# "Caractérisation Perceptive de Champ Sonore à Résolution Spatiale Variable"

Equipe Espaces Acoustiques et Cognitifs IRCAM – CNRS UMR 9912 STMS 1 place Igor-Stravinksy 75004 PARIS France



### Mémoire de Stage Master 2 - SDI spécialité Acoustique - Parcours ATIAM Mars - Juillet 2013

Alexandre Panizzoli

<u>Encadrants</u>: Olivier Warusfel Thibaut Carpentier Markus Noisternig











### Résumé

Ce travail présente une étude perceptive sur la caractérisation de champ sonore enregistré par microphone sphérique restitué sur système de diffusion ambisonics et en binaural avec une résolution spatiale variable. Les nombreuses études objectives existantes sur les méthodes d'analyse-synthèse dans le domaine des harmoniques sphériques ont rendu possible l'utilisation de système de captation et de diffusion ambisonics avec un haut niveau de résolution spatiale. Notre étude s'est consacrée à la comparaison perceptive de deux approches: par encodage ambisonics et Formation de Voies (Beamforming). Dans ce test, nous nous sommes concentrés sur la partie diffuse de l'effet de salle en utilisant des stimuli convolués avec les réponses impulsionnelles multicanales enregistrées dans l'ESPRO. Ce test vise à déterminer quel ordre d'encodage minimum est requis pour la partie tardive de l'effet de salle. En outre, notre expérimentation a pour but d'identifier le seuil où l'ordre HOA n'a plus (ou peu) d'influence sur la perception sonore des paramètres de transcodage. La deuxième question est de savoir si les sujets peuvent faire la distinction entre la Formation des voies et l'Ambisonics. L'analyse ANOVA sur les réponses des auditeurs nous a permis de vérifier que le type d'encodage a un effet sur la perception que les auditeurs ont de la salle. Plus particulièrement, l'analyse post-hoc par test de comparaison multiple nous a permis de déterminer que l'ordre 3 ambisonics n'avait pas de différence audible avec l'ordre HOA-4. Ce travail pose les bases pour un futur développement du logiciel Spat, visant à manipuler paramétriquement les DRIR dans un moteur de réverbération hybride.





## Remerciements

Je souhaiterais particulièrement remercier Mr. Olivier Warusfel, Mr. Markus Noisternig, Mr. Thibaut Carpentier et Mr. Tal Szpruchpour leur soutien sur le plan scientifique, technique aussi bien que méthodologique et pour m'avoir accueilli au sein de l'équipe Espaces Acoustiques et Cognitifs. Nous remercions également Orange Labs pour le prêt de l'EigenMike <sup>®</sup>.

Je tiens également à remercier les divers membres de l'équipe pédagogique du Master ATIAM pour m'avoir permis de passer une année très enrichissante au sein de la formation.

Enfin je tiens également à remercier mes collègues stagiaires de l'équipe EAC et mes camarades de promotion ATIAM, pour avoir créé une ambiance de travail motivante et agréable et pour leur amitié tout au long de l'année.





## Table des Matières

N	otatio	ons		V
1	Intr	oductio	n	2
2	Con	texte &	enjeux	4
	2.1	FDN &	& Convolution	4
	2.2	Motiv	ations	4
	2.3	Etat de	e l'art	5
	2.4	Appro	oches proposées	8
		2.4.1	НОА	9
		2.4.2	Beamforming ou Formation de voies	10
		2.4.3	Echantillonnage de la sphère de mesure	10
	2.5	Object	ifs du stage	11
3	Cap	ture de	champs sonores par microphone sphérique	12
	3.1	Caract	téristiques d'un champ réverbérant	12
	3.2	Analy	se de DRIR dans le domaine des harmoniques sphérique	18
		3.2.1	Décomposition en ondes planes	18
		3.2.2	Phénomènes limitants la représentation exacte du champ sonore	19
	3.3	Protoc	ole de Mesure	21
	3.4	Vérific	cation des données expérimentales	22
		3.4.1	Variance des capsules	22
		3.4.2	Comportement global	24
	3.5	Débru	itage des mesures	27
		3.5.1	Cas monophonique	27
		3.5.2	Extension au débruitage de microphone sphérique	28
4	Ana	lyse-Sy	onthèse de réponses impulsionnelles	32
	4.1	Appro	oche par transcodage HOA	32
	4.2	Appro	oche par formation de voies	35
	4.3	Transc	codage binaural	37





5	Eval	luation perceptive des approches	38
	5.1	Enjeux du test ABX	38
	5.2	Procédure	40
		5.2.1 Préparation des stimuli	40
		5.2.2 Dispositif expérimental	42
	5.3	Analyse des résultats	44
		5.3.1 ANOVA	44
		5.3.2 Multiple Comparison Test	46
_	~ •		
6	Crit	iques et perspectives	50
	6.1	Développements possibles du test perceptif	50
	6.2	Modèle de convolution hybride	50
7	Con	nclusion	51
8	ANI	NEXES	53
	8.1	Décomposition en Harmoniques Sphériques	53
	8.2	Discrétisation de la Sphère	54
	8.3	Min et Max des Variances des Capsules	55





