordre inférieur ou égal $2M$.

Le nombre d'échantillons $2(M+1)^2$ reste cependant très élevé et sous optimal. Pour réduire le nombre de capteurs nécessaires en fonction de l'ordre, nous devons envisager un échantillonnage global à la surface de la sphère sans dissocier l'azimut de l'élévation. Il s'agit alors, dans la mesure du possible, de chercher des dispositions régulières ou quasi-régulières de points à la surface de la sphère qui permettent de préserver la propriété d'orthonormalité des harmoniques sphériques jusqu'à l'ordre maximal considéré. Ceci revient à chercher une formule de quadrature sur la sphère avec le moins de points possible, capable d'intégrer exactement chacune des harmoniques sphériques prises en compte.

Échantillonnage polyédrique régulier et semi-régulier

Contrairement au cercle, il n'est malheureusement pas possible de distribuer de façon exactement uniforme des points à la surface d'une sphère, excepté dans quelques cas particuliers, les *solides platoniciens* du nom du célèbre philosophe grec qui les étudia et qui sont des polyèdres convexes dont les faces, des polygones réguliers, sont identiques. Ces polygones réguliers, figures géométriques planes dont tous les côtés et les angles sont égaux, garantissent en effet l'équidistance entre les sommets des polyèdres réguliers. Seulement cinq polyèdres de ce type, c'est-à-dire cinq façons de diviser la surface de la sphère en aires équivalentes, existent [V01] : le tétraèdre, l'hexaèdre, l'octaèdre, l'icosaèdre et le dodécaèdre (cf. Tableau 5.3). Le nombre de faces, de sommets, et la distance entre deux sommets sont donnés pour chacun des polyèdres réguliers dans le Tableau 5.2. Nous pouvons facilement observer les liens entre les couples cube/octaèdre et icosaèdre/dodécaèdre : on inverse le nombre de faces et de sommets, où plus précisément, les centres des faces de l'un deviennent les sommets de l'autre. Ces polyèdres sont dits *duaux* ; le tétraèdre est, lui, *auto dual*. L'échantillonnage constitué par les sommets de ces polyèdres est par conséquent limité à 20 points correspondants aux sommets d'un dodécaèdre.

Contrairement à un échantillonnage satisfaisant le critère de Shannon, les sommets des polyèdres sont définis indépendamment des harmoniques sphériques. Nous devons par conséquent étudier pour chacun d'entre eux l'ordre maximal pour lequel la propriété d'orthonormalité de la base des harmoniques sphériques échantillonnées est vérifiée. Pour cela, nous définissons la matrice erreur d'orthonormalité comme suit :

$$\mathbf{D} = \mathbf{I}_K - \frac{1}{Q} \mathbf{Y}' \mathbf{Y}, \quad (5.34)$$

où \mathbf{I}_K est la matrice identité de dimensions $K \times K$, et les colonnes de la matrice \mathbf{Y} correspondent aux harmoniques sphériques échantillonnées par les Q points de mesure. Le Tableau 5.3 illustre ces matrices pour les cinq polyèdres réguliers. La matrice \mathbf{D} est une matrice carrée avec en abscisses comme en ordonnées les différentes composantes harmoniques sphériques classées par ordre croissant. Pour plus de clarté, ces composantes sont regroupées par ordre au moyen de lignes séparatrices verticales et horizontales. Un défaut d'orthonormalité entre deux composantes est représenté sous la forme d'un petit rectangle d'autant plus foncé que l'erreur est importante. Plus précisément, si ce petit rectangle se situe sur la diagonale de la matrice \mathbf{D} , il révèle une erreur de normalisation de la composante harmonique correspondante ; s'il se situe en dehors de cette diagonale, il traduit un défaut d'orthogonalité entre deux composantes distinctes. Il est ainsi aisé d'observer le phénomène d'*aliasing* spatial décrit plus haut, c'est-à-dire le repliement de certaines composantes harmoniques sur d'autres d'ordre inférieur. Remarquons par exemple que pour un ordre limité à 1, une configuration tétraédrique convient parfaitement : c'est la configuration géométrique choisie pour la réalisation du microphone *Soundfield* [CG77, F79a, F79b]. Nous pouvons également vérifier que l'icosaèdre, proposé par Gerzon [G73] pour la confection d'un microphone du deuxième ordre et étudié notamment par Philip Cotterell [C02], est bien adapté à un enregistrement d'ordre 2. En revanche, bien que le nombre de sommets soit supérieur au nombre de signaux HOA constituant les ordres 0 à 3, le dodécaèdre semble ne pas préserver la propriété d'orthonormalité des harmoniques sphériques au-delà de l'ordre 2.

Nom	Nombre de Faces	Nombre de sommets	Distance entre chaque sommet
Tétraèdre	4	4	$1,6330 \times R$
Hexaèdre	8	6	$1,4142 \times R$
Octaèdre	6	8	$1,1547 \times R$
Icosaèdre	20	12	$1,0515 \times R$
Dodécaèdre	12	20	$0,7136 \times R$

Tableau 5.2 – Polyèdres réguliers convexes inscrits dans une sphère de rayon R .

Une généralisation possible des polyèdres réguliers conduit aux polyèdres semi réguliers, notamment les treize *polyèdres archimédiens* et leurs polyèdres duaux, les *polyèdres Catalans*. Leurs faces sont des polygones réguliers convexes d'au moins deux types différents dont les côtés ont mêmes mesures et présentant en chaque sommet la même configuration. Certains s'obtiennent par exemple en tronquant les polyèdres réguliers, c'est-à-dire en coupant une partie à partir de chacun de ses sommets. Un exemple de polyèdre archimédien et de polyèdre dual Catalan sont respectivement l'icosaèdre tronqué (modèle du ballon de football) à 60 sommets et 32 faces, ces faces étant constituées par 12 pentagones réguliers et 20 hexagones réguliers) et le pentaki-dodécaèdre formé de 60 faces, des triangles isocèles, et de 32 sommets. Le Tableau 5.4 illustre la matrice d'erreur d'orthonormalité \mathbf{D} pour ces deux polyèdres. Notons que le pentaki-dodécaèdre préserve particulièrement bien l'orthonormalité des harmoniques sphériques échantillonnées jusqu'à l'ordre 4 ; les sommets de ce polyèdre définiront l'emplacement angulaire des points de mesure pour la réalisation d'un prototype microphonique HOA d'ordre 4 (cf. Section 5.4).

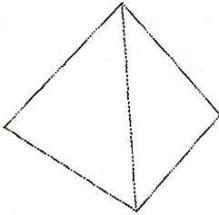
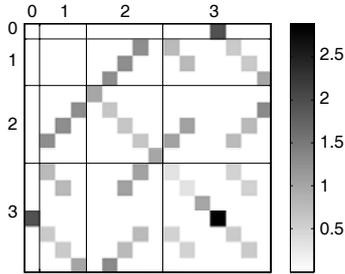
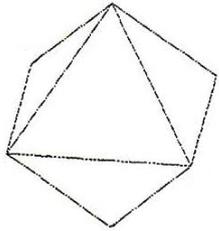
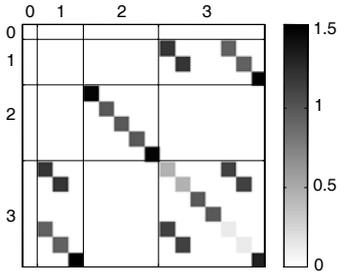
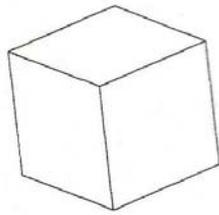
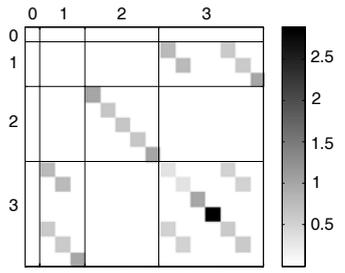
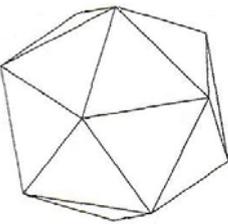
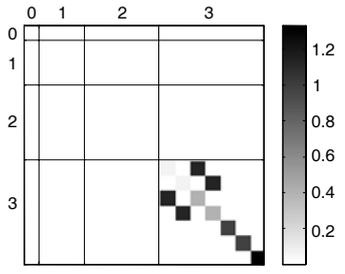
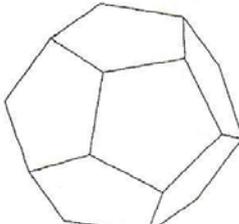
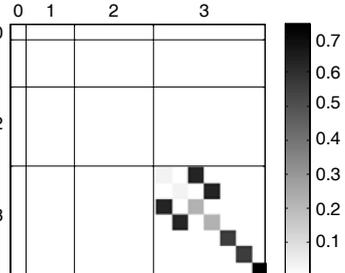
	Schéma du polyèdre régulier	Illustration de l'erreur d'orthonormalité associée
TÉTRAÈDRE		
HEXAÈDRE		
OCTÈDRE		
ICOSAÈDRE		
DODÉCAÈDRE		

Tableau 5.3 – Illustration schématique des polyèdres platoniciens et de la matrice erreur d'orthonormalité associée.

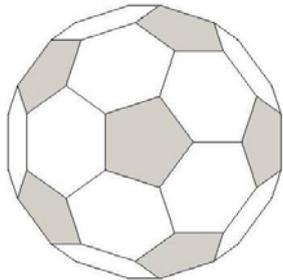
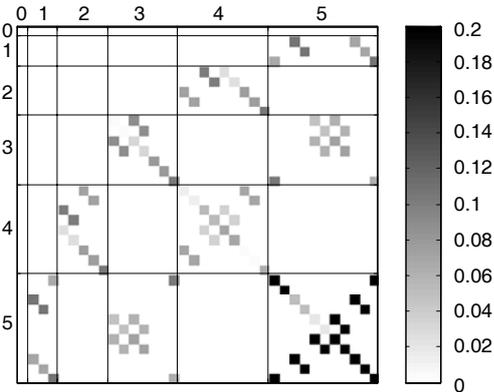
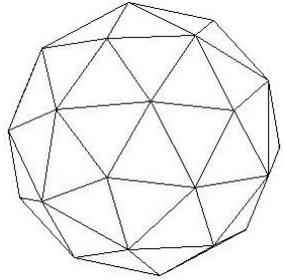
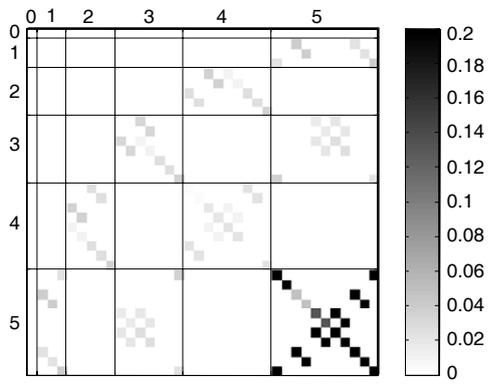
	Schéma du polyèdre semi-régulier	Illustration de l'erreur d'orthonormalité associée
ICOSAÈDRE TRONQUÉ		
PENTAKI-DODÉCAÈDRE		

Tableau 5.4 – Illustration schématique de deux polyèdres semi-réguliers duaux et de la matrice d'erreur d'orthonormalité associée à chacun.

Échantillonnage quasi-régulier à la surface de la sphère

D'une façon générale, nous pouvons chercher à résoudre le problème suivant : *comment placer de façon la plus homogène possible un ensemble de points à la surface d'une sphère ?* Cette question délicate a fait l'objet de nombreuses études dans divers domaines : chimie, géophysique, etc. La résolution de ce problème nécessite avant tout de définir mathématiquement ce que signifie *homogène*. Différentes interprétations existent, qui mènent à des distributions optimales de points légèrement différentes. Chaque interprétation aboutit à la définition d'un critère à partir duquel il suffit de résoudre un problème d'optimisation. Pour un nombre Q de points sur la sphère, un premier critère très répandu consiste à trouver les centres (θ_q, δ_q) de Q cercles identiques qui ne se chevauchent pas, de telle façon que leur rayon commun soit aussi grand que possible (*packing problem*). Une seconde approche, presque semblable, s'attache à trouver les centres de Q cercles identiques qui recouvrent entièrement la sphère de telle façon que leur rayon commun soit le plus petit possible (*covering problem*). Des distributions de points sur la sphère calculées suivant ces deux méthodes sont disponibles sur Internet et peuvent être directement exploitées [S06]. Citons une dernière méthode utilisée notamment par Fliege [F06], dont le but est de minimiser l'énergie potentielle d'électrons à la surface d'une sphère parfaitement conductrice, de telle façon à obtenir

la disposition la plus équilibrée possible (*Thomson's problem*). Dans ce cas, des pondérations associées à chaque point, proches de 1, doivent être calculées. Un certain nombre de configurations incluant les poids associés est également disponible sur Internet [F06]. Si ces différentes méthodes garantissent une relative équidistance des points d'échantillonnage, elles ne sont pas non plus directement reliées aux fonctions harmoniques sphériques et ne préservent donc pas automatiquement leur propriété d'orthonormalité. Nous devons par conséquent observer la matrice d'erreur d'orthonormalité donnée à l'équation (5.34) avant de choisir un échantillonnage adéquat. A titre d'exemple, nous illustrons la matrice **D** pour une distribution de 64 points définie grâce au critère de minimisation de l'énergie potentielle, ainsi que pour une distribution de 24 points définie successivement suivant les critères du *packing problem* et du *covering problem*. L'ensemble des 64 points choisi par Duraiswami et ses collègues [D05] pour réaliser un microphone 3D d'ordre 6 montre en effet que jusqu'à cet ordre la propriété d'orthonormalité des harmoniques sphériques est assez bien respectée. Les deux distributions de 24 points présentent quant à elles une erreur d'orthonormalité comparable, relativement faible jusqu'au 3^{ème} ordre inclus.

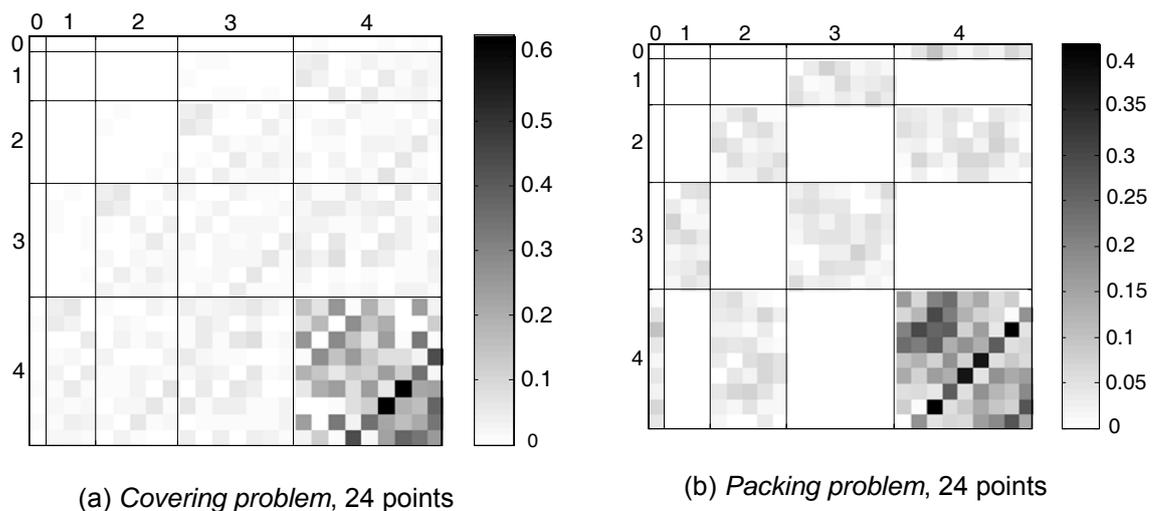


Tableau 5.5 – Illustration de la matrice d'erreur d'orthonormalité des harmoniques sphériques échantillonnées sur 24 points à la surface de la sphère suivant deux méthodes distinctes [S06].

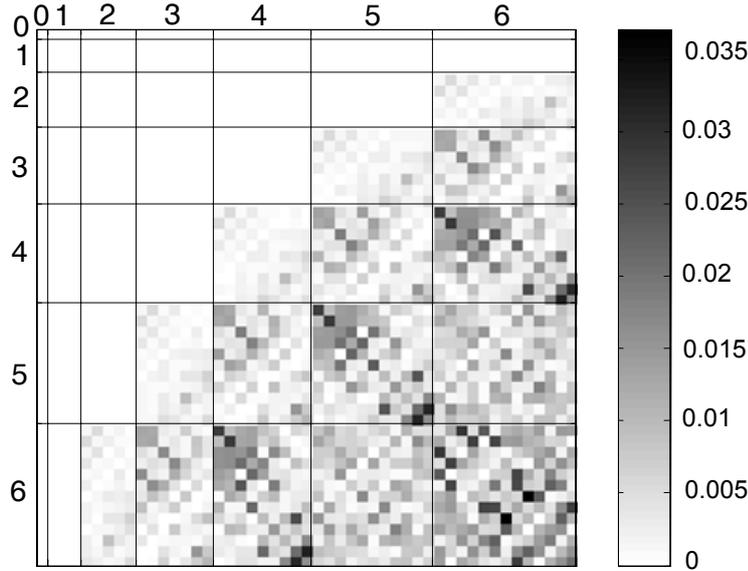


Figure 5.5 – Illustration de la matrice d'erreur d'orthonormalité des harmoniques sphériques échantillonnées sur 64 points d'après une méthode de minimisation de l'énergie potentielle [F06].

5.3.2 Régularisation des signaux HOA estimés

Principe de la régularisation – solution régularisée

Comme nous l'avons précisé précédemment, l'estimation des signaux HOA à partir de l'équation (5.27) est très instable et par conséquent non adaptée. Pour obtenir un résultat sensé, nous devons imposer des contraintes au problème de minimisation (5.25) ; cela peut être réalisé par des méthodes dites de *régularisation*, la plus utilisée étant probablement celle de Tikhonov [TA77]. Sa forme standard consiste à imposer une limite supérieure sur la norme quadratique (l'énergie) de la solution du problème de minimisation. Ce problème s'écrit alors :

$$\min_{\mathbf{b}_{\text{reg}}} \left\{ \left\| \mathbf{s} - \mathbf{M} \cdot \mathbf{b}_{\text{reg}} \right\|_2^2 + \lambda^2 \left\| \mathbf{b}_{\text{reg}} \right\|_2^2 \right\}, \quad \text{avec } \lambda \geq 0. \quad (5.35)$$

λ est le paramètre de régularisation qui détermine, suivant sa valeur, le compromis entre la précision de la solution \mathbf{b}_{REG} et sa norme maximale. Une valeur optimale du paramètre λ aboutit à une solution pour laquelle ce compromis est équilibré. Le problème (5.35) peut s'écrire sous la forme d'un système d'équations linéaires à résoudre :

$$\left(\mathbf{T}^* \cdot \mathbf{T} + \lambda^2 \mathbf{I}_K \right) \cdot \mathbf{b}_{\text{reg}} = \mathbf{T}^* \cdot \mathbf{s}, \quad (5.36)$$

où les notations de l'équation (5.25) sont reprises, et \mathbf{I}_K est la matrice identité de dimensions $K \times K$. Ce système est assez proche du système d'équations normales (5.25) et sa résolution aboutit à la solution :

$$\mathbf{b}_{\text{reg}} = (\mathbf{T}^* \cdot \mathbf{T} + \lambda^2 \mathbf{I}_K)^{-1} \cdot \mathbf{T}^* \cdot \mathbf{s}, \quad (5.37)$$

Définition des filtres de régularisation

En utilisant l'expression factorisée (5.24) de la matrice \mathbf{T} , le processus de régularisation de la solution (5.37) revient à appliquer aux signaux HOA estimés un filtre dont le rôle est d'éliminer les composantes fréquentielles les plus bruitées, c'est-à-dire les composantes basse fréquence des ordres les plus élevés :

$$\mathbf{b}_{\text{reg}} = \text{diag} \left[F_m(kR) \cdot \frac{1}{W_m(kR)} \right] \mathbf{E} \mathbf{s} \quad (5.38)$$

avec

$$F_m(kR) = \frac{|W_m(kR)|^2}{|W_m(kR)|^2 + \lambda^2}. \quad (5.39)$$

Les filtres $F_m(kR)$ sont appelés *filtres de régularisation*, ou filtres de Tikhonov. Lorsque le paramètre de régularisation λ est nul, ces filtres valent 1 et \mathbf{b}_{reg} correspond simplement à \mathbf{b}_{MC} (cf. Équation (5.27)). Quand λ augmente, les filtres de régularisation décroissent et réduisent l'effort d'estimation des composantes HOA correspondant à l'ordre considéré. Remarquons que les filtres $F_m(kR)$ sont proportionnels à $|W_m(kR)|^2$ et par conséquent suffisent à compenser le facteur d'amplification $1/W_m(kR)$. L'égalisation résultant du processus de régularisation est finalement définie par :

$$EQ_m(kR) = \frac{\bar{W}_m(kR)}{|W_m(kR)|^2 + \lambda^2}, \quad (5.40)$$

où $\bar{W}_m(kR)$ désigne le conjugué complexe de $W_m(kR)$. Nous reconnaissons une expression standard du procédé de régularisation des filtres instables par la méthode de Tikhonov [KN96, FU99].

Choix du paramètre de régularisation

Le choix d'un paramètre de régularisation adapté à la situation est primordial pour obtenir une estimation sensée. En effet, un paramètre λ trop grand entraîne une solution trop régularisée et par conséquent trop éloignée de la solution exacte. A l'inverse un paramètre trop petit produit des signaux HOA trop bruités pour les ordres les plus élevés. Un paramètre idéal réalise un compromis entre ces deux extrêmes, en réduisant le bruit au maximum dans les solutions et en préservant au mieux, dans le même temps, les informations utiles.

Le paramètre λ peut être directement relié à une unique amplification maximale a produite par les filtres $EQ_m(kR)$:

$$\lambda = \frac{1 - \sqrt{1 - 1/a^2}}{1 + \sqrt{1 - 1/a^2}}, \quad (5.41)$$

où $a \geq 1$.

L'amplification maximale a peut être déterminée à partir du concept de bande fréquentielle utile (cf. Chapitre 4), en visant un rayon r_{lim} fixe de reconstruction future à toutes les fréquences avec une tolérance d'erreur maximale e_{max} . En effet, nous nous sommes aperçu que pour reproduire un champ acoustique sur une zone de l'espace constante sur toute la bande des fréquences audibles et avec une tolérance d'erreur maximale fixée à l'avance, la contribution des différents modes harmoniques n'est pas égale à toutes les fréquences, en particulier les modes les plus haut sont moins utiles en basse fréquence qu'en haute fréquence. De cette façon, à partir du rayon r_{lim} et de la tolérance d'erreur e_{max} , nous devons dans un premier temps trouver pour chaque ordre m , conformément au concept de bande fréquentielle utile exposé au Chapitre 4, la fréquence limite $f_{\text{lim}}^{(m)}$ en dessous de laquelle cet ordre m devient inutile, ou encore en dessous de laquelle la régularisation peut s'appliquer sans perte d'information essentielle. Par exemple, en choisissant $e_{\text{tol}} = 7\%$, ces fréquences limites s'obtiennent approximativement par la relation (cf. Chapitre 2, Section 2.3.2) :

$$f_{\text{lim}}^{(m)} = \frac{mc}{2\pi r_l}, \quad (5.42)$$

où c est la vitesse de propagation du son dans l'air. À partir des fréquences $f_{\text{lim}}^{(m)}$, les amplitudes maximales a_m nécessaires pour la reproduction sonore dans la zone fixée au préalable se déduisent ensuite de l'expression des fonctions $W_m(kR)$. Chaque paramètre λ_m associé à l'ordre m se calcule finalement grâce à la formule mathématique (5.41). En pratique, le processus de régularisation décale les pics d'amplitude maximale a_m des filtres EQ_m sur des fréquences plus basses que les fréquences $f_{\text{lim}}^{(m)}$ (cf. Figure 5.9). Par conséquent, un petit réajustement peut être nécessaire pour garantir que cette régularisation affecte bien uniquement les fréquences en dessous des fréquences limites $f_{\text{lim}}^{(m)}$; cela se traduit par un léger réhaussement des paramètres λ_m . Cette méthode de choix d'un paramètre de régularisation distinct à chaque ordre possède l'avantage d'être liée à un critère objectif qui permet de prédire la qualité potentielle de reconstruction acoustique.

Le choix du facteur a peut également se faire de façon semblable à tous les ordres, en se basant uniquement sur l'amplification induite du bruit de fond des microphones et en tenant compte notamment de l'amélioration du rapport signal à bruit apportée par l'utilisation de multiples capteurs qui correspond théoriquement à un gain de $10 \log Q$ dans le cas où Q microphones sont utilisés. L'amplification maximale a résultant dans les filtres EQ_m peut ainsi être définie par :

$$a = \sqrt{Q} \cdot 10^{\frac{a_s}{20}} \quad (5.43)$$

où a_s est l'amplification maximale tolérée du bruit exprimée en dB dans les signaux HOA résultants. Plus le nombre de capteurs est grand, plus l'amplification maximale tolérée a peut être importante. Contrairement à la première méthode, l'enregistrement du champ acoustique concerne ici des zones de l'espace de tailles différentes suivant la fréquence. En particulier, la description

acoustique et par conséquent la reconstruction potentielle en basse fréquence est plus large qu'en haute fréquence.

Enfin, en règle générale les deux méthodes présentées ci-dessus peuvent avantageusement être associées à une méthode d'essai-erreur qui consiste à tester plusieurs valeurs du paramètre λ pour obtenir la meilleure estimation possible des signaux HOA.

5.4 Réalisation, mesure et validation objective d'un prototype d'ordre 4

5.4.1 Construction du prototype

Géométrie du dispositif et influence sur la précision d'estimation des signaux HOA

Les microphones sont répartis à la surface d'une sphère rigide selon les 32 sommets d'un polyèdre semi régulier inscrit dans cette sphère, le *pentaki-dodécaèdre*. 12 de ses sommets correspondent à ceux d'un icosaèdre, et les 20 sommets restants sont obtenus en projetant radialement sur la sphère les centres des 20 faces de cet icosaèdre, soit en considérant les sommets du polyèdre dual, le dodécaèdre. L'illustration d'un tel polyèdre et les coordonnées angulaires (azimut et élévation) de ses 32 sommets sont données respectivement à la Figure 5.6 et dans le Tableau 5.6. Cette répartition à la surface de la sphère n'est pas exactement uniforme dans le sens où l'angle entre la position des capteurs n'est pas constant. En effet, l'écart angulaire entre deux sommets du dodécaèdre est d'environ 0,73 radians et celui entre un sommet du dodécaèdre et un sommet de l'icosaèdre vaut approximativement 0,65 radians. Cependant, comme nous avons pu le constater à la Section 2.3.1, cette géométrie angulaire préserve relativement bien la propriété d'orthonormalité des harmoniques sphériques jusqu'à l'ordre 4. En effet, l'erreur maximale d'orthonormalité sur la diagonale de la matrice \mathbf{D} et celle hors de la diagonale atteignent respectivement 3,7 % et 3,73 % (cf. Tableau 5.4). Seul un léger phénomène d'*aliasing* spatial apparaît dans les groupes de composantes harmoniques d'ordre 3 et d'ordre 4, ces dernières se repliant également sur les harmoniques d'ordre 2. Notons également que les harmoniques sous-échantillonnées d'ordre 5 et supérieurs créent, si elles sont significativement présentes au lieu de captation, un effet d'*aliasing* sur les composantes d'ordre inférieur ou égal à 4, c'est-à-dire sur les composantes que nous souhaitons estimer.

Le niveau de présence et donc l'influence néfaste des modes supérieurs à 4 dépendent directement du rayon du dispositif de captation. Rappelons que le choix de la taille du dispositif microphonique résulte d'un compromis difficile entre deux critères contradictoires : la minimisation de l'*aliasing* spatial et la précision d'estimation des signaux HOA dans un domaine basse fréquence. Le rayon choisi pour notre prototype vaut $R=3,5$ cm. Les fréquences théoriques d'*aliasing* correspondant aux deux valeurs distinctes d'écart angulaire entre les capteurs, sont

d'environ 6700 Hz pour 0,73 radians, et 7500 Hz pour 0,67 radians¹⁸. D'autre part, le rayon détermine la contribution théorique des différents modes harmoniques présents dans les mesures, et ainsi le conditionnement de la matrice $\mathbf{T}^*\mathbf{T}$ du système (5.25). Puisque la base des harmoniques sphériques échantillonnées préserve assez bien la propriété d'orthonormalité, la matrice $\mathbf{T}^*\mathbf{T}$ est quasiment diagonale et l'instabilité de l'estimation des signaux HOA principalement causée par les fonctions $W_m(kR)$:

$$\kappa(\mathbf{T}^*\mathbf{T}) = \frac{\max(|W_i(kR)|^2, i=0, \dots, M)}{\min(|W_i(kR)|^2, i=0, \dots, M)}, \quad (5.44)$$

le numérateur et le dénominateur correspondant respectivement à la plus grande et la plus faible valeur singulière de la matrice $\mathbf{T}^*\mathbf{T}$. La Figure 5.7 illustre les fonctions $W_m(kR)$ pour les ordres 0 à 4 parallèlement au nombre de conditionnement $\kappa(\mathbf{T}^*\mathbf{T})$. Nous constatons clairement que ce dernier augmente rapidement en basse fréquence lorsque les différences de marge entre les fonctions $W_m(kR)$ s'amplifient. L'instabilité de l'estimation des signaux HOA apparaît ainsi en basse fréquence et pour les ordres les plus élevés.

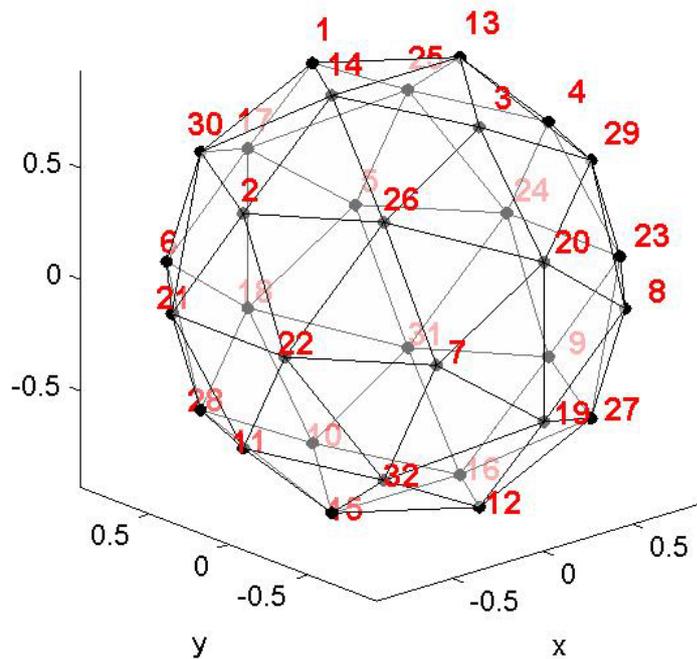


Figure 5.6 – Pentaki-Dodécaèdre, polyèdre semi-régulier à 32 sommets.

¹⁸ Nous supposons que la vitesse de propagation du son dans l'air c vaut 340 m/s.

Sommet	Azimut	Élévation	Sommet	Azimut	Élévation	Sommet	Azimut	Élévation
1	0	1.5708	13	0.6283	0.9184	25	2.5133	-0.1887
2	0	0.4636	14	1.8850	0.9184	26	3.7699	-0.1887
3	1.2566	0.4636	15	3.1416	0.9184	27	5.0265	-0.1887
4	2.5133	0.4636	16	4.3982	0.9184	28	0	-0.9184
5	3.7699	0.4636	17	5.6549	0.9184	29	1.2566	-0.9184
6	5.0265	0.4636	18	0.6283	0.1887	30	2.5133	-0.9184
7	0.6283	-0.4636	19	1.8850	0.1887	31	3.7699	-0.9184
8	1.8850	-0.4636	20	3.1416	0.1887	32	5.0265	-0.9184
9	3.1416	-0.4636	21	4.3982	0.1887			
10	4.3982	-0.4636	22	5.6549	0.1887			
11	5.6549	-0.4636	23	0	-0.1887			
12	0	-1.5708	24	1.2566	-0.1887			

Tableau 5.6 – Coordonnées sphériques angulaires des sommets d'un pentaki-dodécaèdre. Ce polyèdre semi régulier peut être interprété comme l'association d'un icosaèdre et d'un dodécaèdre.

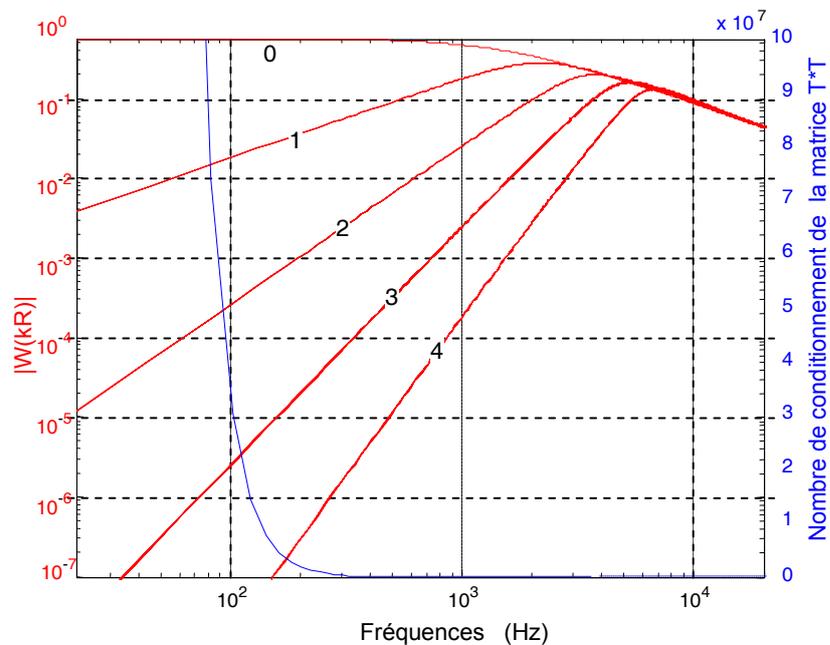


Figure 5.7 – Fonctions $|W_m(kR)|$ pour les ordres 0 à 4 (en rouge), et nombre de conditionnement de la matrice T^*T (en bleu) en fonction de la fréquence ($R = 3,5$ cm).

Réalisation et mise en œuvre des traitements relatifs au prototype

Le prototype est construit avec 32 capteurs Panasonic WM-62PC de 6 mm de diamètre, de type électret et de directivité omnidirectionnelle. Le rapport signal à bruit revendiqué par le constructeur atteint 58 dB. Les capteurs ont été mesurés individuellement en chambre anéchoïque et sélectionnés pour leurs courbes de réponse voisines. Ils sont montés sur une sphère en plastique creuse supposée acoustiquement rigide (cf. Figure 5.8), hypothèse vérifiée expérimentalement à la Section 2.4.2. Les 32 signaux audio sont enregistrés sur un ordinateur PC puissant mais ordinaire (2 GHz, 1 Go de mémoire RAM) via 2 interfaces MOTU 24 entrées/sorties synchronisées reliées à une carte MOTU PCI424¹⁹. Enfin, deux possibilités d'effectuer le traitement des 32 signaux enregistrés délivrant les signaux HOA estimés jusqu'à l'ordre 4 sont possibles suivant l'application. Une première possibilité consiste à utiliser une librairie de routines écrites pour le logiciel Matlab [Mat06]. Pour une utilisation du microphone *en temps réel*, les traitements des 32 signaux ainsi que le décodage spatial pour divers dispositifs de restitution ont été implantés sous la forme de plugins Steinberg VST (*Virtual Studio Technology*) exploitables sous un logiciel hôte tel que *Plogue Bidule*. Ces traitements consistent en une étape de matriçage définie par l'équation (5.28) en fonction de la position des capteurs, suivie d'une étape d'égalisation spectrale par les filtres $EQ_m(kR)$ définis à l'équation (5.40) en fonction du rayon du dispositif microphonique et implantée sous la forme de filtres FIR de longueurs variables. Les modules de ces filtres sont illustrés à la Figure 5.9 pour trois paramètres de régularisation distincts : $\lambda = 7,9 \cdot 10^{-2}$, $\lambda = 2 \cdot 10^{-3}$, et $\lambda = 7,81 \cdot 10^{-5}$, correspondant respectivement à une amplification du bruit de fond des capteurs de 0 dB, +6 dB et +20 dB. Notons que l'amplitude maximale de ces filtres peut dépasser ces derniers niveaux d'amplification puisque le rapport signal à bruit s'améliore de façon proportionnelle à la racine carrée du nombre de capteurs utilisés, soit dans notre cas 32 (cf. Section 5.3.2).

La Figure 10 illustre un mode d'utilisation temps réel de ce microphone HOA à partir de *plugins* VST fonctionnant sous le logiciel hôte *Plogue Bidule*, et développés par l'équipe Son3D de France Télécom R&D (Lannion). Les signaux enregistrés par les capteurs passent par le *plugin* *AmbiMicProcessor* réalisant les opérations de matriçage et de filtrage permettant d'aboutir aux signaux HOA. Deux *plugins* distincts de décodage, précédés chacun par un module *AmbiTransformer* de manipulation de la scène sonore HOA, sont ensuite proposés : un *plugin* *AmbiBinDecoder25in* de décodage binaural au casque d'écoute combiné avec un *head-tracker*, les paramètres de ce dernier pilotant le module *AmbiTransformer* pour corriger les mouvements rotatifs latéraux de la tête de l'auditeur ; et un *plugin* *AmbisonicSpkDecoder25x32* de décodage 2D pour un dispositif circulaire de 12 haut-parleurs. Tous ces traitements fonctionnent simultanément en temps réel.

¹⁹ MOTU 24I/O, accédé le 28 mai 2006, <http://www.motu.com/products/pciaudio/24IO/>



Figure 5.8 – Prototype de microphone HOA d'ordre 4 constitué de 32 capteurs omnidirectionnels répartis sur une sphère en plastique rigide suivant les 32 sommets d'un pentaki-dodécaèdre.

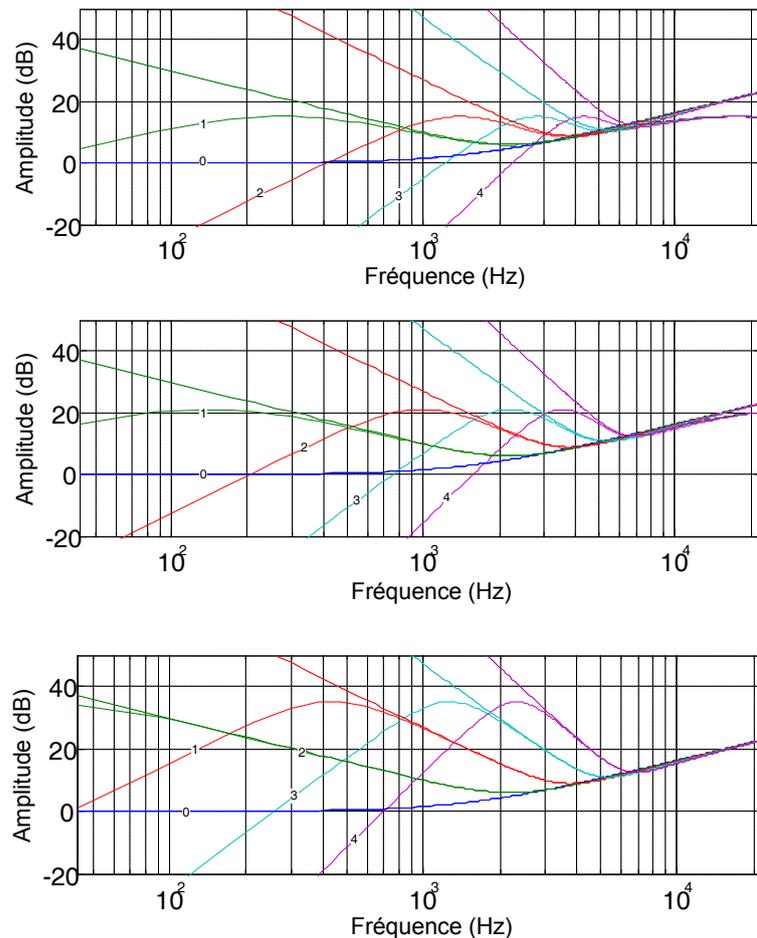


Figure 5.9 – Modules des filtres d'égalisation EQ_m théoriques (lignes en pointillés) et régularisés (lignes continues) correspondant à trois paramètres de régularisation distincts $\lambda = 7,9 \cdot 10^{-2}$, $\lambda = 2 \cdot 10^{-3}$, et $\lambda = 7,81 \cdot 10^{-5}$, associés respectivement à une amplification du bruit des capteurs de 0 dB, +6 dB, et +20 dB (de haut en bas).