## Liste des Figures

1	Diagramme de DirAC - Méthode basée sur une reproduction d'un signal en B-Format	6
2	Schéma d'Analyse-Synthèse des approches Ambisonics et par Formation de Voies	8
3	Illustration du problème intérieur	9
4	Réponse impulsionnelle temporelle simplifiée d'un espace réverbérant et sa distribution	
	temps-fréquence	12
5	Illustration du profil Gaussien de la queue de réverbération	15
6	Vecteur intensité en champs diffus (gauche) et non-diffus (droite)	15
7	Exemple de représentation espace-temps obtenue à partir de décomposition en onde plane	
	d'un enregistrement avec EigenMike®	19
8	Amplitude-Fréquence de la Fonction Radiale pour des sphères ouverte et fermée	20
9	Artefact rémanent sur la réponse impulsionnelle après déconvolution	21
10	Paramètres EDR pour l'analyse des RiR en configuration Réfléchissante (HP 1)	23
11	Valeur Moyenne Globale des paramètres EDR	25
12	Ratio par rapport à la valeur moyenne globale du Po pour un microphone donné	26
13	EDR avant (à gauche) et après (à droite) débruitage des RIR	27
14	Cohérence d'un Champ Diffus	29
15	Cohérence partie diffuse d'une RiR (comparaison des microphones Adjacents et Opposés	
	sur la sphère)	31
16	Diagramme d'amplitude de filtres modal radial inverse	33
17	Gap apparaissant sur la partie débruitée cohérente après transcodage HOA des IR (canal 25)	34
18	Exemple de lobes spatiaux hypercardioïde d'une formation de voie	36
19	Méthode de reproduction ambisonics par reproduction sur haut-parleur virtuel et simula-	
	tion binaural	37
20	Visualisation Géométrique des ordres HOA et du nombre de Voies	38
21	Energie de chaque section ambisonics décodée en Binaural. En abscisse les 4 premiers cor-	
	respondent aux signaux HOA, puis les numéros 5 6 et 7 correspondent aux formations de	
	voies avec 4, 9 et 16 directions.	41
22	Interface du Test ABX communiquant à Spat via OSC et MaxMsp	43
23	Analyse N-Way ANOVA sur la moyenne des résultats du test en studio	45
24	Analyse N-Way ANOVA sur la moyenne des résultats du test en binaural	46
25	Analyse de Comparaison Multiple sur le type d'encodage pour les résultats du test en Studio	47
26	Analyse de Comparaison Multiple à deux facteurs (type d'encodage et "room") pour les	
	résultats du test en Studio	48





27	Analyse MCP du test Binaural)	49
28	Représentation des Harmoniques Sphériques Réelles pour $n = 0, 1, 2$ et $m = -n,, n$	54
29	Affichage des occurrences des valeurs extrêmes (min et max) pour le calcul de la variance	
	des capsules en fonction de la fréquence	55

## Liste des Tableaux

1	Récapitulatif des ordres des stimuli de test ABX	 ł2
-	neeuphunum des ordres des stimum de test mbri	 







## Notations

С	vitesse du son dans l'air (sauf spécifié $c = 343.2 \text{ m/s}$ )
$k=\omega/c=2\pi f/c$	nombre d'onde ( $f$ étant la fréquence)
i	nombre imaginaire, tq. $\mathbf{i}^2 = -1$
$\Re\{.\}$	partie réelle
$\Im\{.\}$	partie imaginaire
(.)!	factorielle
(.)!!	double factorielle
$(.)^*$	complexe conjugué
$\mathbf{x}^T = [x \ y \ z]$	T dénote la transposition du vecteur <b>x</b>
$\mathbf{x}^H = [x \ y \ z]$	$^{H}$ dénote la transposée hermitienne du vecteur <b>x</b>
$\mathbf{r} \equiv (r,\theta,\phi)$	position dans le système de coordonnée sphérique
	avec $\phi$ l'angle azimut sur $[0 - 2\pi]$ et $\theta$ l'élévation (ou site) sur $[0 - \pi]$ .
	sauf spécifié, la convention utilisée est celle décrite dans [Wil99]
	où $\theta$ est mesuré à partir de l'axe polaire <b>z</b>
$\Omega \equiv (\theta,\phi)$	position angulaire en coordonnée sphérique
$P_{n}\left(x\right)$	polynome de LEGENDRE d'ordre $n$
$P_{n}^{m}\left( x ight)$	Fonction de LEGENDRE associée d'ordre $n$ et degré $m$
$Y_{n}^{m}\left(\Omega\right)=Y_{n}^{m}\left(\theta,\phi\right)$	valeur complexe des harmoniques sphériques d'ordre $n$ et degré $m$ à $\Omega$
$\Upsilon_{n}^{m}\left(\Omega\right)=\Upsilon_{n}^{m}\left(\theta,\phi\right)$	valeur réelle des harmoniques sphériques d'ordre $n$ et degré $m$ à $\Omega$
$J_{lpha}\left(x ight)$	Fonction de BESSEL ordinaire de première espèce
$Y_{lpha}\left(x ight)$	Fonction de BESSEL ordinaire de deuxième espèce
$j_{n}\left(x ight)$	Fonction de BESSEL sphérique de première espèce d'ordre $n$
$y_{n}\left(x ight)$	Fonction de BESSEL sphérique de deuxième espèce d'ordre $n$
$h_{n}^{\left(1\right)}\left(x\right)$	Fonction de HANKEL sphérique de première espèce d'ordre $n$
$h_{n}^{\left(2\right)}\left(x\right)$	Fonction de HANKEL sphérique de deuxième espèce d'ordre $n$
$p\left(\mathbf{r},t ight)$	pression acoustique instantanée à la position r, dans le domaine temporel
$p\left(\mathbf{r},k ight)$	pression acoustique complexe à la position <b>r</b> , dans le domaine fréquentiel (à la fréquence <i>k</i> )
$p_{nm}\left(\mathbf{r},k ight)$	coefficients d'expansion en harmoniques sphériques de
	la pression acoustique $p(\mathbf{r}, k)$ , pour l'ordre $n$ et le degré $m$
$d\Omega$	élément de surface infinitésimal











## 1 Introduction

La capture de champs sonores par réponse impulsionnelle et les méthodes d'analyse-synthèse et restitution dans le domaine ambisonics ont fait l'objet d'études récentes [Raf05], [Bra10]. L'un des moteurs de la recherche actuelle dans le domaine est la captation et la diffusion de champs sonores avec un haut niveau de résolution spatiale pour un maximum de fidélité sonore. Les seules techniques permettant à la fois une capture microphonique et une restitution sonore en 3D sont la technique binaurale et la famille ambisonicss. Dans le premier cas l'enregistrement est réalisé par tête artificielle et la restitution se fait sur casque. On peut procéder par synthèse à partir de l'échantillonnage de la fonction de directivité de la tête (HRTFs) [RT13]. Il s'agit donc d'une reproduction individuelle. Dans le second cas, le champ sonore capté par des réseaux sphériques de microphones fait l'objet d'une décomposition dans le domaine modal et la reproduction est généralement réalisée sur un ensemble de haut-parleurs régulièrement répartis autour de la zone publique. Lorsque la décomposition modale est menée à un ordre supérieur à 2, ce qui exige un nombre important de transducteurs (microphones ou haut-parleurs) on parle d'ambisonicss aux ordres supérieurs (HOA pour High Order ambisonics). Une reproduction HOA est également possible sur casque en exploitant soit le paradigme des haut-parleurs virtuels soit en exploitant une décomposition modale de la fonction de directivité de la tête.

Reproduit par le biais des HRTF<sup>1</sup> et les méthodes de reproduction binaurales ou sur des systèmes de diffusion ambisonics, les caractéristiques perceptives des différents niveaux d'encodage HOA<sup>2</sup> n'ont été que peu étudiées. Ce rapport vise à étudier la perception d'un champ sonore préalablement caractérisé par une réponse impulsionnelle multicanale enregistrée avec un microphone sphérique et reproduit au travers d'un casque ou de haut-parleurs.

La perception de l'espace est un attribut très important pour la qualité d'écoute dans une salle de concert. Des études visant à déterminer les qualités subjectives et objectives d'une salle ont fait l'objet de recherche et d'une thèse au sein de l'équipe EAC [Kah95], [Lav89] et permis le développement de critères relatifs à l'audition humaine et sa perception de l'espace sonore. Morimito *et al.* [MJN07], [MNL08] ont notamment trouvé des relations entre les propriétés perceptives et les caractéristiques des réflexions précoces et tardives.

Des méthodes objectives de restitutions basées sur ces critères perceptifs ont également été développées dans le cadre de la thèse de J-M Jot [Jot92]. Cette thèse a contribué au développement du logiciel *Spat* créé par l'équipe EAC. Ce logiciel<sup>3</sup> fonctionnant entre autre sous Max/MSP est utilisé par des RIM<sup>4</sup> de l'IRCAM pour quasiment tous les concerts ou ils travaillent sur la spatialisation sonore. L'ESPRO a pour avantage d'une résolution spatiale exceptionnelle (HOA-3D à l'ordre N=9). Ce dispositif permet à de nombreux compositeurs d'étudier et de travailler sur les relations entre son et espace et proposer une diffusion spatialisée de leurs oeuvres.

<sup>&</sup>lt;sup>1</sup> *Head Related Transfer Functions* - Fonction de Transfert Spatiale relative à la tête humaine et capable de reproduire l'écoute humaine binaural

<sup>&</sup>lt;sup>2</sup> *High Order ambisonics -* ambisonics d'Ordre Elevé

<sup>&</sup>lt;sup>3</sup> Le *Spat* est une suite logicielle dédiée au traitement de spatialisation en temps réel de signaux sonores dans les contextes de création musicale, de post-production et dans les situations de concert

<sup>&</sup>lt;sup>4</sup> Réalisateur en Informatique Musicale







Cependant, jusqu'à présent aucune étude perceptive visant à établir des relations entre propriétés physiques et qualité de reproduction du champ sonore suivant différents niveaux d'encodage n'ont été mené au sein de l'équipe. En conséquence ce travail vise à présenter les méthodes utilisées pour déterminer si ces critères qualitatifs et quantitatifs sont reliés et ont un effet audible sur la qualité du système ambisonics à l'état actuel.

Le travail commencera donc par une étude bibliographique des critères pertinents pour l'évaluation de la qualité acoustique d'une salle. Il se poursuivra par l'élaboration d'un protocole expérimental permettant d'évaluer différents niveaux de précision de la restitution. La base de données de réponses de salles mesurées fera l'objet de procédures d'analyse-resynthèse (segmentation, transcodage, convolution) afin de générer les stimuli de test. L'expérience sera conduite dans l'un des studios de l'Ircam équipé d'un dôme de restitution sonore tridimensionnelle et les données seront ensuite analysées en vue d'établir un ensemble de spécifications techniques pour le moteur de spatialisation ambisonics.





## 2 Contexte & enjeux

#### 2.1 FDN & Convolution

La plupart des réverbérateurs commerciaux développés dans les années 90 utilisent des réseaux de retards rebouclés (Feedback Delay Network) qui simulent de manière efficace les propriétés statistiques de la réverbération dans les salles [JC91],[SP82]. Leur architecture algorithmique permet un contrôle flexible et varié au travers de nombreux descripteurs bas-niveaux (retard initial, densité d'écho, enveloppe de décroissance) et haut-niveau (présence de la source, enveloppement, réverbérance). L'Ircam a développé un outil de spatialisation en temps-réel basé sur ce principe auquel est associé un ensemble de paramètres de contrôle basés sur les résultats d'études perceptives.

Les réverbérateurs plus récents privilégient l'approche par convolution, qui consiste à convoluer les signaux à spatialiser par une réponse impulsionnelle mesurée dans une salle. Elle offre un haut degré de réalisme mais souffre d'un manque de flexibilité avec des paramètres de contrôle qui se concentrent surtout sur l'enveloppe temporelle en négligeant les paramètres spatiaux de la réponse.