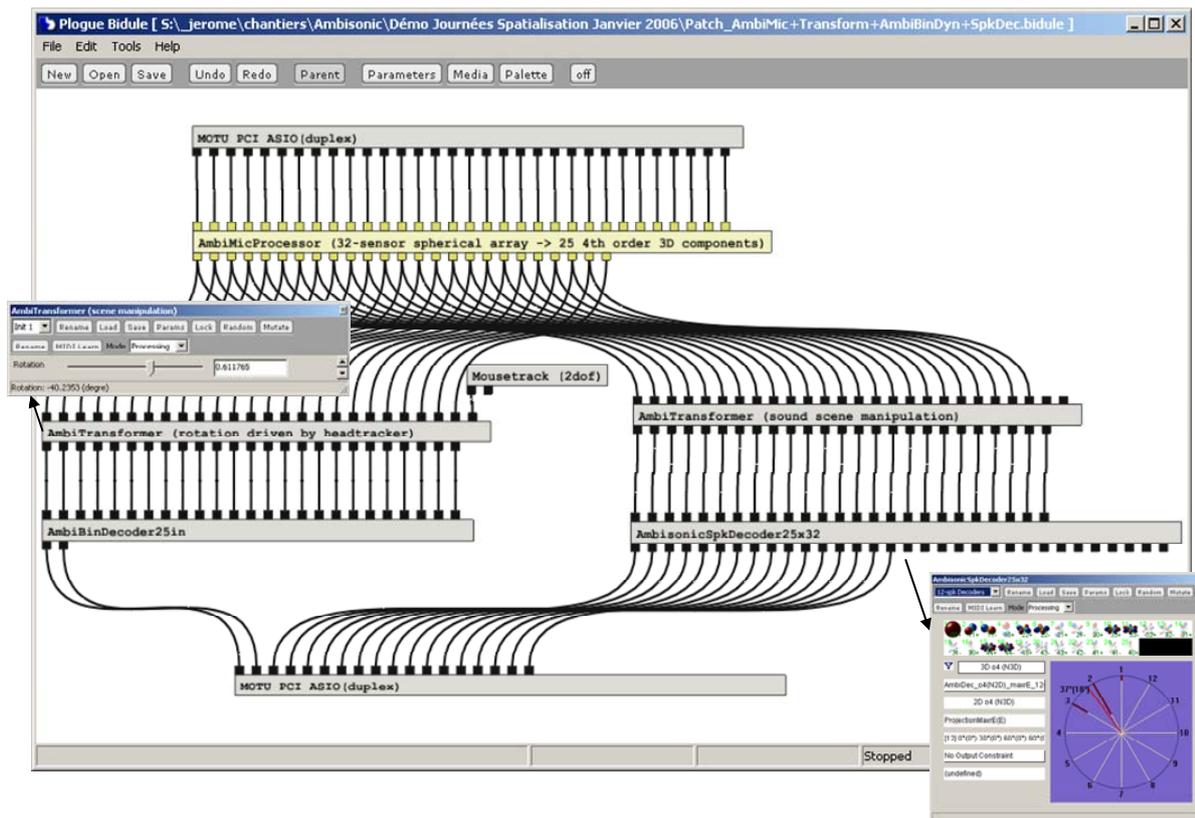


Figure 10 – Mode d'utilisation en temps réel du microphone HOA d'ordre 4 à partir d'une librairie de plug-ins VST exploitée sous le logiciel hôte Plogue Bidule.

5.4.2 Confrontation du prototype au modèle théorique de la sphère rigide

La réalisation pratique du microphone HOA entraîne inévitablement un certain nombre d'erreurs par rapport au modèle théorique, notamment dues au positionnement des capteurs, à la réponse non idéale de ces capteurs, ou encore à la présence d'un pied qui perturbe le champ sonore. D'autre part, nous ne pouvons être assurés que la sphère en plastique creuse se comporte véritablement comme une sphère acoustiquement rigide. Par conséquent, il est utile de vérifier avant toute chose que notre prototype ne s'écarte pas trop de son modèle théorique.

Description du dispositif de mesure

Notre prototype de microphone HOA a fait l'objet d'une campagne de mesures réalisée dans la chambre anéchoïque de l'IRCAM (Institut de Recherche et Coordination Acoustique /Musique) à Paris. L'installation, constituée par une table tournante Brüel et Kjør placée au milieu de la salle et un bras mécanique mobile pouvant s'élever sur lequel est fixé un haut-parleur, permet de mesurer dans les trois dimensions de l'espace (cf. Figure 5.11). Plus exactement, le microphone a été mesuré par pas de 5 degrés en azimut (de 0 à 360 degrés) et en élévation (de -40 à 90 degrés). Un nombre total de 1971 mesures a ainsi été effectué. Les mesures ont été réalisées avec une méthode de sinus logarithmique (*sine sweep*), à partir d'une séquence d'ordre 16. Ces mesures ont été par ailleurs déconvoluées par la réponse du haut-parleur.

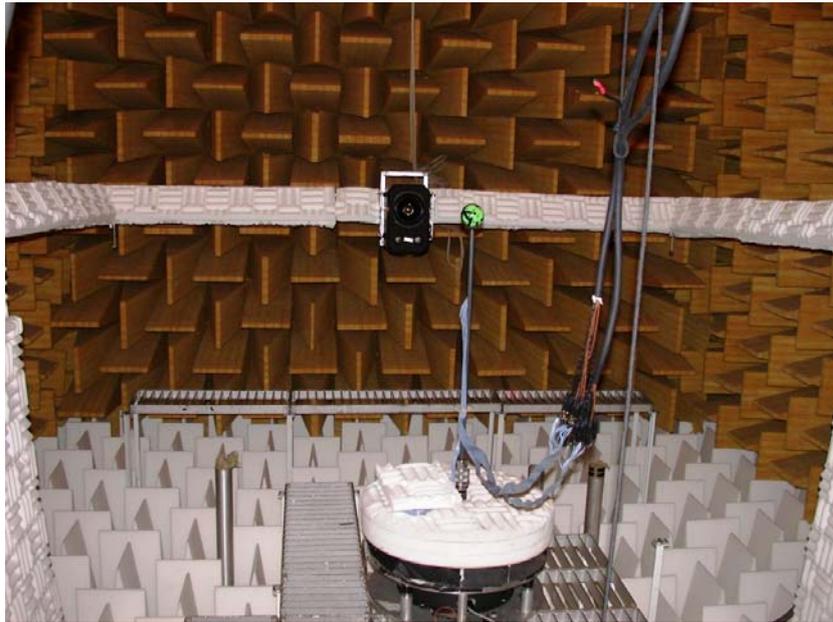


Figure 5.11 – Dispositif de mesure en champ libre du prototype microphonique dans la chambre anéchoïque de l'IRCAM (Paris).

Comparaison visuelle des directivités mesurées et modélisées

Nous comparons tout d'abord visuellement la directivité mesurée des microphones de notre prototype et la directivité simulée grâce au modèle théorique correspondant. La Figure 5.12 illustre ces directivités pour les quatre capteurs 17, 4, 10 et 16 (cf. Figure 5.6) en parallèle à la directivité théorique correspondante, modélisée selon les caractéristiques du prototype et pour quatre fréquences différentes : 1 kHz, 2 kHz, 4 kHz et 8 kHz. Les directivités mesurées expérimentalement sont représentées par une surface maillée, celles modélisées par une surface transparente colorée. Pour les capteurs 17 et 4 les mesures semblent coller relativement bien au modèle jusqu'en haute fréquence, bien au-delà de la fréquence d'*aliasing*. Cependant, pour les capteurs 10 et 16 placés à proximité du pied du microphone, si la directivité est relativement bien respectée jusqu'à 1 kHz, elle se dégrade au fur et à mesure que la fréquence augmente. Cette dégradation est vraisemblablement causée par la perturbation du champ sonore engendrée par la présence du pied.

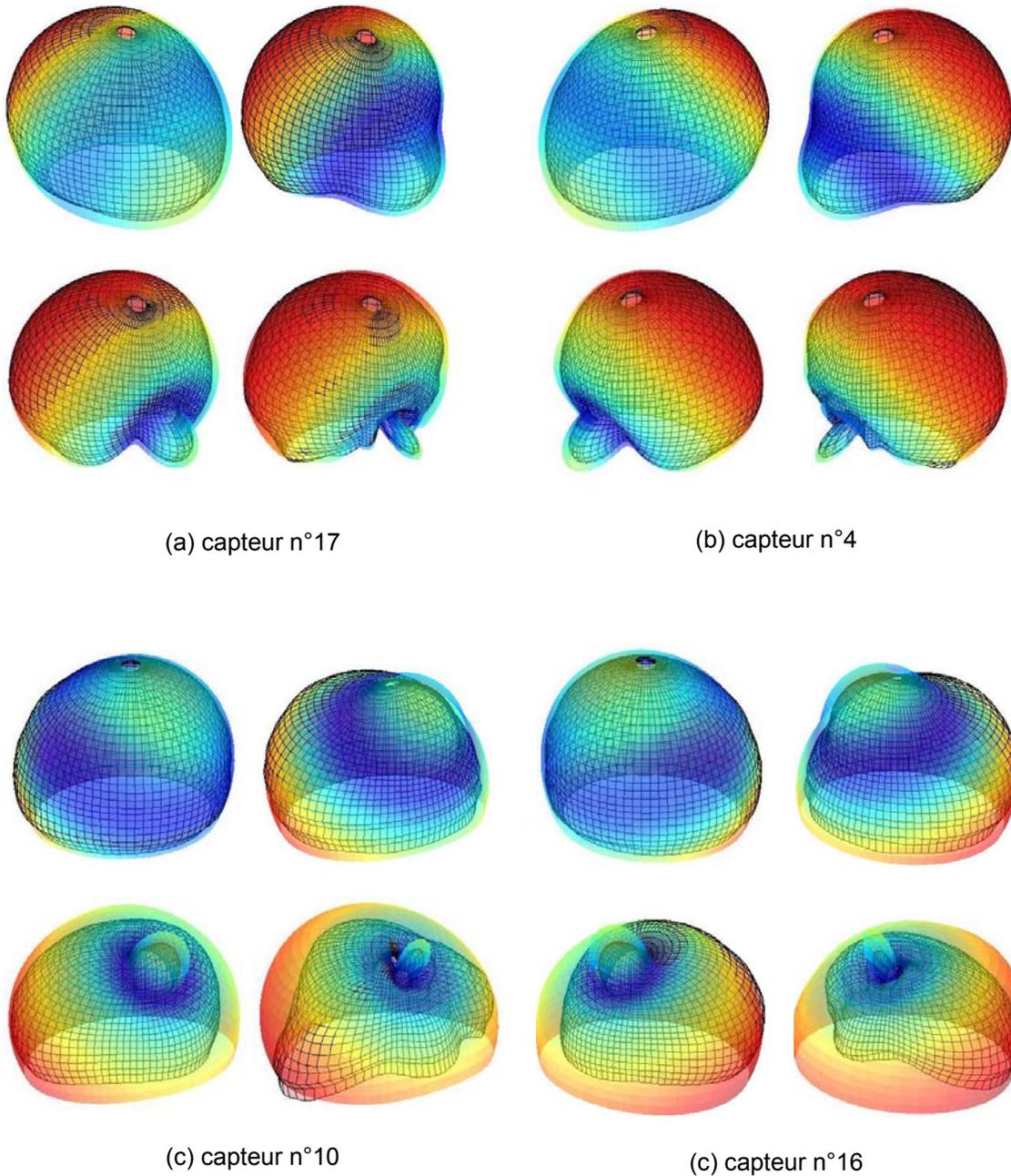


Figure 5.12 – Représentation 3D des directivités mesurées (maillage) et simulées (transparence colorée) pour les capteurs 17 (a), 4 (b), 10 (c) et 16 (d), et les fréquences 1 kHz, 2 kHz, 4 kHz et 8 kHz (respectivement de gauche à droite et de haut en bas).

Comparaison objective : analyse de la corrélation spatiale entre la mesure et le modèle

Pour évaluer de façon plus objective et globale la conformité du prototype avec le modèle physique qu'il est censé reproduire, nous analysons premièrement la corrélation spatiale entre les modules des signaux des capteurs modélisés et mesurés, calculée sur l'ensemble des points de mesure et définies par :

$$C(\omega) = \frac{\langle |\mathbf{s}_{\text{mes}}(\omega)|, |\mathbf{s}_{\text{mod}}(\omega)| \rangle}{\|\mathbf{s}_{\text{mes}}(\omega)\| \|\mathbf{s}_{\text{mod}}(\omega)\|}, \quad (5.45)$$

où nous avons introduit le produit scalaire et la norme associée suivants :

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^* \cdot \text{diag}(\boldsymbol{\gamma}) \cdot \mathbf{y}, \text{ et } \|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}, \quad (5.46)$$

et où les vecteurs \mathbf{s}_{mes} et \mathbf{s}_{mod} contiennent respectivement les signaux mesurés et modélisés des capteurs dans les 1971 directions permises par le dispositif de mesure décrit plus haut, et enfin le vecteur $\boldsymbol{\gamma}$ est constitué par un ensemble de coefficients réels dont la somme vaut 1 et dont le rôle est de compenser la différence de densité des points de mesure au voisinage des pôles et dans la zone équatoriale. Une corrélation parfaite se traduit par une valeur du coefficient C égale à 1. La Figure 5.13(haut) représente la corrélation spatiale des modules des spectres de \mathbf{s}_{mes} et \mathbf{s}_{mod} en fonction de la fréquence pour chacun des 32 capteurs du dispositif microphonique. Nous constatons que la corrélation spatiale est très proche de 1, c'est-à-dire très bonne sur toute la bande des fréquences, excepté pour les quatre capteurs n^{os} 10, 12, 16 et 15. Ces derniers sont en effet placés à proximité du pied et leur mauvaise corrélation en moyenne et haute fréquence s'explique vraisemblablement par la perturbation acoustique engendrée par ce pied.

Pour conclure à une bonne corrélation globale, il est impératif de prendre en compte les différences de phase entre les signaux mesurés et les signaux modélisés, très importantes pour une reconstruction correcte des harmoniques sphériques. Pour cela, nous calculons la corrélation spatiale d'après les *spectres complexes* \mathbf{s}_{mes} et \mathbf{s}_{mod} , plus représentative de la conformité entre notre prototype réel et le modèle physique dont il est issu :

$$C(\omega) = \frac{\langle \mathbf{s}_{\text{mes}}(\omega), \mathbf{s}_{\text{mod}}(\omega) \rangle}{\|\mathbf{s}_{\text{mes}}(\omega)\| \|\mathbf{s}_{\text{mod}}(\omega)\|}, \quad (5.47)$$

où les notations restent semblables à celles utilisées dans l'équation (5.45). La Figure 5.13(bas) représente la corrélation spatiale entre le spectre complexe mesuré et modélisé. Cette corrélation reste logiquement la plus faible pour les capteurs placés à proximité du pied. Mais nous remarquons qu'elle se dégrade également en haute fréquence pour l'ensemble des capteurs. Cette dégradation est probablement due aux diverses erreurs de phase introduites dans le modèle : erreurs de positionnement des capteurs, défauts de la courbe de réponse des capteurs, etc. La corrélation spatiale reste tout de même globalement très satisfaisante, au moins jusqu'à la fréquence d'*aliasing*, et nous pouvons à présent étudier les capacités du microphone HOA en termes de reconstruction des harmoniques sphériques.

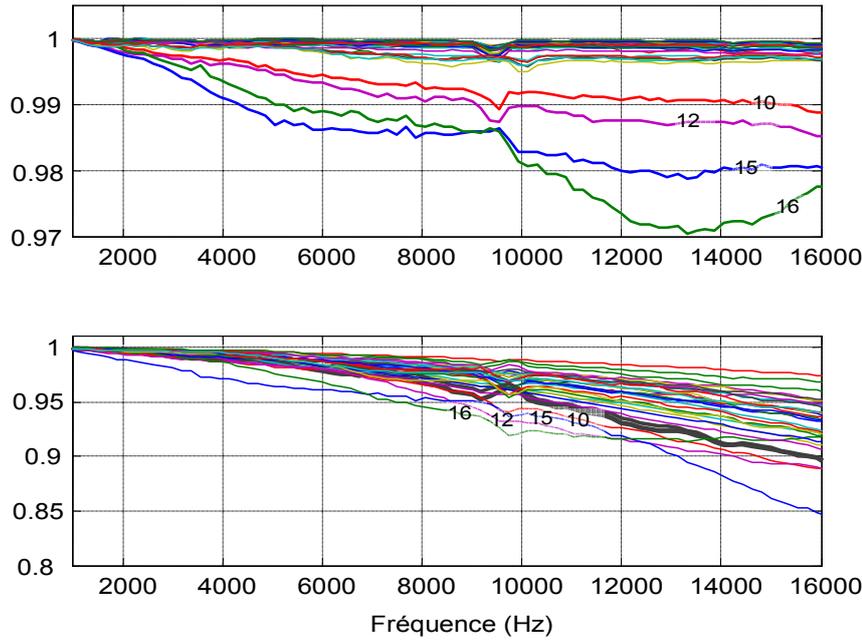


Figure 5.13 – Corrélation spatiale entre la directivité des mesures des capteurs et celles du modèle correspondant ; (haut) : corrélation spatiale calculée sur les modules des spectres, (bas) : corrélation spatiale calculée sur les spectres complexes.

5.4.3 Reconstruction des directivités associées aux harmoniques sphériques

Comparaison visuelle entre les directivités théoriques et celles reconstruites d'après les mesures

Un premier aperçu des performances du microphone HOA peut être donné en visualisant la reconstruction des directivités associées aux harmoniques sphériques d'ordres 0 à 4. Cette reconstruction est réalisée par une première étape de matricage suivie d'une seconde étape de filtrage des signaux enregistrés, comme décrit à la Section 2.4.1. Le paramètre de régularisation λ choisi pour la définition des filtres vaut $2 \cdot 10^{-3}$, ce qui correspond à une amplification maximale du bruit de fond des microphones de +6 dB (cf. Section 2.3.2). La Figure 5.14 illustre la reconstruction des directivités associées à un échantillon d'harmoniques sphériques appartenant aux ordres 1 à 4 et pour deux fréquences distinctes, 5 000 Hz et 10 000 Hz. Nous constatons que ces directivités sont relativement bien reconstruites à la fréquence 5 kHz, et que seule une légère différence d'amplitude apparaît par endroits. En revanche, à 10 kHz, c'est-à-dire à une fréquence supérieure à la fréquence d'*aliasing*, la reconstruction n'est plus assurée et les illustrations mettent en évidence le phénomène de recouvrement spatial des harmoniques d'ordres supérieurs décrit à la Section 2.2.2, et qui pollue ainsi l'estimation des signaux HOA.

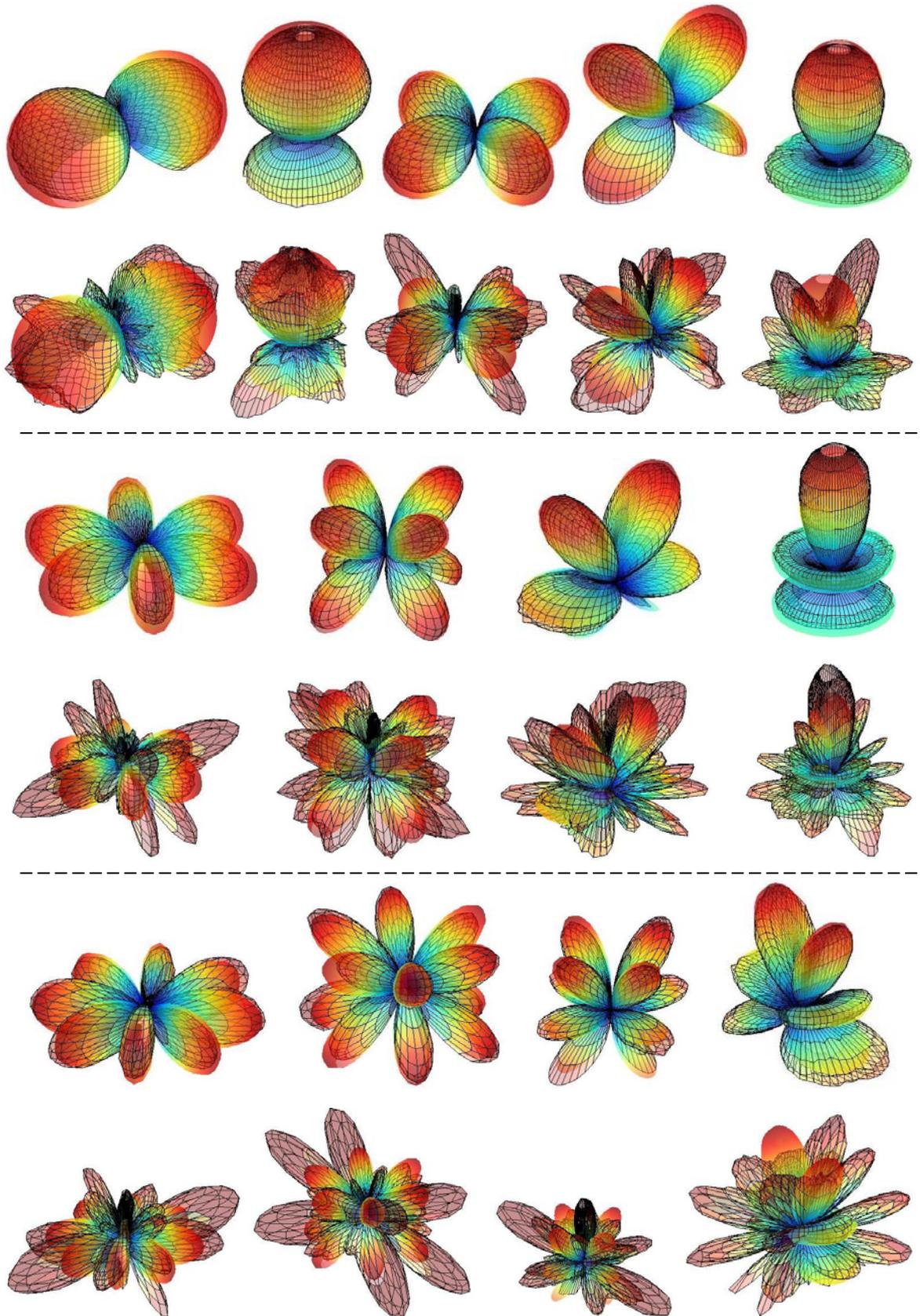


Figure 5.14 – Harmoniques sphériques reconstruites (maillage) et harmoniques sphériques théoriques correspondantes (transparence colorée) d'ordres 1 à 4, à 5kHz (haut) et 10kHz (bas)

La Figure 5.15 propose une vue 2D de la reconstruction des harmoniques sphériques sur une large bande fréquentielle. Pour chaque ordre de 0 à 4, cette figure représente la reconstruction dans le plan horizontal d'une harmonique en fonction de la fréquence. Nous observons clairement le phénomène d'*aliasing* spatial dans le domaine haute fréquence, qui commence à apparaître dans une bande fréquentielle de transition comprise entre les fréquences d'*aliasing* 6700 Hz et 7500 Hz calculées pour les deux écarts angulaires existant entre les capteurs constituant notre microphone HOA (cf. Section 2.4.1). L'effet de la régularisation, caractérisé par une atténuation en basse et moyenne fréquence qui apparaît d'autant plus haut dans le spectre que l'ordre est élevé (cf. Figure 5.9), est également visible en basse et moyenne fréquence. Comme nous l'avons précisé auparavant, cette atténuation intervient dans une zone fréquentielle où les informations du champ sonore portées par les composantes harmoniques sphériques concernent une zone de l'espace beaucoup trop éloignée comparativement à la taille du microphone, et par conséquent « inaccessibles » pour lui.

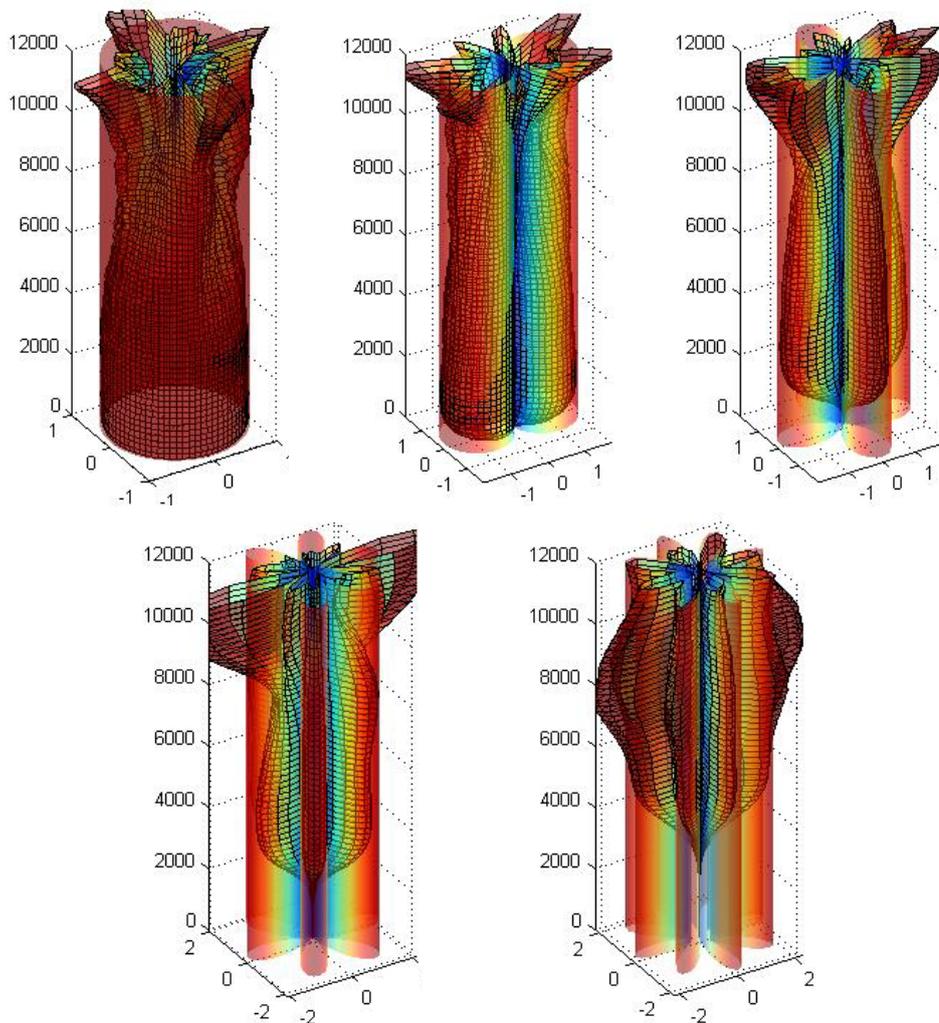


Figure 5.15 – Illustration 2D des harmoniques sphériques reconstruites (maillage) et théoriques (transparence) dans le plan horizontal pour les ordres 0 à 4 et en fonction de la fréquence.

Caractérisation objective de la reconstruction des harmoniques sphériques : analyse de la corrélation spatiale

Pour caractériser de façon objective la reconstruction des directivités associées aux harmoniques sphériques d'ordre 0 à 4, nous étudions premièrement la corrélation spatiale entre ces directivités reconstruites et les directivités théoriques attendues, définie de façon analogue à l'équation (5.45) :

$$C(\omega) = \frac{\langle \mathbf{b}_{\text{mes}}(\omega), \mathbf{b}_{\text{théo}}(\omega) \rangle}{\|\mathbf{b}_{\text{mes}}(\omega)\| \cdot \|\mathbf{b}_{\text{théo}}(\omega)\|}, \quad (5.48)$$

où les vecteurs \mathbf{b}_{mes} et $\mathbf{b}_{\text{théo}}$ contiennent respectivement les harmoniques sphériques reconstruites à partir des mesures effectuées sur notre prototype, et les harmoniques sphériques théoriques échantillonnées à ces mêmes points de mesure. Puisque la corrélation spatiale ne caractérise que la ressemblance entre les variations angulaires des composantes harmoniques mesurées et théoriques, nous la complétons par un critère de niveau moyen des réponses estimées relativement aux réponses théorique idéales :

$$N_{\text{moy}}(\omega) = \frac{\|\mathbf{b}_{\text{mes}}(\omega)\|}{\|\mathbf{b}_{\text{théo}}(\omega)\|}, \quad (5.49)$$

le produit scalaire et la norme employés dans les relations mathématiques (5.48) et (5.49) ayant été définis à l'équation (5.46).

La Figure 5.16 illustre la corrélation spatiale entre les directivités harmoniques sphériques issues du traitement des mesures et les directivités théoriques correspondantes, ainsi que le niveau relatif entre elles pour deux paramètres distincts de régularisation : $\lambda = 7,9 \cdot 10^{-2}$ et $\lambda = 2,10 \cdot 10^{-3}$. Pour plus de lisibilité, chaque courbe est issue du calcul de la moyenne des résultats obtenus pour les composantes d'un même ordre. Le paramètre de régularisation influe seulement sur le niveau moyen des composantes harmoniques reconstruites et non sur leur forme ; leur corrélation spatiale avec les harmoniques théoriques ne change donc pas. Nous remarquons tout d'abord qu'immédiatement au-dessus de la bande fréquentielle définie par les deux fréquences d'*aliasing* 6 700 Hz et 7 500 Hz théoriques associées aux écarts angulaires entre les microphones du prototype et indiquées sur les figures par deux traits verticaux en pointillés, la corrélation spatiale chute brutalement et le niveau moyen relatif augmente. La dégradation de la corrélation spatiale en haute fréquence illustre parfaitement bien les effets néfastes, attendus, de l'*aliasing* spatial. Quant à l'augmentation du niveau moyen, il s'explique par l'utilisation involontaire dans le processus d'estimation de composantes harmoniques d'ordres supérieurs qui se replient sur celles estimées. Nous remarquons par ailleurs qu'en moyenne et basse fréquence, la corrélation spatiale comme le niveau moyen baisse à partir d'une fréquence qui augmente avec l'ordre et de façon inversement proportionnelle au paramètre de régularisation. Ces phénomènes résultent à la fois des erreurs d'estimation et du processus de régularisation qui « coupe » les bandes fréquentielles les plus bruitées. Dans le cas où le paramètre λ vaut $2,10 \cdot 10^{-3}$ (courbes en traits pleins), la dégradation de la corrélation spatiale correspond assez bien à la baisse du niveau moyen des composantes harmoniques sphériques, ce qui atteste que le choix d'un tel paramètre est cohérent : le compromis

entre la préservation des informations utiles et l'élimination des informations erronées est quasiment optimal. Lorsque le paramètre λ est plus élevé (traits en pointillés), les composantes harmoniques de même ordre sont atténuées à partir d'une fréquence plus haute dans le spectre, ce qui écarte un certain nombre d'informations utiles enregistrées par les capteurs. À l'inverse, un paramètre λ trop grand (non représenté ici) préserve toutes les informations utiles mais n'élimine pas assez les bandes fréquentielles basses et moyennes bruitées, qui se retrouvent finalement très amplifiées et rendent les composantes harmoniques résultantes inexploitable (cf. Section 2.2).

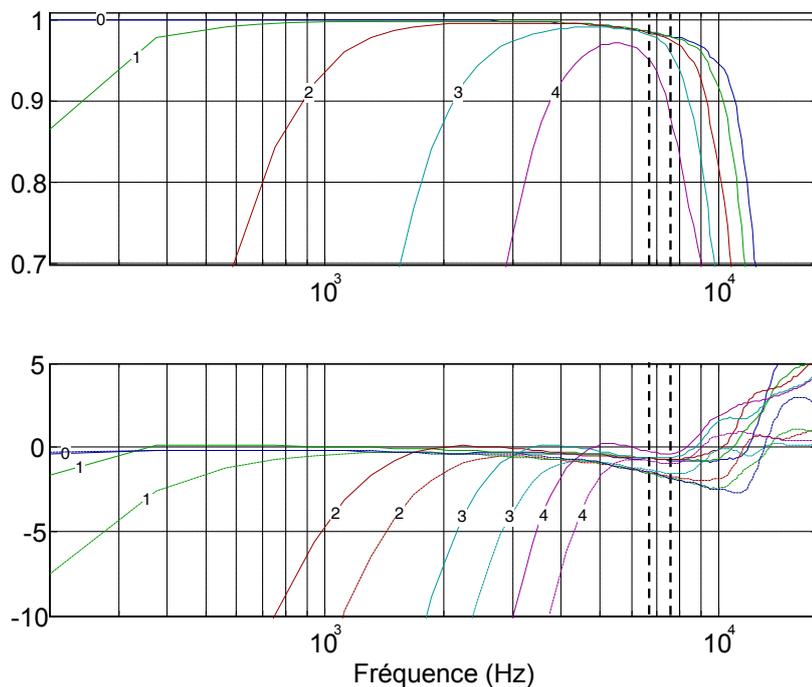


Figure 5.16 – Corrélation spatiale (haut) entre les directivités harmoniques sphériques mesurées et théoriques, et niveau moyen relatif (bas) pour deux valeurs distinctes du paramètre de régularisation : $\lambda = 2.10^{-3}$ (traits pleins) et $\lambda = 7,9.10^{-2}$ (en pointillés). Notons que ce paramètre influe seulement sur le niveau moyen relatif.

Références bibliographiques du Chapitre 5

- [CG77] CRAVEN, P. et GERZON, M.A. *Coincident Microphone Simulation Covering Three Dimensional Space and Yielding Various Directionnal Outputs*. U.S. Patent 4,042,779, août 1977.
- [F79a] FARRAR, K. *Soundfield Microphone*. *Wireless World*, 1979, octobre, p. 48-50.
- [F79b] FARRAR, K. *Soundfield Microphone - 2*. *Wireless World*, 1979, novembre, p. 99-103.
- [M01a] MEYER, J. *Beamforming for a circular microphone array mounted on spherically shaped objects*. *J. Acoust. Soc. Am.*, 2001, Vol. 109, n° 1, p. 185-193.
- [ME02] MEYER, J. et ELKO, G. *A Highly Scalable Spherical Microphone Array Based on an Orthonormal Decomposition of the Soundfield*. *IEEE ICASSP '02*, Vol. 2, p. 1781-1784, mai 2002.
- [BR04] BALMAGES, I. et RAFAELY, B. *Room Acoustics Measurements by Microphone Arrays*. 23rd IEEE convention of Electrical and Electronics Engineers in Israel, p. 420-423, Israel, septembre 2004.
- [G01] GOVER, B.N. *Development and Use of a Spherical Microphone Array for Measurement of Spatial Properties of Reverberant Sound Fields*. Ph. D. Thesis, Waterloo: University of Waterloo, 2001.
- [G05] GOVER, B.N. *Directional Measurement of Airborne Sound Transmission Paths Using a Spherical Microphone Array*. *J. Audio Eng. Soc.*, 2005, Vol. 53, n° 9, p. 787-195.
- [GRS02] GOVER, B.N., RYAN, J.G., et STINSON, M.R. *Microphone array measurement system for analysis of directional and spatial variations of sound fields*. *J. Acoust. Soc. Am.*, 2002, Vol. 112, n° 5, Pt. 1, p. 1980-1991.
- [GRS03] GOVER, B.N., RYAN, J.G., et STINSON, M.R. *Designing a spherical microphone array for the directional analysis of reflections and reverberation*. AES 115th Convention, preprint 5862, New York, NY, USA, 10-13 octobre 2003.
- [GRS04] GOVER, B.N., RYAN, J.G., et STINSON, M.R. *Measurements of directional properties of reverberant sound fields in rooms using a spherical microphone array*. *J. Acoust. Soc. Am.*, 2004, Vol. 116, n° 4, Pt. 1, p. 2138-2148.
- [BA05a] BETLEHEM, T. et ABHAYAPALA, T.D. *Theory and design of sound field reproduction in reverberant rooms*. *J. Acoust. Soc. Am.*, 2005, Vol. 117, n° 4, p. 2100-2111.
- [BA05b] BETLEHEM, T. et ABHAYAPALA, T.D. *A Modal Approach to Soundfield Reproduction in Reverberant Rooms*. *IEEE ICASSP '05*, Vol. 3, p. 289-292, Philadelphie, PA, USA, 18-23 mars 2005.

- [L03] LABORIE, A., BRUNO, R., et MONTROYA, S. *A New Comprehensive Approach of Surround Sound Recording*. AES 114th Convention, preprint 5717, Amsterdam, The Netherlands, 22-25 mars 2003.
- [E00a] ELKO, G., *Chap. 10. Superdirectional Microphone Arrays*, in *Acoustic Signal Processing for Telecommunication*. 2000. Boston: Kluwer Academic Publishers. p. 181-238.
- [DK99] DICKINS, G. et KENNEDY, R.A. *Towards Optimal Sound Field Representation*. 106th AES Convention, Munich, 1999.
- [ENP97] ELKO, G. et NGUYEN PONG, A.-T. *A steerable and variable first-order differential microphone*. Proc. IEEE ICASSP'97, Vol. 1, p. 223-226, Munich, Allemagne, 21-24 avril 1997.
- [MA03] MEYER, J. et AGNELLO, T. *Spherical microphone array for spatial sound recording*. AES 115th Convention, preprint 5975, New York, NY, USA, 10-13 octobre 2003.
- [LD05a] LI, Z. et DURAISWAMI, R. *A Robust and Self-Reconfigurable Design of Spherical Microphone Array for Multi-Resolution Beamforming*. Proc. IEEE ICASSP '05, Vol. 4, p. 1137-1140, Philadelphie, PA, USA, 18-23 mars 2005.
- [MI68] MORSE, P.M. et INGARD, K.U. *Theoretical Acoustics*. New York: Mc Graw-Hill, 1968, 927 p.
- [B98] BRUNEAU, M. *Manuel d'acoustique fondamentale*. Hermès, 1998p.
- [GVL96] GOLUB, G. et VAN LOAN, C. *Matrix Computations, 3rd edition*. London: The Johns Hopkins University Press, 1996, 694 p.
- [DH94] DRISCOLL, J.R. et HEALY, D.M. *Computing Fourier Transforms and Convolutions on the 2-Sphere*. Adv. Appl. Math., 1994, Vol. 15, p. 202-250.
- [A85] ARFKEN, G. *Mathematical Methods for Physicists, 3rd ed.* San Diego, CA: Academic Press, 1985, 985 p.
- ARFKEN, G. et WEBER, H.J. *Mathematical Methods for Physicists, 5th ed.* San Diego, CA: Academic, 2001p.
- [V01] VERHAEVERT, J., VAN LIL, E., et VAN DE CAPELLE, A. *Uniform Spherical Distributions For Adaptive Array Applications*. Vehicular Technology Conference (VTC'01), Vol. 1, p. 98-102, Rhodes, Grèce, 6-9 mai 2001.
- [G73] GERZON, M.A. *Periphony: With-Height Sound Reproduction*. J. Audio Eng. Soc., 1973, Vol. 21, n° 1, p. 2-10.
- [C02] COTTERELL, P.S. *On the Theory of the Second-Order Soundfield Microphone*. Ph. D. Thesis, Reading, UK: The University of Reading, 2002.

-
- [S06] SLOANE, N.J.A., et al. *Spherical Codes*. <http://www.research.att.com/~njas/packings/> (dernier accès le 26 mai 2006)
- [F06] FLIEGE, J. *Integration nodes for the sphere*. <http://web.mat.bham.ac.uk/J.Fliege/nodes/nodes.html> (dernier accès le 26 mai 2006)
- [D05] DURAISWAMI, R., et al. *High Order Spatial Audio Capture and its Binaural Head-Tracked Playback over Headphones with HRTF Cues*. AES 119th Convention, preprint 6540, New York, NY, USA, 7-10 octobre 2005.
- [TA77] TIKHONOV, A.N. et ARSENIN, V.A. *Solution of Ill-posed Problems*. Washington: Winston & Sons, 1977p.
- [KN96] KIRKEBY, O. et NELSON, P.A. *Fast Deconvolution of Multi-Channel Systems using Regularisation*. Institute of Sound and Vibration Research (ISVR) Technical Report n° 255, Southampton, UK, 1996.
- [FU99] FARINA, A. et UGOLOTTI, E. *Subjective comparison between Stereo Dipole and 3D Ambisonics surround systems for automotive applications*. AES 16th International Conference on Spatial Sound Reproduction, Rovaniemi, Finlande, 10-12 avril 1999.
- [Mat06] THE MATH WORKS INC. *Matlab, version 7.2*. 2006.

Chapitre 6

Conclusions et perspectives d'avenir

Les objectifs fixés au cours de ces travaux de thèse concernaient la conception d'outils d'encodage spatial associé à la technologie *Higher Order Ambisonics* dans le but de favoriser le développement de cette technique de spatialisation sonore, selon nous prometteuse dans un secteur multimédia en plein essor.

Nous avons dans ce sens proposé une méthode de conception de filtres permettant la reconstruction du champ proche des sources virtuelles, potentiellement utiles dans les applications de réalité virtuelle ou les jeux vidéo par exemple. Cette méthode se base sur le concept de bande fréquentielle utile qui relativise l'intérêt des signaux HOA dans certaines bandes de fréquences pour une zone de reconstruction visée. Les « bandes inutiles » dont la prise en compte est problématique car nécessitant d'être énormément amplifiées, peuvent ainsi être supprimées sans dégrader la représentation de façon apparente ou excessive. Nous avons pour cela proposé un critère dit de *focalisation d'énergie* pour définir objectivement ces bandes inutiles, et qui aboutit à un bon compromis entre l'effet de champ proche restitué et l'énergie déployée pour y arriver.

Dans un second temps, nous avons étudié l'enregistrement HOA en 3D d'ordres supérieurs à 1 à partir d'un dispositif sphérique de microphones. Les signaux HOA sont ainsi obtenus après *matriçage* (combinaison) des signaux microphoniques de façon à reconstruire les directivités associées aux harmoniques sphériques, et *égalisation* des signaux résultants pour compenser la directivité intrinsèque du dispositif de captation (directivité des capsules et diffraction du son par la structure). Les points critiques de l'échantillonnage spatial et surtout de l'égalisation spectrale citée précédemment ont fait l'objet d'une attention particulière. Cette dernière égalisation implique théoriquement une amplification dans un domaine basse fréquence des signaux HOA qui s'avère être excessive notamment à cause du bruit de fond des capsules microphoniques, et qui traduit l'effort fourni par le microphone HOA pour estimer des informations du champ acoustique hors de sa portée. Pour résoudre ce problème, nous avons proposé une méthode de *régularisation* des

filtres en question dont le but est de préserver au maximum les informations utiles du champ sonore enregistrées tout en éliminant les données bruitées qui en compromettent l'exploitation. Cette étude a conduit à la réalisation d'un microphone HOA d'ordre 4, construit à partir d'un ensemble de 32 capteurs à directivité omnidirectionnelle répartis à la surface d'une sphère rigide. Ce microphone a fait l'objet d'une campagne de mesures effectuées en chambre anéchoïque et en trois dimensions. Nous avons ainsi pu évaluer de façon objective ses performances réelles en termes d'estimation des composantes harmoniques sphériques, que nous avons par ailleurs confrontées aux performances théoriques du modèle mathématique associé.