#### 2.2 Motivations

Le stage s'inscrit dans un projet de développement d'une méthode de spatialisation hybride permettant de préserver le caractère réaliste des réponses impulsionnelles mesurées dans une salle tout en autorisant des opérations de transformation selon différentes dimensions objectives ou perceptives. Par exemple, on peut vouloir moduler les impressions de réverbérance et d'enveloppement en modifiant la distribution temporelle, spatiale et fréquentielle de la réponse mesurée. Des réponses impulsionnelles multicanales ont été acquises dans différentes salles de concert (Cité de la Musique, Ircam) à l'aide d'un microphone à très grande résolution spatiale [Nic10]. L'exploitation de ces mesures dans un moteur de réverbération convolutif requiert des ressources de calcul très importantes en raison du nombre important de données à filtrer. Le but du stage est d'élaborer et conduire une étude perceptive permettant de dimensionner au mieux le moteur de rendu hybride, c'est-à-dire permettant de déterminer d'un point de vue perceptif la résolution spatiale nécessaire/suffisante à la resynthèse convaincante de l'effet de salle. L'étude distinguera notamment les réflexions précoces - jouant un rôle dans la sensation d'enveloppement - du champ diffus de la partie tardive des réponses de salle. En outre l'étude comparera, sur une échelle perceptive, différentes approches du moteur de synthèse par une méthode purement convolutive dans le domaine ambisonics.





#### 2.3 Etat de l'art

Les méthodes de reproduction de champ sonore multicanal à partir de réponse impulsionnelle ont été discutées depuis plusieurs années et quelques méthodes se sont attachées à reproduire de manière flexible ce type de signaux pour diverses configurations (Surround 5.1 [IR10], VBAP<sup>5</sup>, ambisonics [Ger75], [Dan01], [ZPN12]) [PM06]. Les critères perceptifs pertinents pour l'analyse de champs sonore étant très nombreux et consultables dans de nombreux travaux [Kah95], nous baserons notre étude sur un domaine plus bas niveau, où l'on cherchera avant tout à savoir quelle est la méthode et le niveau d'encodage nécessaire pour la reconstruction d'un champ diffus.

Merima et Pulkki [MP05], [PM06], [Pul07] ont développé deux méthodes assez proches afin d'effectuer un rendu de réponse impulsionnelle spatiale du champ sonore applicable sur différents systèmes de diffusion. SIRR<sup>6</sup> et DirAC<sup>7</sup> se basent sur l'analyse temporelle, fréquentielle (pour DirAC) et spatiale de la réponse impulsionnelle afin d'estimer la direction d'arrivée des ondes et la séparer ainsi de la zone du champ diffus. Cette technique permet ainsi de recréer un champ sonore où un auditeur humain est capable de percevoir et de localiser des sources de manière très précise, notamment dans la direction frontale, en recréant la partie des réflexions précoces et en prenant en compte les indices de retard interauraux<sup>8</sup>. Ils prennent également en compte les indices de localisation monauraux, la cohérence interaurale et préserve le timbre sur les différentes dimensions spectrales, temporelles et spatiales.

En effectuant une analyse d'intensité et de l'énergie sur le signal enregistré la méthode SIRR propose une estimation de la diffusion du champ sonore par la relation suivante :

$$\psi = 1 - \frac{||\langle \mathbf{I}(t)/c \rangle||}{\langle E(t) \rangle} = 1 - \frac{2Z_0||\langle p(t)\mathbf{u}(t) \rangle||}{\langle p^2(t) \rangle + Z_0^2 \langle \mathbf{u}^2(t) \rangle}$$
(1)

Où I(t) représente le vecteur intensité moyénné sur le temps t (définit comme le produit pression vitesse des particules noté u(t)). La densité d'énergie instantanée est représentée par E(t). La quantité  $1 - \psi$ est égale à la vitesse du transfert d'énergie divisée par la vitesse du son c.  $Z_0$  est l'impédance acoustique du milieu définie comme  $Z_0 = \rho c$  avec  $\rho$  la masse volumique de l'air. La proportion de l'énergie sonore oscillante localement peut donc être utilisée pour estimer la diffusivité du son. Où une valeur de  $\psi = 1$ indique un champ diffus idéal et une valeur de  $\psi = 0$  indique une absence totale d'oscillation énergétique. Pour augmenter la rapidité de calcul on peut également effectuer ce calcul dans le domaine fréquentiel.

Leur méthode de synthèse pour la diffusion se base sur une banque de prise de son par microphone Soundfield (B-Format équivalent à l'ambisonics ordre 1 [MPL05]) en utilisant seulement le canal W om-

<sup>&</sup>lt;sup>5</sup> Vector Based Amplitude Panning - Positionnement de source par panoramique vectoriel d'amplitude [Pul97]

<sup>&</sup>lt;sup>6</sup> Spatial Impulse Response Rendering

<sup>&</sup>lt;sup>7</sup> Directional Audio Coding

<sup>&</sup>lt;sup>8</sup> Pour déterminer la direction d'arrivée des ondes sonore vers la tête, le système auditif humain analyse l'information du signal transmis par le biais de plusieurs indices, dont: - La Différence de Temps Interaurale (*Interaural Time Difference*). Le son du côté droit atteint l'oreille droite plus tôt que l'oreille gauche. Le système auditif évalue donc des ITD par les retards de phase à basse fréquence et retards de groupe à des fréquences élevées. - Les différences de niveau interaurales (*Interaural Level Difference*). Le son du côté droit a un niveau supérieur à l'oreille droite comparé à l'oreille gauche, dû masquage de la tête sur l'oreille gauche. Ces différences de niveau sont très dépendants de la fréquence et ils augmentent proportionnellement avec la fréquence.





nidirectionnel afin d'effectuer l'estimation du champ diffus. Pour reconstruire la partie directionnelle une technique par VBAP a été utilisée à partir des paramètres de localisation (azimut et élévation) estimés par calcul d'intensité.



**Figure 1:** Diagramme de DirAC - Méthode basée sur une reproduction d'un signal en B-Format
[Pul07]

Cet article présente également une étude perceptive sur la qualité de localisation et la qualité globale du champ sonore reproduit à partir d'enregistrement en B-Format. A la suite de ces travaux, les résultats indiquaient que cette méthode donnait de meilleurs résultats que la reproduction ambisonics d'ordre 1 sur un système de reproduction multicanal 5.1.

Ces résultats intéressants montrent qu'une reproduction séparée des différentes sections temporelle d'une réponse impulsionnelle permet de recréer une cohérence interaurale de qualité raisonnable. Leur article se basait en premier lieu sur des signaux simulés à l'aide du logiciel DIVA [SJHL99] et restitués en chambre anéchoïque. Cependant la seconde partie de leur expérimentation a montré qu'une reproduction en conditions réelles entraînait des résultats dégradés sur l'ensemble des attributs perceptifs testés. Notre étude s'attachera donc à employer des outils propres à la technologie ambisonics existant au sein de l'équipe EAC. Nous aborderons une nouvelle approche recommandée par Merima et Pulkki [PM06] en augmentant la résolution spatiale afin de vérifier quel est l'ordre optimal à restituer pour aborder les problèmes de paramétrisation et d'analyse automatique des sections temporelles des réponses.

Pour revenir justement sur le but de notre étude il faut mettre en avant le fait que nous utilisons des enregistrements réels de réponses de salle pour faire la convolution et qui seront ensuite utilisés pour simuler des réverbérations artificielles (ou virtuelles). Pour mesurer ces salles, la solution retenue par de nombreux travaux [BF12] peut-être d'utiliser des micros sphériques afin que ces réponses révèlent les dimensions spatiales de la salle et pas seulement le contenu temporel et fréquentiel (méthode utilisée dans DirAC [Pul07] et SIRR [MP05]).





L'état actuel de la recherche présente beaucoup d'études sur le microphone sphérique mais basées sur des méthodes objectives de signaux et de principes acoustiques. Ces dernières années, on commence tout de même à voir apparaître des études perceptives sur le domaine ambisonics ou bien sur l'utilisation des ces réponses spatiales dans un but d'auralisation.

Avant de présenter notre approche nous pouvons étudier les manques existants dans la littérature afin d'orienter notre étude et pallier les défauts de notre système. Par exemple, Bertet et al. [BDPW13] ne regardent que le son direct, pour lequel ils montrent que l'ordre 4 est nécessaire pour une bonne reproduction. Polack et Figueiredo [PF12] n'utilisent qu'un Soundfield (ordre 1), tout comme Merima et Pulkki. Avni et al. [AAG<sup>+</sup>13] qui comparent différents ordres HOA de réponses mais ne distinguent pas les différentes sections temporelles entres elles (Onde Directe, réflexions précoces, queue de réverbération). On ne peut donc pas conclure sur l'origine des différences entre les différents stimuli proposés, les origines peuvent être multiples et l'analyse est sans doute trop globale. De plus il s'agit de réponses très courtes (avec un temps de réverbération inférieur à 0.6 secondes), et cela correspond donc à des réponses de salle très particulières.

Notre but est de faire une analyse plus détaillée. On commencera par la partie diffuse, en prenant par exemple la queue de réverbération. Bien que la partie tardive puisse aisément être simulée par un réseau de retards rebouclés, il est tout de même important de savoir quel ordre il faut pour encoder/décoder cette partie diffuse (que celle-ci soit exponentiellement décroissante ou non) dans des conditions optimales. Une précédente étude menée par Warusfel *et al.* [Bea07] à montré par ailleurs que des microphones sphériques d'ordres supérieurs permettaient une meilleure localisation des sources en présence lors de la restitution. En partant sur les bases de cette étude nous développerons un test perceptif prenant en compte la chaîne globale d'analyse-synthèse d'un signal dans le domaine ambisonics.





#### 2.4 Approches proposées

Pour essayer de palier le fait que les méthodes DirAC et SIRR n'utilisent pas des ordres ambisonics élevés, on propose deux méthodes d'analyse/resynthèse dans le domaine sphérique. Les deux approches reposent sur un encodage/décodage dans le format Ambisonics ou sur un découpage spatial de la réponse impulsionnelle selon différents faisceaux directifs (formation de voies)<sup>9</sup>. Le schéma de principe représentant les différentes étapes nécessaires à la synthèse, à partir d'un enregistrement réalisé via le microphone sphérique est visible ci-dessous (figure 2).



**Figure 2:** Schéma d'Analyse-Synthèse des approches Ambisonics et par Formation de Voies Ce diagramme décrit les étapes de captation par microphone sphérique multicanal et de restitution de champs sonores à résolution spatiale variable dont l'ordre sera dicté par le nombre de capsules et le rayon de la sphère du micro. Il indique enfin le choix de l'ordre pour la restitution des différentes sections de la réponse impulsionnelle sur le système de diffusion ambisonics ou binaural.

La comparaison et l'évaluation perceptive de la résolution spatiale nécessaire à une reconstruction réaliste du champs diffus est liée à la partie centrale de ce diagramme, c'est-à-dire la partie transcodage qui nous permet d'agir sur les paramètres de résolution spatiale et de changer l'ordre d'encodage, d'étudier les différences entre champ direct et diffus et subséquemment d'évaluer la différence du coût de calcul (en nombre de canaux pour la convolution) pour le temps-réel. Il est utile de noter que l'ordre d'encodage/décodage de chaque section peut être paramétré. Mais pour simplifier l'étude, nous nous limiterons à la paramétrisation de R3. Les autres sections étant conservées à l'ordre maximal (fonction des dispositifs de mesure et reproduction) Nous allons donc présenter dans les sections suivantes le principe de ces deux techniques et leurs possibles applications dans notre approche.

<sup>&</sup>lt;sup>9</sup> *Beamforming* - Formation de Voies ou de Faisceaux (cf. explication ci-dessous et approche détaillée chapitre IV.)





#### 2.4.1 HOA

L'ambisonics est une technique développée par Gerzon [Ger85], [Ger75] comme une approche d'encodage spatiale de champs sonores et de restitution sur des systèmes de haut-parleurs 3D. La représentation mathématique de l'ambisonics est basée sur la décomposition en harmoniques sphériques du champ sonore dans le système de coordonnées sphériques. Les détails de calcul du passage dans le domaine des harmoniques sphériques est donné en Annexe section 8.1.