Les travaux de thèse rapportés dans le présent document méritent d'être poursuivis dans plusieurs directions. En ce qui concerne l'encodage de source virtuelle à proximité de l'auditeur, il resterait notamment à implanter efficacement les filtres de champ proche puis à mener une campagne de tests subjectifs pour valider définitivement les résultats et simulations théoriques présentées au Chapitre 4, et prouver l'utilité de la reproduction du champ proche des sources sonores dans les jeux vidéo ou les applications de réalité virtuelles par exemple.

Quant à l'enregistrement HOA naturel, les perspectives d'évolution de la méthode proposée pour l'estimation des signaux HOA concernent notamment le choix du paramètre de régularisation. En effet, lorsque la diffusion de l'enregistrement ne se fait pas en temps réel comme la production d'un disque par exemple, il est probablement possible d'affiner avec une méthode de type *L-Curve* ce choix du paramètre à partir de la connaissance préalable des signaux captés par les microphones. Lorsque le paramètre de régularisation est optimal, la limitation des performances d'estimation des signaux HOA s'explique essentiellement par des contraintes matérielles : taille réduite du microphone sphérique, bruit de fond et nombre insuffisant des capteurs. L'amélioration de l'enregistrement HOA passe désormais par la réalisation d'une structure microphonique plus large dans le but d'améliorer les performances d'estimation basse fréquence, et avec plus de capteurs pour limiter le phénomène d'*aliasing* spatial. D'autres stratégies d'enregistrement peuvent également être utilisées comme par exemple la construction de multiples sphères microphoniques concentriques, actives dans des bandes de fréquences distinctes.

Enfin, des écoutes informelles et comparatives nous ont permis de confirmer l'intérêt d'utiliser des signaux HOA d'ordres supérieurs à 1 pour permettre à un auditeur, et à plus forte raison plusieurs, de percevoir l'espace sonore reconstruit avec toujours plus de précision. Ces écoutes ont notamment permis de mesurer l'apport des ordres 2 à 4 de notre prototype microphonique en comparant ses performances avec celles du microphone *Soundfield* (ordre 1). Il resterait cependant à réaliser une série de tests subjectifs pour évaluer de façon plus précise et rigoureuse l'impact des différents ordres sur la précision du rendu spatial et apporter des éléments de réponse à des questions d'ordre pratique : par exemple, quel ordre maximal est-il nécessaire de prendre en compte pour une restitution HOA binaurale au casque d'écoute ? Ou encore dans le but de lier la perception auditive de l'espace à un critère objectif d'erreur de reconstruction du champ

acoustique : quelle valeur de la tolérance d'erreur de reconstruction du champ sonore pouvons-nous accepter sans trop dégrader la perception spatiale ?

D'un point de vue plus général, la technologie HOA nous semble désormais assez « mûre » pour être utilisée dans les applications grand public. Ses avantages, notamment sur la stéréophonie, ne sont plus à démontrer : description rationnelle du champ acoustique, manipulable, relativement concise, compatible avec tous les systèmes de diffusion existants et tournée vers l'avenir. Cependant, le succès d'une technologie n'est pas seulement déterminé par sa supériorité technique mais également par d'autres facteurs tels que le contexte, la facilité d'emploi, le coût de production et d'équipement des usagers, la quantité et la qualité des contenus proposés. Si la technologie HOA pêche actuellement sur ce dernier point, le contexte de développement du secteur audiovisuel, l'engouement actuel pour la spatialisation sonore et la demande croissante des utilisateurs de plus en plus habitués à l'interactivité offerte par Internet en contenus toujours plus souples et mobiles semblent en revanche favorables à son développement. D'autre part, l'intégration de décodeurs HOA dans les appareils multimédias (chaînes hi-fi, lecteurs de DVD, postes de télévision, autoradios, etc.) sous la forme de puce électronique ou de DSP (*Digital Signal Processor*) à la manière des décodeurs *Dolby Prologic*, *Dolby Digital* ou *DTS (Digital Theatre System)* paraît relativement aisée, le décodage consistant simplement à effectuer des opérations simples de matricage, voire de filtrage sur des signaux audio en fonction du dispositif de restitution choisi. Enfin, pour tenter d'émerger dans un paysage audiovisuel dominé quasi-exclusivement par la stéréophonie, la technologie HOA doit à présent proposer des contenus attrayants, ce qui sous-entend le développement en amont d'outils diversifiés, simples à utiliser et largement diffusables de création et de mixage sonore. À bon entendre...

Bibliographie

- [AS64] ABRAMOWITZ, M. et STEGUN, I. A. *Handbook of Mathematical Functions*. Washington: National Bureau of Standards, 1964, 1046 p.
- [A01] ALGAZI, V. R., et al. *The CIPIC HRTF Database*. IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics, New York, USA, 21-24 octobre 2001.
- [A85] ARFKEN, G. *Mathematical Methods for Physicists, 3rd ed.* San Diego, CA: Academic Press, 1985, 985 p.
- [BR04] BALMAGES, I. et RAFAELY, B. *Room Acoustics Measurements by Microphone Arrays*. 23rd IEEE convention of Electrical and Electronics Engineers in Israel, p. 420-423, Israel, septembre 2004.
- [B95] BAMFORD, J. S. *An Analysis of Ambisonics Sound Systems of First and Second Order*. M. Sc. Thesis, Waterloo: University of Waterloo, 1995.
- [BA05a] BETLEHEM, T. et ABHAYAPALA, T.D. *Theory and design of sound field reproduction in reverberant rooms*. J. Acoust. Soc. Am., 2005, Vol. 117, n° 4, p. 2100-2111.
- [BA05b] BETLEHEM, T. et ABHAYAPALA, T.D. *A Modal Approach to Soundfield Reproduction in Reverberant Rooms*. IEEE ICASSP '05, Vol. 3, p. 289-292, Philadelphie, PA, USA, 18-23 mars 2005.
- [BVV93] BERKHOUT, A.J., DE VRIES, D., et VOGEL, P. *Acoustic control by wave field synthesis*. J. Acoust. Soc. Am., 1993, Vol. 93, n° 5, p. 2764-2778.
- [B97] BLAUERT, J. *Spatial Hearing. The Psychophysics of Human Sound Localization*. revised ed., Cambridge, Massachusetts: The MIT Press, 1997, 494 p.
- [B98] BRUNEAU, M. *Manuel d'acoustique fondamentale*. Paris: Hermès, 1998, 576 p.
- [Bg98] BRUNGART, D. S. *Near-Field Auditory Localisation*. Ph. D. Thesis, Cambridge, MA: Massachusetts Institute of Technology, 1998.

- [BR99] BRUNGART, D. S. et RABINOWITZ, W. M. *Auditory localization of nearby sources. Head-related transfer functions*. J. Acoust. Soc. Am., 1999, Vol. 106, n° 3, pt. 1, p. 1465-1479.
- [BDR99] BRUNGART, D. S., DURLACH, N. I., et RABINOWITZ, W. M. *Auditory localization of nearby sources. II. Localization of a broadband source*. J. Acoust. Soc. Am., 1999, Vol. 106, n° 4, (Pt. 1), p. 1956-1968.
- [C90] CANÉVET, G., Chap. 3. Audition binaurale et localisation auditive. Aspects physiques et psychoacoustiques, in *Psychoacoustique et perception auditive*. 1989, Paris : INSERM. p. 83-122.
- [C78] CONDAMINES, R. *Stéréophonie - Cours de relief sonore théorique et appliqué*. Paris : Masson, 1978, 301 p.
- [CS71] COOPER, D. H. et SHIGA, T. *Discrete-Matrix Multichannel Stereo*. J. Audio Eng. Soc., 1971, Vol. 20, n° 5, p. 346-360.
- [C02] COTTERELL, P.S. *On the Theory of the Second-Order Soundfield Microphone*. Ph. D. Thesis, Reading, UK: The University of Reading, 2002.
- [CG77] CRAVEN, P. et GERZON, M. A. *Coincident Microphone Simulation Covering Three Dimensional Space and Yielding Various Directional Outputs*. U.S. Patent 4,042,779, août 1977.
- [C03] CRAVEN, P. *Continuous Surround Panning for 5-Speaker Reproduction*. AES 24th International Conference on Multichannel Audio, Banff, Canada, mai 2003.
- [DRP98] DANIEL, J., RAULT, J.-B., et POLACK, J.-D. *Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions*. AES 105th Convention, preprint 4795, 26-29 septembre 1998.
- [DRP99] DANIEL, J., RAULT, J.-B., et POLACK, J.-D. *Acoustic Properties and Perceptives Implications of Stereophonic Phenomena*. AES 16th International Conference on Spatial Sound Reproduction, Rovaniemi, Finlande, 10-12 avril 1999.
- [D00] DANIEL, J. *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. Thèse de Doctorat, Paris: Université de Paris 6, 2000.
- [DNM03] DANIEL, J., NICOL, R., et MOREAU, S. *Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging*. 114th AES Convention, preprint 5788, Amsterdam, Pays-Bas, 22-25 mars 2003.
- [D03] DANIEL, J. *Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format*. AES 23rd International Conference, Helsingor, Danemark, 23-25 mai 2003.
- [DM04] DANIEL, J. et MOREAU, S. *Further Study of Sound Field Coding with Higher Order Ambisonics*. AES 116th Convention, preprint 6017, Berlin, Allemagne, 8-11 mai 2004.

- [DK99] DICKINS, G. et KENNEDY, R. A. *Towards Optimal Sound Field Representation*. 106th AES Convention, Munich, 1999.
- [DH94] DRISCOLL, J. R. et HEALY, D. M. *Computing Fourier Transforms and Convolutions on the 2-Sphere*. Adv. Appl. Math., 1994, Vol. 15, p. 202-250.
- [D05] DURAISWAMI, R., et al. *High Order Spatial Audio Capture and its Binaural Head-Tracked Playback over Headphones with HRTF Cues*. AES 119th Convention, preprint 6540, New York, NY, USA, 7-10 octobre 2005.
- [ENP97] ELKO, G. et NGUYEN PONG, A.-T. *A steerable and variable first-order differential microphone*. Proc. IEEE ICASSP'97, Vol. 1, p. 223-226, Munich, Allemagne, 21-24 avril 1997.
- [E00a] ELKO, G., Chap. 10. *Superdirectional Microphone Arrays*, in *Acoustic Signal Processing for Telecommunication*. 2000. Boston: Kluwer Academic Publishers. p. 181-238.
- [FU98] FARINA, A. et UGOLOTTI, E. *Software Implementation of B-Format Encoding And Decoding*. AES 104th Convention, preprint 4691, Amsterdam, Pays-Bas, 15-20 mai 1998.
- [FU99] FARINA, A. et UGOLOTTI, E. *Subjective comparison between Stereo Dipole and 3D Ambisonics surround systems for automotive applications*. AES 16th International Conference on Spatial Sound Reproduction, Rovaniemi, Finlande, 10-12 avril 1999.
- [F79a] FARRAR, K. *Soundfield Microphone*. Wireless World, 1979, octobre, p. 48-50.
- [F79b] FARRAR, K. *Soundfield Microphone – 2*. Wireless World, 1979, novembre, p. 99-103.
- [F74] FELLGETT, P. B. *Ambisonic reproduction of directionality in surround-sound systems*. Nature, 1974, Vol. 252, p. 534-538.
- [F94] FILIPPI, P.J.T. *Acoustique générale*. Paris: Les éd. de physique, Société française d'acoustique, 1994, 382 p.
- [F06] FLIEGE, J. *Integration nodes for the sphere*. <http://web.mat.bham.ac.uk/J.Fliege/nodes/nodes.html> (dernier accès le 26 mai 2006)
- [F99] FURSE, R. *3D Audio Links and Information*. 1999. <http://www.muse.demon.co.uk/3daudio.html> (dernier accès le 23 mai 2006)
- [GM95] GARDNER, B. et MARTIN, K. *HRTF Measurements of a KEMAR*. J. Acoust. Soc. Am., 1995, Vol. 97, p. 3907-3908.
- [G73] GERZON, M. A. *Periphony: With-Height Sound Reproduction*. J. Audio Eng. Soc., 1973, Vol. 21, n° 1, p. 2-10.
- [G77] GERZON, M.A. *The Optimum Choice of Surround Sound Encoding Specification*. AES 56th Convention, preprint 1199, Paris, France, 1977.

- [G85] GERZON, M. A. *Ambisonics in Multichannel Broadcasting and Video*. J. Audio Eng. Soc., 1985, Vol. 33, n° 11, p. 859-871.
- [GB92] GERZON, M. A. et BARTON, G. J. *Ambisonic Decoders for HDTV*. AES 92nd Convention, preprint n° 3345, Vienne, Autriche, 24-27 mars 1992.
- [G92b] GERZON, M. A. *Psychoacoustic Decoders for Multispeaker Stereo and Surround-Sound*. AES 93rd Convention, preprint 3406, San Francisco, USA, octobre 1992.
- [G92d] GERZON, M. A. *General Metatheory of Auditory Localisation*. AES 92nd Convention, preprint 3306, Vienne, Autriche, 23-29 mars 1992.
- [GI03] GERZONIC INDUSTRIES. *3D-Audio Software*. 2003. <http://www.gerzonic.net> (dernier accès le 28 mai 2006)
- [GVL96] GOLUB, G. et VAN LOAN, C. *Matrix Computations, 3rd edition*. London: The Johns Hopkins University Press, 1996, 694 p.
- [G01] GOVER, B. N. Development and Use of a Spherical Microphone Array for Measurement of Spatial Properties of Reverberant Sound Fields. Ph. D. Thesis, Waterloo: University of Waterloo, 2001.
- [GRS02] GOVER, B. N., RYAN, J. G., et STINSON, M. R. *Microphone array measurement system for analysis of directional and spatial variations of sound fields*. J. Acoust. Soc. Am., 2002, Vol. 112, n° 5, Pt. 1, p. 1980-1991.
- [GRS03] GOVER, B. N., RYAN, J. G., et STINSON, M. R. *Designing a spherical microphone array for the directional analysis of reflections and reverberation*. AES 115th Convention, preprint 5862, New York, NY, USA, 10-13 octobre 2003.
- [GRS04] GOVER, B. N., RYAN, J.G., et STINSON, M.R. *Measurements of directional properties of reverberant sound fields in rooms using a spherical microphone array*. J. Acoust. Soc. Am., 2004, Vol. 116, n° 4, Pt. 1, p. 2138-2148.
- [G05] GOVER, B. N. Directional Measurement of Airborne Sound Transmission Paths Using a Spherical Microphone Array. J. Audio Eng. Soc., 2005, Vol. 53, n° 9, p. 787-195.
- [G96] GROEMER, H. *Geometric Applications of Fourier Series and Spherical Harmonics*. Cambridge: Cambridge University Press, 1996, 343 p.
- [JLP99] JOT, J.-M., LARCHER, V., et PERNAUX, J.M. *A Comparative Study of 3-D Audio Encoding and Rendering Techniques*. AES 16th International Conference on Spatial Sound Reproduction, Rovaniemi, Finlande, 10-12 avril 1999.
- [KN96] KIRKEBY, O. et NELSON, P. A. *Fast Deconvolution of Multi-Channel Systems using Regularisation*. Institute of Sound and Vibration Research (ISVR) Technical Report n° 255, Southampton, UK, 1996.

- [L03] LABORIE, A., BRUNO, R., et MONTOYA, S. *A New Comprehensive Approach of Surround Sound Recording*. AES 114th Convention, preprint 5717, Amsterdam, The Netherlands, 22-25 mars 2003.
- [LD05a] LI, Z. et DURAISWAMI, R. *A Robust and Self-Reconfigurable Design of Spherical Microphone Array for Multi-Resolution Beamforming*. Proc. IEEE ICASSP '05, Vol. 4, p. 1137-1140, Philadelphie, PA, USA, 18-23 mars 2005.
- [M62] MAKITA, Y. *On the Directional Localization of Sound in the Stereophonic Sound Field*. E.B.U. Review, 1962, Vol. A, n° 73, p. 102-108.
- [M90] MALHAM, D. G. *Ambisonics - A Technique for Low Cost, High Precision, Three Dimensional Sound Diffusion*. International Computer Music Conference, p. 118-120, Glasgow, Scotland, septembre 1990.
- [M92] MALHAM, D. G. *Experience With Large Area 3-D Ambisonic Sound Systems*. Proc. of the Institute of Acoustics, Vol. 14, p. 209-215, 1992.
- [M95] MALHAM, D. G. *Basic Ambisonics*. 1995. http://www.york.ac.uk/inst/mustech/3d_audio/ambsnlct.htm (dernier accès le 23 mai 2006)
- [MF03] MALHAM, D. et FIELD, A. *VST Ambisonic Tools*. York University Music Technology Group, York, 2003. http://www.york.ac.uk/inst/mustech/3d_audio/vst/welcome.html (dernier accès le 28 mai 2006)
- [M05] MALHAM, D. G. *Second and Third Order Ambisonics - the Furse-Malham Set*. 2002. http://www.york.ac.uk/inst/mustech/3d_audio/seconдор.html#ref2 (dernier accès le 23 mai 2006)
- [Mat06] THE MATH WORKS INC. *Matlab, version 7.2*. 2006, <http://www.mathworks.com> (dernier accès le 26 mai 2006)
- [M01a] MEYER, J. *Beamforming for a circular microphone array mounted on spherically shaped objects*. J. Acoust. Soc. Am., 2001, Vol. 109, n° 1, p. 185-193.
- [ME02] MEYER, J. et ELKO, G. *A Highly Scalable Spherical Microphone Array Based on an Orthonormal Decomposition of the Soundfield*. IEEE ICASSP '02, Vol. 2, p. 1781-1784, mai 2002.
- [MA03] MEYER, J. et AGNELLO, T. *Spherical microphone array for spatial sound recording*. AES 115th Convention, preprint 5975, New York, NY, USA, 10-13 octobre 2003.
- [Mo92] MOLLER, H. *Fundamentals of Binaural Technology*. Applied Acoustics, 1992, Vol. 36, p. 171-218.
- [M00] MONRO, G. *In-phase corrections for Ambisonics*. Proceedings of the International Computer Music Conference (ICMC), p. 292-295, Berlin, Allemagne, 27 août -1er septembre 2000.

- [MD04] MOREAU, S. et DANIEL, J. *Study of Higher Order Ambisonic Microphone*. Congrès commun CFA-DAGA'04, Strasbourg, France, 22-25 mars 2004.
- [MDC04] MOREAU, S., DANIEL, J., et CHRAA, A. Production, transmission et restitution temps réel d'une scène sonore dans un format audio 3D flexible. CORESA 2004, Lille, France, 25-26 mai 2004.
- [MDB06] MOREAU, S., DANIEL, J., et BERTET, S. 3D Sound Field Recording with Higher Order Ambisonics - Objective Measurements and Validation of Spherical Microphone. AES 120th Convention, preprint 6857, Paris, France, 20-23 mai 2006.
- [MI68] MORSE, P. M. et INGARD, K. U. *Theoretical Acoustics*. New York: Mc Graw-Hill, 1968, 927 p.
- [N99] NICOL, R. Restitution sonore spatialisée sur une zone étendue : application à la téléprésence. Thèse de Doctorat, Le Mans: Université du Mans, 1999.
- [P00] PASQUINI, L. Accurate computation of the zeros of the generalized Bessel polynomials. *Numerische Mathematik*, 2000, Vol. 86, p. 507-538.
- [PBJ98] PERNAUX, J.-M., BOUSSARD, P., et JOT, J.M. *Virtual Sound Source Positioning and Mixing in 5.1 Implementation on the Real-Time System Genesis*. DAFX'98, Barcelone, Espagne, 19-21 novembre 1998.
- [P96] POLETTI, M. A. The Design of Encoding Functions for Stereophonic and Polyphonic Sound Systems. *J. Audio Eng. Soc.*, 1996, Vol. 44, n° 11, p. 948-963.
- [P00] POLETTI, M. A. *A Unified Theory of Horizontal Holographic Sound Systems*. *J. Audio Eng. Soc.*, 2000, Vol. 48, n° 12, p. 1155-1182.
- [P97] PULKKI, V. *Virtual Sound Source Positioning Using Vector Base Amplitude Panning*. *J. Audio Eng. Soc.*, 1997, Vol. 45, n° 6, p. 456-466.
- [R89] REINHARD, H. *Cours de mathématique du signal*. Paris: Dunod Université, 1989.
- [S06] SLOANE, N. J. A., et al. *Spherical Codes*. <http://www.research.att.com/~njas/packings/> (dernier accès le 26 mai 2006)
- [SH01] SONTACCHI, A. et HÖLDRICH, R. *Further Investigations on 3D Sound Fields Using Distance Coding*. DAFX-01, Limerick, Irlande, 6-8 décembre 2001.
- [SH02] SONTACCHI, A. et HÖLDRICH, R. *Distance Coding in 3D Sound Field*. AES 21st Conference, St. Petersburg, Russie, 1er-3 juin 2002.
- [TA77] TIKHONOV, A. N. et ARSEININ, V. A. *Solution of Ill-posed Problems*. Washington: Winston & Sons, 1977.
- [V97] VERHEIJEN, E.N.G. *Sound reproduction by Wave Field Synthesis*. Ph. D. Thesis, Delft: Delft University of Technology, 1997.

-
- [V01] VERHAEVERT, J., VAN LIL, E., et VAN DE CAPELLE, A. *Uniform Spherical Distributions For Adaptive Array Applications*. Vehicular Technology Conference (VTC'01), Vol. 1, p. 98-102, Rhodes, Grèce, 6-9 mai 2001.
- [WA01a] WARD, D. B. et ABHAYAPALA, T. D. *Performance Bounds on Sound Field Reproduction Using a Loudspeaker Array*. IEEE WASPAA 2001, New York, octobre 2001.
- [WA01b] WARD, D. B. et ABHAYAPALA, T. D. *Reproduction of a Plane-Wave Sound Field Using an Array of Loudspeakers*. IEEE Transaction on Speech and Audio Processing, 2001, Vol. 9, n° 6, p. 697-707.
- [W02] WARUSFEL, O. *LISTEN HRTF database*. 2002. <http://recherche.ircam.fr/equipes/salles/listen/> (dernier accès le 25 mai 2006)
- [W03] WIGGINS, B., et al. *The Design and Optimisation of Surround Sound Decoders Using Heuristic Methods*. Proceedings of UKSim 2003, Conference of the UK Simulation Society, p. 106-114, 2003.
- [W99] WILLIAMS, E. G. *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. London, UK: Academic Press, 1999, 306 p.

Annexes

Publications

DANIEL, J., NICOL, R., et MOREAU, S. *Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging*. 114th AES Convention, preprint 5788, Amsterdam, Pays-Bas, 22-25 mars 2003.

MOREAU, S. et DANIEL, J. *Study of Higher Order Ambisonic Microphone*. Congrès commun CFA-DAGA'04, Strasbourg, France, 22-25 mars 2004.

DANIEL, J. et MOREAU, S. *Further Study of Sound Field Coding with Higher Order Ambisonics*. AES 116th Convention, preprint 6017, Berlin, Allemagne, 8-11 mai 2004.

MOREAU, S., DANIEL, J., et CHRAA, A. *Production, transmission et restitution temps réel d'une scène sonore dans un format audio 3D flexible*. CORESA 2004, Lille, France, 25-26 mai 2004.

MOREAU, S., DANIEL, J., et BERTET, S. *3D Sound Field Recording with Higher Order Ambisonics - Objective Measurements and Validation of a 4th Spherical Microphone*. AES 120th Convention, preprint 6857, Paris, France, 20-23 mai 2006.



Audio Engineering Society Convention Paper

Presented at the 114th Convention
2003 March 22–25 Amsterdam, The Netherlands

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

Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging

Jérôme Daniel¹, Rozenn Nicol¹, and Sébastien Moreau¹

¹France Telecom R&D, 2 Avenue Pierre Marzin, 22307 Lannion Cedex, France

jerome.daniel@francetelecom.com, rozenn.nicol@francetelecom.com, sebastien.moreau@francetelecom.com

ABSTRACT

Ambisonics and Wavefield Synthesis are two ways of rendering 3D audio, which both aim at physically reconstructing the sound field. Though they derive from distinct theoretical fundamentals, they have already been shown as equivalent under given assumptions. This paper further discusses their relationship by introducing new results regarding the coding and rendering of finite distance and enclosed sources. An updated view of the current knowledge is first given. A unified analysis of sound pickup and reproduction by mean of concentric transducer arrays then provides an insight into the spatial encoding and decoding properties. While merging the analysis tools of both techniques and investigating them on a common ground, general compromises are highlighted in terms of spatial aliasing, error and noise amplification.

1. INTRODUCTION

Among sound spatialisation technologies, both Wavefield Synthesis (WFS) and High Order Ambisonics (HOA) aim at physically reconstructing the sound field, though they historically belong to distinct worlds. Whereas WFS is considered as *the* solution for providing large listening areas, Ambisonics is originally known as dedicated to surround systems having a limited sweet spot. Nevertheless, the latter's extension to higher spatial resolutions (HOA) has known an increasing interest during past years, featuring scalability and flexibility properties in addition to enlarged listening areas.

The present paper provides new insights on WFS and HOA by analysing them on a common ground. It first

supplies an updated state of art (theory and application). A bigger part is dedicated to HOA, to reflect recent progresses that allow it being practically compared with WFS. Then both technologies are investigated side-by-side. After completing the formal connection between their associated sound field representations, we list the reconstruction artefacts expected from practical system limitations. Next, virtual sound imaging simulations help comparing reconstruction properties, and characterising them in terms of spatial information consistency and plausible localisation effect. Enlarging considerations to "real" recording issues (*i.e.* involving microphone arrays), artefacts appear to be shared by both approaches since these obey the same practical limitations. Finally we derive preferences on encoding strategies, and compromises

on technical choices (array size).

To perform the comparison more efficiently, the scope of this paper focuses on concentric (*i.e.* circular or spherical), regular arrays. As a global result, a converging view of the two approaches is given, and a piece of "physical feeling" is offered to intuitively understand underlying phenomena.

2. WAVE FIELD SYNTHESIS (WFS)

2.1. Huygens' Principle

The Wave Field Synthesis is a concept of spatialised sound reproduction that was proposed by Berkhout in the late 80's [1, 2]. It may be identified to the acoustical equivalent to holography, and for this reason, it is sometimes referred to as "holophony" [3]. Indeed, WFS aims at reproducing sound waves (and especially the wave front curvature) by loudspeaker array. Physically, it is derived from the Huygens' Principle, and more precisely, from the idea, that a wave front may be considered as a secondary source distribution. In other words, the wave, which propagates from a given wave front, may be considered as emitted either by the original sound source (the primary source) or by a secondary source distribution along the wave front. As a consequence, the secondary source distribution may be substituted for the primary source, in order to reproduce the primary sound field.

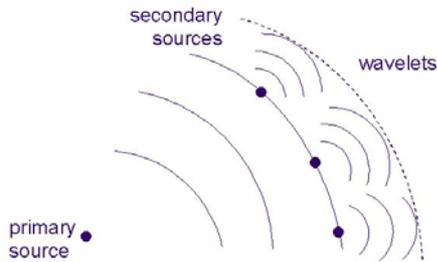


Figure 1 Illustration of the Huygen's Principle

2.2. Kirchhoff-Helmholtz Integral

The Kirchhoff-Helmholtz Integral expresses this idea in a mathematical way. The acoustical pressure p within a given area A is derived from the knowledge of the acoustical pressure p_0 and its gradient $\vec{\nabla}p_0$ over the boundary ∂A of the considered area:

$$\forall \vec{r} \in A, \quad p(\vec{r}) = \iint_{\partial A} \left[\vec{\nabla}p_0 \cdot \vec{n} - \frac{\vec{R}}{R} \cdot \vec{n} (1 + jkR) \frac{p_0}{R} \right] \frac{e^{-jkR}}{4\pi R} dS_0 \quad (1)$$

with wave number k and unitary outside normal \vec{n} . Vector \vec{R} defines the propagation path between a secondary source and the listening point.

The Kirchhoff-Helmholtz Integral may be interpreted as a continuous distribution of secondary sources. Each secondary source is composed of two elementary sources: one monopole, which is fed by the pressure gradient signal, and one dipole, which is fed by the pressure signal.

It should be noticed that the Kirchhoff-Helmholtz Integral, contrary to the Huygens' Principle, does not require that the boundary should be a wave front. The boundary may follow any geometry, which does not depend on the wave front. This remark highlights a noticeable difference between the two formulations: in the Huygens' Principle, the secondary sources are driven only by the magnitude signal, whereas in the Kirchhoff-Helmholtz Integral, they are driven both by the magnitude and the phase signal. Indeed, it should be kept in mind that in the former case, the secondary sources are distributed along a wave front, which is defined as an equal phase surface. To some extent, the Kirchhoff-Helmholtz Integral generalizes the Huygens' Principle by adding one degree of freedom for the secondary source distribution geometry, which is paid by increasing source signal complexity.

2.3. Application to spatialised sound recording and reproduction

The Kirchhoff-Helmholtz Integral gives a straightforward way of reproducing a sound field. At the recording stage, the listening area is surrounded (top of Figure 3) by a microphone array, which is composed of both pressure and velocity microphones, and which records the primary sound field due to external sources (Figure 2).

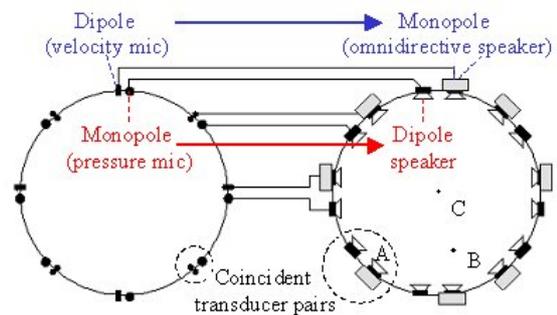


Figure 2 Application of Kirchhoff-Helmholtz Integral for holophonic sound field reconstruction.

For the reproduction stage, loudspeakers are substituted for the microphones, by replacing the

pressure microphones by dipole sources and the velocity microphones by monopole sources. Each loudspeaker is fed by the signal that was previously recorded by its associated microphone (Figure 2). It should be kept in mind that the geometry of the microphone array and the loudspeaker array should be identical. Another setup, which is exactly equivalent, consists in surrounding the primary source area by the microphone array, instead of the listening area (bottom of Figure 3).

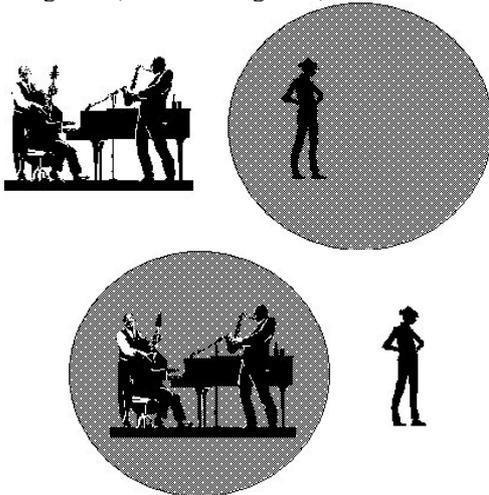


Figure 3 Two equivalent holophonic setups: by surrounding either the listener (top), or the primary sources (bottom).

The key-features of this solution of sound spatialization should now be pointed out.

Firstly, provided that the process is ideally followed (which implies for instance ideal transducers and continuous arrays), the Kirchhoff-Helmholtz Integral ensures that the sound field synthesized by the secondary sources is reproduced identical to the original one, which means that the temporal and spatial properties of the primary sound field are restored. Particularly, the localization of the sound sources is fully rendered and the listener will perceive and localize the sound sources as he would do in a real listening situation.

What's more, the sound field reproduction is valid not only at one point, but at any point within the whole area, which is delimited by the transducer array. An extended listening area is thus provided, which allows the listener to move inside the listening space and also to share this space with other listeners. Finally, it should be mentioned that, theoretically, the process requires no signal processing between the recording and reproduction stage. The all process complexity is managed by the physics, *i.e.* the reconstruction work is handled by wave interference between the secondary sources.

2.4. Practical limitations

Though the Kirchhoff-Helmholtz Integral provides a very attractive solution of spatialised sound recording and reproduction, practical limitations are obvious.

First, it requires continuous, closed-surface transducer arrays, whereas only discontinuous arrays are available¹, which raises the problem of spatial sampling. Indeed, discrete arrays cannot correctly sample incident waves which wavelength is too small with regard to the transducer spacing $\Delta_{\text{transducer}}$. Such spatial aliasing typically occurs above the so-called "spatial aliasing frequency"²:

$$f_{sp} = \frac{c}{2\Delta_{\text{transducer}}}, \quad (2)$$

Moreover linear *i.e.* not surface arrays are usually preferred in order to focus on the horizontal sound scene spatialisation.

Secondly, each secondary source is composed of two elementary transducers (both for the sound recording and reproduction), which should be coincident. This setup is also not strictly feasible.

Thirdly, the quality of the sound field reconstruction depends on the transducer characteristics, which should be the closest to the ideal one.

2.5. From holophony to WFS: approximating the Kirchhoff-Helmholtz Integral

In spite of its practical limitation, sound spatialization by holophony is not so unfeasible as it could seem at first sight. Indeed, research carried on by Berkhout & al at the TUD has shown that holophonic systems are available, provided that some approximations are applied to the Kirchhoff-Helmholtz Integral [1, 2]. These approximations define the Wave Field Synthesis concept, which has been developed by the acoustic laboratory of TUD.

Three main approximations, which are essentially based on physical feeling, have been pointed out.

Stationary phase approximation

First, the stationary phase approximation allows reducing the ideally surface transducer array to a linear horizontal "slice", in order to keep only the most useful secondary sources, according to the primary source and the listener positions.

¹ Nevertheless, the new technology of DML (Distributed Mode Loudspeaker), which are based on large vibrating panel, offers a promising answer to this issue for holophony and WFS [4].

² To be exact, this frequency depends also on the wave incidence with regard to the array.

Single directivity transducer array

Secondly, it has been remarked that the two elementary transducers (monopole and dipole) of the secondary source are highly redundant, so that only one of them is necessary. Thus the WFS concept practically uses monopole loudspeaker array fed by figure-of-eight or even cardioid microphones (Figure 4). Nevertheless, real-life loudspeaker or microphone directivity is neither ideal monopole, nor ideal dipole, but rather cardioid for the medium frequencies with lower directivity at low frequencies and higher directivity at high frequencies.

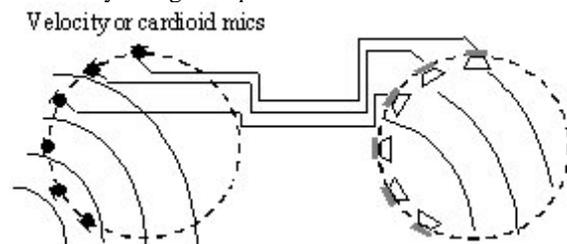


Figure 4 Restriction to single directivity transducer arrays for practical holophonic system.

Notional source encoding

Thirdly, most often, the recording is not made by microphone array, but by close microphones, which pick up the direct sound of each primary source. The microphone signal is then propagated to the virtual microphone array by applying amplitude weight and time delay, as suggested by Figure 4. Each microphone signal is thus identified to one individual primary source and may be considered as a virtual substitute for this source, *i.e.* a notional source. Such "virtual recording" allows also windowing secondary source amplitudes [5], *i.e.* using "unreal" microphone directivities, if needed for a better final rendering. Close miking provides several other advantages. In most cases, the number of microphones (and consequently the number of recorded signals) is greatly reduced in comparison with microphone array. Moreover, for up-mixing purpose, standard monophonic recordings may feed WFS rendering without extra signal processing.

Example of application

As an example, WFS has been implemented and is being experimented at the France Telecom R&D Labs, over either square, polygonal, or circular arrays composed of 48 loudspeakers (Figure 5). Especially, subjective experiments are driven in the context of the Carrouso project, with other European partners (<http://www.emt.iis.fhg.de/projects/carrouso/>).



Figure 5 Quasi-circular 48-speaker array for WFS and HOA rendering experiments at the France Telecom R&D Labs

2.6. Separating the microphone and the loudspeaker arrays

It was previously pointed out that the microphone array and the loudspeaker array must be identical, mainly in terms of geometry and number of transducers. With the notional source concept, this property is already invalidated. As a matter of fact, it can be further stated that it is always possible to circumvent this constraint and to fully dissociate the microphone array from the loudspeaker array. As for the notional source concept, the two transducer arrays can be dissociated by simulating the acoustic propagation between the actual microphone array and the actual loudspeaker array, where should be the theoretical microphone array. This is done by interfacing an extrapolation matrix between the microphones and the loudspeakers [1].

2.7. Synthesizing enclosed sound source

Another constraint of the Kirchhoff-Helmholtz Integral may be overcome. Concerning the position of the primary sources, it should be realised that in theory, the listening area, which is delimited inside the loudspeaker array (Figure 3), should be free from any primary source. In other words, the loudspeaker array is able to synthesize only sound sources, which are outside the loudspeaker array. However, early developments of the WFS concept has pointed out that it is also possible to synthesize sound sources inside the loudspeaker array, merely by inverting the phase of the secondary sources, so that the last fed loudspeaker becomes the first fed and vice versa, in order that a concave wave front, instead of a convex one, is reconstructed [6].

This process may be compared with the time reversal approach applied to sound focusing [7, 8]. Indeed sound focusing by WFS may be identified to time reversal restricted to the direct sound.

3. HIGHER ORDER AMBISONICS (HOA)

Ambisonics was developed several decades ago, mostly by Gerzon, as a spatial sound encoding approach dedicated to surround (2D) and periphonic (3D) multi-speaker systems [9] [10]. For about eight years, its extension to higher spatial resolution systems has been the object of increasingly numerous studies, which promising features are becoming practicable. The following state of art includes recent and relevant progress in this field.

3.1. Mathematic fundamentals

Ambisonic representation is based on the spherical harmonic decomposition of the sound field, which comes from writing the wave equation in the spherical coordinate system (Figure 6) where a point \vec{r} is described by a radius r , an azimuth θ and an elevation δ .

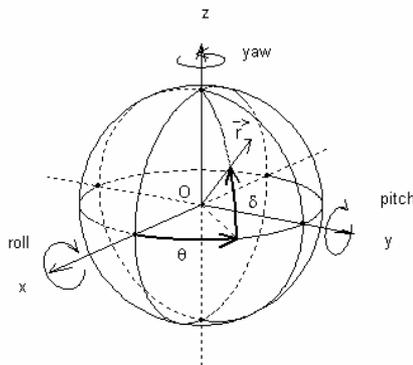


Figure 6 Spherical coordinate system, with the three elementary rotation degrees

Therefore the pressure field can be written as the Fourier-Bessel series (3), which terms are the weighted products of directional functions $Y_{mn}^\sigma(\theta, \delta)$ (called "spherical harmonics") and radial functions:

$$p(\vec{r}) = \sum_{m=0}^{\infty} j_m^m j_m(kr) \sum_{0 \leq n \leq m, \sigma = \pm 1} B_{mn}^\sigma Y_{mn}^\sigma(\theta, \delta) + \sum_{m=0}^{\infty} j_m^m h_m^-(kr) \sum_{0 \leq n \leq m, \sigma = \pm 1} A_{mn}^\sigma Y_{mn}^\sigma(\theta, \delta) \quad (3)$$

with the wave number $k=2\pi f/c$.

This is the general equation for the case of a sphere layer ($R_1 \leq r < R_2$) that is free from sources (Figure 7). The weighting coefficients B_{mn}^σ associated with the spherical Bessel functions $j_m(kr)$ (first series) describe the "through-going" field (due to outside sources), whereas the coefficients A_{mn}^σ associated with the divergent spherical Hankel functions

$h_m^-(kr) = j_m(kr) - j_n(kr)$, describe the "outgoing" field (due to inside sources)³.

As a sound spatialization approach, Ambisonics basically assumes a centered point of view, thus a centered listening area that is free of virtual sources. Thus only a "through-going" field, as represented by the coefficients B_{mn}^σ , is considered, the outgoing field being null ($A_{mn}^\sigma = 0$). Components B_{mn}^σ are the expression in the frequency domain of what we'll call "ambisonic" signals.

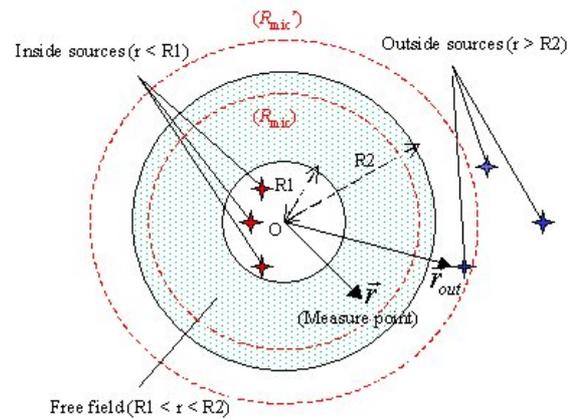


Figure 7 Inter-sphere free field volume where spherical harmonic representation (3) applies

The spherical harmonic functions

The spherical harmonic functions $Y_{mn}^\sigma(\theta, \delta)$ exhibited in (3) are defined as following:

$$Y_{mn}^{\sigma(\text{NSD})}(\theta, \delta) = \sqrt{2m+1} \sqrt{(2-\delta_{0,n})} \frac{(m-n)!}{(m+n)!} P_{mn}(\sin \delta) \times \begin{cases} \cos n\theta & \text{if } \sigma = +1 \\ \sin n\theta & \text{if } \sigma = -1 \text{ (ignored if } n=0) \end{cases} \quad (4)$$

with the $P_{mn}(\zeta)$ being the associated Legendre functions⁴ of degree m and order n , and where δ_{pq} equals to 1 if $p=q$ and 0 otherwise (Kronecker symbol). They form an orthonormal base, *i.e.* $\langle Y_{mn}^\sigma | Y_{m'n'}^{\sigma'} \rangle_{4\pi} = \delta_{mm'} \delta_{nn'} \delta_{\sigma\sigma'}$, in the sense of the spherical scalar product $\langle F | G \rangle_{4\pi} = \frac{1}{4\pi} \iint F(\theta, \delta) G(\theta, \delta) d\Omega$.