En tant qu'approche pour la spatialisation du son dans un domaine fermé, cette technique implique un "point de vue" (ou d'écoute) libre de source virtuelle à l'intérieur du domaine [DM04], [Ple09]. Il s'agit du problème intérieur représenté par la figure 3. Dans le cadre de ce problème, les sources sonores sont situées à l'extérieur d'une sphère de rayon  $r_{max}$ . Nous cherchons à caractériser le champ sonore à l'intérieur de cette sphère qui constitue le domaine où la pression acoustique reconstituée par la contribution de chaque haut-parleur est valable.



Figure 3: Illustration du problème intérieur [Col11]

Si nous connaissons la pression acoustique sur l'une des sphère de rayon  $r_d \leq r_{max}$  (qui correspond dans notre cas à la sphère de mesure), nous pouvons donc calculer les coefficients de Fourier spatiaux et les réinjecter par la suite dans l'équation de propagation dans le domaine sphérique. Une fois les coefficients spatiaux déterminés, il est donc possible d'obtenir la pression acoustique en tous point de notre domaine de validité.





#### 2.4.2 Beamforming ou Formation de voies

La formation de voies est une technique en traitement du signal utilisant des réseaux de transducteurs (ou dans notre cas de microphones) et qui permet de fournir un outil de grande précision pour le filtrage spatial [ea10]. Le but est de sélectionner une ou plusieurs régions spécifiques de l'espace en atténuant les interférences et le bruit provenant des autres directions. Dans le cadre de notre étude nous procédons en plusieurs étapes. Le but est d'opérer un découpage spatial du champ sonore selon un nombre fini de directions. On suppose que la restitution peut-être réalisée en simulant chacun de ces faisceaux par une onde plane restituée par le système de diffusion selon la direction du faisceau. En pratique ces faisceaux sont synthétisés dans le domaine des harmoniques sphériques ce qui permet d'établir un cadre commun aux deux approches pour notre étude. La dérivation des filtres nécessaires à l'opération de formation de voies est réalisée dans le domaine modal. En effet la symétrie sphérique de notre réseau de transducteurs permet de calculer efficacement les filtres paramétriques pour la formation de voies. [Raf04], [Raf05].

L'ordre d'encodage de chaque rayon sera équivalent au niveau de résolution du microphone utilisé pour la mesure. Le paramètre que nous ferons varier étant celui du nombre de directions relatif au degré d'échantillonnage de la sphère (cf. partie du bas - *Steering Direction* - de la Figure 20).

#### 2.4.3 Echantillonnage de la sphère de mesure

Le nombre et la répartition spatiale des cellules microphoniques (reposant sur une grille d'échantillonnage spécifique de la sphère) induisent un ordre maximum auquel pourra être menée la décomposition en harmoniques sphériques. De même cette grille d'échantillonnage microphonique conditionne la finesse de découpage spatial accessible en changeant la largeur des lobes spatiaux.

Par ailleurs, la grille d'échantillonnage de la sphère de mesure de l'EigenMike® utilisée dans nos travaux est un icosaèdre tronqué. D'autres méthodes d'échantillonnage de la sphère existent pour la formation de voie, notamment la grille d'hyperinterpolation se basant sur la théorie des systèmes extrêmes [SW03]. Le détail des méthodes d'échantillonnage de la sphère est disponible en Annexe section 8.2 (cf. également chapitre 2.2 et 2.3 [Col11].





#### 2.5 Objectifs du stage

L'objectif du stage est donc d'évaluer les deux approches proposées, à savoir celle de la formation de voies et de l'ambisonics. Nous nous intéresserons en particulier aux ordres de reproductions des différentes sections temporelles de la réponse impulsionnelle en prenant comme référence l'ordre ambisonics maximum autorisé par l'EigenMike®.

L'idée sera d'utiliser ce formalisme ambisonics afin de le relier à l'approche purement paramétrique (reposant sur le formalisme du Spatialisateur) et d'aborder l'approche hybride. En effet, l'équipe a pour volonté à plus long terme de vouloir manipuler paramétriquement les DRIR<sup>10</sup>. En essayant de combler les lacunes des méthodes actuelles (DirAC et SIRR), il s'agira donc de créer un moteur de convolution permettant de restituer en temps-réel les caractéristiques spatiales des réponses impulsionnelles enregistrées par microphone sphérique avec un haut niveau de définition.

Une étude récente [AAG<sup>+</sup>13] à montré que de nombreux paramètres utilisés dans le domaine des harmoniques sphériques ont une influence sur la perception que l'on a du champ sonore reconstitué. Cette article effectue une reconstitution totale des sections temporelles de la réponse de la salle et ceci ne permet pas de distinguer les éléments de la RIR responsables de jugements perceptifs des participants. De plus, leur méthode étant basée sur une expérimentation à l'aide d'une restitution binaurale, nous avons souhaité mener nos recherches en utilisant des méthodes similaires mais avec les outils propres de l'équipe EAC et en étendant le test perceptif à une configuration réelle dans un studio avec un dôme de reproduction ambisonics.

<sup>&</sup>lt;sup>10</sup> *Directional Room Impulse Response* - Réponse impulsionnelle de salle présentant les caractéristiques directionnelles spatiales des ondes acoustiques de pression arrivant au niveau du microphone.



## 3 Capture de champs sonores par microphone sphérique

Le prochain chapitre décrit l'étape d'analyse de champs diffus dans le cadre de mesures par microphone sphérique. Elle se base sur l'étude des paramètres calculés à partir de l'analyse d'*Energy Decay Relief* (définition donnée section suivante) et de la cohérence spatiale entre les différentes capsules. L'étude des réponses impulsionnelles spatiales captées par le microphone sphérique passe en premier lieu par l'observation de chacune des sections temporelles de la réponse. Ces sections correspondent au trajet des ondes entre la source (haut-parleur) et le récepteur (microphone sphérique) dans un espace clos. De plus, le phénomène de réflexion des ondes et des directions d'arrivée sur le récepteur peut également être simulé à partir de la décomposition en onde plane du signal capté. Cette décomposition est utile pour l'analyse du champs diffus ainsi que la reconstruction du signal débruité qui sera plus tard transformé et encodé au divers ordres ambisonics.