³ Note that Hulsebos [11] combines outgoing field and ingoing (rather than "through-going") field.

⁴ For computational application, the values of these functions can be derived using recurrence relations (see appendix A.2.2 of [12]).

Figure 8 provides a 3D view of spherical harmonics. There are $(2m+1)$ components, including 2 horizontal components (those with $n=m$), per order $m \geq 1$.

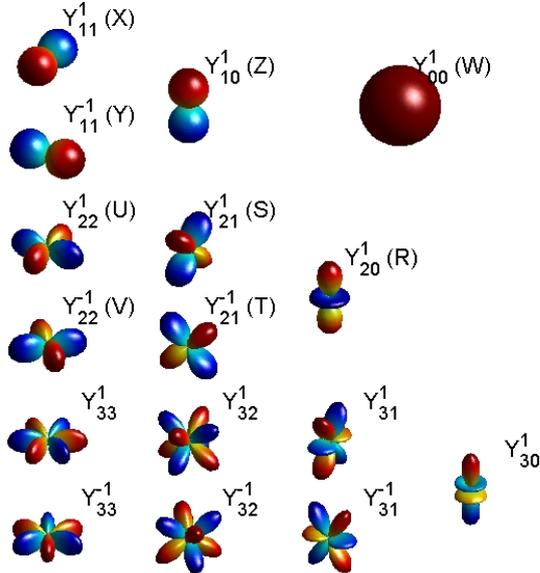


Figure 8 3D view (with respect to Figure 6) of spherical harmonic functions with usual designation of associated ambisonic components.

Interpretation: directional information and radial approximation

Spherical harmonic components B_{mn}^σ are closely related to the pressure field and its derivatives (or momentums) of successive orders around the origin O . The first four components are well known: $B_{00}^{+1} = W$ is the pressure, and $B_{11}^{+1} = X$, $B_{11}^{-1} = Y$, $B_{10}^{+1} = Z$ are related to its gradient or also the acoustic velocity. Each additional group of higher order components or momentums provides an approximation of the sound field over a larger neighbourhood of the origin with regard to the wavelength (Figure 10).

In practice, only a finite number of components (up to a given order M) can be transmitted and exploited, and even estimated. Thus the represented field is typically approximated by the truncated series:

$$p(\vec{r}) = \sum_{m=0}^M j_m^m(kr) \sum_{0 \leq n \leq m, \sigma = \pm 1} B_{mn}^\sigma Y_{mn}^\sigma(\theta, \delta), \quad (5)$$

involving $K_{3D} = (M+1)^2$ components. An interesting interpretation comes from commenting Figure 8 and Figure 9 with respect to each other: harmonics $Y_{mn}^\sigma(\theta, \delta)$ with higher angular variability are associated with radial functions $j_m(kr)$ which first maximum occur at larger distances kr . To

summarize: *higher directional resolution goes with greater radial expansion and vice-versa.*

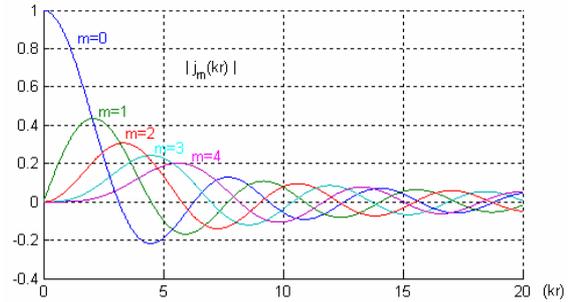


Figure 9 Spherical Bessel functions $j_m(kr)$

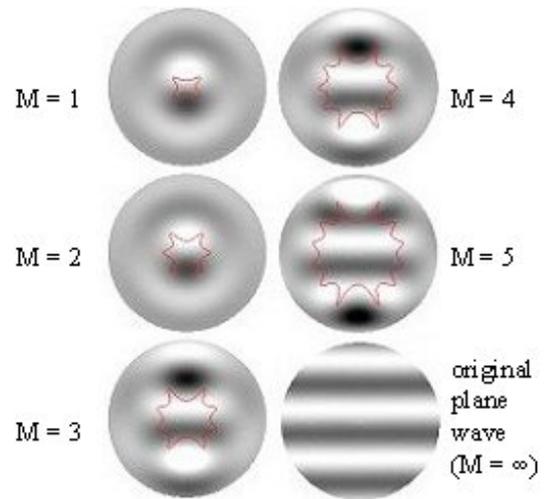


Figure 10 Monochromatic plane wave sound field and its approximation by the truncated Fourier-Bessel series (5) for several orders M

The plane wave case: directional encoding equations

The spherical harmonic decomposition of a plane wave of incidence (θ_s, δ_s) conveying a signal S leads to the simple expression of the ambisonic component.

$$B_{mn}^\sigma = S Y_{mn}^\sigma(\theta_s, \delta_s) \quad (6)$$

Thus a far field source is encoded by simply applying *real gains* to the received pressure signal S .

2D-restricted formalism: cylindrical decomposition

The cylindrical coordinate system has often been used in the literature when dealing with horizontal-only reproduction system and virtual sources [5, 12-14]. This leads to the Fourier-Bessel series:

$$p(r, \theta) = B_{00}^{+1(N2D)} J_0(kr) + \sum_{m=1}^{\infty} (B_{mm}^{+1(N2D)} \frac{\sqrt{2} \cos m\theta}{Y_{mm}^{+1(N2D)}(\theta, 0)} + B_{mm}^{-1(N2D)} \frac{\sqrt{2} \sin m\theta}{Y_{mm}^{-1(N2D)}(\theta, 0)}) J_m(kr) \quad (7)$$

This way the 2D (horizontal) ambisonic components B_{mm}^σ derive from a kind of circular Fourier Transform of the sound field involving angular functions $Y_{mm}^{\sigma(N2D)}(\theta, 0)$. They form an orthonormal base in the sense of the circular scalar product:

$$\langle F|G \rangle_{2\pi} = \frac{1}{2\pi} \int_0^{2\pi} F(\theta)G(\theta)d\theta$$

One can unify the two formalisms by considering the circular (thus horizontal) harmonics as a subset of the spherical ones (4), modulo a weighting factor [12]:

$$Y_{mm}^{\sigma(N2D)}(\theta, \delta) = \sqrt{\frac{2^{2m} m!^2}{(2m+1)!}} Y_{mm}^{\sigma(N3D)}(\theta, \delta) \quad (8)$$

The cylindrical formalism, which encoding functions (7) are simpler and less numerous ($K_{2D}=2M+1$) than the spherical ones, is useful to design the decoding for horizontal-only loudspeaker arrays (see next section).

3.2. The reproduction step: decoding design

The re-encoding principle

The design of ambisonic decoding basically relies on what could be called "the re-encoding principle" [12, 15]: the aim is to acoustically recompose encoded ambisonic components (pressure field and its "derivatives") \tilde{B}_{mm}^σ (9) at the centre of the array.

Assuming that loudspeakers are far enough from the listening centre point, their signals S_i are encoded as plane waves with coefficient vectors \mathbf{c}_i :

$$\mathbf{c}_i = \begin{bmatrix} Y_{00}^{+1}(\theta_i, \delta_i) \\ Y_{11}^{+1}(\theta_i, \delta_i) \\ Y_{11}^{-1}(\theta_i, \delta_i) \\ \dots \\ Y_{mm}^\sigma(\theta_i, \delta_i) \\ \dots \end{bmatrix} \quad \tilde{\mathbf{B}} = \begin{bmatrix} \tilde{B}_{00}^{+1} \\ \tilde{B}_{11}^{+1} \\ \tilde{B}_{11}^{-1} \\ \dots \\ \tilde{B}_{mm}^\sigma \\ \dots \end{bmatrix} \quad \mathbf{S} = \begin{bmatrix} S_1 \\ S_2 \\ \dots \\ S_N \end{bmatrix} \quad (9)$$

Thus the re-encoding principle can be written in the matrix form (10), with $\mathbf{C}=[\mathbf{c}_1 \dots \mathbf{c}_N]$ being the "re-encoding matrix":

$$\tilde{\mathbf{B}} = \mathbf{C}\mathbf{S}, \quad (10)$$

The decoding operation aims at deriving signals \mathbf{S} from matrixing original ambisonic signals $\tilde{\mathbf{B}}$:

$$\mathbf{S} = \mathbf{D}\tilde{\mathbf{B}} \quad (11)$$

To ensure $\tilde{\mathbf{B}}=\mathbf{B}$, system (10) must be inverted. Therefore decoding matrix \mathbf{D} is typically defined as:

$$\mathbf{D} = \text{pinv}(\mathbf{C}) = \mathbf{C}^T \cdot (\mathbf{C}\mathbf{C}^T)^{-1}, \quad (12)$$

provided that there are enough loudspeakers: *i.e.* $N \geq K_{2D}$ or $N \geq K_{3D}$. The case of regular layouts⁷ simplifies the expression of the decoding matrix \mathbf{D} . Indeed, by choosing the appropriate normalized encoding convention (either (4) for full-3D or (8) for

horizontal-only reproduction), one easily shows [12, 15] that:

$$\mathbf{D} = \frac{1}{N} \mathbf{C}^T \quad (13)$$

Outside the reconstructed domain (HF/off-center)

The reconstruction over a given listening area is achieved only up a frequency that depends on the area size. Above this frequency, other decoding criteria and solutions (called "max r_E " and "in-phase") are preferably applied to optimize the perceived spatial rendering [12, 15]. Such decoding optimization is simply done by applying gains g_m on the appropriate frequency-bands to the components B_{mm}^σ before processing the "basic" decoding:

$$\mathbf{D} = \frac{1}{N} \mathbf{C}^T \cdot \text{Diag}([\dots \ g_m \ \dots]) \quad (14)$$

Generic formulae for g_m are fully defined in [12].

Equivalent panning functions and directivities

By combining encoding equation (6) and decoding matrix (14), one derives an equivalent panning function $G(\gamma)$ [12, 15]:

$$G(\gamma) = \frac{1}{N} \left(g_0 + 2 \sum_{m=1}^M g_m \cos(m\gamma) \right) \quad \text{for 2D}, \quad (15)$$

$$G(\gamma) = \frac{1}{N} \sum_{m=0}^M (2m+1) g_m P_m(\cos \gamma) \quad \text{for 3D}, \quad (16)$$

such that loudspeaker i is fed with $S_i = S \cdot G(\gamma)$, with $\gamma = \arccos(\vec{u}_i \cdot \vec{u}_s)$ being the angle between the speaker and source directions \vec{u}_i and \vec{u}_s . Figure 11 shows that higher orders help using loudspeakers with a finest angular selectivity for sound imaging, thus benefiting of their angular density.

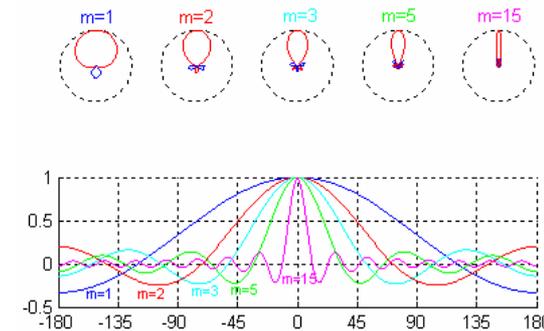


Figure 11 Equivalent directivities and panning laws associated to basic 2D decoding with various orders m (normalized regarding their max values).

This will be a helpful tool for interpreting rendering properties (section 4.2). Another helpful interpretation can be derived in terms of equivalent

recording setup: the polar diagram of a given order m (top of Figure 11) describes the directivity pattern of coincident microphones pointing to the loudspeakers that they would respectively feed.

3.3. Recent progress: supporting near field

Previous literature only rarely addressed the modelling of spherical waves, radiated by finite distance sources [12]. Nevertheless, correct encoding and reconstruction of realistic sound fields require it, and couldn't satisfy themselves with the usual plane wave approximation.

Spherical wave encoding (finite distance sources):

From the decomposition of a spherical wave given in [16], one derive [12] the formulae (17) (18) describing source encoding at a finite distance ρ :

$$B_{mn}^\sigma = S.F_m^{(\rho/c)}(\omega)Y_{mn}^\sigma(\theta, \delta) \quad (17)$$

$$F_m^{(\rho/c)}(\omega) = \sum_{n=0}^m \frac{(m+n)!}{(m-n)!n!} \left(\frac{-jc}{\omega\rho} \right)^n, \text{ with } \omega=2\pi f \quad (18)$$

Note that equation (17) involves the pressure field S captured at O , assuming that $1/\rho$ attenuation and delay ρ/c due to finite distance propagation are already modeled.

Filters shown in (18) are typically "integrating filters", which are unstable by nature (infinite bass-boost, see Figure 12). This means also that the currently adopted HOA encoding format is unable to physically represent (*i.e.* by finite amplitude signals) natural or realistic sound fields, since these always include more or less near field sources.

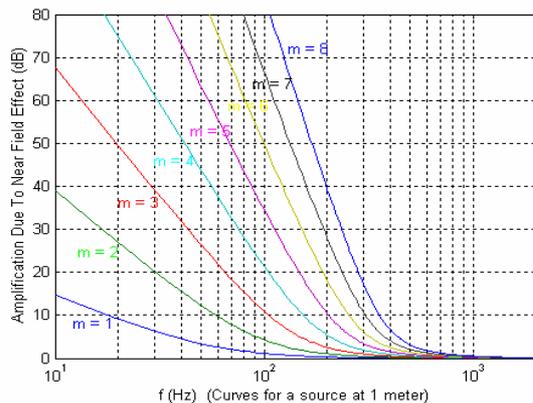


Figure 12 Low frequency infinite boost ($m \times 6$ dB/octave) of ambisonic components due to near field effect

Corrected decoding for a proper reconstruction

In order to truly satisfy the "re-encoding principle" and re-compose the encoded sound field, the near field effect of the loudspeakers has to be considered,

i.e. compensated, as already illustrated in [12]. The corrected decoding operation is thus:

$$S = D.\text{Diag} \left(\left[\dots \frac{1}{F_m^{(R/c)}(\omega)} \dots \right] \right).B \quad (19)$$

This decoding is practicable since inverse filters $1/F_m^{(R/c)}(\omega)$ are stable, and actually manages to preserve wave fronts original shape⁵ (Figure 17).

Distance coding filters

The solution for practicable, finite distance source encoding, is to introduce the near field compensation (19) from the encoding stage and no longer at the decoding. Its combination with the virtual source near field effect (17) leads to the definition of stable "Distance (or Near Field) Coding filters"⁶:

$$H_m^{\text{NFC}(\rho/c, R/c)}(\omega) = \frac{F_m^{(\rho/c)}(\omega)}{F_m^{(R/c)}(\omega)} \quad (20)$$

They are characterized by a finite, low frequency amplification $m \times 20 \log_{10}(R/\rho)$ (in dB), which is positive for enclosed sources ($\rho < R$) and negative for outside sources ($\rho > R$), as shown Figure 13.

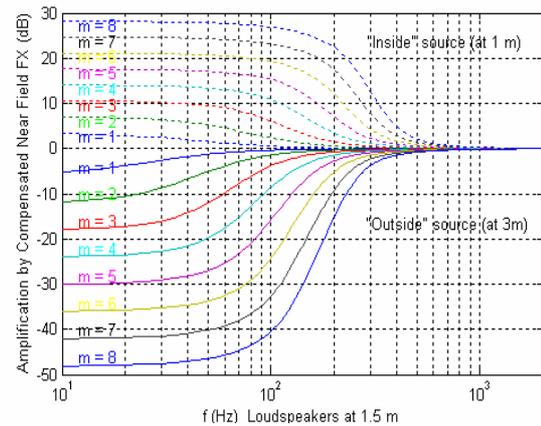


Figure 13 Finite amplification of ambisonic components from pre-compensated Near Field Effect (dashed lines: $\rho/R=2/3$; cont. lines: $\rho/R=2$).

A viable, new ambisonic format

Thus filters (20) advantageously replace filters (18) in encoding equation (17). At the same time, we have to consider a new encoding format called "Near Field Compensated Higher Order Ambisonics" format (NFC-HOA), and defined by the relation:

⁵ Without near field compensation, an encoded plane wave is reconstructed as a spherical one coming from the loudspeaker array.

⁶ Efficient, parametric digital filters⁵ (for practical use) are described in [17].

$$\tilde{B}_{mn}^{\sigma \text{ NFC}(R/c)} = \frac{1}{F_m^{(R/c)}(\omega)} B_{mn}^{\sigma} \quad (21)$$

It can represent any realistic sound field by finite amplitude signals and only requires the decoding operation (13), while implying a reference parameter: the reproduction loudspeaker distance R . Nevertheless, adaptation to any other array radius R' is possible by applying filters defined in (20) (replace ρ by R and R by R') before decoding. Finally, finite distance source encoding equation (17) becomes:

$$\tilde{B}_{mn}^{\sigma \text{ NFC}(R/c)} = S.H_m^{\text{NFC}(\rho/c, R/c)}(\omega) Y_{mn}^{\sigma}(\theta, \delta) \quad (22)$$

This direct and rational way of encoding distance is an advantageous alternative to the WFS+HOA coupling scheme suggested in [18].

Equivalent pickup directivity or panning law

The same way as in 3.2, an equivalent panning law or pickup directivity can be derived, from replacing or multiplying gains g_m in (15) (for 2D case) by the frequency dependent complex gains (20):

$$G^{\text{NFC}(R,c)}(\rho, \gamma, \omega) = \frac{1}{N} \left(g_0 + 2 \sum_{m=1}^M g_m H_m^{\text{NFC}(\rho/c, R/c)}(\omega) \cdot \cos(m\gamma) \right) \quad (23)$$

Figure 14 shows the case of far virtual source (plane wave) at different frequencies and for $g_m=1$ (basic decoding).

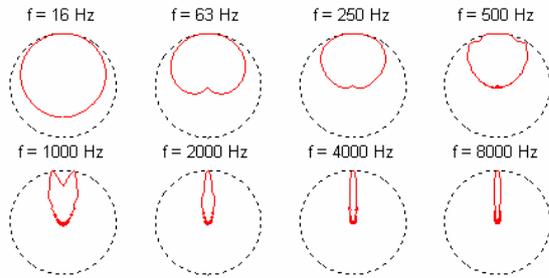


Figure 14 Equivalent pickup directivities or panning laws (normalized absolute values) for $\rho=\infty$ (plane wave) and $R_{\text{spk}}=1.5\text{m}$, and with "basic" ($g_m=1$), 2D, 15th order rendering with Near Field pre-Compensation.

It's worth noticing that high frequency equivalent directivity tends to be the same, thus as high, as without near field coding (Figure 11, $m=15$), whereas a lower directivity (down to cardio or even omni) is observed at low frequencies.

3.4. Natural sound field recording

Up to this section, only virtual source encoding has been addressed. But natural sound field recording is also a possible and important feature of Higher Order

Ambisonics, although up to very recently, Ambisonics recording possibilities have been restricted to the 1st order "Soundfield" microphone [19]. Indeed, theoretical studies regarding higher orders addressed recording systems based horizontal circular microphone arrays [5, 14] and more recently spherical arrays [12, 20-22].

The reader interested in further issues of ambisonic recording systems will find a full study in [23]. The following, lighter description aims at bringing out basic issues of practical systems, which are *spatial aliasing* and *noise amplification*.

Basic principle

The basic idea is to process a discrete spherical Fourier Transform of the sound field, based on its spherical sampling. For this purpose, one considers an array of N microphones distributed over a sphere (or an horizontal circle, in more restricted systems) of radius R_{mic} and centre O , and positioned and oriented according to the directions \vec{u}_i . From these we can process a directional sampling of the spherical harmonics: $\mathbf{y}_{mn}^{\sigma} = [Y_{mn}^{\sigma}(\vec{u}_1) \ Y_{mn}^{\sigma}(\vec{u}_2) \ \dots \ Y_{mn}^{\sigma}(\vec{u}_N)]$.

For the study needs, let's consider the simple case of cardio-like directivity: $G(\theta) = \alpha + (1-\alpha)\cos(\theta)$ (far field directivity). It combines the pressure value p as expressed in (5) and its radial derivative with respective weighting factors α and $(1-\alpha)$. Therefore the field captured by the microphone i is:

$$p_R(\vec{u}_i) = \sum_{m=0}^{\infty} W_m(\omega) \sum_{0 \leq n \leq m, \sigma = \pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\vec{u}_i) \quad (24)$$

with the weighting factor:

$$W_m(\omega) = j^m (\alpha j_m(kR_{\text{mic}}) - j(1-\alpha)j_m'(kR_{\text{mic}})) \quad (25)$$

Provided that the array geometry verifies some regularity conditions⁷, one can estimate ambisonic components by projecting the spatially sampled sound field $\mathbf{p} = [p(\vec{u}_1) \ \dots \ p(\vec{u}_N)]^T$ onto each sampled spherical harmonic \mathbf{y}_{mn}^{σ} :

$$\hat{B}_{mn}^{\sigma} = \text{EQ}_m(\omega) \langle \mathbf{p} | \mathbf{y}_{mn}^{\sigma} \rangle_N, \quad (26)$$

where the following equalization filters are also applied:

$$\text{EQ}_m(\omega) = \frac{1}{W_m(\omega)} \quad (27)$$

⁷ The underlying condition is that spatial sampling preserves spherical harmonic base orthonormality, i.e. $\langle \mathbf{y}_{mn}^{\sigma} | \mathbf{y}_{m'n'}^{\sigma'} \rangle_N = \frac{1}{N} \mathbf{y}_{mn}^{\sigma} \cdot \mathbf{y}_{m'n'}^{\sigma'}{}^T = \delta_{mm'} \delta_{nn'} \delta_{\sigma\sigma'}$ (for $m, m' \leq M$).

Spherical harmonic spectrum aliasing / spatial aliasing

According to [12], components B_{mn}^σ ($m \leq M$) are estimated with the residual error:

$$\tilde{\mathcal{E}}_{mn}^\sigma = B_{mn}^\sigma - \hat{B}_{mn}^\sigma = \sum_{m' > M} \text{EQ}_m(\omega) W_{m'}(\omega) \sum_{0 \leq n' \leq m', \sigma' = \pm 1} B_{m'n'}^{\sigma'} \langle y_{mn}^\sigma | y_{m'n'}^{\sigma'} \rangle_N \quad (28)$$

which exhibits the projection of higher order, insufficiently sampled components $B_{m'n'}^{\sigma'}$, onto the estimated one \hat{B}_{mn}^σ . This is an aliasing effect on the estimated spherical harmonic spectrum.

The "projection factor" $\langle y_{mn}^\sigma | y_{m'n'}^{\sigma'} \rangle_N$ typically decreases when increasing N , thus the sensor angular density (spatial "oversampling"). The other weighting factor $\text{EQ}_m(\omega) W_{m'}(\omega) = W_{m'}(\omega) / W_m(\omega)$ is an increasing function of the frequency, and it also globally increases with the array radius R_{mic} . It appears finally that the frequency, above which spherical harmonic spectrum aliasing (28) becomes significant, decreases when the distance between acoustic sensors increases. It clearly has to be related to the "spatial aliasing frequency" (2), as introduced with WFS in 2.4! This was also pointed out in [24].

Applying near field pre-compensation (for practicable systems)

As shown in section 3.3, one has to introduce a near field pre-compensation at the encoding stage, in order to be able to represent any sound field with finite amplitude components. Therefore, instead of (27), the required equalization filters become:

$$\text{EQ}_m^{\text{NFC}(R_{\text{mic}}/c, R_{\text{spk}}/c)}(\omega) = \frac{\text{EQ}_m(\omega)}{F_m^{(R_{\text{spk}}/c)}(\omega)} = \frac{1}{F_m^{(R_{\text{spk}}/c)}(\omega) W_m(\omega)} \quad (29)$$

Unlike filters (27), these are now stable (finite low frequency amplification as shown Figure 15) and produce signals \tilde{B}_{mn}^σ that are compliant with the "NFC-HOA" format (21). The reference distance R_{spk} is preferably chosen as the radius of a typical loudspeaker array. Figure 15 shows the case of a "reproduction distance" $R_{\text{spk}}=1\text{m}$ much larger than the microphone radius $R_{\text{mic}}=5\text{cm}$.

Now it's possible to discuss to another important issue of practicable ambisonic microphones, which is the noise and error amplification problem.

Noise and error amplification

Electric signals derived from real life acoustic sensors always include noise. It's important to know what this noise becomes when computing ambisonic components and then, when decoding them and rendering the sound field. For this purpose, let's consider the signals \mathbf{p} as pure, uncorrelated noise

signals of equal energy $|p|^2$. One easily finds that the resulting noisy component has energy⁷:

$$\begin{aligned} \left| \tilde{B}_{mn}^{\sigma \text{ NFC}(R_{\text{spk}}/c)} \right|^2 &= \left| \text{EQ}_m^{\text{NFC}(R_{\text{spk}}/c)}(\omega) \right|^2 \frac{1}{N^2} y_{mn}^\sigma \cdot y_{mn}^{\sigma T} |p|^2, \quad (30) \\ &= \frac{1}{N} \left| \text{EQ}_m^{\text{NFC}(R_{\text{spk}}/c)}(\omega) \right|^2 |p|^2 \end{aligned}$$

Thus noise amplification fits equalization law (29) (Figure 15) lowered by $-10 \cdot \log_{10}(N)$ dB (e.g. -15dB for $N=32$). The quite high amplification (especially for low frequencies and high orders) reveals an important "effort" for extrapolating the sound field knowledge from a little radius R_{mic} to a much larger one R_{spk} .

Figure 15 shows that very low frequency amplification is about "one order lower" with ideal cardioid sensors than with pressure sensors over rigid sphere. Indeed, cardioid sensors already include first order directivity, but in real life they tend to be omni and/or noisy at low frequency!

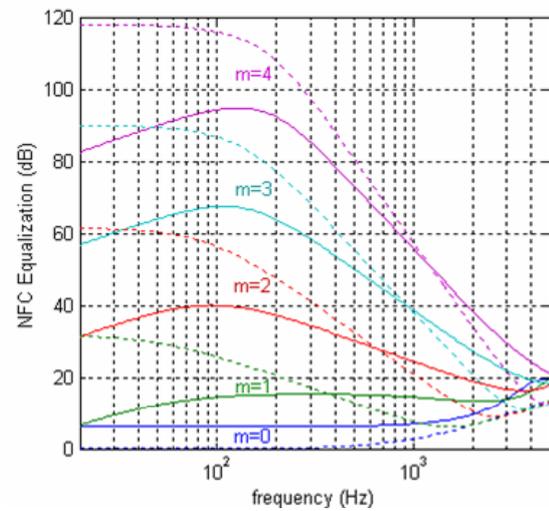


Figure 15 Near Field Compensated Equalization (29) involved in sphere microphone processing ($R_{\text{mic}}=5\text{cm}$, $R_{\text{spk}}=1\text{m}$). Cont. lines: ideal cardioid sensors; dotted lines: pressure sensors over a rigid sphere.

It should be added that even the effective captured signals don't necessarily exactly fit the theoretical modeling (24): that may be because of acoustic disturbance from the mechanical structure or because of bad sensor calibration, etc. This kind of error is also amplified during the processing.

It will be later discussed what this noise becomes after decoding and regarding the reconstructed sound field.

3.5. Applied State of Art

After mostly theoretical studies on High Order Ambisonics, their promising potentialities are becoming reality. Indeed, some essential features have been made practicable by solving the near field problem (as mentioned in sections 3.3 and 3.4). Related work done at the France Telecom R&D Labs is briefly listed below for illustration.

DSP tools (software generic implementation without limitation on system order)

- Encoding tools available are now: directional encoding functions and distance coding filters.
- Decoding tools include matrix and shelf-filters design for loudspeaker presentation. Design and processing of "Ambisonics to Ears" Transfer Functions, are also concerned, for binaural rendering (over headphones).
- Sound field transformations: addressing rotation matrix design (Figure 6 shows basic rotation angles) and focalisation.

Practical embodiments and experimentations

- 2D holophonic configurations are used (48 speaker, circular or dodecagonal arrays), also for comparison with WFS (Figure 5).
- A 4th order 3D microphone (based on 32 capsules placed over a rigid sphere) with associated DSP is being experimented [23].
- A full (or nearly-full) 3D ambisonics configuration (4th order, 32 or 21 loudspeakers) is in project: to be built in our anechoic room.

3D audio multi-channel format:

Original (1st order) "B-format" introduced by Gerzon has been recently extended to a 2nd order encoding format FMH ("Furse-Malham Harmonics": http://www2.york.ac.uk/inst/mustech/3d_audio/secon dor.html), used by e.g. some music composers. There has been also a first attempt by Richard Dobson at handling them as sound files using an extension of the WAV format (WAVE-EX).

Since a viable, new ambisonic format is now mathematically defined (21), specifications are being discussed [17] for handling it as a multi-channel WAV-EX file, and also as a compressed multi-channel AAC stream in MPEG-4 (output document w5386 from the Awaji Meeting, December 2002).

4. COMPARING WFS AND HOA: FROM CONNECTIONS TO COMPROMISES

Up to this point, each approach has been described separately. It is now intended to further investigate them side-by-side, with the hope of offering the

reader an "intuitive feeling" of the underlying physics, while making the views converge. First, a formal connection between their intrinsic representations is completed, and a list of reconstruction artefacts is drawn from examining typical departures from theoretical conditions. Then, some major artefacts are physically interpreted and characterized with the help of visualizations of simulated sound fields. Finally, recommendations and compromises are highlighted regarding virtual sound imaging and natural recording strategies, and also the encoding format.

4.1. Formal connection - Consequences of departures from theory

For a long time, WFS and HOA have been considered as two different, and even opposite, ways of sound spatialization. Nevertheless, it has been recently pointed out that they are closely connected approaches of 3D audio recording and reproduction [3, 5, 12]. Indeed, several analogies can be mentioned. Both WFS and HOA are based on sound recording and reproduction by resp. microphone and loudspeaker arrays. These respectively perform an acoustic encoding and decoding of the spatial sound information, which aim at a physical reconstruction of the primary sound field. While being based on two different representations of the sound field, (Kirchhoff-Helmholtz Integral for WFS and spherical harmonic expansion for HOA), both provide exact solution to the sound wave equation, and, for this reason, are fully equivalent. Moreover, under given assumptions⁸, it has been shown how the ambisonic encoding and decoding equations may be derived from the Kirchhoff-Helmholtz Integral [3, 5, 12].

In this section, the connection between both intrinsic representations is completed and further discussed, regarding finite radius boundary (top of Figure 3). Then departures from theoretical conditions due to practical constraints are considered in terms of reconstruction error. While classifying expected resulting artefacts, difference and convergence are highlighted regarding extreme and intermediate forms of both approach implementations, especially regarding the microphone array radius R_{mic} .

Completing the connection between sound field intrinsic representations

To complete the convergence of views, both descriptions have to be confronted in identical

⁸ These assumptions are: plane wave (infinite distance boundary and secondary sources), in addition to continuous sources distribution.

conditions/situation of recording and reproduction, that means: with the same transducer arrays. At least and even for ambisonics, we have to examine the case of a microphone array having a non-negligible radius R_{mic} , and especially the case $R_{mic}=R_{spk}$. We'll show how these recording considerations address the issue of sound field intrinsic representation.

Let's place the microphone array in a free field sphere strata as shown in Figure 7: we chose $R_1 < R_{mic} < R_2$. It's worth highlighting that both intrinsic representations (derived from spherical harmonic decomposition and Kirchhoff-Helmholtz Integral) are able to distinguish between inside and outside sources, *i.e.* between outgoing and through-going field. This appears directly regarding respectively A_{mn}^σ and B_{mn}^σ components for HOA (3), and indirectly regarding the pressure values $p(\vec{R})$ and normal velocity values

$$v_n(\vec{R}) = \frac{1}{j\omega R} \vec{\nabla} p_0 \cdot \vec{n} = \frac{1}{j\omega R} \frac{\partial p}{\partial r}(\vec{R}) \quad \text{for WFS (1).}$$

A more explicit connection derives from applying spherical Fourier Transform onto the pressure and velocity boundary distributions:

$$\begin{aligned} P_{mn}^\sigma &= \frac{1}{4\pi} \iint_{|\vec{u}|=1} p(R, \vec{u}) Y_{mn}^\sigma(\vec{u}) d\Omega \\ &= j^m j_m(kR) B_{mn}^\sigma + j^m h_m^-(kR) A_{mn}^\sigma \\ V_{mn}^\sigma &= \frac{1}{4\pi} \iint_{|\vec{u}|=1} \frac{1}{j\omega R} \frac{\partial p}{\partial r}(R, \vec{u}) Y_{mn}^\sigma(\vec{u}) d\Omega \\ &= \frac{j^{m-1}}{cR} (j_m'(kR) B_{mn}^\sigma + h_m^-(kR) A_{mn}^\sigma) \end{aligned} \quad (31)$$

where we have used the series (3) and its radial derivative to express p and $\partial p/\partial r$, and also the orthonormality of spherical harmonics Y_{mn}^σ . Therefore "through-going" and "outgoing" field descriptors are:

$$\begin{aligned} B_{mn}^\sigma &= j^{-m} \frac{h_m^-(kR) P_{mn}^\sigma - jcR h_m^-(kR) V_{mn}^\sigma}{j_m(kR) h_m^-(kR) - j_m'(kR) h_m^-(kR)} \\ A_{mn}^\sigma &= j^{-m} \frac{j_m'(kR) h_m^-(kR) P_{mn}^\sigma - jcR j_m'(kR) V_{mn}^\sigma}{j_m'(kR) h_m^-(kR) - j_m(kR) h_m^-(kR)} \end{aligned} \quad (32)$$

Inversely, one can recompose the pressure and velocity fields at the boundary using the series (3) and its radial derivative. This is the transposition, in terms of 3D representation, of the relationship established by Huselbos³ [11] for the horizontal case. Finally, this completes the connection previously stated for an infinite boundary radius R [3, 5, 12].

It's worth recalling and highlighting here that the exact representation intrinsically depends on R_{mic} : if we move R_{mic} to a distance R_{mic}' beyond one or several "outside" sources (Figure 7), these become "inside" sources and then the spherical harmonic representation changes in terms of A_{mn}^σ and B_{mn}^σ ! On

the opposite, there are only "outside" sources ($A_{mn}^\sigma=0$) from the centre point of view, *i.e.* for $R_{mic}=0$. From similar considerations, the extrapolation of the Kirchhoff-Helmholtz Integral to $R_{mic} \neq R_{spk}$, may become mathematically invalid.

Introducing practical limitations: Classification of expected artefacts

Reconstruction errors arise from departures from theory, when obeying constraints of practical system embodiment (limited number of loudspeakers, use of a single kind of transducer directivity, restriction to 2D), or even from fundamental limitations ("inside" sources). A qualitative and restricted comparison between WFS and HOA was already given in [25], in terms of artefacts and compromises related to the reproduction constraints. The following list rationally classifies expected artefacts by referring to previous sections. Points 1, 3 and 4 address encoding issues whereas artefact 2 arise from reproduction array restrictions.

1. Restriction to single directivity microphone arrays (section 2.4) disables inside/outside dissociation and may cause encoding confusions.

Indeed, equation (32) clearly shows that if the captured signals are just a combination of pressure and velocity, "inside" and "outside" descriptors A_{mn}^σ and B_{mn}^σ cannot be unambiguously derived. If we assume $A_{mn}^\sigma=0$ (free field enclosed area) even though there are enclosed sources, then these are rendered with an inverted wave front curvature (spatial mirroring with regard to the centre C), in addition to the time reversal effect (see also point 5 below) as explained in 2.7.

2. Restriction to single directivity loudspeaker arrays (section 2.4) implies a reconstruction error along the array border.

Figure 2 helps understanding that the combination of monopole-dipole contributions (from the array point A) is not the same from a central viewpoint C , as from a viewpoint B along the border of the listening area, *i.e.* closer to the array. The single directivity approximation is only acceptable for the centre C , and is no longer valid along the border, which implies that reconstruction may be affected.

3. Spatial aliasing arises from the microphone spacing.

This artefact has been first introduced in 2.4 (with WFS) as depending on the spatial sampling of secondary source array, and then further identified in 3.4 (with HOA) as occurring at the recording stage. Therefore, it is a sound field *encoding* issue, which depends on both the array radius R_{mic} and the number of transducers N . That's why the extreme, "ideal" encoding form (22) of HOA (virtual source with

$R_{mic}=0$) doesn't suffer from spatial aliasing, as illustrated in the next section.

4. Vertical/horizontal aliasing occurs when using circular arrays for recording.

This is another kind of spatial sampling artefact. Even if a horizontal restricted reproduction is targeted, only a spherical microphone array allows discarding vertical sound field components from the horizontal sound field representation and prevents these unwanted components from spoiling the audio rendering as completely unrelated acoustic phenomena. This could be proved by extending the aliasing error computation (28) of section 3.4.

5. Enclosed sources: exact sound field reproduction over the enclosed area is physically impossible.

Nevertheless, both approaches can even do something for reproducing sources inside the reproduction area: WFS' trick consists in temporally reversing the wave front propagation (section 2.7), whereas HOA may just extrapolate the wave front description, as seen from the centre ($A_{mn}^{\sigma}=0$), up to the source distance.

4.2. Characterizing and interpreting artefacts

This section illustrates rendering properties and artefacts (as listed in 4.1) of each system, through visualizations of simulated sound field reconstruction (restricted to the horizontal plane for convenience). This will help understanding and interpreting them physically. This also leads to characterize the rendering in terms of listening area wideness, or in terms of plausibly perceived effect or annoyance.

Note that the two systems are shown in their respective basic and extreme forms, *i.e.* considering virtually $R_{mic}=0$ for HOA and $R_{mic}=R_{spk}$ for WFS. Sound imaging relies respectively on "virtual source" or "notional source" encoding. For the latter, WFS involve cardioid microphones that point outwards for outside sources and inwards for inside sources. For reproduction, loudspeakers are supposed to be omnidirectional (no dipole "secondary source"). They are placed at a distance $R_{spk}=1.5m$ from the centre. We consider of course a limited number ($N=32$), thus a limited ambisonic order ($M=15$).

In all figures, instantaneous pressure amplitude is represented in grey scale. Regarding the case of monochromatic fields (which show one frequency at once), this is also the real part of the complex pressure value in the frequency domain. The length of red wide arrows represents the signal amplitude of associated loudspeakers. Reconstruction error err is then computed (for each position) as the absolute, normalized difference between the synthesized field p_{syn} and the reference (original) field p_{ref} :

$$err = abs\left(\frac{p_{ref} - p_{syn}}{p_{ref}}\right) \quad (33)$$

To better comment the artefacts, three listening positions are symbolized with small heads and are referred to in the following as: *position C* (at the Centre), *position U* (in the Upper half of the disk, and also "Upstream" regarding the wave propagation) and *position D* (in the lower half-disk, and also "Downstream").

Low frequency: border error

According to the discussion of section 3.1 and considering the restriction to a given order M , the radial expansion of the achievable sound field approximation is proportional to the wavelength. This is also true for the reproduction with infinite distance loudspeakers. For example, a 15th order approximation of a 200Hz plane wave would be achieved over more than a 4 m radius area. Nevertheless, reconstruction shown Figure 16 doesn't even reach the area boundary ($R_{spk}=1.5m$), with both HOA and WFS. It has been moreover verified that increasing the order M (and N) doesn't improve it. This illustrates the "border error" expected in point 2 of section 4.1 and explained by the restriction to a single directivity transducer array (instead of using monopole + dipole pairs).

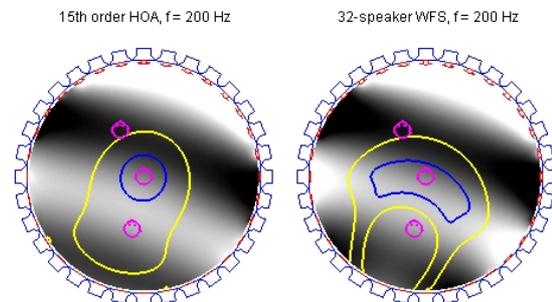


Figure 16 Reconstruction of a low frequency plane wave ($f=200$ Hz) with HOA and WFS ($R_{spk}=1.5m$). Blue/dark and yellow/bright contours enclose well-reconstructed areas with error tolerance of resp. 5% and 20%.

Nevertheless, this isn't a very damaging error (a low error tolerance is chosen for Figure 16) since it only causes a slight wave front shape distortion. Moreover, it concerns only the border where higher frequency artefacts are much more annoying, as it is illustrated below. By the way, to rightly discuss the reconstruction extent as a function of the frequency, one has to compare yellow/bright contours of Figure 16 with blue/dark ones in Figure 17 (20% error tolerance).

Higher frequency: spatial aliasing versus decreasing radial expansion

Besides the latter "border error" and according to 2.4, WFS is expected to provide a good reconstruction over the enclosed area up to the so-called "spatial aliasing frequency" f_{sp} (2). In the present case, its value is about 586Hz (let's say 600Hz). Indeed, the top of Figure 17 shows that the shape of a 600Hz plane wave is rather well preserved even outside error contours. It is noticeable that HOA provides a quite similar reconstruction quality. Moreover, in both cases the 20% error contour is about the same as for the 200Hz wave shown Figure 16.

WFS and HOA begin to distinguish from each other in terms of plane wave reproduction only above the spatial aliasing frequency, as shown Figure 17.

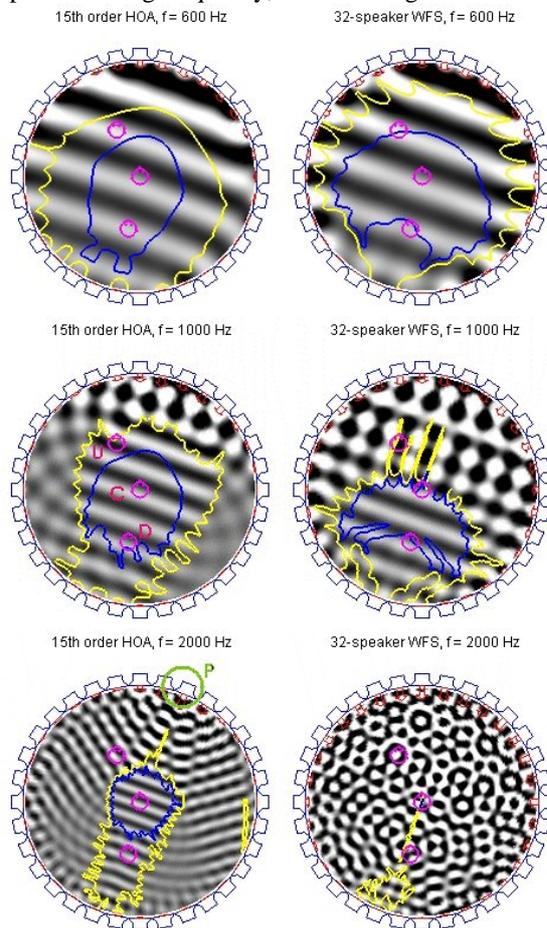


Figure 17 Reconstruction of monochromatic plane waves with HOA and WFS. Blue/dark and yellow/bright contours enclose well-reconstructed areas with error tolerance of resp. 20% and 50%.

HOA simulations clearly show that the reconstruction area progressively narrows around the centre position C when the frequency inversely

increases, as expected from discussion of section 3.1. By the way, the centred listener C is especially favoured, since reconstruction will be perfect for him up to about 10kHz and for any sound incidence. But it's worth highlighting that at other listener positions, wave front shape remains quite consistent, even if their apparent origin progressively moves to a fixed point " P " on the loudspeaker array.

At the same time but with WFS, a strong interference effect rapidly spreads over the area as the frequency increases. Correct reconstruction is still observed over the quarter of area opposite to the virtual sound incidence for $f=1000\text{Hz}$, then not at all for $f=2000\text{Hz}$ ("honeycomb" interference pattern).

To summarize: unlike with WFS, there's no spatial aliasing effect with HOA virtual imaging. A first explanation is that *spatial aliasing is related to the transducer spacing at the recording stage* (point 3 of section 4.1)... which is virtually null in the case of HOA virtual source encoding ($R_{mic}=0$).

Another explanation comes from interpreting HOA and WFS renderings in terms of *equivalent panning functions or sound pickup directivity*, observed as a function of the frequency. Indeed one understands that out of exact reconstruction conditions, *wave interferences at a given position are the stronger and the more damaging, as significant contributions come from widely spread directions and therefore are contradictory*⁹. This is very well shown with WFS, which relies on a quite low (cardioid) directivity for all frequencies. The case $f=1000\text{Hz}$ is especially instructive: sound field is highly disturbed at the listener position U (U pstream), which is surrounded by the most contributing loudspeakers; but there's no damaging interference effect at the "remote" listener position D (D ownstream), which "sees" the contributing loudspeakers as being less angularly spread. HOA has a fully different behaviour: Figure 14 means that loudspeaker contributions are used with a finest angular selectivity around the virtual source direction as the frequency increases. As a consequence, only quite slight sound field disturbances appear off-centre and at relatively high frequencies. In the end (high frequency tendency), panning law fits the "old style" rendering (*i.e.* without Near Field Control), and the sound image tends to be "projected" over the loudspeaker array⁵ (point " P ", bottom-left of Figure 17). At intermediary frequencies (*e.g.* 1000Hz),

⁹ This angular spread could be concisely characterised by the so-called "*energy vector*", this being computed locally (*i.e.* for given position and frequency). Its modulus varies from 1 (single contribution) to 0 (fully contradictory contributions).

interference patterns similar to WFS' ones appear at a distance and upstream from the centre, because of the less angular selectivity.

Consequences of artefacts in terms of audible effects

With HOA: localisation cues (especially ITD, *i.e.* Interaural Time Difference, and ILD, *i.e.* Interaural Level Difference) remain quite consistent along all the frequencies, though being progressively distorted for off-axis positions. Future listening experiments should precise the actual subjective effect.

With WFS: localisation relies essentially on cues up to the spatial frequency (thus mainly on the low frequency ITD) or a little higher, depending on where the listener is placed. Spatial information is objectively poor at higher frequency. Moreover, interference effects due to spatial aliasing are perceived as coloration effects (according to experiments done at the TUD). A slight decorrelation of loudspeaker signals, or an additional room effect, can reduce this coloration.

"Inside" (enclosed) sources

Simulating sources inside the reproduction area is a very special case of acoustic field reconstruction. Indeed, a full, true reconstruction is physically impossible in this case. Nevertheless, the following illustrates that partial (with HOA) or time inverted (with WFS) reconstruction is achievable.

Figure 18 shows the case of an inside source at a distance $\rho = 1\text{m}$ from the centre.

It is first noticeable that the spherical wave front shape seems correctly synthesized by WFS and over the whole area, up to the spatial aliasing frequency (about 600Hz). For the same frequencies, HOA only reproduces the shape over a disk of radius $\rho = 1\text{m}$, just excluding the virtual source. This may be linked to the intrinsic limitation of HOA representation (section 3.1), which validity is limited to a free field sphere ($A_{mn}^{\sigma=0}$).

A further viewing reveals that the sound field phase is inverted with WFS, *i.e.* that the synthesized wave propagates *towards* the virtual source (*time-reversing*). At the same time, HOA involves a *great energy to restore the proper direction* of propagation, especially at low frequencies (see loudspeaker feedings shown by red wide arrows, top-left of Figure 18; see also the NFC-amplification shown in Figure 13). This reconstruction effort causes strong interferences in the periphery beyond the virtual source ($\rho < r < R_{\text{spk}}$): the interference angular frequency is directly linked to the highest and most amplified spherical harmonic mode ($M=15$ periods per 2π).

Above the spatial aliasing frequency, spatial aliasing affects WFS reconstruction once again, although a

centred area is still preserved at $f=1000\text{Hz}$. With HOA, the reconstruction effort is progressively relaxed and peripheral interferences tend to be reorganised as wave fronts coming from the array projection point "P", like for the plane wave case (bottom-left of Figure 17).

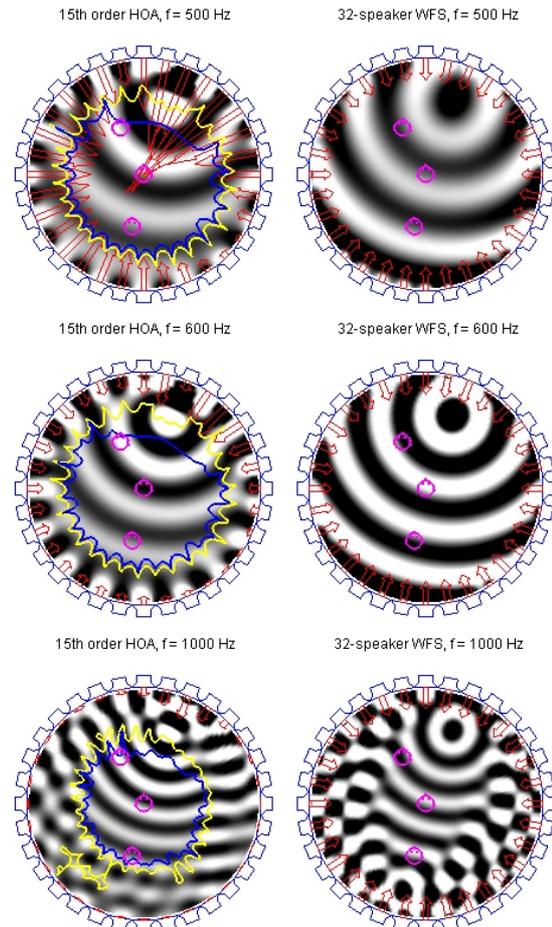


Figure 18 HOA and WFS rendering of an inside source (monochromatic spherical wave). Reconstructed spherical wave is propagating from the virtual source with HOA, but towards the same convergence point with WFS.

Consequences of artefacts in terms of audible effect

With WFS: where the time-reversed spherical wave is reconstructed, a correct ILD is expected since the spherical wave energy gradient is restored, but the ITD is inverted because of the reversed propagation. Thus these two primary localisation cues are contradictory. This leads to an amazing, but consistent effect, already noticed at the TUD.

With HOA: future listening experiments should teach us about the actual subjective effect in the presence

of the strong interference patterns, as seen beyond the source radius and for relatively low frequencies.

4.3. Recommendations and compromises

The previous section has begun pointing out some preferences and compromises on virtual sound imaging strategies. They are further discussed in the following, and then enlarged to natural sound field recording as well as encoding format issues.

Virtual sound imaging strategies: compromise for inside sources

Objective considerations would encourage using HOA as a preferred sound imaging strategies, at least for "outside" virtual sources. Indeed, HOA reconstruction quality is similar to WFS up to the spatial aliasing frequency f_{sp} (2), and above it, HOA is more robust and affected by less damaging artefacts.

A more intriguing question concerns *enclosed virtual sources*, especially *at low frequencies, below f_{sp}* . Here, there's a compromise to find between a reconstruction that preserves the propagation but is energy demanding and which radial extension is limited by the source (HOA), and wave shape preserving, but time reversing reconstruction over the whole area (WFS). Future subjective experiments are expected to bring some answers.

Real recording system: noise/error amplification versus spatial aliasing

Unlike mathematic encoding equations (4)(20)(22) for HOA virtual source imaging, natural sound field recording systems (section 3.4) must obey physical constraints. More particularly, estimating pressure field derivatives (*i.e.* ambisonic components) from capture points that are close to each other with respect to the wavelength, is all the harder, especially considering high order components. Therefore, a too small microphone array radius R_{mic} may imply a strong noise and error amplification, as stated in 3.4. On the opposite, a too large radius (*e.g.* $R_{mic}=R_{spk}$ in the extreme) causes spatial aliasing artefacts, as shown for WFS. One could envisage an intermediary radius ($0 < R_{mic} < R_{spk}$): then the artefacts observed for WFS (Figure 17) would be rescaled down according to radius R_{mic} instead of R_{spk} . Nevertheless, if one considers spherical, instead of circular, microphone arrays (while keeping fixed number N and radius R_{mic}) in order to avoid vertical aliasing (point 4 of section 4.1), then microphone spacing increases, thus spatial aliasing frequency decreases.

Noise and error issues may also be examined regarding the reconstructed field. Figure 9 shows how ambisonic components, including their

"measurement noise", participate to the sound field reconstruction as a function of kr . One notices for example that the highest order, and at the same time the noisiest components have a negligible presence at a small distance kr from the centre. From combining curves of Figure 9 and Figure 15, one could further state that the noise "recomposed" at a given listener position is directly linked to the "effort" for extrapolating the sound field knowledge from the radius R_{mic} to the listener distance $R_{listener}$. Therefore, the noise/error issues are less damageable for moderated sizes of listening area.

A last aspect is concerned with the choice of the radius R_{mic} : it is not desirable that the microphone array encloses real sources, since unlike with WFS "notional source encoding", it cannot naturally operate time reversing in order to avoid wave front curvature inversion (see 2.7, and points 1 and 5 of 4.1), and neither apply microphone inward pointing. The compromises between noise/error amplification, spatial aliasing, and listening area won't be further and more quantitatively discussed here. At least we have highlighted how HOA and WFS approaches begin to share their originally own characteristic artefacts when dealing with practicable recording systems.

3D audio encoding format

Slightly different spatial encoding formats may derive from either HOA encoding equations or WFS-like "notional source encoding" scheme, or even from their coupling [18], which would consist of the components P_{mn}^σ of equation (31), but relying on a discrete spherical integration.

Until next discussions, the HOA encoding scheme (section 3.3) is preferred as being exact, efficient and scalable at once, and is further described in [17].

5. CONCLUSION

WFS and HOA approaches have been reviewed regarding their mathematic fundamentals *and* their practical application (*i.e.* usability and efficiency), on the basis of an updated state of art. Recent and relevant progresses regarding HOA have to be noticed. The first addresses near field modelling, which allows: preserving original wave front curvatures even when considering finite distance loudspeakers; deriving distance coding filters; and defining a viable "Near Field Compensated HOA" format. The second addresses feasible, higher order microphone systems.

A formal connection has been given between both intrinsic spatial sound field representations. In addition, it has been shown that when regarding practical (recording and reproduction) systems, both

approaches begin to share their own characteristic encoding and reconstruction artefacts, and especially the spatial aliasing.

Objective artefact characterisation relying on sound field simulations, has led to globally *prefer HOA as a more robust and efficient strategies for virtual sound imaging* (virtual source encoding). Indeed, it isn't affected by spatial aliasing. Not only reconstruction is achieved beyond the spatial aliasing frequency (though over a narrowing, centred area), but also off-centre distorted wave fronts (at high frequencies) keep consistent spatial information, unlike WFS "aliased" sound field.

Nevertheless *HOA and WFS meet similar limitations and compromises when dealing with real recording systems*. Indeed, both may suffer from noise/error amplification and/or spatial aliasing, which depend on the size of the microphone array, but in opposite ways. Therefore, the array size has to be defined regarding specific constraints and priorities, like the listening area extent. Finally, the established convergence of view may help refining the design of one technique by benefiting from the knowledge of the other.

Experiments are in preparation in the France Telecom R&D Labs and have to be conducted to subjectively characterise some of the discussed artefacts in terms of degree of annoyance. An interesting comparison is expected regarding the case of sources enclosed by the loudspeaker array. This may lead to further recommendations on the virtual sound imaging (WFS or HOA) in this case.

6. REFERENCES

- [1] A.J. Berkhout, A Holographic Approach to Acoustic Control, J. Audio Eng. Soc., pp. 977-995, 1988.
- [2] A.J. Berkhout, D.d. Vries, and P. Vogel, Acoustic Control by Wave Field Synthesis, J. Acoust. Soc. Am., vol. 93, pp. 2764-2778, 1993.
- [3] R. Nicol and M. Emerit, 3D-Sound Reproduction over an Extensive Listening Area: A Hybrid Method Derived from Holophony and Ambisonic, presented at the AES 16th Int. Conference on Spatial Sound Reproduction, Rovaniemi, Finland, 1999.
- [4] E. Horbach, E. Corteel, and D.d. Vries, Spatial Audio Reproduction using Distributed Mode Loudspeaker Arrays, presented at the AES 21st Int. Conference, St Petersburg, Russie, 2002.
- [5] R. Nicol, Restitution Sonore Spatialisée sur une Zone Étendue : Application à la Téléprésence, Ph. D. Thesis, Université du Maine, Le Mans, France, 1999, http://gyronymo.free.fr/audio3D/Guests/RozennNicol_PhD.html.
- [6] E. Verheijen, Sound Reproduction by Wave Field Synthesis, Ph. D. Thesis, Faculty of Applied Physics, Delft, The Netherlands, 1996.
- [7] S. Yon, M. Tanter, and M. Fink, Sound Focusing in Rooms : The Spatio-Temporal Inverse Filter, J. Acoust. Soc. Am., 2002.
- [8] S. Yon, M. Tanter, and M. Fink, Sound Focusing in Rooms: The Time Reversal Approach, J. Acoust. Soc. Am., 2002.
- [9] M.A. Gerzon, Ambisonics in Multichannel Broadcasting and Video, J. Audio Eng. Soc., vol. 33(11), pp. 859-871, 1985 Nov.
- [10] M.A. Gerzon, Periphony : With-Height Sound Reproduction, J. Audio Eng. Soc., vol. 21(1), pp. 2-10, 1973.
- [11] E. Hulsebos, D.d. Vries, and E. Bourdillat, Improved Microphone Array Configurations for Auralization of Sound Fields by Wave Field Synthesis, preprint 5337 presented at the AES 110th Convention, Amsterdam, The Netherlands, 2001 May 12-15.
- [12] J. Daniel, Représentation de Champs Acoustiques, Application à la Transmission et à la Reproduction de Scènes Sonores Complexes dans un Contexte Multimédia, Ph.D. Thesis, University of Paris 6, Paris, France, 2000, http://gyronymo.free.fr/audio3D/download_Thesis_PwPt.html.
- [13] J.S. Bamford, An Analysis of Ambisonics Sound Systems of First and Second Order, M. Sc. Thesis, University of Waterloo, Waterloo, Ont., Canada, 1995.
- [14] M.A. Poletti, A Unified Theory of Horizontal Holographic Sound Systems, J. Audio Eng. Soc., vol. 48(12), pp. 1155-1182, 2000 Dec.
- [15] J. Daniel, J.-B. Rault, and J.-D. Polack, Ambisonic Encoding of Other Audio Formats for Multiple Listening Conditions, preprint 4795 presented at the AES 105th Convention, San-Francisco, USA, 1998 Sept.
- [16] P.M. Morse and K.U. Ingard, Theoretical Acoustics, McGraw-Hill ed, 1968.

- [17] J. Daniel, Spatial Sound Encoding Including Near Field Effect : Introducing Distance Coding Filters and a Viable, New Ambisonic Format, presented at the AES 23rd International Conference, 2003 23-25 May.
- [18] A. Sontacchi and R. Höldrich, Further investigations on 3D sound fields using distance coding, presented at the DAFX-01, Limerick, Ireland, 2001 Dec. 6-8.
- [19] P.G. Craven and M.A. Gerzon, Coincident Microphone Simulation Covering Three Dimensional Space and Yielding Various Directional Outputs, US Patent 4,042,779, filed July 7, 1975, issued Aug. 16, 1977.
- [20] T.D. Abhayapala and D.B. Ward, Theory and Design of High Order Sound Field Microphones Using Spherical Microphone Array, presented at the IEEE ICASSP-02, Orlando, Florida, USA, 2002 May 13-17.
- [21] J. Meyer and G. Elko, A Highly Scalable Spherical Microphone Array Based on an Orthonormal Decomposition of the Soundfield, presented at the IEEE ICASSP-02, Orlando, Florida, USA, 2002 May 13-17.
- [22] P. Cotterell, On The Theory of the Second-Order Soundfield Microphone, Ph. D. thesis, University of Reading, UK, 2002, <http://www.personal.rdg.ac.uk/~shr97psc/Thesis.html>
- [23] J. Daniel and S. Moreau, Theory and Design Refinement of High Order Ambisonic Microphones - Experiments with a 4th Order Prototype, presented at the AES 23rd International Conference, 2003 May 23-25.
- [24] D. Malham, Higher order Ambisonic systems for the spatialisation of sound, presented at the ICMC99, Beijing, 1999 Oct.
- [25] J. Daniel, Position Paper, presented at the ACM-SIGGRAPH and EUROGRAPHICS Campfire on Acoustic Rendering for Virtual Environments, Snowbird, Utah, USA, 2001 <http://www.bell-labs.com/topic/conferences/campfire/abstracts/daniel.pdf>.

Study of Higher Order Ambisonic Microphone

Sébastien Moreau¹ and Jérôme Daniel¹

¹ France Telecom R&D, 2 avenue Pierre Marzin, 22307 Lannion Cedex, France
sebastien.moreau@francetelecom.com, jerome.daniel@francetelecom.com

Introduction

Ambisonics is a sound spatialization technology which provides a rational and flexible way to encode and reproduce 3D sound fields on the basis of their spherical harmonics decomposition. Formerly restricted to order 1 (B-Format), it was more recently extended to higher spatial resolutions (High Order Ambisonics or HOA), thus allowing to enlarge the area of acoustic reconstruction [1]. The present paper proposes a way to design Higher Order Ambisonic microphone with a rigid sphere microphone array. Due to small array size, low frequency estimation may imply an excessive bass-boost. To limit it, we introduce a criterion based on the targeted size of the reproduction area.

Natural sound field encoding with HOA

Harmonic decomposition of a 3D sound field

In the spherical coordinate system (azimuth θ , elevation δ , radius r), solving the Helmholtz equation $(\Delta+k^2)p=0$ (with wave number $k=2\pi f/c$, frequency f and sound speed c) in a source-free region of space leads to the so-called Fourier-Bessel decomposition:

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

where $Y_{mn}^{\sigma}(\theta, \delta)$ are angular functions called *spherical harmonics*, $j_m(kr)$ are radial functions called *spherical Bessel functions*, and B_{mn}^{σ} are HOA components that describe entirely the sound field (for more details, see [1]). Pressure sound $p(\theta, \delta, kr)$ is expressed from a reference point: the listener point.

To record HOA components, we introduce in the sound field a rigid sphere centered on the reference point, and measure the sound pressure at its surface. For a given radius a , the pressure at the surface can also be expressed in terms of spherical harmonics:

$$p_a(\theta, \phi) = \sum_{m=0}^{\infty} \sum_{0 \leq n \leq m, \sigma = \pm 1} W_m(ka) B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \phi) \quad (2)$$

with

$$W_m(ka) = \frac{i^{m-1}}{(ka)^2 h_m^{-\prime}(ka)}$$

Sound field approximation by truncation

An approximation of a 3D sound field can be achieved by truncating the Fourier-Bessel series (1):

$$\hat{p}_M(\theta, \delta, kr) = \sum_{m=0}^M j_m^m(kr) \sum_{0 \leq n \leq m, \sigma = \pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \delta) \quad (3)$$

where M is the truncation order. Therefore the number of retained HOA components B_{mn}^{σ} is $K=(M+1)^2$. The spatial area of correct approximation depends on the order M and the frequency (wave number).

To characterize the approximation accuracy, we define the normalized quadratic truncation error ε_M associated to a finite order of truncation M :

$$\varepsilon_M(kr) = \frac{\int_S |p(\theta, \delta, kr) - \hat{p}_M(\theta, \delta, kr)|^2 dS}{\int_S |p(\theta, \delta, kr)|^2 dS}, \quad (4)$$

where integrals apply over a sphere S of radius kr .

Encoding natural sound field HOA components

Practically, to encode a 3D natural sound field, we need to *sample spatially the surface of the sphere* and make pressure measures only at a finite number of points Q . Each recorded signal contains a "portion" of spatial components B_{mn}^{σ} . Equation (2) can be rewritten for order of truncation M and for each measure q ($1 \leq q \leq Q$):

$$p_a(\theta_q, \phi_q) = \sum_{m=0}^M W_m(ka) \sum_{0 \leq n \leq m, \sigma = \pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta_q, \phi_q) \quad (5)$$

For correct HOA components estimation, we set $Q \geq K$ and arrange the Q microphones as uniformly as possible over the surface of the sphere.

In a global matrix formulation, (5) becomes:

$$\begin{pmatrix} p_a(\theta_1, \phi_1) \\ p_a(\theta_2, \phi_2) \\ \vdots \\ p_a(\theta_Q, \phi_Q) \end{pmatrix} = \begin{pmatrix} Y_{00}^1(\theta_1, \phi_1) & Y_{11}^1(\theta_1, \phi_1) & \dots & Y_{mn}^{\sigma}(\theta_1, \phi_1) \\ Y_{00}^1(\theta_2, \phi_2) & Y_{11}^1(\theta_2, \phi_2) & \dots & Y_{mn}^{\sigma}(\theta_2, \phi_2) \\ \vdots & \vdots & \vdots & \vdots \\ Y_{00}^1(\theta_Q, \phi_Q) & Y_{11}^1(\theta_Q, \phi_Q) & \dots & Y_{mn}^{\sigma}(\theta_Q, \phi_Q) \end{pmatrix} \times \begin{pmatrix} W_0(ka) & 0 & \dots & 0 \\ 0 & W_1(ka) & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & W_m(ka) \end{pmatrix} \begin{pmatrix} B_{00}^1 \\ B_{11}^1 \\ \vdots \\ B_{mn}^{\sigma} \end{pmatrix}$$

Or, by shortly naming matrixes and vectors:

$$p_a = Y.W.B \quad (6)$$

By inverting this spatial sampling equation, it is possible to estimate HOA components B_{mn}^σ from Q microphone signals p_a up to finite order M :

$$\mathbf{B} = \mathbf{W}^{-1} \cdot (\mathbf{Y}^t \mathbf{Y})^{-1} \mathbf{Y}^t \cdot \mathbf{p}_a \quad (7)$$

In other words, we can record HOA components of a natural sound field by matrixing microphone signals then applying equalizers $\text{EQ}_m = \mathbf{W}_m^{-1}$.

The choice of the radius of the spherical array is a compromise. Indeed, distance between microphones introduces high frequency estimation artefacts called *spatial aliasing*. Array radius should be small to confine this artefact above a frequency limit as high as possible. On the other hand, the radius should be large enough to reduce estimation effort that results in excessive bass boost of high order components (dashed curves of Figure 1). Additionally, the next method introduces considerations on reconstruction properties to limit low frequency estimation effort and therefore the bass-boost.

Limiting excessive bass-boost

Excessive bass-boost involved in sound field recording with relatively small microphone array can be interpreted as follows: especially for higher orders and low frequencies, the system tries to catch spatial information that is very poor at the measurement points and is substantial only at a distance from the microphone array. Now it appears from spherical Bessel functions $j_m(kr)$ in equation (1) that, considering the sound field reconstruction up to a given radius, higher order B_{mn}^σ become unnecessary at relatively low frequencies. Therefore we introduce a criterion to characterize useless low frequency band of HOA components depending on the order and targeted reconstruction area dimensions. This criterion is based on the quadratic error \mathcal{E}_M . Excessive amplification in EQ_m can be limited in these frequency bands thanks to high-pass filters.

Characterizing HOA components usefulness for a given radius of representation area

We assume that microphone array dimensions are small compared to the distance of sources and thus, that recorded waves are plane. Therefore, the normalized error \mathcal{E}_M becomes [2]:

$$\mathcal{E}_M(kr) = 1 - \sum_{m=0}^M (2m+1) (j_m(kr))^2 \quad (8)$$

Note that \mathcal{E}_M is independent of the source direction. By considering a targeted radius r_t of reproduction area, \mathcal{E}_M can be expressed as a function of frequency. For a given error \mathcal{E} , we can deduce from (8) a corresponding limit frequency $f_{\text{lim}}^{(M)}$ for each order. Then, we consider that low frequency band with upper limit $f_{\text{lim}}^{(M)}$ of M order component don't significantly contribute to the sound field in the targeted reproduction area compared to lower orders.

Application to sound field recording by high-passing HOA components

Low frequency bands defined previously don't need excessive amplification regarding to their usefulness. We can therefore associate high-pass filters H_m to EQ_m (Fig. 1). These filters must compensate (or even inverse) EQ_m slopes in useless low frequency bands. Another essential requirement is that the filters have to preserve EQ_m phase in the pass-band, *i.e.* that they must have a linear phase response. If the resulting amplification is judged still excessive, notably regarding to noise amplification, ambitions of estimation must be reduced by choosing a smaller target radius or a bigger error \mathcal{E} . Note that noise amplification is not only determined by EQ_m but also by matrixing operations.

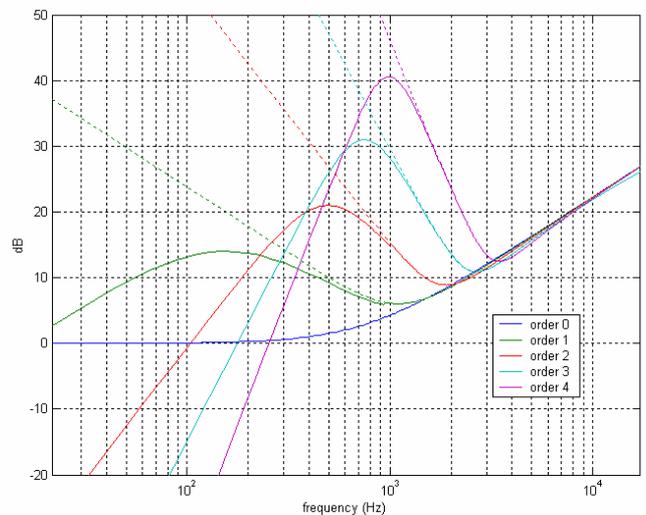


Figure 1: Order 0 to 4 HOA components equalization EQ_m (dashed lines) associated to high-pass filters (cont. lines) for $a = 0,25\text{m}$, $r_t = 0.25\text{m}$, $\mathcal{E} = 0.25$.

Conclusion

The present paper introduced a practicable way to encode natural 3D sound fields with Higher Order Ambisonics. It consists in recording signals at the surface of a rigid sphere and applying on them operations of matrixing and filtering that theoretically contain excessive bass-boost due to small microphone array dimensions. This excessive amplification is compensated thanks to high-pass filters which discard useless frequency bands of HOA components. These useless frequency bands are determined by a targeted radius of reproduction area and an error criterion. A 4th order prototype is actually under validation in our laboratories.

References

- [1] J. Daniel, R. Nicol, and S. Moreau (2003) "Further Investigations of Higher Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging", *AES 114th Convention*, Amsterdam
- [2] D.B. Ward, T.D. Abhayapala (2001) "Reproduction of a Plane-Wave Sound Field Using an Array of Loudspeakers", *IEEE Trans. on Speech and Audio Proc.*, vol. 9, n°6, pp 697-707



Audio Engineering Society Convention Paper

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

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

Further Study of Sound Field Coding with Higher Order Ambisonics

Jérôme Daniel and Sébastien Moreau

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

ABSTRACT

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

1. INTRODUCTION

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

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

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

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

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

2. PREVIOUS WORK ON HOA

2.1. Higher Order Ambisonics Basics

2.1.1. Sound field decomposition

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

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

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

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

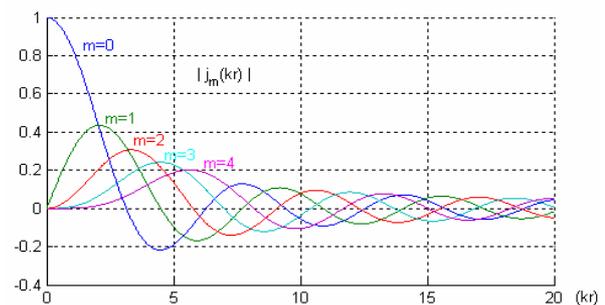


Figure 1 Spherical Bessel functions $j_m(kr)$

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

2.1.2. Encoding equations (for individual sources)

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

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

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

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

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

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

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

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

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

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

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

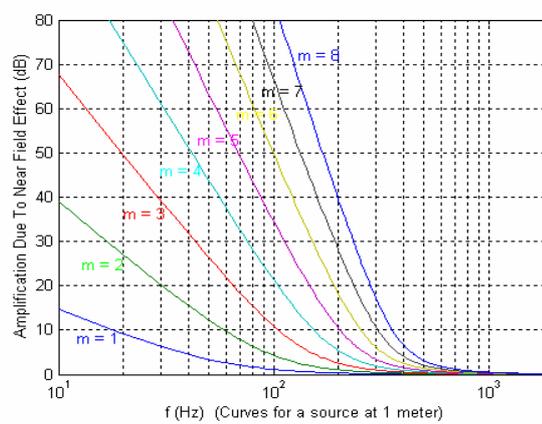


Figure 2 Low frequency infinite boost of ambisonic components due to near field effect $F_m(k\rho)$. Curves are shifted to the right (resp. the left) when the source distance ρ decreases (resp. increases).

2.1.3. Natural sound field encoding with a microphone array

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

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

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

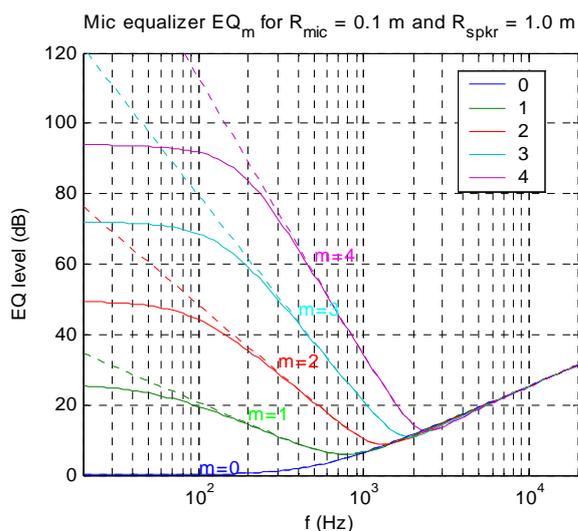


Figure 3 Equalization (EQ_m) required in microphone array processing to theoretically retrieve ambisonics components for pressure microphones distributed over a rigid sphere (radius $R_{mic}=10$ cm). Without (dotted lines) and with compensation of loudspeaker near field (cont. lines, with $R_{spr}=1$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

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

developed in 2.2 introduces finite low frequency amplification.

2.1.4. Decoding and rendering

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

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

2.2. Near-Field Compensated HOA

2.2.1. Introducing Near-Field pre-Compensation

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

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

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

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

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

2.2.2. Distance coding filters

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

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

featuring distance coding transfer functions:

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

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

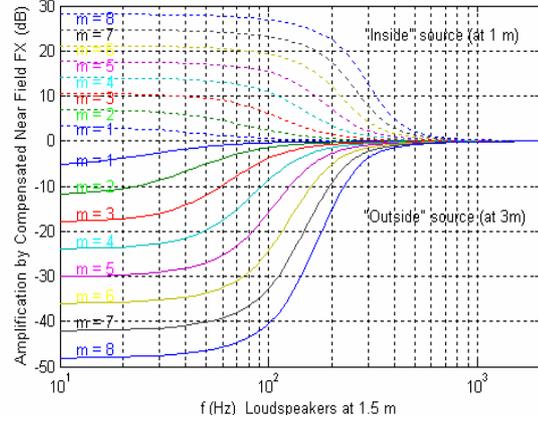


Figure 4 NFC filters frequency responses: finite amplification of ambisonic components from pre-compensated Near Field Effect (dashed lines: $\rho/R=2/3$; cont. lines: $\rho/R=2$).

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

2.2.3. Equalizers for microphone processing

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

2.3. Problematic cases in terms of bass-boost

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

2.3.1. Sound field recording with small size microphone array

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

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

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

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

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

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

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

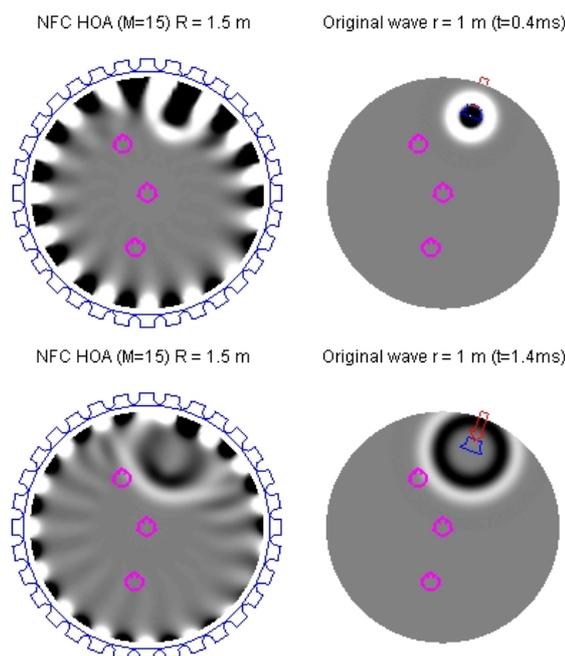


Figure 5 Two snapshots (time domain) of spherical wave synthesis with NFC-HOA for an enclosed virtual source ($r=1\text{m} < R=1.5\text{m}$). Left: 15th order rendering over a 32-loudspeaker circular array. Right: reference wave from a single source.

3. A SAFE ENCODING STRATEGY

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

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

3.1. Characterizing representation usefulness

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

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

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

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

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

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

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

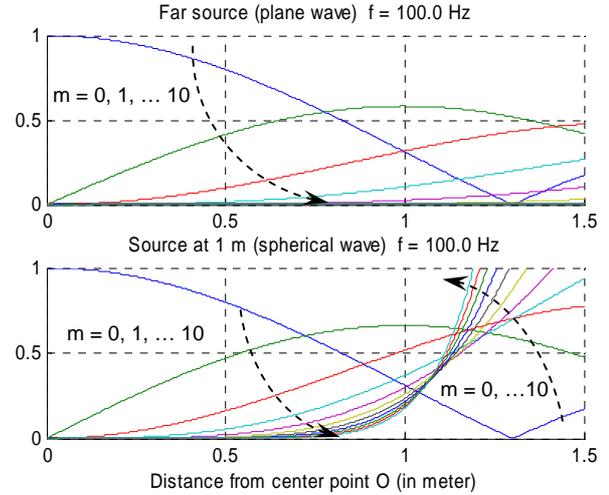


Figure 6 Radial contribution of spatial components B_{mm}^σ of orders $m=0$ to 10 and for a low frequency $f=100$ Hz in the cases of a plane wave (top) and a spherical wave (bottom, with $\rho=1$ m), as quantified by functions $|J_m(kr)|$ and $|J_m(kr)F_m(k\rho)|$ respectively.

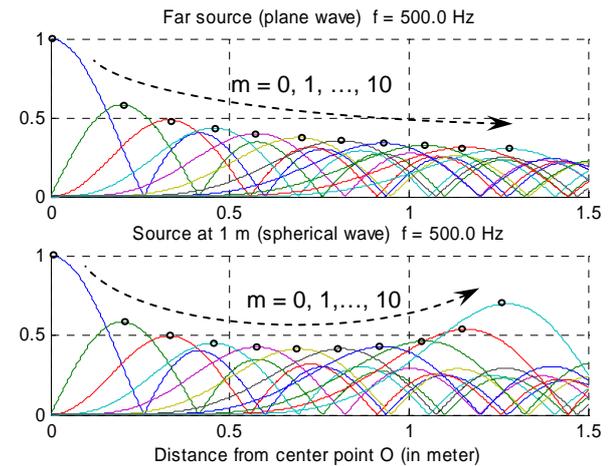


Figure 7 Same as Figure 6 but for a frequency $f=500$ Hz

3.1.2. Considerations on the plane wave case

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

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

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

3.1.3. Special considerations on spherical wave

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

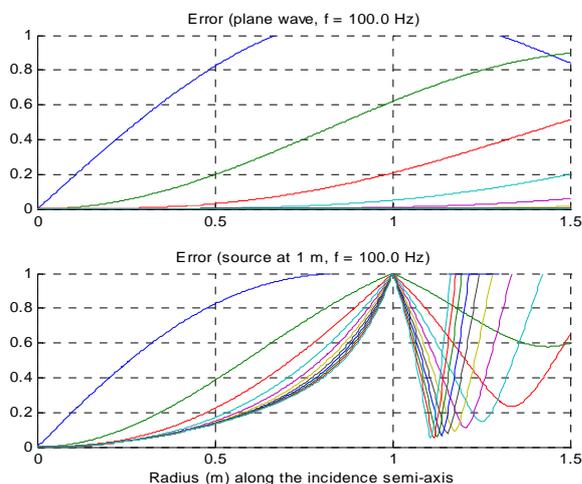


Figure 8 Relative error of sound field reconstruction along the semi-axis oriented towards the wave incidence ($\gamma=0$), in conditions similar to Figure 6.

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

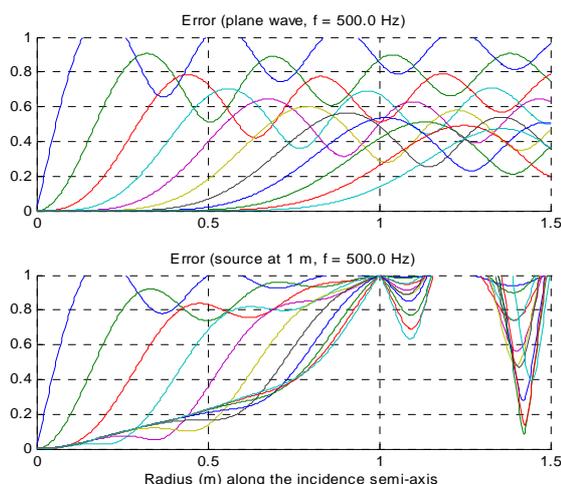


Figure 9 Same as Figure 8 for conditions similar to Figure 7 ($f=500\text{Hz}$)

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

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

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

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

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

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

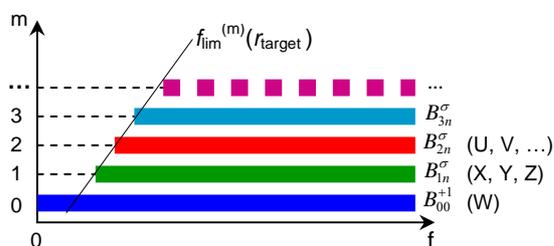


Figure 10 Schematic view of "useful" high frequency bands for the first few groups (colored horizontal bars) of spatial components (mentioned on the right part) according to their order m .