#### 3.1 Caractéristiques d'un champ réverbérant



**Figure 4:** Réponse impulsionnelle temporelle simplifiée d'un espace réverbérant et sa distribution temps-fréquence Modèle implémenté dans *Spat* pour le contrôle paramétrique d'une salle [CSNW13]

Lorsqu'un son se propage dans un espace fermé, l'auditeur positionné dans la salle est confronté à plusieurs événements liés au front d'onde acoustique. Pour simplifier la représentation (influence de l'humidité, de la température), le chemin entre la source et le récepteur est considéré comme un système linéaire et invariant dans le temps, entièrement caractérisé par sa réponse impulsionnelle.

Comme on peut le voir Fig. 4, cette réponse est généralement décomposée en trois ou quatre sections temporelles distinctes: le premier évènement est celui du son direct, puis on considère les premières réflexions provenant des surfaces alentours; vient enfin la réverbération tardive ou queue de réverbération, perçue comme un son diffus avec une enveloppe décroissante. Ces trois sections sont évidemment conditionnées par les paramètres acoustiques de la salle et vont influencer la réponse impulsionnelle suivant de nombreux paramètres tels que la position de la source, son orientation ou sa directivité, la géométrie de la salle et les conditions de propagation. Si on considère l'audition humaine, des facteurs psycho-acoustiques et perceptifs sont évidemment à prendre en compte pour le récepteur et des relations entre paramètres physiques et qualités subjectives doivent être étudiées. Les prochaines sections présentent en détail les termes relatifs aux sections temporelles des RIR<sup>11</sup> et effectuent une confrontation entre divers critères acoustiques objectifs et leur correspondance en termes perceptifs.

<sup>&</sup>lt;sup>11</sup> *Room Impulse Response* - Réponse Impulsionnelle de Salle





#### **Onde** Directe

Le son direct est le premier front d'onde qui atteint l'auditeur. Il est équivalent à une onde sonore se propageant en champ libre sans conditions limites de réflexion. Si celui-ci est plus faible ou possède un timbre coloré différemment de la queue de réverbération, la source est perçue comme étant placée derrière un obstacle. Pour être plus précis, le timbre du son direct est la plupart du temps différent de celui de la queue de réverbération. Mais en général il contient plus d'aigus (compte tenu de l'absorption des parois, et de l'absorption en très haute fréquence au cours de la propagation dans l'air). C'est donc l'effet de filtrage passe-bas (absence d'aigus) dans le son direct qui indique que la source est masquée par un obstacle.

Le logiciel *Spat* offre la possibilité d'agir sur la direction de la source par rapport au récepteur. L'intensité et la différence temporelle entre le son direct et les réflexions qui suivent donnent une indication à l'auditeur de sa distance à la source et exercent une grande influence sur la perception que l'on a de la taille de la salle.

#### **Réflexions Précoces**

Les réflexions précoces désignent un régime de réflexions discrètes provenant des murs environnants. Les différences de timbre et de dynamique avec le son direct sont déterminées par les matériaux qui composent l'espace. De manière générale on considère que ces premières réflexions arrivent entre 20ms et 80ms [Jot99]. Ces réflexions jouent un rôle sur l'effet de distance apparente ainsi que de la présence de la source et de sa largeur apparente [Ste06] [Kah95].

#### **Réflexions Tardives**

Le champ de réverbération tardif est la partie diffuse de la réponse impulsionnelle. On peut considérer un "paquet" d'ondes distinct de R1 (précoce) que l'on nommera R2 et faisant partie du champ diffus. Puis la queue de réverbération (R3) avec son enveloppe temporelle décroissante [CSNW13]. D'un point de vue perceptif cette section a un effet direct sur l'impression de la taille de la salle.

#### Détermination des limites des sections temporelles

Même si nous avons étudié les temps moyens d'arrivée des différentes sections temporelles et que nous avons fixé ces paramètres pour la génération des stimuli de notre étude, il est possible d'utiliser des méthodes de détection automatique à l'aide de méthode par traitement du signal. Dans le cadre de l'analyse des salles de concert, la frontière entre réflexions précoces susceptibles de s'intégrer au son direct et réverbération tardive est généralement considérée autour de 80ms, comme pour l'indice de Clarté C80. Les études subjectives rapportées par E.Kahle [Kah95] et exploitées dans le *Spat* considèrent une transition à 100ms [Pol93]. Cependant d'un point de vue objectif, la transition vers un régime diffus dépend du volume de la salle. En l'absence de connaissance sur la géométrie de la salle il peut donc être intéressant de détecter cette transition directement sur le signal de la réponse impulsionnelle.





#### Approches statistiques : Densité d'échos ou Kurtosis

La zone des réflexions tardives suit une loi normale et correspond à une distribution gaussienne des échantillons. Si on affiche un histogramme des valeurs des échantillons (valeurs situées entre 0 et 1) et que l'on regarde leur distribution temporelle, ils suivront au bout d'un moment une forme de distribution gaussienne caractéristique du champ diffus qui équivaut à une répartition aléatoire des échantillons. Le principe de la mesure donné par [AH06] est de compter les échantillons de la réponse impulsionnelle qui ont des valeurs supérieures à l'écart-type de la fenêtre temporelle considérée et de la normaliser par la valeur de l'écart-type attendue pour une fenêtre de même taille d'un bruit Gaussien centré de moyenne nulle.