3.1.5. Frequency dependent truncation by high-passing HOA components

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

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

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

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

3.2. Application to sound field recording

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

3.3. Application to virtual source encoding

3.3.1. Toward a relevant criterion for truncation

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

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

3.3.2. Energy focalization on the virtual source location

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

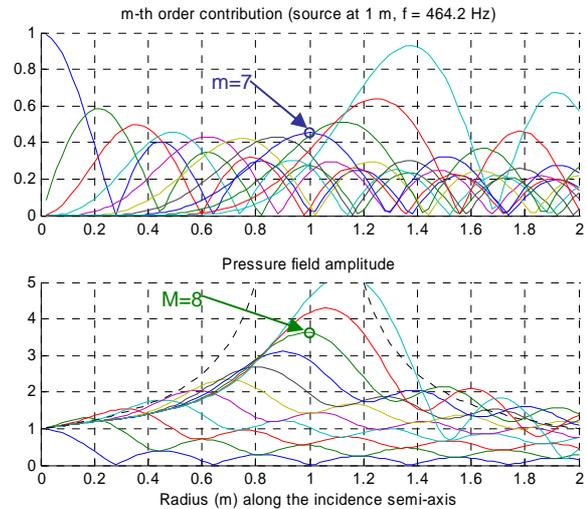


Figure 11 Top: radial contribution $J_m(kr) \cdot F_m(k\rho)$ of spatial components of various orders m for a source at distance ρ . Bottom: associated sound field approximations (truncated series in absolute value) on the semi-axis passing through the virtual source. Dotted curve indicates $\rho/|r-\rho|$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

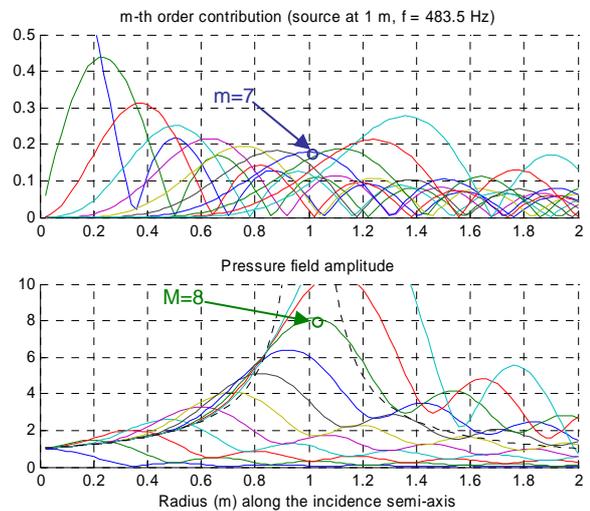


Figure 12 Same comments as for Figure 11, but with spherical Bessel functions and "3D" sound field approximation.

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

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

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

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

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

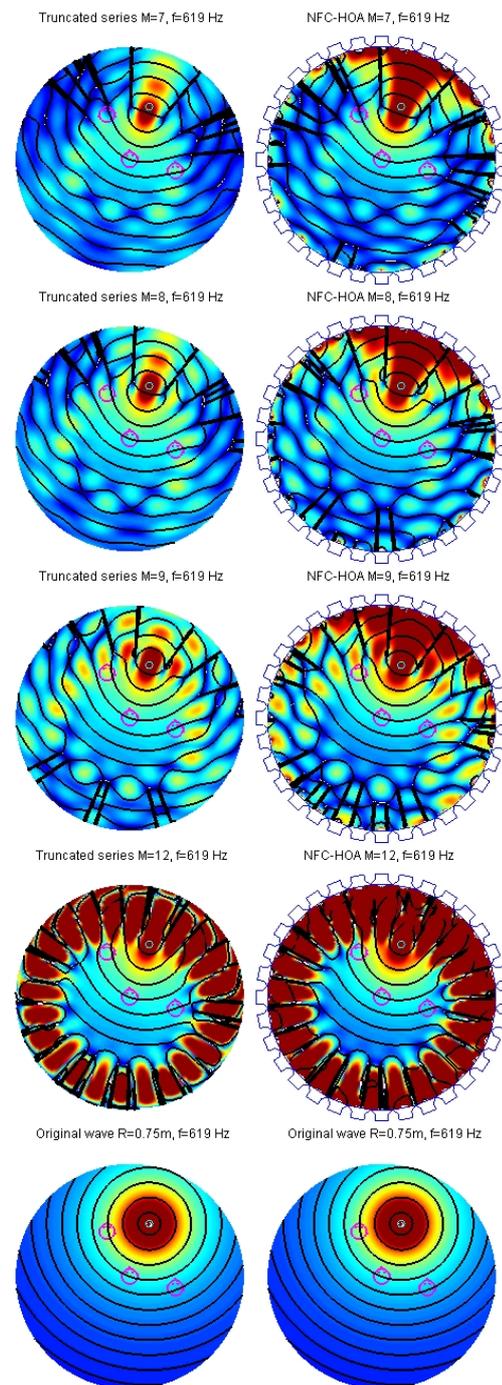


Figure 13 Top view of 2D sound field approximations (truncated Fourier-Bessel series (11): left) and reconstructions (by a 32-loudspeaker array: right), for a source simulated at 75cm from the center (bottom views). Orders $M=7, 8, 9, 12$ from top to bottom.

3.3.3. Refining high-pass filters specifications

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

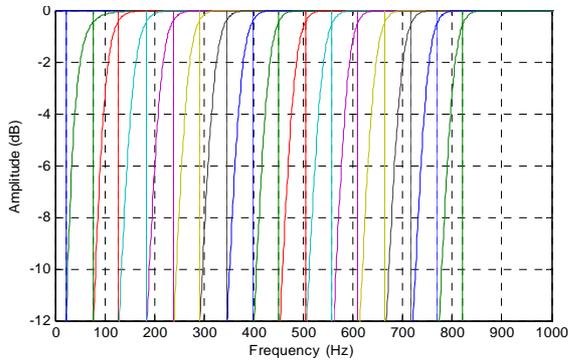


Figure 14 Spectra of high-pass filters $H^{(m)}$ designed to limit functions $F_m(k\rho)$ for orders $m=1$ to 15 (from left to right) and a source distance $\rho=1\text{m}$. For each curve (of $H^{(m)}$), pass-band frequency limit is marked by a vertical bar (with the same color as the curve). The same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

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

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

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

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

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

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

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

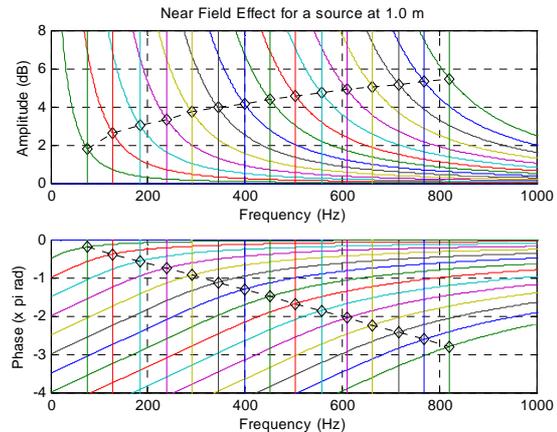


Figure 15 Amplitude and phase of Near Field Functions $F_m(k\rho)$ with $\rho=1\text{m}$ and for orders $m=1$ to 15 (from left to right). Vertical bars mark frequency limits $f_{\text{lim}}^{(m)}$.

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

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

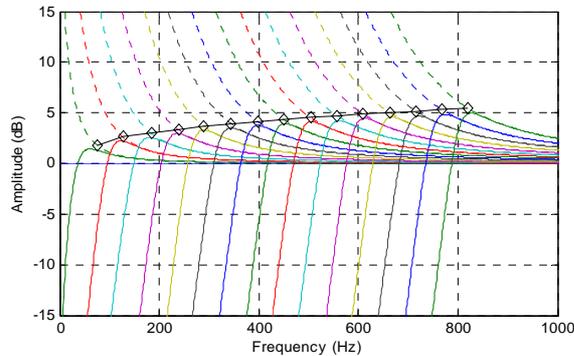


Figure 16 Spectra of high-passed Near Field functions $H^{(m)}(\omega).F_m(\omega)$ for orders $m=1$ to 15 (from left to right) and a source distance $\rho=1\text{m}$. Original Near Field functions $F_m(\omega)$ are shown in dotted line for comparison. Diamonds show for each order m the highest "useful"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

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

3.4. Distributing bass rendering effort amongst loudspeakers

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

4. CONCLUSION

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

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

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

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

5. REFERENCES

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

Production, transmission et restitution temps réel d'une scène sonore dans un format audio 3D flexible

Sébastien Moreau¹ Jérôme Daniel¹ Adil Chraa¹

¹ France Télécom R&D - DIH / IPS

2, avenue Pierre Marzin, 22 307 Lannion cedex – France

{sebastien.moreau, jerome.daniel, adil.chraa}@francetelecom.com

Résumé

Cet article présente une plateforme de production, transmission et restitution temps réel d'une scène sonore 3D. Parmi les formats audio existants, le format High Order Ambisonics (HOA) a été retenu pour sa souplesse d'utilisation. Des outils d'exploitation, inexistant pour ce format encore récent, ont été élaborés pour développer la plateforme temps réel. Cette dernière est constituée d'un microphone haute résolution, d'un module de transmission/réception par réseau IP, et d'un décodeur qui permet la restitution sur divers systèmes de diffusion. L'implantation a été effectuée avec la technologie VST de Steinberg et s'adapte à beaucoup de logiciels audio grand public. La plateforme réalisée peut concerner la diffusion musicale, radiophonique, etc.

Mots clefs

High Order Ambisonics, spatialisation sonore, format audio.

1 Introduction

Nous présentons ici un système temps réel de spatialisation sonore dont le but est de transmettre et de reconstruire un champ sonore 3D naturel dans un lieu distant du champ original. Un format de représentation du champ acoustique a été choisi parmi ceux existants en fonction de critères objectifs tels que la souplesse d'utilisation et la quantité d'information à transmettre.

2 Représentation d'une scène sonore 3D : quel format ?

Une technique de spatialisation sonore permet de représenter les informations spatiales et temporelles d'un champ acoustique dans le but de les stocker, transmettre, ou reproduire. Ceci pose le problème de la définition d'un format de représentation du champ acoustique.

2.1 Considérations préliminaires

La Figure 1 représente les principes de bases de la spatialisation sonore : l'encodage spatial d'une scène

sonore dans un format de représentation, puis le décodage de la scène formatée permettant sa reproduction.

Le format de représentation contient les informations spatiales et temporelles du champ sonore nécessaires à sa reconstruction. Il décrit donc, dans les limites de ses capacités, un champ sonore. Il est constitué d'un ensemble de signaux temporels auxquels peuvent être associés des paramètres de spatialisation.

L'encodage spatial désigne le processus qui permet d'acquérir, de traduire une scène sonore dans un format spécifique de représentation.

Le décodage spatial concerne quant à lui l'adaptation des données formatées au système de restitution (casque, 5.1, etc.). Si le format est constitué de signaux destinés à alimenter directement un système de diffusion particulier, le décodage n'est pas nécessaire (binaural, stéréophonie).



Figure 1 – Principe de base de la spatialisation sonore

2.2 Choix du format

Le choix du format de représentation d'une scène sonore est un choix décisif puisque de lui dépend la qualité potentielle de la reconstruction spatiale des sources sonores. Pour le développement de notre application temps réel, le choix s'est effectué selon trois critères :

- la possibilité de manipuler la scène sonore encodée ;
- la flexibilité par rapport au système de restitution ;
- la concision de la scène sonore encodée.

On entend par manipulation d'une scène sonore encodée le fait de modifier la position spatiale des sources sonores déjà présentes.

D'autre part, un format est dit flexible s'il peut s'adapter à divers systèmes de restitution pour la reconstruction de la scène sonore.

Enfin, la concision de la représentation est une contrainte imposée par les ressources matérielles limitées. Il est donc avantageux qu'un format de représentation contienne peu d'informations redondantes ou qu'il puisse être compressé

sans entraîner une dégradation de la qualité (du moins perceptive) de reproduction.

2.3 Comparaison des principaux formats existants

Le Tableau 1 permet de comparer les formats associés aux principales techniques de spatialisation existantes suivant les trois critères définis plus haut. Ces principaux formats sont : la *stéréophonie* (deux canaux ou plus) ; le *binaural* ; *High Order Ambisonics* (HOA) ; et le *tout paramétrique*.

	Manipulation de la scène sonore	Dispositif de restitution	Taille de la scène encodée
Stéréophonie (2.0, 5.1, etc.)	Impossible, scène sonore figée	fixe, lié au format	1 signal par HP
Binaural			2 signaux
HOA (ordre M)	globale (rotation, etc.)	adaptable, indépendant du format	3D : $(M+1)^2$ 2D : $(2M+1)$ signaux
tout paramétrique	chaque source de façon indépendante		1 signal par source + paramètres de position

Tableau 1 - Comparaison des principaux formats de représentation de scène sonore

Le format *stéréophonique* tel que défini ici, est constitué de signaux destinés à alimenter directement les haut-parleurs d'une configuration prédéfinie (2.0, 4.0, 5.1, etc.). Le format *binaural* est composé de deux signaux, appelés *signaux binauraux*, contenant les transformations subies par le son lors de son interaction avec le corps de l'auditeur (pavillons, tête et torse) et devant être reproduit au niveau des tympons au moyen d'un casque.

Le format associé à la technologie *High Order Ambisonics* (HOA) est constitué de signaux appelés *signaux HOA* qui résultent de la décomposition en harmoniques sphériques d'un champ sonore en un point. Le nombre de signaux dépend de l'ordre M de la décomposition : $(M+1)^2$ en 3D et $(2M+1)$ en 2D. De l'ordre dépend également la résolution spatiale de la description.

Enfin, un format *tout paramétrique* contient autant de signaux que de sources sonores, auxquels sont associés des paramètres de position spatiale. La reconstruction est ensuite effectuée lors du décodage selon la technique de spatialisation choisie (*binaural*, *HOA*, *Wave Field Synthesis*, etc.) et le dispositif de restitution.

2.4 HOA : un format avantageux

Les formats *HOA* et *tout paramétrique* sont particulièrement intéressants car ils permettent de

représenter une scène sonore 3D indépendamment du système de restitution. Bien que le second paraisse offrir plus de possibilités en terme de manipulation de scène sonore, il comporte un nombre important de signaux lorsque le champ acoustique encodé est complexe. De plus il suppose la prise de son isolée des sources sonores et s'adapte donc mal à l'enregistrement naturel en 3D.

Le format HOA semble réaliser un bon compromis entre taille et souplesse d'utilisation. Il a donc logiquement été choisi pour développer notre application temps réel. Cependant, les outils de production qui lui sont associés sont à l'heure actuelle relativement limités. Nous en avons nous-même développés sur les bases théoriques suivantes.

3 High Order Ambisonics

Cette section introduit les éléments théoriques qui nous permettrons d'exploiter le format HOA dans notre application temps réel. Seuls certains aspects correspondant aux fonctions requises seront abordés : l'encodage de champ sonore naturel, les rotations du champ sonore, et le décodage pour divers systèmes de restitution.

3.1 Série de Fourier-Bessel

La technologie *High Order Ambisonics* (HOA) se base sur une représentation harmonique spatiale du champ sonore, solution de l'équation d'onde exprimée en coordonnées sphériques (azimut θ , élévation δ , rayon r). Sur une zone exempte de source sonore, cette solution est la *série de Fourier-Bessel* :

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

Les fonctions radiales $j_m(kr)$ sont les *fonctions de Bessel sphériques* de première espèce d'ordre m (k étant la longueur d'onde). Les fonctions angulaires $Y_{mn}^{\sigma}(\theta, \delta)$ sont les *fonctions harmoniques sphériques*. Et enfin, les signaux B_{mn}^{σ} exprimés ici dans le domaine fréquentiel, sont les *signaux HOA* qui décrivent le champ sonore sur la zone considérée [1].

Notons que les harmoniques sphériques sont orthogonales deux à deux, ce qui se traduit mathématiquement par l'équation suivante :

$$\int_S Y_{mn}^{\sigma}(\theta, \delta) Y_{m'n'}^{\sigma'}(\theta, \delta) dS = \delta_{mm'} \delta_{nn'} \delta_{\sigma\sigma'} \quad (2)$$

où $\delta_{mm'}$ est le symbole de Kronecker, l'intégration se faisant sur la sphère unité.

En tronquant la série (1) à un ordre fini M ($m \leq M$), on obtient une approximation du champ sonore sur une zone de l'espace d'autant plus grande que M est élevé.

3.2 Encodage de champ sonore naturel 3D

L'encodage de champ naturel au format HOA consiste à extraire en un point d'un champ acoustique les signaux

B_{mn}^σ . La stratégie adoptée consiste à mesurer la pression sonore à la surface d'une sphère rigide de rayon a fixé. Celle-ci peut s'exprimer de la façon suivante :

$$p_a(\theta, \delta) = \sum_{m=0}^{\infty} W_m(ka) \sum_{0 \leq n \leq m, \sigma = \pm 1} B_{mn}^\sigma Y_{nm}^\sigma(\theta_q, \delta_q) \quad (3)$$

avec
$$W_m(ka) = \frac{i^{m-1}}{(ka)^2 h_m^-(ka)}$$

En utilisant la *propriété d'orthogonalité* des harmoniques sphériques (2), on peut calculer d'après l'équation (3) les signaux HOA (projection sur la base des harmoniques sphériques) :

$$B_{mn}^\sigma = \frac{1}{W_m(ka)} \int_S p_a(\theta_q, \delta_q) Y_{nm}^\sigma(\theta_q, \delta_q) dS \quad (4)$$

En pratique, la pression sonore n'est mesurée qu'en un nombre fini Q de positions à la surface de la sphère. On effectue alors une approximation de la décomposition harmonique jusqu'à un ordre M maximal restreint :

$$B_{mn}^\sigma = \frac{1}{W_m(ka)} \sum_{q=1}^Q p_a(\theta_q, \delta_q) Y_{nm}^\sigma(\theta_q, \delta_q), \quad m \leq M. \quad (5)$$

Le nombre de signaux HOA estimé est alors de $K=(M+1)^2$. Pour une bonne estimation, on choisit Q tel que $Q \geq K$. D'autre part, l'équation (5) suppose que la base échantillonnée des harmoniques sphériques (représentée par les positions des microphones) est orthogonale. Dans des conditions plus générales, on préférera à cette *méthode de projection*, une méthode d'estimation basée sur une *approximation aux moindres carrés* [2].

Quelque soit la méthode choisie, les signaux HOA sont obtenus par *matricage* puis *égalisation* $EQ_m = 1/W_m(ka)$ des signaux microphoniques. Cette égalisation nécessite, notamment en basse fréquence, d'être limitée en amplitude pour sa mise en œuvre [2].

3.3 Rotation de la scène sonore

Lorsque la scène sonore pivote (suivant trois degrés de liberté), le groupe formé par les $(2m+1)$ signaux HOA de même ordre m est transformé de la façon suivante :

$$b'_m = R_m \cdot b_m, \quad (6)$$

avec $b'_m = [b'_{mm}, b'_{m,m-1}, \dots, b'_{m,1}, b'_{m,0}]^t$ et

$$b_m = [b_{mm}, b_{m,m-1}, \dots, b_{m,1}, b_{m,0}]^t.$$

b'_m est le vecteur contenant les $(2m+1)$ signaux HOA, exprimés dans le domaine temporel, résultant de la manipulation, R_m la matrice de dimensions $(2m+1) \times (2m+1)$ caractérisant la rotation, et b_m le vecteur contenant les $(2m+1)$ signaux HOA de départ, exprimés également dans le domaine temporel.

La rotation de scène sonore est ici destinée à être utilisée en association avec un *head-tracker* pour corriger les mouvements de tête en écoute binaurale (avec casque). Nous ne considérons donc ici que les rotations dans le

plan azimutal (rotation gauche-droite). Dans ce cas particulier, chaque couple de signaux (b_{mn}^l, b_{mn}^{l-1}) , m et n étant fixés, s'obtiennent pour une rotation d'angle θ' d'après la relation :

$$\begin{bmatrix} b_{mn}^l \\ b_{mn}^{l-1} \end{bmatrix} = \begin{bmatrix} \cos(n\theta') & -\sin(n\theta') \\ \sin(n\theta') & \cos(n\theta') \end{bmatrix} \begin{bmatrix} b_{mn}^1 \\ b_{mn}^{-1} \end{bmatrix}. \quad (7)$$

Nous pouvons ainsi calculer la matrice de rotation en fonction de l'ordre m considéré. Par exemple, pour l'ordre $m = 2$, la matrice caractérisant une rotation θ' en azimut est la suivante :

$$R_2 = \begin{bmatrix} \cos 2\theta' & -\sin 2\theta' & 0 & 0 & 0 \\ \sin 2\theta' & \cos 2\theta' & 0 & 0 & 0 \\ 0 & 0 & \cos \theta' & -\sin \theta' & 0 \\ 0 & 0 & \sin \theta' & \cos \theta' & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (8)$$

3.4 Décodage

La *décodage* consiste à adapter une scène sonore encodée au format HOA à un dispositif particulier de L haut-parleurs, que l'on choisit concentrique. Les haut-parleurs sont au moins aussi nombreux que les K signaux HOA.

On considère que les haut-parleurs émettent des ondes planes (champ lointain) qui contribuent à la reconstruction des signaux HOA suivant l'équation matricielle :

$$b = C \cdot s, \quad (9)$$

où $b = (b_{00}^1, b_{11}^1, b_{11}^{-1}, \dots, b_{mm}^\sigma)^t$ est le vecteur constitué par les signaux HOA (domaine temporel), C est la matrice contenant les gains $Y_{mn}^\sigma(\theta_l, \delta_l)$ associés à la direction l de chaque haut-parleur, et enfin $s = (s_1, s_2, \dots, s_L)^t$ est le vecteur contenant les signaux temporels émis par les haut-parleurs.

Ces signaux sont obtenus en inversant l'équation (9), soit par combinaison (matricage) des signaux HOA :

$$s = D \cdot b \quad (10)$$

avec

$$D = C^t (C C^t)^{-1}.$$

D est appelé la *matrice de décodage*, c'est la matrice pseudo-inverse de C .

Si on désire effectuer un décodage pour le casque d'écoute, il est nécessaire d'ajouter une étape de simulation binaurale des haut-parleurs : on applique aux signaux issue de (10) des filtres binauraux $h_G(\theta, \delta)$ et $h_D(\theta, \delta)$ caractérisant le canal acoustique entre les haut-parleurs et les oreilles gauche et droite.

Pour plus d'efficacité, on synthétise les opérations de matricage et de filtrage binaural sous la forme d'un ensemble f de filtres s'appliquant directement sur les signaux HOA. On obtient :

$$f_G = h_G \cdot D \text{ et } f_D = h_D \cdot D, \quad (11)$$

avec $f_G = [f_{G00}^1, \dots, f_{Gmm}^\sigma]$, $f_D = [f_{D00}^1, \dots, f_{Dmm}^\sigma]$,

$$h_G = [h_G(\theta_1, \delta_1), \dots, h_G(\theta_L, \delta_L)], \quad h_D = [h_D(\theta_1, \delta_1), \dots, h_D(\theta_L, \delta_L)].$$

Le décodage consiste alors à effectuer les opérations de filtrage suivantes (* désigne l'opération de convolution) :

$$g = \sum_{m,n,\sigma} b_{mn}^{\sigma} * f_{Gmn}^{\sigma} \quad \text{et} \quad d = \sum_{m,n,\sigma} b_{mn}^{\sigma} * f_{Dmn}^{\sigma}, \quad (12)$$

où g et d sont les signaux binauraux gauche et droit.

4 Plateforme temps réel

La technologie HOA a donc été choisie pour réaliser un système d'acquisition, transmission, et restitution temps réel d'une scène sonore 3D naturelle. Plusieurs applications sont visées : diffusion sur Internet d'émissions radiophoniques spatialisées, de concerts musicaux ; partage d'ambiance sonore, etc.

4.1 Description

La Figure 2 montre les différents éléments constituant la plateforme développée. Un *Encodeur* encode un champ sonore naturel 3D au *format HOA*. La scène encodée est ensuite transmise au *décodeur* qui permet de l'adapter à divers systèmes de restitution.

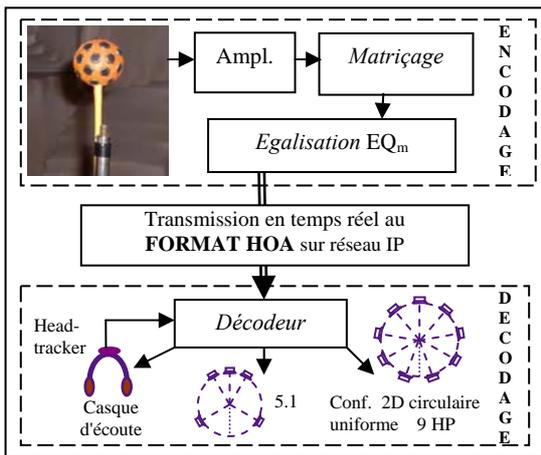


Figure 2 - Description de la plateforme temps réel

L'encodage du champ sonore 3D s'effectue grâce à un microphone pouvant estimer les signaux HOA jusqu'à l'ordre 4. Il est constitué de 32 capsules disposées à la surface d'une sphère rigide de rayon 2,6 cm. Les 32 signaux acquis sont matricés puis égalisés (section 3.2) de façon à obtenir un champ sonore encodé au format HOA indépendant du système microphonique.

Le décodage adapte la scène sonore encodée à divers systèmes de restitution : 5.1, configuration 2D circulaire uniforme de 9 HP, casque d'écoute (simulation binaurale d'une configuration 3D de 12, 20 ou 32 HP suivant l'ordre considéré) avec *head-tracker*, donnant au décodeur en temps réel les informations de position de tête.

4.2 Implantation

La technologie VST (*Virtual Studio Technology*) de Steinberg a été choisie pour l'implantation des différents

modules de la plateforme décrite ci-dessus. Cette technologie permet d'intégrer dans un grand nombre de logiciels audio professionnels et grand public (Cubase, Wavelab, Nuendo, etc.), sous forme de *plugins*, les traitements relatifs à la technologie HOA.

Quatre *plugins VST* ont été développés :

- Un premier concerne les traitements de matriçage/égalisation des signaux microphoniques.
- Un *plugin serveur* permet ensuite l'envoi de la scène encodée sur le réseau vers une adresse IP.
- Un *plugin client* a en charge la réception, sur un ordinateur distant, de la scène transmise.
- Enfin, un dernier *plugin* effectue le décodage de la scène sonore encodée. Il prend comme paramètre la nature du dispositif de restitution choisi.

4.3 Transmission de la scène sonore 3D

La transmission de la scène sonore 3D encodée est une étape délicate puisqu'elle doit se faire en temps réel et concerne une quantité d'information importante : à l'ordre 4, 25 signaux HOA sont nécessaires en 3D.

Il peut donc s'avérer indispensable de réduire cette quantité d'information. Plusieurs stratégies sont étudiées. La première suppose que l'on connaisse à l'avance le dispositif de restitution et consiste à effectuer le décodage avant transmission, réduisant ainsi le nombre de canaux au nombre de haut-parleurs ou à deux pour le casque d'écoute. Il est également possible de réduire la résolution spatiale (ordre $M < 4$, restriction 3D → 2D). On peut enfin avoir recours à la compression audio (MP3, AAC).

Différents tests de transmission sur un LAN et sur Internet sont actuellement menés. Ils permettront d'évaluer, en fonction du débit disponible, l'impact de la réduction des informations transmises sur la qualité de restitution.

5 Conclusion

Un système d'acquisition, transmission et restitution en temps réel d'une scène sonore 3D a été développé et est actuellement en cours de test. Il est basé sur la technologie *High Order Ambisonics* dont l'intérêt réside dans la souplesse d'utilisation, notamment pour la restitution. Un certain nombre d'outils, encore inexistant pour cette technologie récente, ont été conçus pour mener à bien ce projet. Les applications possibles sont multiples : diffusion radiophonique, musicale, partage d'ambiance sonore, etc.

Références

- [1] J. Daniel, R. Nicol, et S. Moreau. Further Investigations of Higher Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging, 114^{ème} convention de l'AES, Amsterdam, 2003.
- [2] S. Moreau et J. Daniel. Study of Higher Order Ambisonic Microphone. Dans *Actes de la conférence CFA/DAGA'04*, Strasbourg, Mars 2004.



Audio Engineering Society

Convention Paper

Presented at the 120th Convention
2006 May 20–23 Paris, France

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

3D Sound Field Recording with Higher Order Ambisonics – Objective Measurements and Validation of a 4th Order Spherical Microphone

Sébastien Moreau¹, Jérôme Daniel², and Stéphanie Bertet²

¹ 23/25 boulevard Edgard Quinet, 92700 Colombes, France
moreau-sebastien@ifrance.com

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

ABSTRACT

Higher Order Ambisonics (HOA) is a flexible approach for representing and rendering 3D sound fields. Nevertheless, lack of effective microphone systems limited its use until recently. As a result of authors' previous work on the theory and design of spherical microphone arrays, a 4th order HOA microphone has been built, measured and used for natural recording. The present paper first discusses theoretical aspects and physical limitations proper to discrete, relatively small arrays (spatial aliasing, low-frequency estimation). Then it focuses on the objective validation of such microphones. HOA directivities reconstructed from simulated and measured 3D responses are compared to the expected spherical harmonics. Criteria like spatial correlation help characterizing the encoding artifacts due to the model limitations and the prototype imperfections. Impacts on localisation criteria are evaluated.

1. INTRODUCTION

Ambisonics is a 3D sound spatialisation technology developed mostly by Michael Gerzon [1] in the early 1970s, which permits to render 3D sound fields in a flexible way from the knowledge of their 1st order directive information at one point: omnidirective (W) and 3D bidirective (X, Y, Z) components which constitute the so-called B-Format. Nevertheless, its low spatial resolution limits the correct sound field reconstruction to a small listening area, especially for

high frequencies. Higher Order Ambisonics (HOA) [2-4] extends the B-Format to higher resolution by means of spherical harmonic decomposition of the sound field, which results in enlarging the reproduction area (*i.e.* the sweet spot).

A practical 1st order HOA recording system was described from the very beginning by Craven and Gerzon [5], from which the so-called SoundField microphone was built. Nevertheless, until recently physical limitations of small discrete microphone arrays didn't make possible the implementation of HOA pickup systems with higher spatial resolution. Authors'

previous work and other recent studies brought practical solutions to design such microphone arrays. Different strategies have been proposed, including spherical microphone arrays [6-12] and arrays with multiple radii [13].

After reviewing HOA basics, the present paper focuses on the objective validation of spherical microphones arrays for HOA recording. Theoretical aspects of discrete sensor arrays are discussed and highlight physical limitations proper to relatively small discrete arrays, especially concerning spatial aliasing and low frequency estimation. The latter involves filters with excessive bass-boost which need to be reduced for practical purpose. This is done by introducing regularization filters in the estimation process. Finally, a 4th order prototype have been built, measured and used for natural recording for objective validation of such microphones. HOA directivities reconstructed from simulated and measured 3D responses are compared to the expected spherical harmonics. Criteria like spatial correlation help characterizing the encoding artifacts due to the model limitations and the prototype imperfections.

2. HIGHER ORDER AMBISONICS BASICS

This section draws a brief overview of basic concepts and properties of HOA. Before introducing further acoustics fundamentals in 2.2, we first restrain to horizontal systems description in 2.1. All concepts introduced in this section are more thoroughly discussed and extended to 3D in [3, 14]. They are useful as a starting point for the problematic of HOA recording. They also provide complementary characterisation criteria to asses the microphone performance from the reproduction point of view.

2.1. Spatial encoding and decoding for virtual sound imaging

Ambisonics can be categorized in the family of amplitude panning techniques, when dealing with virtual source spatialisation, and/or coincident microphone techniques when dealing with sound field recording. Among them Ambisonics has the particularity of splitting the spatialisation process into spatial encoding and spatial decoding.

2.1.1. Spatial encoding

Spatial encoding itself is a kind of amplitude pan-pot but results in an intermediary format that is not dedicated to any loudspeaker setup for sound reproduction. Equations system (1) shows encoding laws for a source in azimuth direction θ_v and emitting a signal S_v :

$$\begin{aligned} W &= B_{00}^{+1} = 1 \cdot S_v \\ X &= B_{11}^{+1} = \sqrt{2} \cos \theta_v \cdot S_v \\ Y &= B_{11}^{-1} = \sqrt{2} \sin \theta_v \cdot S_v \\ &\vdots \\ B_{mm}^{+1} &= \sqrt{2} \cos m\theta_v \cdot S_v \\ B_{mm}^{-1} &= \sqrt{2} \sin m\theta_v \cdot S_v \\ &\vdots \end{aligned} \quad (1)$$

The first three lines define the so-called “1st order horizontal encoding” and yield the horizontal components (W,X,Y) of the well-known B-Format introduced by Gerzon [15]. For each new "order" $m > 1$ up to a maximum order M , a pair of additional lines exhibits Higher Order Ambisonics encoding functions and associated components B_{mm}^{σ} . Note that the order m corresponds to the angular frequency. An M^{th} order 2D sound field representation comprises $(2M+1)$ components.

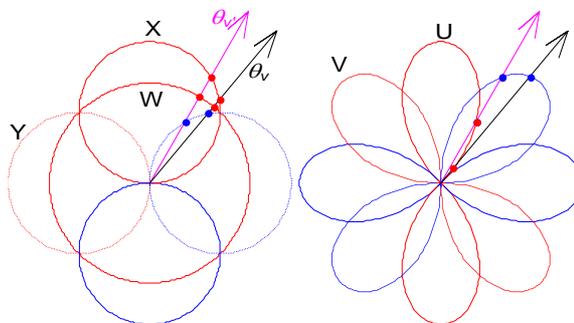


Figure 1 Horizontal encoding using 0th, 1st order (left) and 2nd order (right) directivity patterns. Positive and negative gains are resp. in red and blue. For each source (in arrow's direction) and each directivity, the encoding gain is given by the intersection point between arrow and pattern (more precisely its distance from the centre).

As illustrated on the left part of Figure 1, 1st order encoding offers a quite limited angular discrimination for virtual sources that are relatively close to each other. This is probably a first cause of the lack of sound image

precision and robustness criticized by detractors of Ambisonics (and of coincident microphone techniques more generally). By introducing “higher order ambisonic” encoding functions (1), *i.e.* functions which variation is greater for the same angular variation, a better angular discrimination is possible (right part of Figure 1 illustrates encoding with 2nd order functions).

Left column of Figure 3 shows polar patterns of ambisonics directivities up to the 4th order. Since 1st order recording requires omni and bidirectional microphones, or a combination of *e.g.* cardioids, it easily makes sense for sound engineers. The latter might be dubious regarding practical recording using higher order directivities since no existing single microphone has one of such directivities. It is the object of this paper to prove that an efficient solution exists.

2.1.2. Spatial decoding

Generally speaking, spatial decoding consists in combining ambisonic signals B_{mm}^σ to produce signals S_n radiated by loudspeakers, with the aim at creating phantom sound images accordingly to the encoded sound scene. More formally, this is done by a matrix operation:

$$\mathbf{s} = \mathbf{D}\mathbf{b}, \quad (2)$$

exhibiting decoding matrix \mathbf{D} and vectors $\mathbf{s} = [S_1 \dots S_N]$ and $\mathbf{b} = [B_{00}^{+1} B_{11}^{+1} B_{11}^{-1} \dots B_{MM}^{+1} B_{MM}^{-1}]$. We qualify as “basic” the decoding such that the ambisonic sound field that would be recorded in the centre of loudspeakers setup is the same as the originally encoded sound field [16]. In the case where loudspeakers form a regular polygon and with encoding convention of (1), decoding matrix is defined as:

$$\mathbf{D} = \frac{1}{N} \mathbf{C}^T, \quad (3)$$

where columns of matrix $\mathbf{C} = [\mathbf{c}_1 \dots \mathbf{c}_N]$ are vectors $\mathbf{c}_n = [1 \dots \sqrt{2} \cos(M\theta_n) \sqrt{2} \sin(M\theta_n)]^T$ of encoding gains associated to loudspeaker azimuth θ_n . For an M^{th} order encoding, at least $N=2M+2$ loudspeakers are recommended for a homogeneous reproduction.

Another way of understanding ambisonic spatialization is to merge spatial encoding and decoding, so that encoding directivities are themselves combined to form directivities of virtual microphones associated to

loudspeakers individually, as if these virtual microphones would directly capture the sound field and feed the loudspeakers (Figure 2).

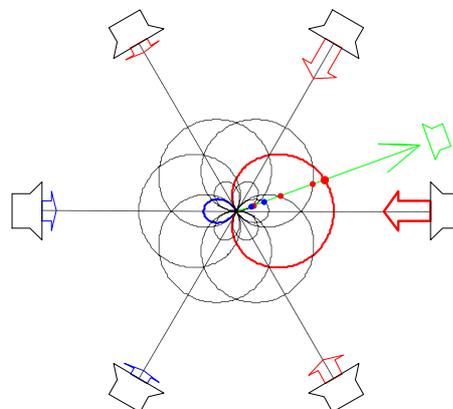


Figure 2 Virtual recording equivalent to 1st order encoding and basic decoding over a hexagonal layout, with a sub-cardioid pointing to each loudspeaker (only one is coloured for clarity). For each loudspeaker and for a given direction virtual source (in green), the panning gain is characterized by the length and colour of the large arrow, and also related to the intersection point between the directivity pattern and the green arrow.

Two observations arise from Figure 2. First, the main lobe of the directivity pattern is quite large regarding the angular spacing between loudspeakers. With a 2nd order encoding and decoding, one gets a thinner virtual directivity (2nd column of Figure 3, $M=2$) that helps making a more selective use of loudspeakers and creating more precise and robust sound images. With higher orders, even thinner directivities are involved instead, while requiring more loudspeakers in order to avoid “holes” between loudspeakers. The second comment is about the presence of non negligible secondary lobe(s). These are useful at low frequency for a proper reproduction of wave propagation properties (see also next section), but should be reduced regarding other criteria. This can be done by changing the proportion of the various elementary encoding patterns when combining them. In practice, this involves a modified decoding that consists in weighting ambisonic signals B_{mm}^σ by order-dependent gains g_m before applying the basic decoding matrix [16]. The so-called “max r_E ” decoding [16] aims at “concentrating” the energy contributions towards the expected direction (see also in next section). As shown in Figure 3, it reduces secondary lobes without substantially enlarging the

main one. It is recommended for medium and high frequencies. The so-called “in-phase” [17] (or “controlled-opposite” according to Richard Furse’s terminology) decoding completely removes secondary lobes, but enlarge quite significantly the main one. It is recommended for large listening area to limit the perception of loudspeakers in directions opposite to the expected one at off-centred listening positions.

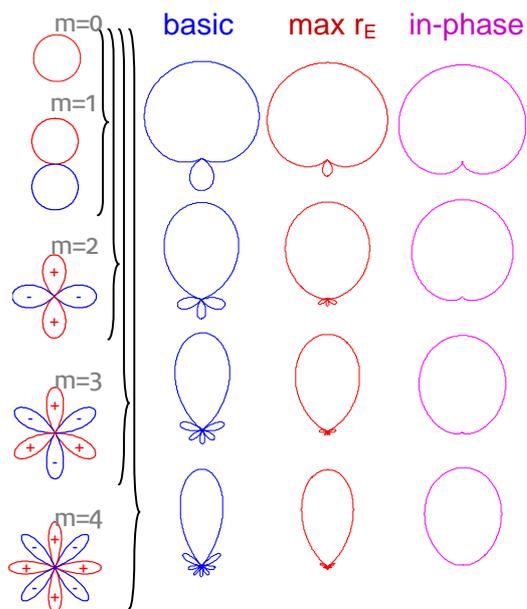


Figure 3 – Typical examples of recombination of encoding directivities (left column) of orders $m=0$ to 4, resulting in equivalent virtual microphone directivities, for three kinds of decoding. Getting from a virtual mic directivity to the next order’s one, the lobes of the additional encoding pattern help (from front to rear) reinforcing the main lobe, thinning it, and reducing secondary lobes (while creating two others).

2.1.3. Characterization of spatial rendering

As explained just above, the distribution of the energy of signals emitted by loudspeakers gets confined in a thinner angular sector around the virtual source direction as higher order components are involved. As a consequence, the acoustic interference occurs mainly between waves from loudspeakers closer to each other, and therefore it gets larger. Finally, the wave front synthesized by wave interference is reconstructed over an area that extends as the order increases (Figure 4).

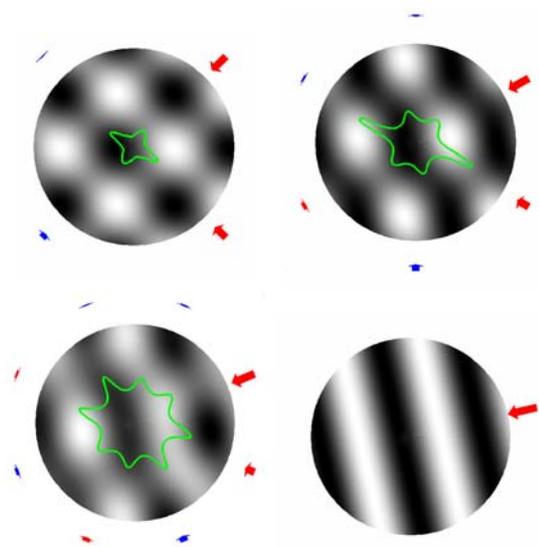


Figure 4 Progressive reconstruction of a wave front (right-bottom) with ambisonic encoding and decoding of orders $M=1$ to 3 (left to right, top to bottom), by interference of waves coming from resp. 2, 4, 6 loudspeakers (see conventions of Figure 2 for arrows). The boundary of well reconstructed area is shown as a constant-error contour.

So the area of well reconstructed sound field is bounded by the interference width. To get an idea of it, let’s consider the case where the number N of loudspeakers fits the best the angular selectivity, *i.e.* $N=2M+2$. For some incidences, interference mainly occurs between waves with angular gap $2\pi/N=2\pi/(2M+2)$, and therefore its width is:

$$\Lambda(f) = \frac{c/f}{2 \sin(\pi/(2M+2))} \quad (4)$$

Considering a centred listening area of radius R , and a “safety margin” R_m to avoid interference side effects, the sound field might be correctly rendered at the listener’s ears up to a frequency f_{lim} such that $\Lambda(f_{lim}) = D = 2(R+R_m)$. It has been numerically observed that the necessary safety margin decreases in inverse proportion to increasing order M , so that it can be approximately replaced by $R_m=R/M$:

$$f_{lim} \approx \frac{cM}{4R(M+1) \sin(\pi/(2M+2))} \approx \frac{cM}{2\pi R} \quad (5)$$

A useful example is the case of a single, centred listener. Fixing $R=8.8\text{cm}$ as a typical head radius, rounded values of f_{lim} are listed in Table 1 and used as reference in the analysis of microphone performances in section 6.

Above the frequency f_{lim} , a decoding that better “concentrates” the energetic contributions in the expected direction is recommended. That is the so-called “max r_E ” decoding [16], which maximizes the norm r_E of the energy vector \vec{E} :

$$\vec{E} = \frac{\sum_{n=1}^N G_n^2 \vec{u}_n}{\sum_{n=1}^N G_n^2}, \text{ with } \vec{u}_n = \begin{pmatrix} \cos \theta_n \\ \sin \theta_n \end{pmatrix} \quad (6)$$

where G_n is the signal gain associated to the loudspeaker n in the case of a single virtual source. In terms of perceived localisation, energy vector’s norm r_E ($r_E \leq 1$) can be interpreted as a reduction factor of lateralisation effect [3]. That might result in an elevation effect of angle $\alpha_E = \arccos(r_E)$ when the listener turns the head. More generally, one could roughly consider that angle α_E characterizes the blur width of the created phantom image since in a particular, but representative case, the latter is produced by two loudspeakers at $+\alpha_E$ and $-\alpha_E$ around the virtual source direction. Values given in Table 1 are also used as reference in section 6. More generally $\alpha_E = \pi/(2M+2)$ for a M^{th} “max r_E ” decoding.

Order M	1	2	3	4
f_{lim}	700 Hz	1300 Hz	1900 Hz	2500 Hz
α_E	45°	30°	22.5°	18°

Table 1 Limit frequencies f_{lim} of the acoustic reconstruction at a centred listener ears. Predicted angle α_E of the blur width of the phantom image.

2.2. Further acoustics basics for HOA

2.2.1. Fourier-Bessel series

HOA is grounded on spatial harmonic representation of 3D sound field, the so-called Fourier-Bessel series, which comes from the resolution of the homogeneous Helmholtz equation $(\Delta + k^2)p=0$ (with wave number $k=2\pi f/c$, frequency f and sound speed c) in a source-free region of space. In the spherical coordinate system shown in Figure 5 (radius r , azimuth θ , elevation δ), Fourier-Bessel series is defined as follow:

$$p(kr, \theta, \delta) = \sum_{m=0}^{\infty} i^m j_m(kr) \sum_{n=0}^m \sum_{\sigma=\pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \delta) \quad (7)$$

where

$$B_{mn}^{\sigma} = \frac{1}{i^m j_m(kr)} \iint_S p(r, \theta, \delta) Y_{mn}^{\sigma}(\theta, \delta) dS, \quad (8)$$

if $j_m(kr) \neq 0$, S is the unit sphere.

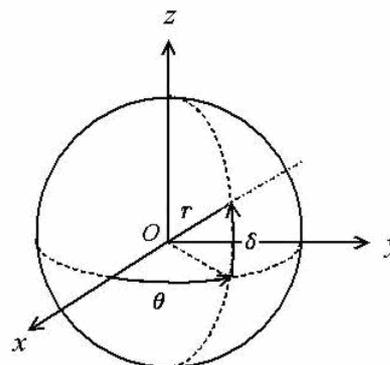


Figure 5 Spherical coordinate system in which any point of 3D Euclidean space is described by its radius r , its azimuth θ , and its elevation δ .

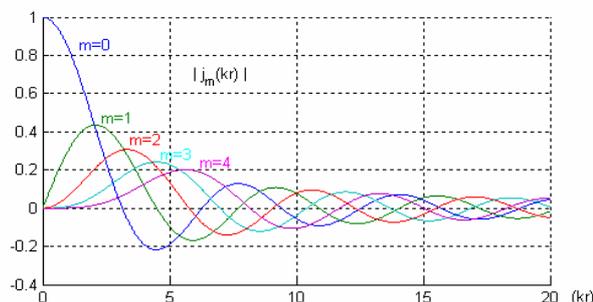


Figure 6 Illustration of spherical Bessel function up to order 4.

Radial¹ functions $j_m(kr)$ are spherical Bessel functions and are illustrated in Figure 6. Angular functions Y_{mn}^{σ} are the so-called spherical harmonics (Figure 7), with $m \geq 0$, $0 \leq n \leq m$ and $\sigma = \pm 1$. Various definitions of these functions exist and we use real spherical harmonics defined according to [18]:

¹ Note that the radial “distance” from origin actually depends on both radius and frequency.

$$Y_{mn}^\sigma(\theta, \delta) = \sqrt{(2m+1)(2-\delta_{0,n}) \frac{(m-n)!}{(m+n)!}} P_m(\sin \delta) \times \begin{cases} \cos n\theta & \text{if } \sigma = +1 \\ \sin n\theta & \text{if } \sigma = -1 \text{ (ignored if } n=0) \end{cases} \quad (9)$$

where $P_m(\sin \delta)$ are the associated Legendre functions.

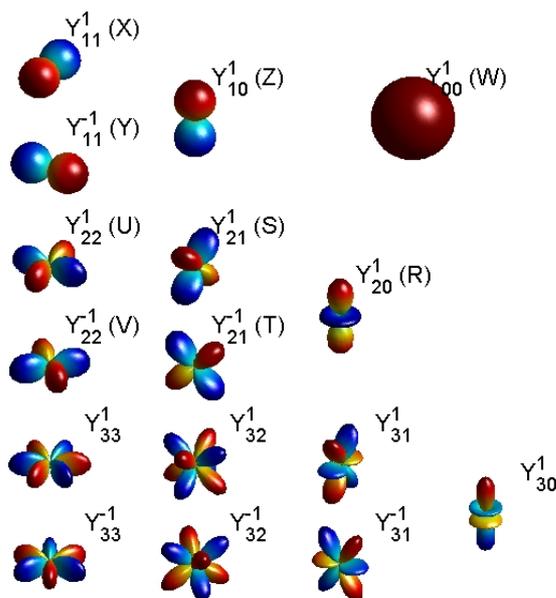


Figure 7 3D view of spherical harmonics up to order 3 with usual designation of associated HOA components.

The case of a sound field consisting of a plane wave is of particular interest. Considering a wave incidence (θ_s, δ_s) and a conveyed signal S leads to the following expression of ambisonic components:

$$B_{mn}^\sigma = S Y_{mn}^\sigma(\theta_s, \delta_s) \quad (10)$$

This equation provides new, 3D encoding laws that complete equations (1) or can be used instead. They yield $(M+1)^2$ components for an M^{th} order 3D representation. The spatial decoding process is very similar as in 2.1.2, but requires "periphonic" (3D) loudspeaker setups (e.g. according to polyhedral geometries). For further explanation, refer to [3].

2.2.2. HOA representation and sound field approximation

The M -order truncation of Fourier-Bessel series provides an approximate representation of the sound field:

$$p_M(kr, \theta, \delta) = \sum_{m=0}^M i^m j_m(kr) \sum_{n=0}^m \sum_{\sigma=\pm 1} B_{mn}^\sigma Y_{mn}^\sigma(\theta, \delta) \quad (11)$$

This approximation requires $K=(M+1)^2$ HOA signals. In order to objectively characterize the approximation accuracy, we use the Normalized Mean Square Error (NMSE) associated to a finite order of truncation M :

$$e(kr) = \frac{\iint_S |p(kr, \theta, \delta) - p_M(kr, \theta, \delta)|^2 dS}{\iint_S |p(kr, \theta, \delta)|^2 dS}, \quad (12)$$

where S is the unit sphere. For the particular case of a sound field only made by plane waves, Equation (12) can be simplified [19]:

$$e(kr) = 1 - \sum_{m=0}^M (2m+1) (j_m(kr))^2 \quad (13)$$

In this case, the error $e(kr)$ is independent of angular position of the source. Figure 8 illustrates the NMSE as a function of kr for truncation orders going from 0 to 15.

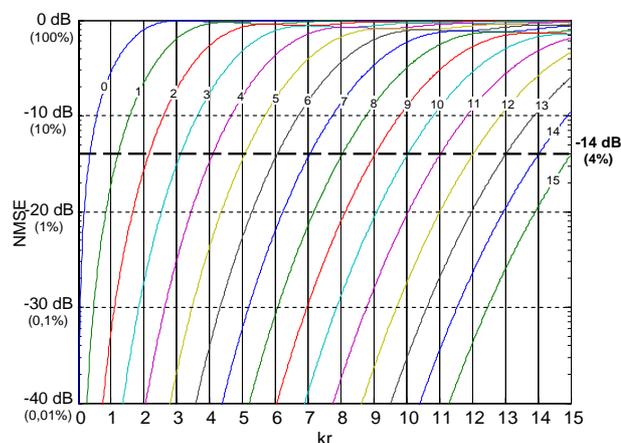


Figure 8 Normalized Mean Square Error for the plane wave case and order 0 to 15.

Clearly, it shows that for a constant value of error (e.g. 4% represented by the dotted line in Figure 8), i.e. for the same approximation accuracy, the higher kr , the higher the truncation order M should be. Moreover, the dotted line in Figure 8 corresponding to a 4% error cut

the curves such that $M = kr$ approximately. Thus, for a given wave number K (or a corresponding frequency) and a given radius R of the approximation sphere, a simple rule can be defined to determine the order that allow a reasonably accurate approximation of the sound field [19]:

$$M = \lceil KR \rceil, \quad (14)$$

where $\lceil \cdot \rceil$ denotes rounding up to the nearest integer. For example if we wish to represent a 1kHz plane-wave field within a sphere of radius 0.1m with a 4% error, this rule indicates a sufficient truncation order of value 2. But for a 5kHz plane-field within the same sphere of radius 0.1m and with 4% error, the truncation order must be at least equal to 9.

It's worth noticing that equation (11) fully complies with estimation of frequency limit given by (5). Such considerations can provide a criterion for the design of HOA microphone system and for the assessment of its performance in term of potential spatial rendering (see section 6.3).

3. THEORETICAL BASES OF SPHERICAL MICROPHONE FOR HOA RECORDING

Direct recording of HOA signals that describe an actual sound field using of directional microphones located at one single point is not feasible. An alternative method consists in deriving HOA signals from a set of acoustic measurements made by usual microphones and a corresponding mathematical model which relates these quantities to HOA signals. In this section, we first study continuous modeling of the problem of HOA signals estimation. Then, we transpose the continuous models into their discrete counterparts and discuss induced spatial aliasing.

3.1. Mathematical modeling of a continuous spherical HOA microphone

3.1.1. Spherical microphone derived from spherical Fourier transform

The spherical Fourier transform (8) suggests a theoretical method of estimating HOA signals from the recording of acoustic pressure over a continuous spherical surface of radius R . This method illustrated in Figure 9 consists in projecting the continuous spherical measurement onto the spherical harmonic basis and equalizing the resulting signals according to their order by means of filters $EQ_m(kR) = 1/i^m j_m(kR)$.

Nevertheless, this method is valid only if $j_m(kR) \neq 0$. Unfortunately, the Figure 6 shows that for all m orders, $j_m(kR)$ takes the zero value regularly over the kr range. Moreover, when $EQ_m(kR)$ is defined, it yields a huge signals amplification for kR around the zeros of $j_m(kR)$. Such amplification is impracticable.

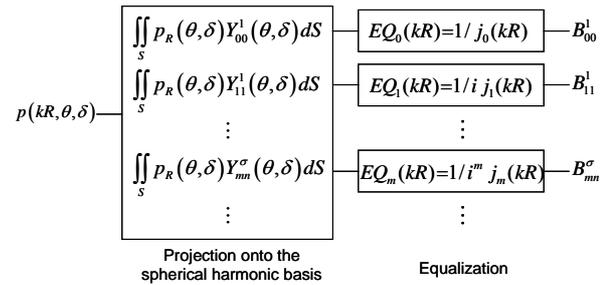


Figure 9 Schematic view of a theoretical spherical HOA microphone.

The huge amplification reflects difficulties encountered by the spherical microphone to estimate HOA signals from measures. These difficulties are partly explained by the lack of intrinsic directivity of the spherical microphone. Indeed, only phase differences exist between each measurement point in order to distinguish all spherical harmonics components. This creates indeterminate frequencies that are functions of the radius and the order, and correspond to the zeros of the spherical Bessel functions. For example, the zero order HOA signal B_0^1 in the Fourier Bessel series can't be estimated with a spherical sensor array at frequencies $f_a = ac/2R$, where a is a positive integer, c the celerity of sound, and R the radius of the sphere, that is at wave numbers $k = 2R$. A solution to avoid undetermined frequencies is to make the sensors measurements directional.

3.1.2. Improving the spherical microphone directivity

Usual directional sensors can be used to improve the microphone directivity. A 1st order directional sensor records a signal proportional to a combination of acoustic pressure and particle velocity:

$$s(kR, \theta, \delta) = \alpha p(kR, \theta, \delta) + (1 - \alpha) \rho c \frac{\partial p(kR, \theta, \delta)}{\partial kR} \quad (15)$$

where ρc characterizes acoustic impedance. Typically $0 \leq \alpha \leq 1$ in order to obtain the maximum value 1 at the

incidence ($\theta = 0^\circ, \delta = 0^\circ$). By introducing Equation (15) in the Fourier Bessel series (7), we obtain the spherical harmonic expansion of a 1st order directional measure:

$$s(kR, \theta, \delta) = \sum_{m=0}^{\infty} i^m (\alpha j_m(kR) - (1-\alpha)j'_m(kR)) \times \sum_{n=0}^m \sum_{\sigma=\pm 1} B_{mn}^\sigma Y_{mn}^\sigma(\theta, \delta) \quad (16)$$

where $j'_m(kR)$ is the first derivative of $j_m(kR)$ according to kR . Orthogonal projection of $s(kR, \theta, \delta)$ onto the spherical harmonic basis leads to the following expression of HOA signals:

$$B_{mn}^\sigma = \frac{1}{i^m (\alpha j_m(kR) - (1-\alpha)j'_m(kR))} \times \iint_S s(kR, \theta, \delta) Y_{mn}^\sigma dS \quad (17)$$

allowing us to deduce HOA signals in a similar way as shown in Figure 9, the equalization process being defined as follow:

$$EQ_m(kR) = \frac{1}{(\alpha j_m(kR) - (1-\alpha)j'_m(kR))}$$

With $\alpha j_m(kR) - (1-\alpha)j'_m(kR) \neq 0$. This radial expression is illustrated in Figure 10 for the particular case of cardioid sensors, i.e. for $\alpha = 0,5$. The figure shows that this function equals the zero value only for $kR = 0$ (and orders $m \geq 1$), and then seems to be more adapted to HOA signal recording over a large frequency band.

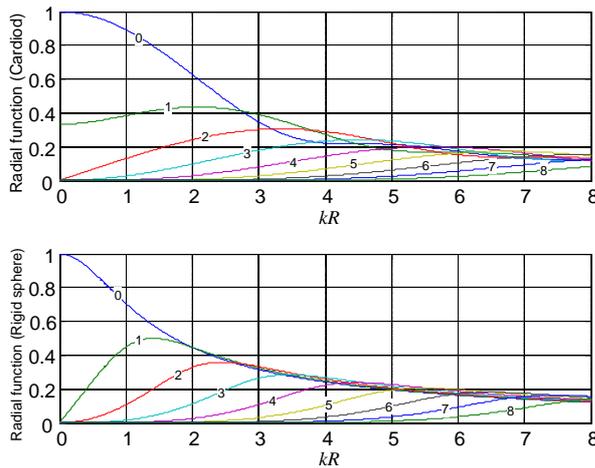


Figure 10 Radial functions for directional HOA microphone: cardioid sensors (top) and omnidirectional sensors at the surface of a rigid sphere (bottom).

A different approach for improving the directivity of continuous spherical microphone consists in introducing into the sound field a diffracting structure, e.g. a rigid sphere. Such a rigid sphere has already been used by Gary Elko [20] for the construction of a 1st order differential microphone with 3D adjustable orientation and more recently for spatial sound recording and beamforming [9, 21] in parallel to authors' work. In presence of a rigid sphere, the sound field is defined as a combination of direct and diffracted sound pressure [18]:

$$p_{\text{rig}}(kr, \theta, \delta) = \sum_{m=0}^{+\infty} \left(j_m(kr) - \frac{j'_m(kr)}{h_m^-(kr)} h_m^-(kr) \right) \times \sum_{n=0}^m \sum_{\sigma=\pm 1} B_{mn}^\sigma Y_{mn}^\sigma(\theta, \delta) \quad (18)$$

where R is the radius of the sphere, $h_m^-(kr)$ the divergent Hankel function, and $h_m^-(kr)$ its 1st derivative according to kR . A particularly interesting case is the recording of acoustic pressure at the surface of the rigid sphere ($r=R$):

$$p_{\text{rig}}(kR, \theta, \delta) = \sum_{m=0}^{+\infty} \left(\frac{i^{m-1}}{(kR)^2 h_m^-(kR)} \right) \times \sum_{n=0}^m \sum_{\sigma=\pm 1} B_{mn}^\sigma Y_{mn}^\sigma(\theta, \delta) \quad (19)$$

The radial dependence between brackets is shown in Figure 10. As for the cardioid sensor case, this radial function equals zero only for $kr = 0$ (and $m > 0$). Orthogonal projection of Equation (19) onto the spherical harmonic basis leads to the following expression of HOA signals:

$$B_{mn}^\sigma = \left(i^{-m+1} (kR)^2 h_m^-(kR) \right) \times \iint_S p_{\text{rig}}(kR, \theta, \delta) Y_{mn}^\sigma(\theta, \delta) dS \quad (20)$$

which is defined for $i^{-m+1} (kR)^2 h_m^-(kR) \neq 0$, i.e. for $m > 0$ and $kR > 0$ as we can see in (Figure 4). HOA signal can therefore be estimated according to the method describe in Figure 9 by defining equalisation filters as follow: $EQ_m(kR) = i^{-m+1} (kR)^2 h_m^-(kR)$.

In conclusion, directional spherical microphone seem to be more adapted than omnidirectional one to HOA recording since equalisation filters are defined for all non-null frequencies. Nevertheless, we can see in Figure 10 that an excessive amplification remains below a limit value of kR that increases with order m . Indeed,

especially for higher orders and low frequencies, the microphone tries to catch spatial information that is very poor at the measurement point and is substantial only at a distance from the recording system. Consequently, the microphone will be in this case very sensitive to estimation errors (*e.g.* sensor self-noise, position errors).

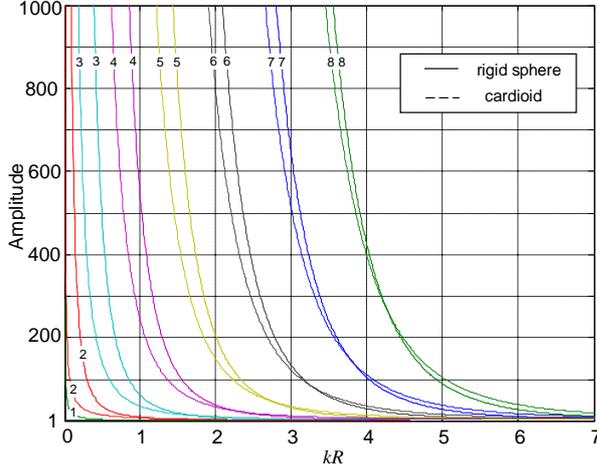


Figure 11 Inverse of radial functions $|EQ_m(kR)|$ for cardioid and rigid sphere cases.

3.2. Discretization of the continuous spherical microphone

3.2.1. Discrete formulation of the estimation problem

In order to realize a spherical HOA microphone, we must use a finite number of sensors. On a spherical surface of radius R , we define a set of Q points at locations given by the angles (θ_q, δ_q) , $1 \leq q \leq Q$. We only take into account the cases of spherical directional microphone and omnidirectional sensors at the surface of a rigid sphere. In these two cases, the q^{th} sensor picks up an acoustic signal which can be expressed as follow:

$$S_q = \sum_{m=0}^{+\infty} W_m(kR) \sum_{n=0}^m \sum_{\sigma=\pm 1} B_{nm}^\sigma Y_{nm}^\sigma(\theta_q, \delta_q) \quad (21)$$

where

$$W_m(kR) = \begin{cases} \alpha j_m(kR) + i(\alpha - 1)j_m'(kR) & \text{directional microphones} \\ i^{-m+1}(kR)^2 h_m'(kR) & \text{rigid sphere} \end{cases} \quad (22)$$

Actually, we can estimate HOA signals up to a restricted order M such that the total number $K = (M+1)^2$ of components does not exceed the number Q of sensors ($K \leq Q$). Thus, we consider the M order truncation of Equation (21). The resulting system of linear equations can be written in discrete matrix form:

$$\mathbf{T} \cdot \mathbf{b} = \mathbf{s}, \quad (23)$$

where the $Q \times K$ "transfer matrix" \mathbf{T} , the K -length column vector \mathbf{b} and the Q -length column vector \mathbf{s} are defined as follow, respectively:

$$\begin{aligned} \mathbf{T} &= \begin{pmatrix} Y_{00}^1(\theta_1, \delta_1) & \cdots & Y_{M0}^1(\theta_1, \delta_1) \\ \vdots & & \vdots \\ Y_{00}^1(\theta_Q, \delta_Q) & \cdots & Y_{M0}^1(\theta_Q, \delta_Q) \end{pmatrix} \cdot \text{diag}[W_m(kR)] \\ &= \mathbf{Y} \cdot \text{diag}[W_m(kR)], \quad 0 \leq m \leq M \end{aligned}$$

$$\mathbf{s} = (S_1, S_2, \dots, S_Q)^t, \quad \mathbf{b} = (B_{00}^1, B_{11}^1, B_{11}^{-1}, B_{10}^1, \dots, B_{M0}^1)^t.$$

Due to the spherical geometry of the microphone array, the matrix \mathbf{T} is a product of a real matrix \mathbf{Y} whose columns are sampled spherical harmonics \mathbf{y}_{mm}^σ by a diagonal matrix of radial-dependent filters. The vector \mathbf{s} contains the signals recorded by the Q sensors and \mathbf{b} is the vector whose elements are the unknown HOA signals up to a finite order M . These two last vectors are connected by a discrete mathematical model, the matrix \mathbf{T} .

3.2.2. "Naive" least-squares resolution and instability of the solution

Since the matrix \mathbf{T} is a $Q \times K$ rectangular matrix with more rows than columns ($Q \geq K$), the system (23) is overdetermined. In general, such a linear system cannot be solved exactly and the problem amounts to find an approximate solution which minimizes the square norm of the residual [22]:

$$\min_{\mathbf{b}} \|\mathbf{s} - \mathbf{T}\mathbf{b}\|_2^2.$$

The solution \mathbf{b}_{LS} to this minimization problem satisfies a new system of linear equations called normal equations [22]:

$$(\mathbf{T}^* \mathbf{T}) \mathbf{b}_{LS} = \mathbf{T}^* \mathbf{s}. \quad (24)$$

where \mathbf{T}^* denotes the transpose conjugate of the complex matrix \mathbf{T} . As the columns of the matrix \mathbf{T} are linearly independent, the definite positive matrix $\mathbf{T}^* \mathbf{T}$ is invertible and the solution of the system (24) can be expressed as follow [22]:

$$\mathbf{b}_{LS} = (\mathbf{T}^* \mathbf{T})^{-1} \mathbf{T}^* \mathbf{s}. \quad (25)$$

\mathbf{b}_{LS} is the least-squares solution of the system (23), i.e. the solution which corresponds to projecting \mathbf{b} onto the columns space of \mathbf{T} . By introducing the factorized expression of the matrix \mathbf{T} in Equation (25), we finally obtain:

$$\mathbf{b}_{LS} = \text{diag} \left[\frac{1}{W_m(kR)} \right] \mathbf{E} \mathbf{s}, \quad (26)$$

where

$$\mathbf{E} = (\mathbf{Y}^t \mathbf{Y})^{-1} \mathbf{Y}^t.$$

The real matrix \mathbf{E} is usually called the Moore-Penrose inverse of the matrix \mathbf{Y} [22].

The “naïve” least-squares resolution of the system (23) provides an easy way to compute its best approximate solution \mathbf{b}_{LS} in the least-squares sense. Nevertheless, even if the residual is theoretically minimized, the solution becomes meaningless when $W_m(kR)$ is very small, i.e. corresponding to low frequencies and highest estimate order. Indeed, the system (24) is in this case particularly instable so that small errors in the measure vector \mathbf{s} , e.g. due to sensor self-noise or sensor position errors, are magnified in the resulting estimation of HOA signals \mathbf{b}_{LS} . A useful objective indicator of non-stability of the system (24) is the condition number of the matrix $(\mathbf{T}^* \mathbf{T})$ defined by [22]:

$$\kappa(\mathbf{T}^* \mathbf{T}) = \|\mathbf{T}^* \mathbf{T}\|_2 \left\| (\mathbf{T}^* \mathbf{T})^{-1} \right\|_2. \quad (27)$$

The ideal condition number is worth 1 and means that all columns in the matrix are linearly independent and of balanced norm. When the level difference between columns vectors become too large, $\kappa(\mathbf{T}^* \mathbf{T})$ increases. The matrix $\mathbf{T}^* \mathbf{T}$ is said ill-conditioned if $\kappa(\mathbf{T}^* \mathbf{T}) \gg 1$. The resulting relative error of the least squares solution can be as large as the product of the relative residual by the condition number [22]:

$$\frac{\|\mathbf{b} - \mathbf{b}_{LS}\|_2}{\|\mathbf{b}_{LS}\|_2} \leq \kappa(\mathbf{T}^* \mathbf{T}) \frac{\|\mathbf{T}^* \mathbf{s} - \mathbf{T}^* \mathbf{T} \mathbf{b}_{LS}\|_2}{\|\mathbf{T}^* \mathbf{s}\|_2}. \quad (28)$$

Consequently, a poorly conditioned matrix \mathbf{T} can lead to a physically meaningless solution. Hence, this solution must be replaced by an approximated one where the noisy components are filtered out (cf. Section 4.1).

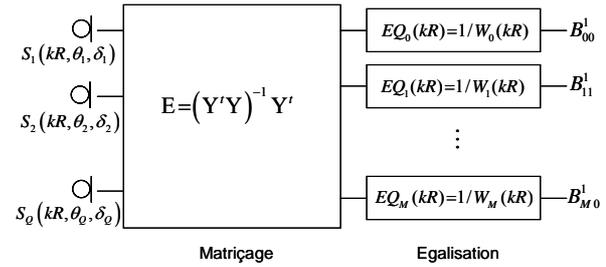


Figure 12 Discrete spherical HOA microphone processing.

3.2.3. Limitations introduced by discretization: spatial aliasing

The discrete formulation of the problem of HOA recording introduces some limitations in the estimation process. These limitations concern first the frequency band of HOA signals: according to the so-called Shannon criterion, the larger distance d between sensors defines a limit frequency above which spatial aliasing occurs:

$$f_{al} = \frac{c}{2d} = \frac{c}{2R\gamma}, \quad (29)$$

where c is the sound speed, R is the radius of the microphone array, and γ is the maximum angle between two sensors.

Furthermore, spatial aliasing also concerns undersampled spherical harmonic functions. The number of distinguishable harmonics by the microphone array depends on both the number of sensors and their positions on the sphere. For example, in order to estimate K different spatial harmonics up to a maximal order M , at least K sensors are needed to sample them correctly. Moreover, the sensor positions must satisfy the discrete orthonormality properties for sampled spherical harmonics:

$$\frac{1}{Q} \mathbf{y}_{mm'}^{\sigma \prime t} \cdot \mathbf{y}_{m'm''}^{\sigma'} = \delta_{mm'} \delta_{m''} \delta_{\sigma \sigma'}, \quad (30)$$

where $0 \leq m, m' \leq M$, $0 \leq n, n' \leq m$, $\sigma, \sigma' = \pm 1$, and the Kronecker delta δ_{ij} equals 1 if $i=j$ and 0 otherwise. Finally, spatial aliasing due to higher order harmonics must be taken into account. Indeed, when undersampled spherical harmonics of higher order than M are present in the sound field, they fold up on the lower components and can severely corrupt the estimation of HOA signals. To suitably minimize spatial aliasing effects, the sensors placed on the sphere must be in general more numerous than the HOA signals that we seek to estimate, i.e. $Q > K$ in the system (23).

4. PRACTICAL DESIGN OF SPHERICAL HOA MICROPHONE

This section aims at deriving a meaningful solution to HOA signal estimation by choosing a suitable sampling scheme on the sphere and introducing a regularization process based on Tikhonov filters [23].

4.1. Practical equalization involving regularization filters

4.1.1. Filtering approach of regularization

The regularization process aims at damping the noisy frequency components of HOA signals resulting from the least-squares resolution of the system (24). As explained above, this filtering concerns essentially low frequency bands of HOA signals of highest orders. As we assume that the sampled spherical harmonic matrix \mathbf{Y} is near to be orthonormal, the regularization can be realized by introducing a filter in the equalization:

$$\mathbf{b}_{REG} = \text{diag} \left[F_m(kR) \cdot \frac{1}{W_m(kR)} \right] \mathbf{E} \mathbf{s} \quad (31)$$

where $F_m(kR)$ is the regularization filter, and the vector \mathbf{b}_{REG} contain the regularized HOA signals. An optimal regularization filter $F_m(kR)$ suppresses noise in the data vector \mathbf{s} while keeping at the same time all relevant information. Various definition of the filter coefficients can be obtained by imposing additionally constraints on the least-square problem (24). The most commonly used method of regularization is probably the Tikhonov method [23]. Standard form of Tikhonov regularization imposes a maximal energy for the solution, i.e. resulting HOA signals. This method can be interpreted by defining the filters as:

$$F_m(kR) = \frac{|W_m(kR)|^2}{|W_m(kR)|^2 + \lambda^2}. \quad (32)$$

λ is the regularization parameter. For λ equals to 0, $F_m(kR) = 1$ and \mathbf{b}_{REG} simply correspond to HOA signals obtained by the direct least squares resolution of the system. As λ increases, $F_m(kR)$ decay to zero, implying that an λ too large will result in over regularized solution. Note that filters $F_m(kR)$ are proportional to $|W_m(kR)|^2$ and therefore they decay fast enough to compensate for the increasing factor $1/W_m(kR)$. Figure 13 shows $F_m(kR)$ as a function of $|W_m(kR)|$. Note that Tikhonov filters (cont. line) can be interpreted as a 1st order approximation of an exponential attenuation (dashed line). Equalization filters are finally defined by:

$$EQ_m(kR) = \frac{\bar{W}_m(kR)}{|W_m(kR)|^2 + \lambda^2}, \quad (33)$$

where $\bar{W}_m(kR)$ is the complex conjugate of $W_m(kR)$. We can recognise the standard expression of Tikhonov regularization used for instable filters [24].

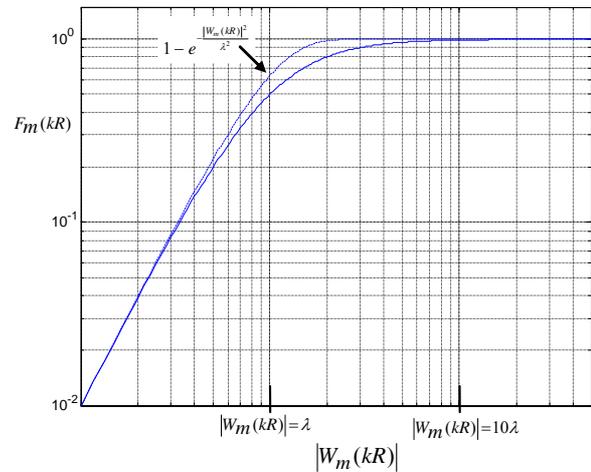


Figure 13 Illustration of the regularization filter $F_m(kR)$ as a function of $|W_m(kR)|$.

4.1.2. Choice of the regularization parameter

A suitable choice of regularization parameter λ is not trivial. Indeed, too large values of λ yield over-filtered solution. On the other hand, a too small parameter results in not enough filtered one. An ideal parameter must avoid noise in the solution while preserving at the

same time all of its relevant information. The regularization parameter λ can be related to a unique maximal amplification factor a induced by the filters $F_m(kR)$:

$$\lambda = \frac{1 - \sqrt{1 - 1/a^2}}{1 + \sqrt{1 - 1/a^2}}. \quad (34)$$

The choice of the maximal amplification factor a can advantageously take into account improvement in the signal to noise ratio due to the use of multiple sensors. In this case, a is directly connected to sensor noise amplification by:

$$a = \sqrt{Q} 10^{\frac{a_s}{20}}.$$

where a_s is the maximum sensor noise amplification expressed in dB, and Q is the total number of sensors used. A large number of sensors improves the signal to noise ratio of the spherical microphone array and yield to a larger amplification factor a (for a constant maximal noise amplification a_s), and so an accurate estimation of HOA signals. In general, the regularization parameter λ can be set by trial and error method.

4.2. Spatial sampling on the sphere

4.2.1. Equiangular sampling

In a way similar to the standard sampling theorem in one dimensional space, Driscoll and Healy [25] stated that a band limited function defined on a sphere, *i.e.* a function for which $B_{mn}^\sigma = 0$ if $m > M$ can be exactly recovered from a finite sum of samples of this function. This requires to sample uniformly and independently the angles θ and δ at $2M + 2$ locations:

$$\begin{aligned} \theta_i &= \frac{\pi i}{2M + 2}, & i &= 0, \dots, 2M + 1 \\ \delta_j &= \frac{2\pi j}{2M + 2}, & j &= 0, \dots, 2M + 1 \end{aligned} \quad (35)$$

Exact formulation of HOA signals up to order M is then expressed as follow:

$$B_{mn}^\sigma = \frac{1}{W_m(kR)} \frac{\sqrt{2\pi}}{2M + 2} \sum_{i=0}^{2M+1} \sum_{j=0}^{2M+1} \alpha_i s_R(\theta_i, \delta_j) Y_{mn}^\sigma(\theta_i, \delta_j) \quad (36)$$

where the coefficients α_i given in [25] have been introduced to compensate for the difference between the density of points near the poles and that in the equatorial

zone. Equation (36) requires $4(M+1)^2$ sensors arranged on the sphere to compute HOA signals up to M order.

It is however possible to reduce by half this number by using a Gaussian quadrature [26]. δ is then sampled at unequally spaced angles according to the $M+1$ zeros of Legendre polynomials, so that $P_{M+1}(\sin \delta_j) = 0$, $j = 0, \dots, M$. Equation (36) becomes:

$$B_{mn}^\sigma = \frac{1}{W_m(kR)} \sum_{i=0}^{M+1} \sum_{j=0}^{2M+1} \alpha_j s_R(\theta_i, \delta_j) Y_{mn}^\sigma(\theta_i, \delta_j) \quad (37)$$

where the coefficients α_j are the associated weights to Gauss sampling [26]. Nevertheless, the total number $2(M+1)^2$ of sensors required remains very high and not optimal. In order to reduce this number, we need to find more regular sampling schemes.

4.2.2. Regular and semi-regular polyhedral sampling

Unfortunately, it is not possible to uniformly sample the surface of a sphere, except in five particular cases, according to the vertices of Plato's polyhedrons [27]: tetrahedron, hexahedron, octahedron, dodecahedron and icosahedron. As these regular point distributions are not defined directly according to spherical harmonics, we need to examine the orthonormality properties of induced sampled spherical harmonics. This can be done by considering the orthonormality error matrix defined by:

$$\mathbf{D} = \mathbf{I}_K - \frac{1}{Q} \mathbf{Y}^t \cdot \mathbf{Y}, \quad (38)$$

where \mathbf{I}_{K^2} is the $K \times K$ identity matrix with $K = (M+1)^2$, Q is the number of sampling points, and the columns of the matrix \mathbf{Y} are sampled spherical harmonics. Figure 14 shows the matrix \mathbf{D} associated to all five regular polyhedrons. Orthonormality error between two sampled spherical harmonics is represented by a small gray square. For more clarity, harmonics of different orders are separated by lines. We can see that the tetrahedron, the hexahedron, and the dodecahedron yield arrangements which satisfy exactly the orthonormality condition up to order 1. Note that the so-called SoundField microphone [5] used for Ambisonics recording (1st order HOA) is based on a tetrahedral sampling scheme. The icosahedron and the dodecahedron vertices preserve the orthonormality properties of sampled spherical harmonics up to order 2.

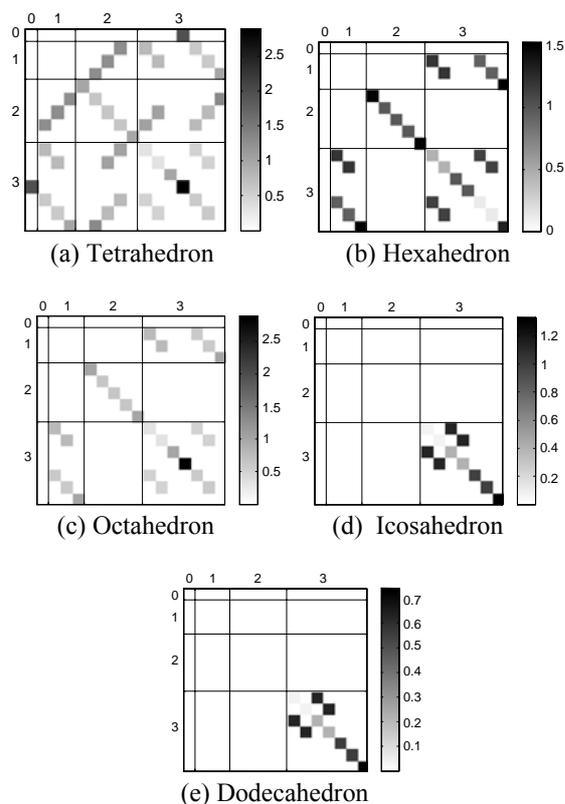


Figure 14 Matrices of orthonormality error for the five regular polyhedrons. Gray level is related to error level.

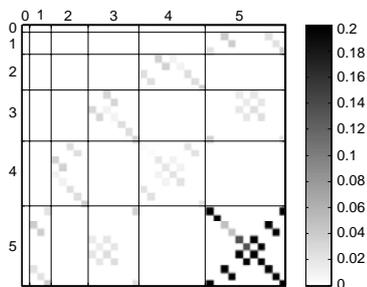


Figure 15 Matrix of orthonormality error associated to the pentakis-dodecahedron. Gray level is related to error level.

In order to record higher order HOA signals, sampling schemes based on semi-regular polyhedrons, *i.e.* convex polyhedrons for which faces are regular polygons but at least of two different types, can be investigated. We have constructed a 4th order spherical microphone based on the vertices of a pentakis-dodecahedron which

consists of a combination of an icosahedron and a Dodecahedron (cf. Figure 16). The matrix of orthonormality error of such a polyhedron is represented in Figure 15 and will be discussed in Section 5.

4.2.3. Quasi-uniform sampling

Now the question is: how to arrange in a homogeneous way a set of points at the surface of a sphere? The answer depends on the mathematical definition of “homogeneous” that we choose. Indeed, different existing interpretations lead to slightly different arrangements. For example, a first approach consists in finding the centres of non-overlapping identical circles such that their radius is maximized (packing problem). Another very similar method aims at defining the centres of overlapping identical circles which entirely cover the surface of the sphere such that their radius is minimized (covering problem). A last example to define point positions on a sphere derives from solving the problem of finding the lowest energy configuration of point charges on a conducting sphere (Thomson problem). Results of all these methods are available on Internet for various numbers of points, e.g. [28, 29]. Although these methods yield a relative equidistance between sensors, they are not directly linked to spherical harmonics and don’t guarantee the orthonormality condition for sampled spherical harmonics. The properties must be analyzed in terms of the orthonormality error matrix \mathbf{D} define by Equation (38).

5. CONSTRUCTION OF A 4TH ORDER PROTOTYPE

5.1. Geometrical description and associated characteristics

5.1.1. The choice of sensor positions

The distribution of sensors on the rigid sphere is based on the 32 vertices of a pentakis-dodecahedron shown in Figure 16. Such a polyhedron can be seen as the association of a dodecahedron (12 vertices) and an icosahedron (20 vertices). This sampling scheme is not exactly uniform since angles between sensors are not equal. Indeed, angular gap between two vertices of the dodecahedron and between one vertex of the dodecahedron and one of the icosahedron are worth 0.73 and 0.65 radians, respectively. Nevertheless, as we can see in Figure 15 the discrete spherical harmonics associated to this sampling scheme is very close to be

orthonormal up to order 4 although the maximum diagonal term and the maximum off-diagonal term correspond to an error of 3.7% and 3.73%, respectively. Also, we can see in Figure 15 that spatial aliasing could potentially occur in the 1st and 3rd orders if the 5th order is strongly present in the sound field near the sensors.

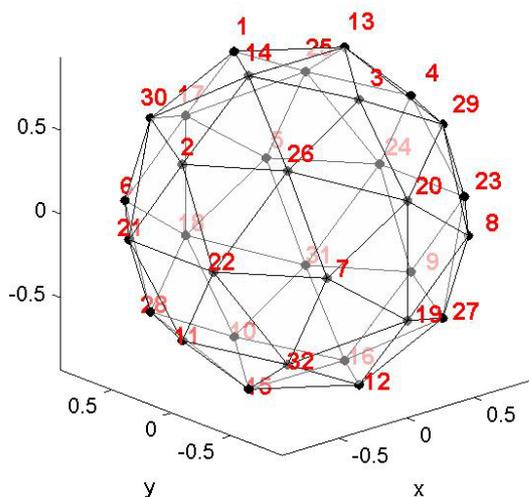


Figure 16 Spherical distribution of sensors according to the 32 vertices of a pentaki-dodecahedron.

5.1.2. The choice of the radius

The choice of the radius results from a compromise between spatial aliasing at high frequencies and consistence of estimation at low frequencies for highest order. The radius of our prototype is worth $R=3,5\text{cm}$ which corresponds to approximates aliasing frequencies of 6700Hz and 7500Hz for the largest angular gap and the smallest one, respectively. The size of the spherical microphone array also determines the “presence level” of HOA signals according to order by means of the functions $W_m(kR)$. Since the sampled spherical harmonic matrix \mathbf{Y} is approximately orthonormal, the condition number of the matrix $\mathbf{T}^*\mathbf{T}$ in Equation (24), *i.e.* the stability of the estimation process, depends only on the functions $W_m(kR)$ (and so the radius R):

$$\kappa(\mathbf{T}^*\mathbf{T}) = \frac{\max(|W_m(kR)|^2, m=0\dots M)}{\min(|W_m(kR)|^2, m=0\dots M)} \quad (39)$$

Figure 17 illustrates the modulus of the functions $W_m(kR)$ for orders $m=0$ to 4 in parallel to the condition number of the matrix $\mathbf{T}^*\mathbf{T}$. We can easily see that $\kappa(\mathbf{T}^*\mathbf{T})$ rapidly increases while the difference between $W_m(kR)$ of distinct orders increases. Instability occurs in highest orders and at low frequencies.

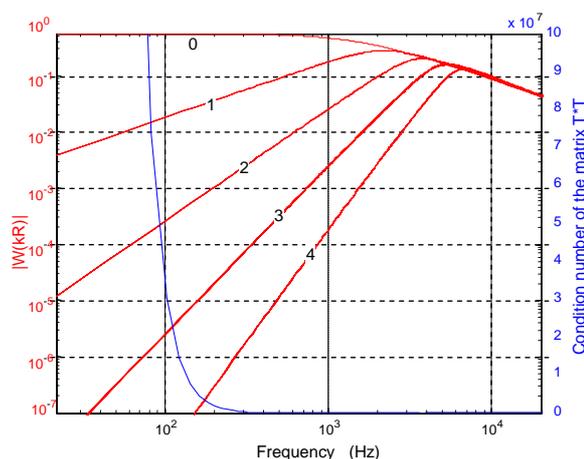


Figure 17 Functions $|W_m(kR)|$ and the condition number of the transfer matrix $\mathbf{T}^*\mathbf{T}$ as a function of frequency ($r=3,5\text{cm}$).

5.2. Realization and implementation of the prototype

The 4th order HOA microphone prototype illustrated in Figure 18 is made of an empty plastic ball that is acoustically non-strictly rigid. Holes are distributed over the sphere to incorporate a regular configuration of HOA microphone network. Small placement errors may occur and could be corrected with post signal processing. The sphere is cut into two parts to insert the 32 sensors from inside. The sensors wires are pool together to a plastic sheath fixed to the ball. Sensors close to the sheath will have directivity errors compared to theoretical model discussed in 6.3.



Figure 18 32 sensors compose the 4th order HOA microphone prototype.

Figure 19 shows regularization filters associated to the prototype for different maximal sensor noise amplification: 0dB, 6dB and 20dB.

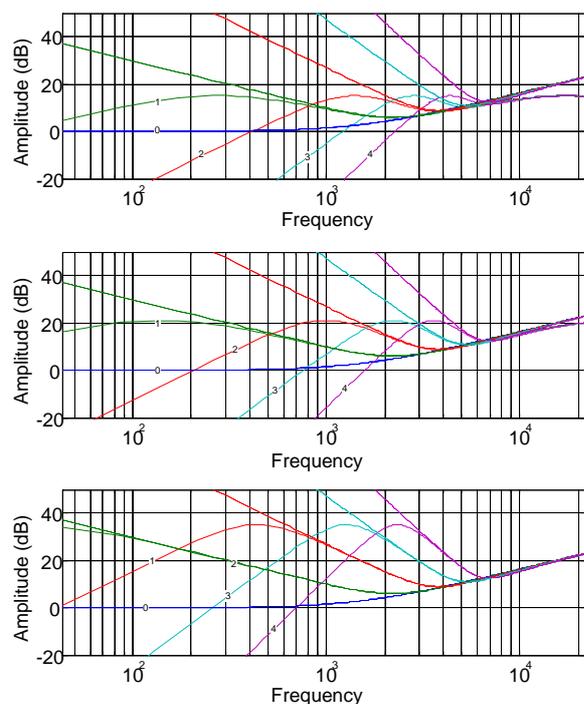


Figure 19 Amplitude of theoretical and regularized equalization filters $EQ_m(kR)$ (dashed lines and cont. lines resp.) with 3 different regularization parameters corresponding to 3 max. *noise amplification* levels: 0dB, 6dB, and 20dB. (Note that maximal gains of equalization filters go +15dB beyond these limit values, accordingly to the number of sensors used.)

Acquisition system is composed of a 32 channel home-made preamplifier, two MOTU 24I/O AD converters plugged to a MOTU PCI424 card in a computer. The software Plogue Bidule is used in combination with VST plug-ins to make real-time recording and rendering over a predefined configuration of loudspeakers or over headphones.

6. OBJECTIVE MEASUREMENTS AND VALIDATION

6.1. Measurements

6.1.1. Free field measurements

The 4th order, 32 sensor prototype microphone has been measured in IRCAM anechoic chamber. It has been mounted on a Brüel and Kjær turn table situated in the middle of the room allowing azimuth rotation. Two laser lights installed above the system and at 0° elevation respectively indicate the centre position of the microphone. A Tannoy loudspeaker mounted on a moving arm allowing elevation measurements was used. The setup is shown Figure 20.

The spherical microphone array has been measured every 5 degrees from 0 to 360 degrees azimuth and from -40 to 90 degrees elevation, totalizing $N=1971$ measures for each microphone. The correlation between the 0 degree and the 360 degrees azimuth sensors response has been calculated verifying the measures repeatability.

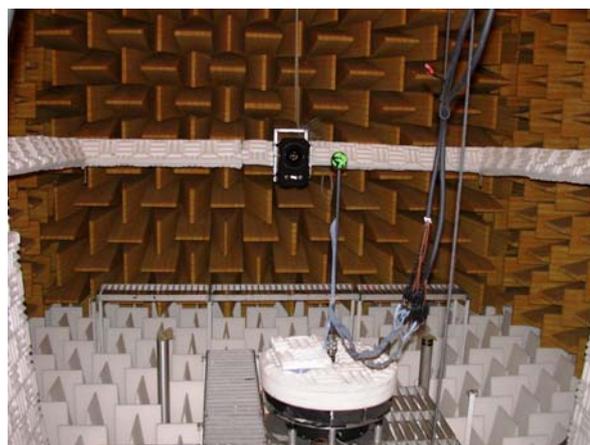


Figure 20 Free field measurements setup for the 32 sensors microphone in IRCAM anechoic chamber.

A Studer 24 channel preamplifier has been used for the acquisition measurements. Therefore measurement procedure has been performed twice to get the responses of 24 sensors at once. 12 sensors have been measured twice allowing us to verify the measures.

We used logarithm sine sweep method with a 16th order sequence. We placed the signal level in order to have at least 30dB dynamic range to the measured impulse response.

During the measurements it was not sure that the centre of the sphere microphone strictly remains on the turntable rotation axis, nor on the loudspeaker rotation axis. To estimate the subsequent time shift in measured responses and compensate for it, we calculated for each incidence the barycentre of the arrival time of all sensors responses. The maximal time shift we found corresponds to a distance of 7mm, which has negligible impact in terms of angle error but had to be taken into account for phase correction of responses.

The loudspeaker response has been measured with a 1/4 inch B&K microphone. The sensors measures were deconvolved by the loudspeaker response in order to remove its influence.

6.1.2. Diffuse field measurements

The directional sphere was not completely sampled during the free field measurements: -90 to -40 degrees elevation measurements were left apart due to system limitation. In order to evaluate the individual sensor response discrepancy complementary diffuse field measurements have been done in IRCAM "Espace de Projection" room.

Sensors have been measured with the same equipment than free field measurements. The logarithm sine sweep sequence was longer (18th order) to have a diffuse field in the room.

We used the Matlab "Room Analysis Toolbox" (RAT) developed by IRCAM and based on Early Decay Curve analyses. Diffuse field sensors responses were estimated on the basis on the initial power spectrum response, which is estimated by linear regression of the retrograde integration of the energy decay curve for each frequency band (short time Fourier transform on each band).

6.2. Comparison of sensors responses with the sphere diffraction model

The prototypes we built are hoped to be conform enough to the theoretical model of sphere microphone of 5 for which we've been able to define an appropriate processing. Nevertheless, such real prototypes potentially present various questionable features: sphericity, acoustic rigidity, acoustic disturbance (by the bottom sheath), radius, positioning errors, individual responses discrepancy. In the following, we present some first methods to identify and characterize deviations from the model.

For this purpose, sensors' responses have been simulated for the same wave incidences as used in measurements, accordingly to the prototype description (radius, sensors positions) of section 5. Simulations were performed from (19). Sound speed has been fixed to $c=340\text{m/s}$. These simulations are used as a reference basis for comparison with the measured directivity. This helps to analyze how well the measures match the model, what are the non modelled disturbances, and to detect possible positioning errors.

From this point we still use the notation introduced in 3 to describe "sphere related" *transfer functions* (instead of *signals*) with the assumption of sound fields consisting in a single plane wave. For each sensor q , we get a set of N transfer functions $s_n^q(\theta_n, \delta_n)$ associated to the N directions (θ_n, δ_n) of measurements. These compose the vector \mathbf{s}_q (which might be merely noted \mathbf{s}). It is implicitly understood that such vectors and their elements are functions of the frequency f . We'll denote respectively by $\mathbf{s}_q^{\text{meas}}$ and $\mathbf{s}_q^{\text{model}}$ the data derived from measurement and simulation.

6.2.1. Qualitative comparison of directivities

Figure 21 shows 3D plots of sensors directivity for one sensor in the upper hemisphere (#4) and for one sensor near the sheath, below the sphere (#16). In the first case, the global directivity shape looks the same for both measured and simulated responses, even in relatively high frequency although a slight change can be observed at low elevation angles. For sensor #16, the shape is roughly respected at relatively low frequency, but the plot at 8kHz makes clearly appear the disturbance of the measured directivity caused by the sheath underneath the sphere (loss of axial symmetry compared with the model).

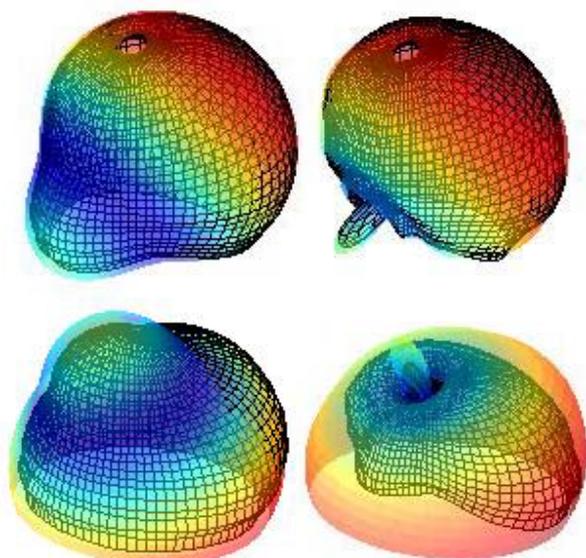


Figure 21 3D plots of measured (meshed plot) and simulated (in transparency) directivity for sensors #4 (top) and #16 (bottom) and at frequencies $f=1\text{kHz}$ (left) and $f=8\text{kHz}$ (right), with respect to sensor and axis positioning of Figure 18.

Figure 22 shows polar patterns at elevation= 0° and for sensor #5, which is located in the horizontal plane and shouldn't be too much affected by acoustic disturbance of the bottom sheath. The global shapes of measured and simulated patterns fit rather well together. We notice that with the model, secondary lobes are always a bit greater than with the measure, while the troughs mostly coincide². That means that the real prototype is slightly more directive than the model. At this point, it is hard to state whether the difference is mostly due to sheath disturbance or non strict acoustic rigidity of the sphere.

² We tried simulations with several values of sound speed c without getting a significantly better matching.

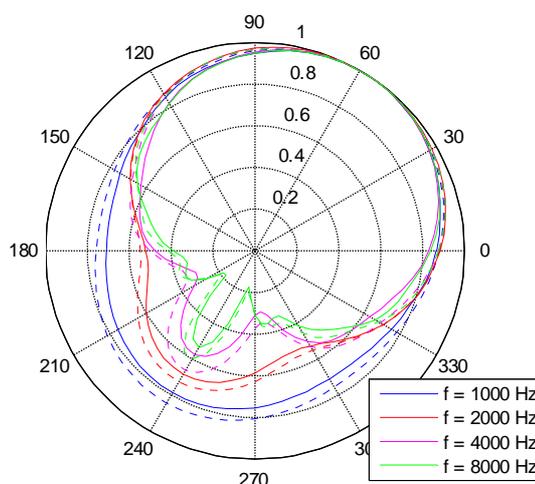


Figure 22 Polar patterns of measured (cont. lines) and simulated (dashed lines) sphere directivity for sensor #5. Directivities normalized with respect to the max values.

6.2.2. Quantitative analysis using spatial correlation

A more quantitative and global assessment of the sphere directivity model can be achieved by calculating spatial correlation between model and measure over the entire set of measurement incidences. We first compute it on the spectra modulus of sensors responses (40):

$$R_{|\mathbf{s}^{\text{meas}}|, |\mathbf{s}^{\text{model}}|} = \frac{\langle |\mathbf{s}^{\text{meas}}|, |\mathbf{s}^{\text{model}}| \rangle}{\| |\mathbf{s}^{\text{meas}}| \| \| |\mathbf{s}^{\text{model}}| \|}, \quad (40)$$

where we introduce the spatial scalar product and norm

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^* \cdot \text{diag}(\boldsymbol{\gamma}) \cdot \mathbf{y} \quad \text{and} \quad \|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}, \quad (41)$$

with the vector $\boldsymbol{\gamma}$ of spatial weightings (which sum is equal to 1) to compensate for the non homogeneous distribution of measurement angles around the sphere.

The top of Figure 23 shows that values are very close (if not equal) to 1 especially at low and mid frequencies. They decrease at high frequencies especially for a few sensors. Although it's hard to state a quality judgment from absolute values, it's worth noticing that the worst values concern the sensors [#10,12,15,16] near the sheath underneath the sphere, which causes a disturbing and non modelled diffraction effect.

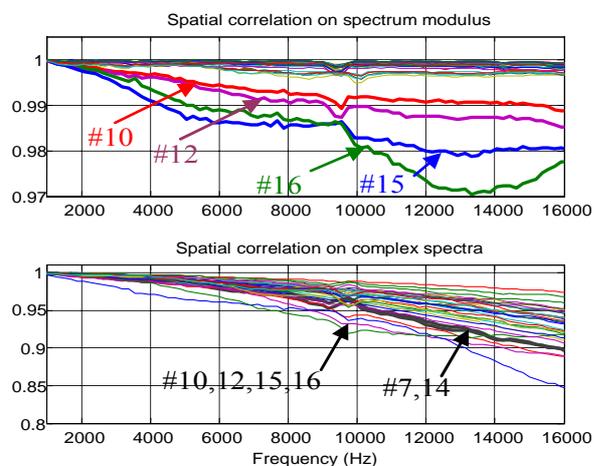


Figure 23 – Spatial correlation on spectra modulus (top) and complex spectra (bottom) between measured and simulated sensors responses.

As phase relationships are important when recombining sensors responses to form HOA directivities, it is relevant to compute spatial correlation on complex spectra (42), which includes the quality of phase matching between both modeled and measured directivity responses. For this purpose, the temporal realignment of measures as described in 6.1.1 is essential.

$$R_{\mathbf{s}_{\text{meas}}, \mathbf{s}_{\text{model}}} = \frac{\langle \mathbf{s}_{\text{meas}}, \mathbf{s}_{\text{model}} \rangle}{\|\mathbf{s}_{\text{meas}}\| \cdot \|\mathbf{s}_{\text{model}}\|} \quad (42)$$

Bottom of Figure 23 shows that sensors #10,12,15,16 still present the worst values. It also shows *relatively* bad values for few sensors (eg #7,14) that presented a quite good directivity shape (from correlation on modulus). This is probably not due to the sheath disturbance (all the more that sensor #14 is quite elevated and far away from the sheath), but rather to a slight positioning error on the sphere that causes a shift of phase information (related to time arrival of impinging waves) over the entire sphere. As a matter of fact, a position error of roughly 1 or 1.5 mm could be visually observed for these sensors.

To summarize, spatial correlation values are globally good or very good. Nevertheless, it would be worth converting phase errors into phase delay errors in order to better identify positioning errors. Moreover, little phase errors (and therefore little complex response

errors) at low frequency might be greatly amplified by microphone processing and might result in significant HOA reconstruction errors. This will be discussed in 6.3.3.

6.2.3. Extraction of individual sensors responses

In addition to directivity properties, another potential deviation of our prototypes from the theoretical model is the discrepancy between individual sensor responses. Equivalent free field response of each sensor (*i.e.* as omnidirectional sensor in the absence of the diffracting sphere) can be estimated together with the spatial correlation above, from the same measurements and simulations. Indeed, the measured responses can be factorized into the directivity effect and the sensor response. Provided that the measured directivity fits the model well enough (which is globally the case), equivalent free field response of a given sensor is estimated as:

$$|\mathbf{s}_{\text{meas}}| = |\hat{\mathcal{S}}^{\text{freefield}}| \cdot |\mathbf{s}_{\text{model}}| \quad (43)$$

$$|\hat{\mathcal{S}}^{\text{freefield}}| = \frac{\langle |\mathbf{s}_{\text{meas}}|, |\mathbf{s}_{\text{model}}| \rangle}{\|\mathbf{s}_{\text{model}}\|^2} \quad (44)$$

This uses the diffraction model to spatially weight the measures and compensate for their absence in the bottom part of the directional sphere, where the bottom sensors should have their maximum response. In principle the estimation is as reliable as the calculated spatial correlation is good, which is presently the case.

We find that sensors responses are mostly parallel with a maximum level difference less than 3dB. Relative levels found between these estimations are compliant with diffuse field responses estimations discussed in 6.1.2.

6.3. Objective characterization of the HOA microphone prototype performances

6.3.1. Processing methods and computation parameters

To objectively analyse the performance of our HOA microphone prototype, we process the measured responses by means of the matrix and filters defined in 5.2, with regularisation parameters corresponding to a maximum noise amplification of resp. +0dB, +6dB and

+20dB. In order to distinguish between the limitations inherent to the model and the artefacts due to the “imperfection” of the prototype regarding the model, we process the same way the simulated responses already involved in 6.2. We’ll note $\mathbf{S} = [\mathbf{s}_1 \dots \mathbf{s}_Q]^T$ the matrix of measured (or simulated) responses and $\hat{\mathbf{B}} = [\hat{\mathbf{b}}_1 \dots \hat{\mathbf{b}}_K]^T$ the matrix of estimated HOA responses, $\hat{\mathbf{b}}_k$ being the vector of responses for all measurement directions for the of the k^{th} HOA component. HOA responses are obtained from $\hat{\mathbf{B}} = \mathbf{P}\mathbf{S}$, with the global processing matrix:

$$\mathbf{P} = \text{diag}(EQ_m)\mathbf{E} \quad (45)$$

Finally, to investigate the real potential of the prototype after removing the limitations caused by its discrepancy from a known model, we introduce and try another kind of processing matrix \mathbf{P} . This is such that when applied on the whole set \mathbf{S} of measured responses, the set of estimated HOA responses $\hat{\mathbf{B}} = \mathbf{P}\mathbf{S}$ fits the best the HOA gains \mathbf{B} expected regarding the measurement directions. This leads to a classical minimisation problem that can be solved in practice by system inversion with regularisation, as it was discussed in previous sections. The “optimal” processing matrix is computed according to the following equation:

$$\mathbf{P} = \mathbf{B} \cdot \text{diag}(\gamma) \cdot \mathbf{S}^* \cdot (\mathbf{S} \cdot \text{diag}(\gamma) \cdot \mathbf{S}^* + \lambda \mathbf{I})^{-1}, \quad (46)$$

where regularisation parameter λ is tuned in the same way as in 4.1.2 (but without the factor \sqrt{Q}) to induce the same maximum value of noise amplification. Nevertheless, parameter λ and equation (46) have to be slightly modified to ensure both a little value of noise amplification (esp. 0dB) and a correct response level. Note that this matrix is frequency dependent, *i.e.* it is a matrix of filters that is much more consuming³ than the factorized solution in terms of digital signal processing.

In the following sections, we’ll denote by \mathbf{b} and $\hat{\mathbf{b}}$ the vector of resp. theoretical and estimated HOA responses. We’ll use labelling (fMo+XdB) or (iMe+XdB) when referring resp. to factorized processing (45) or the solution (46), with a max noise amplification of X dB.

For clearer illustrations, sensors responses have been pre-equalized to compensate their mean “free-field” response, without changing their relative levels.

6.3.2. Qualitative comparison of reconstructed and expected HOA directivities

With the aim to give a first qualitative insight into what happens in terms of HOA directivity reconstruction, we first focus on the result of the factorized processing (fMo+0dB).

Figure 24 shows 3D plots of some reconstructed directivities for 2 symptomatic frequencies. At 5 kHz, the patterns are mostly very well reconstructed (the slight amplitude mismatch will be discussed later). At 10 kHz, *i.e.* beyond the so-called spatial aliasing frequency (estimated as 7.5kHz in 5), patterns appear quite disturbed. One easily observes the presence of spherical harmonics of higher orders than the estimated ones, with sometimes a common azimuth (or elevation) dependence, presenting numerous additional lobes (higher angular frequencies). This is namely the effect of spatial aliasing.

To better assess the quality of reconstruction as a function of the frequency, let’s focus on the horizontal section (0 degree elevation angle) of directivity patterns and to “horizontal” HOA components. Figure 25 shows “cylindrical views” formed by concatenating these horizontal sections along a vertical axis that represents the frequency scale. It clearly makes appear a transition frequency [band] (about 7 or 8kHz or even above depending on the order m) where patterns begin getting disturbed in terms of proportion between lobes and/or number of lobes. On the other side the low frequency response, which is very good for the 0th component, progressively thins out over a low frequency band that enlarges as the order m increases. This is partly due to the quite moderated equalization effort involved in the processing here. The next section introduces quantitative criteria to further explore the reconstruction quality depending on the processing options.

³ Nevertheless we have verified that such a 32x25 filter matrix, with 256-tap FIR, can process in real time on an ordinary computer, if it is appropriately implemented.

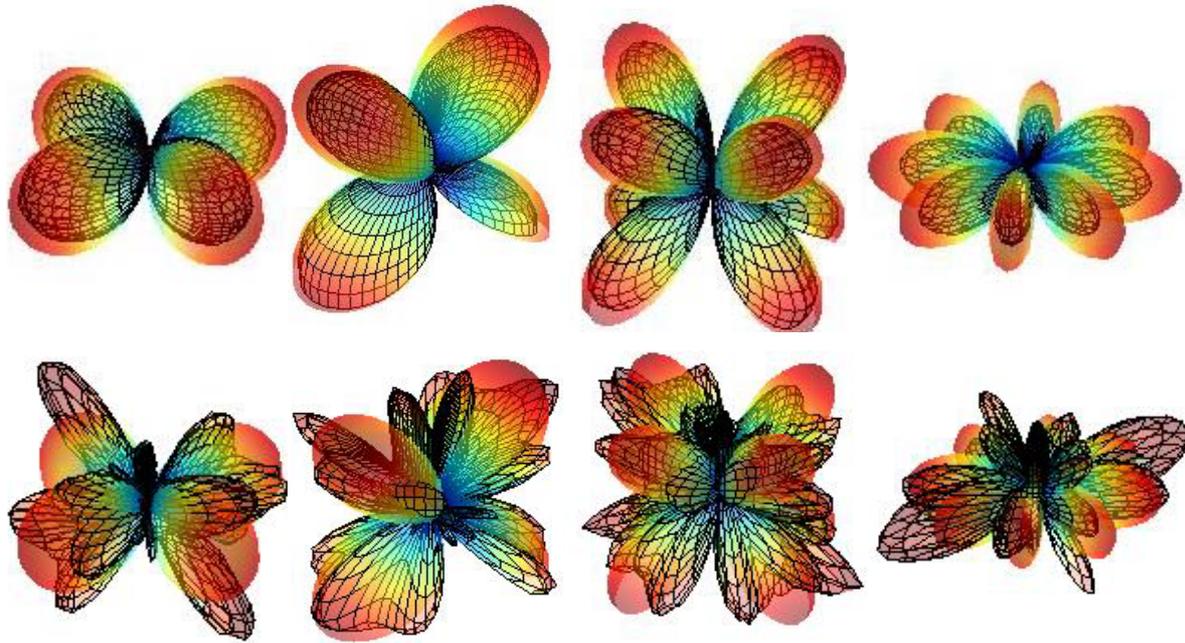


Figure 24 Reconstruction of a some specimens of HOA directivity at 5 kHz (top) and at 10 kHz (bottom). From left to right: components B_{22}^{+1} , B_{21}^{+1} , B_{32}^{+1} , B_{44}^{+1} .

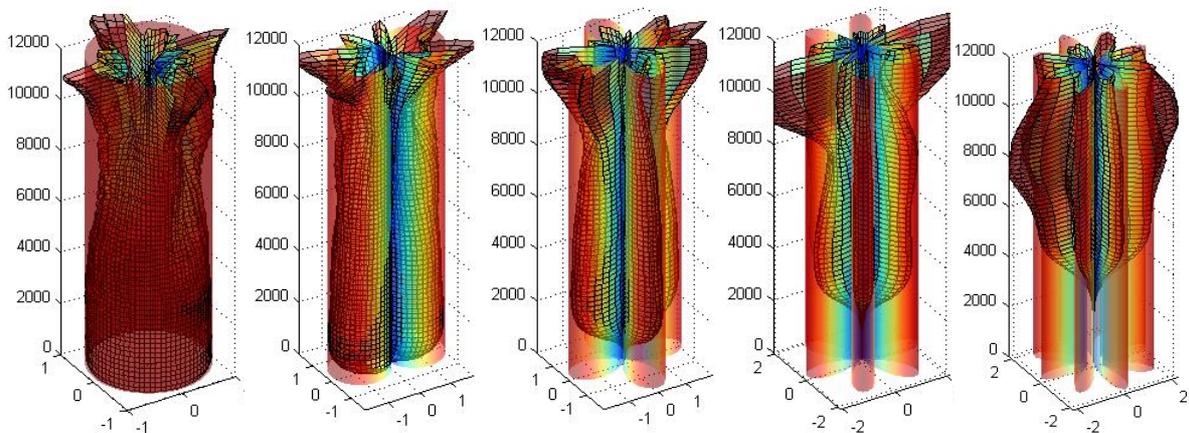


Figure 25 Cylindrical plots of reconstructed (meshed plots) and expected (in transparency) directivities for incidences in the horizontal plane, and for horizontal components B_{mm}^{+1} with $m=0$ to 4 (from left to right). Vertical axis is for frequency (in Hz).

6.3.3. Quantitative analysis

Similarly as in 6.2.2, we compute spatial correlation between reconstructed HOA directivities vectors $\hat{\mathbf{b}}$ and theoretical ones \mathbf{b} :

$$R_{\hat{\mathbf{b}},\mathbf{b}} = \frac{\langle \hat{\mathbf{b}}, \mathbf{b} \rangle}{\|\hat{\mathbf{b}}\| \cdot \|\mathbf{b}\|} \quad (47)$$

We complete this by a characterization of the mean level of estimated responses, relatively to the expected one:

$$L_{\hat{\mathbf{b}},\mathbf{b}} = \frac{\|\hat{\mathbf{b}}\|}{\|\mathbf{b}\|} \quad (48)$$

These equations indicates that in the case of factorized processing, the post-equalization has an impact of the component level (*i.e.* the scale of directivity response), but none on spatial correlation (*i.e.* the respect of the directivity "shape").

Both a good spatial correlation and a correct level of estimated HOA components take part into the quality of the captured sound field representation, and therefore of its reconstruction at the final rendering stage (after spatial decoding) 2.1.3.

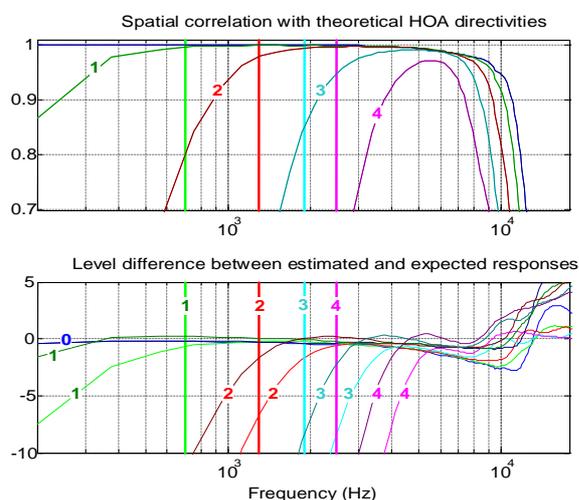


Figure 26 Quality of reconstructed directivity when processing is applied to measures. For more readability, each group of m^{th} order components is represented by a single, mean curve. Vertical bars correspond to limit frequencies (700, 1300, 1900, 2500 Hz) corresponding

to 1st, 2nd, 3rd, and 4th order rendering, as computed in 2.1.3. For bars and curves, 0th, 1st, 2nd, 3rd, and 4th orders are respectively represented in blue, green, red, cyan, and magenta, and appear from left to right. Component levels (bottom) corresponding to amplification parameters +0dB, +6dB resp. appear from right to left while darkening in color.

Figure 26 shows curves computed for measures processed with (fMo+0dB) (fMo+6dB) options. The case (fMo+20dB) is not plotted here to avoid confusion. In the present case, correct spatial correlation is bounded by both low and high frequency limits, the former being as high as the order. For comparison, Figure 27 shows the case of responses obtained from simulation. Since no errors are introduced in simulations with regard to the model, spatial correlation is perfect from 0Hz to the spatial aliasing frequency 7.5 kHz and even above (as the order decreases).

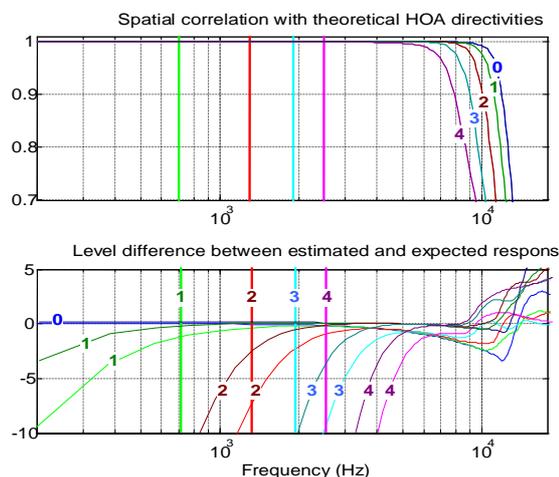


Figure 27 Same as Figure 26 but for processing applied to simulations.

Relative levels of estimated components look very similar in both cases (measures and simulations). They are approximately flat over a frequency band (also bounded by the spatial aliasing frequency) that extends towards low frequencies as much as the order is low and as the equalisation effort is high. With a +20dB allowed noise amplification, the lower bound of the "correct level frequency band" reaches the limit frequencies shown as vertical bars, even for 3rd and 4th orders. Nevertheless, there is no use to invest such an equalization effort (at least for the 4th order components) when dealing with the real prototype, for which the frequency band of potentially good reconstruction is

bounded by the low frequency slope of spatial correlation curves: Top of Figure 26 shows that good correlation is not reached at the limit frequency for the 4th order. As a general rule, there's no use that the frequency band of correct level exceeds the frequency band of correct spatial correlation.

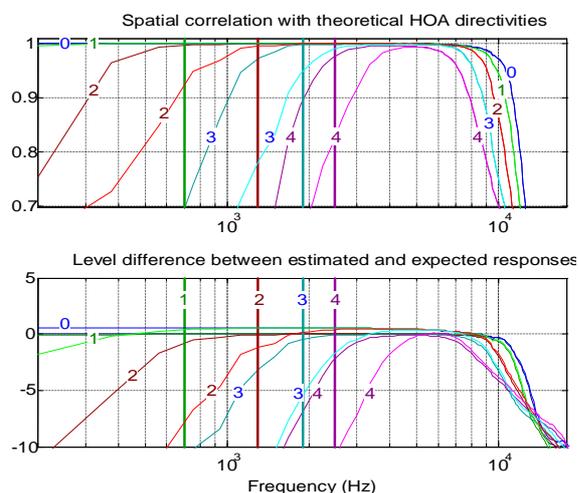


Figure 28 Same as previous figures, but with optimised processing (iMe+6dB) and (iMe+20dB) applied to measures.

Another useful comparison can be done with the HOA estimation derived from the processing labelled (iMeXdB) (Figure 28), which is optimized from the whole set of measurements (46). With the same given estimation effort, it presents roughly the same features as above in terms of maximal noise amplification and relative level of estimated components (except above the spatial aliasing frequency). It presents a spatial correlation that improves with higher estimation effort similarly to the component level. That means that with a more complete knowledge of the prototype's real directivity, one is able to reach HOA reconstruction properties as good as with model-based simulations.

6.3.4. Energy vector analysis

To complete the analysis of the prototype microphone performances, we anticipate now another feature of the spatialisation effect observed at the final rendering stage: that is a localisation criterion called "energy vector", introduced and discussed in 2.1.3. We compute it by operating a 4th order, "max r_E " decoding over estimated HOA frequency responses. It is considered to be a relevant localisation criterion in the medium-high

frequency band where acoustic reconstruction at the listener scale is no longer satisfied.

Energy vector characteristics are drawn as functions of the frequency in Figure 29 (bottom part): its mean angle error regarding the expected directions (*i.e.* the measurement directions) and the "blur width angle" α_E computed from its mean norm.

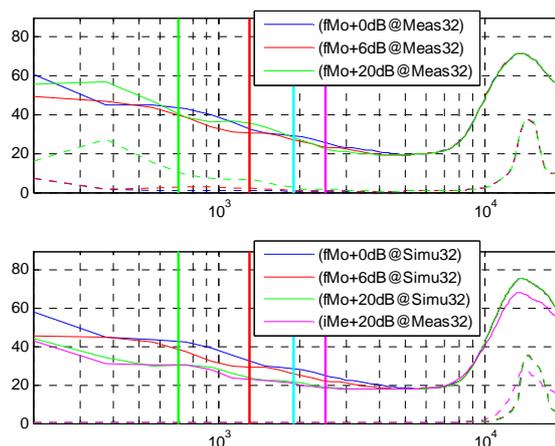


Figure 29 Global quality of energy vector at the reproduction stage: mean angle error regarding the expected directions (dashed lines); "blur width angle" α_E (cont. lines). All angles are in degrees.

In almost all cases, the angle error is approximately null up to the supposed spatial aliasing frequency and even beyond (about 10 kHz). The "blur angle" (attached to the vector norm) is quite important at low frequencies. It reaches its minimal value at about 5-6 kHz: that is also nearly the "optimal" value given by Table 1 for $M=4$, *i.e.* 18 degrees. It increases again from the spatial aliasing frequency and reaches a maximum of 70 degrees above 10 kHz.

In the case of factorized processing of measures (Top of the figure), it appears that a +6dB estimation effort reduces the "blur angle" just a bit (from 5 to 10 degrees). Curves confirm that a +20dB effort is counterproductive: blur angle as well angle error are greater at low frequency. On the contrary, increasing efforts are fruitful when applied to the simulated responses, as the bottom figure shows it (about 15 degrees blur angle reduction). Finally, optimised measures processing competes with processing on simulated responses for the same effort (+20dB on the figure).

7. CONCLUSION

This study completes an investigation work done by present authors as well as concurrent teams for some years on 3D sound field recording with microphone arrays, especially in relation with the spatialisation approach called "Higher Order Ambisonics". Focussing on spherical arrays, this work has led to the construction of a 4th order, 32 sensor microphone prototype and the design of appropriate signal processing methods grounded on both theoretical and practical considerations.

To objectively assess the performance of this HOA microphone as an efficient tool for spatially encoding natural sound fields, we performed complete directivity measurements and analysis. We proposed and discussed objective criteria like the spatial correlation between estimated HOA directivities with the expected spherical harmonics. The relevance of the encoding with regard to the sound field reproduction stage has also been discussed using the energy vector criterion. As a result, the prototype presents a nice potential as a 3D sound field recording means with moderate amplification effort. Further improvements are expected from this study.

Now a complete HOA spatialisation chain can be demonstrated, featuring 3D recording, virtual source encoding, sound field manipulations and spatial decoding, with signal processing modules implemented e.g. as VST plug-ins. As a particular example, real time HOA recording, processing and decoding for ITU loudspeaker setup as well as for a head-tracked binaural rendering, have been successfully demonstrated at the "Journées d'Etudes sur la Spatialisation" (JES2006) that took place at IRCAM and ENST-Paris in January 2006.

8. ACKNOWLEDGEMENTS

The authors would like to thank Olivier Warusfel for his valuable help in the objective microphones measurements done in IRCAM and for the room analysis toolbox.

9. REFERENCES

- [1]. GERZON, M.A., *Periphony: With-Height Sound Reproduction*. J. Audio Eng. Soc., 1973. **21**(1): p. 2-10.
- [2]. BAMFORD, J.S., *An Analysis of Ambisonics Sound Systems of First and Second Order*. 1995, Université de Waterloo: Waterloo.
- [3]. DANIEL, J., *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. 2001, Paris 6: Paris.
- [4]. DANIEL, J., NICOL, R., and MOREAU, S. *Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging*. in *114th AES Conv.* 2003. Amsterdam, The Netherlands.
- [5]. CRAVEN, P. and GERZON, M.A., *Coincident Microphone Simulation Covering Three Dimensional Space and Yielding Various Directional Outputs*. 1977: U.S.
- [6]. ABHAYAPALA, T.D. and WARD, D.B. *Theory and Design of Higher Order Sound Field Microphones Using Spherical Microphone Array*. in *IEEE ICASSP '02*. 2002.
- [7]. DURAISWAMI, R., et al. *High Order Spatial Audio Capture and its Binaural Head-Tracked Playback over Headphones with HRTF Cues*. in *AES 119th Convention*. 2005. New York, NY, USA.
- [8]. GOVER, B.N., RYAN, J.G., and STINSON, M.R., *Measurements of directional properties of reverberant sound fields in rooms using a spherical microphone array*. J. Acoust. Soc. Am., 2004. **116**(4, Pt. 1): p. 2138-2148.
- [9]. MEYER, J. and AGNELLO, T. *Spherical microphone array for spatial sound recording*. in *AES 115th Conv.* 2003. New York, NY, USA.
- [10]. MOREAU, S. and DANIEL, J. *Study of Higher Order Ambisonic Microphone*. in *joint meeting cfa-daga '04*. 2004. Strasbourg, France.
- [11]. MOREAU, S., DANIEL, J., and CHRAA, A. *Production, transmission et restitution temps réel d'une scène sonore dans un format audio 3D flexible*. in *CORESA 2004*. 2004. Lille.

- [12]. RAFAELY, B., *Analysis and Design of Spherical Microphone Arrays*. IEEE Transaction on Speech and Audio Processing, 2005. **13**(1): p. 135-143.
- [13]. LABORIE, A., BRUNO, R., and MONTOYA, S. *A New Comprehensive Approach of Surround Sound Recording*. in *AES 114th Conv.* 2003. Amsterdam, The Netherlands: AES.
- [14]. DANIEL, J. and MOREAU, S. *Further Study of Sound Field Coding with Higher Order Ambisonics*. in *AES 116th Conv.* 2004. Berlin, Germany: AES.
- [15]. GERZON, M.A., *Ambisonics in Multichannel Broadcasting and Video*. J. Audio Eng. Soc., 1985. **33**(11): p. 859-871.
- [16]. DANIEL, J., RAULT, J.B., and POLACK, J.D. *Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions*. in *AES 105th Conv.* 1998.
- [17]. MALHAM, D. *Experience With Large Area 3-D Ambisonic Sound Systems*. in *Proc. of the Institute of Acoustics*. 1992.
- [18]. MORSE, P.M. and INGARD, K.U., *Theoretical Acoustics*. Mc Graw-Hill ed. 1968, New York.
- [19]. WARD, D.B. and ABHAYAPALA, T.D., *Reproduction of a Plane-Wave Sound Field Using an Array of Loudspeakers*. IEEE Transaction on Speech and Audio Processing, 2001. **9**(6): p. 697-707.
- [20]. ELKO, G. and NGUYEN PONG, A.-T. *A steerable and variable first-order differential microphone*. in *Proc. 1997 IEEE ICASSP*. 1997.
- [21]. LI, Z. and DURAISWAMI, R. *A Robust and Self-Reconfigurable Design of Spherical Microphone Array for Multi-Resolution Beamforming*. in *IEEE ICASSP '05*. 2005.
- [22]. GOLUB, G. and VAN LOAN, C., *Matrix Computations, third edition*. 1996, London: The Johns Hopkins University Press.
- [23]. TIKHONOV, A.N. and ARSEININ, V.A., *Solution of Ill-posed Problems*. 1977, Washington: Winston & Sons.
- [24]. KIRKEBY, O. and NELSON, P.A., *Fast Deconvolution of Multi-Channel Systems using Regularisation*. 1996, ISVR Technical Report no.255: Southampton, UK.
- [25]. DRISCOLL, J.R. and HEALY, D.M., *Computing Fourier Transforms and Convolutions on the 2-Sphere*. Adv. Appl. Math., 1994. **15**: p. 202-250.
- [26]. ARFKEN, G., *Mathematical Methods for Physicists, 3rd ed.* 1985, San Diego, CA: Academic Press.
- [27]. VERHAEVERT, J., VAN LIL, E., and VAN DE CAPELLE, A. *Uniform Spherical Distributions For Adaptive Array Applications*. in *VTC'01*. 2001: IEEE.
- [28]. SLOANE, N.J.A., and al., *Spherical Codes*. <http://www.research.att.com/~njas/packings/>
- [29]. FLIEGE, J., *Integration nodes for the sphere*. <http://web.mat.bham.ac.uk/J.Fliege/nodes/nodes.html>