$$\eta(t) = \frac{1}{erfc(1/\sqrt{2})} \sum_{\tau = t-\delta}^{t+\delta} w(\tau) \, \mathbf{1}\{ |h(\tau)| > \sigma \}$$
(2)

où  $\eta(t)$  représente le profil de densité d'échos. La fonction  $erfc(1/\sqrt{2}) = 0.3173$  est la valeur de seuil pour les échantillons hors de la valeur d'écart-type de la fenêtre de distribution Gaussienne. 1{.} est la fonction caractéristique donnant 1 quand son argument est vrai et 0 ailleurs. w(t) une fenêtre temporelle de pondération préservant les valeurs non-nulles sur les bords et de somme unité.  $\delta$  la taille de la fenêtre, h(t) la réponse impulsionnelle et l'écart-type calculé comme :

$$\sigma = \left[\sum_{\tau = t-\delta}^{t+\delta} w(\tau) h^2(\tau)\right]^{\frac{1}{2}}$$
(3)

Ainsi en fixant un seuil de détection nous pouvons déterminer la zone temporelle où commence le champ diffus et les sections R2-R3 correspondantes. La figure 5 présente une illustration du profil Gaussien de la queue de réverbération en affichant sous la réponse impulsionnelle d'une salle l'histogramme de distribution des échantillons au cours du temps.

Une approche alternative proposée par ( [Ste06], [SS07] ) consiste à évaluer le Kurtosis (moment statistique d'ordre 4) pour mesurer la disposition des masses de probabilité autour de leur centre donné par l'espérance mathématique. En d'autres termes elle représente le regroupement des échantillons proches ou lointains du centre de probabilité [Ste06], [SS07]. Dans le cas de réponses impulsionnelles captées par un microphone sphérique, le calcul de densité d'écho ou de Kurtosis peut - par exemple - être conduit sur le canal W véhiculant l'information omnidirectionnelle.





Figure 5: Illustration du profil Gaussien de la queue de réverbération

La courbe en rouge du haut représente la réponse impulsionnelle mesurée (temps en échelle logarithmique). Les points bleus donnent le centre de la fenêtre temporelle utilisée pour les histogrammes du bas. Sur les histogrammes du bas, les courbes vertes donnent une idée du profil Gaussien de la distribution et les points rouges indiquent l'occurrence des différentes valeurs des points mesurées sur la réponse impulsionnelle [AB04]

<u>Approche "spatiale" par intensimétrie :</u> Pulkki et Ahonen (2009) ont fournis une méthode d'estimation de la partie diffuse du champ sonore en analysant la variation temporelle du vecteur intensité. En calculant le vecteur intensité à partir du calcul de densité d'énergie ils arrivent à estimer les instants correspondants aux réflexions précoces et au champs diffus. En effet, la direction d'arrivée du premier front d'onde est censée être très localisée, tandis que le champs diffus présente une répartition aléatoire des directions du vecteur intensité (cf. Figure 6).



Figure 6: Vecteur intensité en champs diffus (gauche) et non-diffus (droite)
[AP09]





#### Autre(s) approche(s) :

Hormis les méthodes statistiques et approches spatiales de distinction du champ diffus, certaines étude ont abordées des méthodes liées à la relation entre la source du signal représentée par l'onde directe et sa corrélation avec le reste de la réponse impulsionnelle. Ont peu notamment citer la méthode numériques de "matching pursuit" : [DDP08], [DDP09].

#### Energy Decay Curve - Courbe de Décroissance Energétique

Schroeder a développé une méthode pour estimer le Temps de Réverbération d'une salle [Sch54]. Le calcul de l'énergie restante dans la réponse impulsionnelle après le temps t permet d'estimer le temps de réverbération en effectuant une régression linéaire des moindres carrés entre deux points sur la courbe de décroissance énergétique ( $EDC_h$ ), avec h(t) la réponse impulsionnelle.

$$EDC_{h}(t) = \int_{t}^{\infty} h^{2}(\tau)d\tau$$
(4)

#### Energy Decay Relief - Relief de Décroissance Energétique

L'analyse EDC de Schroeder peut être définie sur la base de représentation temps-fréquence  $\rho_h(t, f)$  obtenue de la réponse h(t) par la fonction de distribution temps-fréquence  $\rho$ . Une représentation spectrale fine de l'EDC peut ainsi être obtenue par calcul de Transformée de Fourier à Court Terme<sup>12</sup>. L'analyse EDR est donc une extension multibandes de l'EDC.

D'après [WJC97], si  $\rho(\tau, f)$  est une représentation temps-fréquence-énergie du signal h(t), l'EDR est défini par :

$$EDR_{h}(t,f) = \int_{t}^{\infty} \rho^{2}(\tau,f)d\tau$$
(5)

Des artefacts de bruit peuvent être présents dans des mesures de réponses impulsionnelles de salle. En extraire les paramètres de Temps de Réverbération permet d'obtenir des résultats plus aptes à être analysés et dont la composante de bruit à été filtrée. L'enveloppe de la réponse impulsionnelle étant exponentiellement décroissante, nous pouvons déterminer facilement le spectre de puissance initial ainsi que le temps de réverbération en effectuant un ajustement de courbe par régression linéaire sur la pente du signal EDR affiché en échelle logarithmique.

<sup>&</sup>lt;sup>12</sup> STFT - Short Time Fourier Transform







L'équation de la droite permettant d'estimer la pente de décroissance énergétique est donnée par la relation suivante dans le domaine fréquentiel :

$$P(f) = ENV_h(0, f) = \frac{6\ln(10)}{Tr(f)} EDR_h(0, f) = \frac{6\ln(10)}{Tr(f)} |H(f)|^2$$
(6)

avec P(f) le spectre de puissance initial, Tr(f) le temps de réverbération et H(f) la réponse en fréquence du signal h(t). Estimer les valeurs de Tr(f) et P(f) nous permet d'extrapoler la courbe de décroissance énergétique au-delà du bruit de mesure, dont on a calculé la limite temporelle  $T_{lim}$  par un calcul de rapport signal/bruit. Le but de cette méthode est de filtrer le signal sur une plus grande longueur et de réduire ainsi le bruit de mesure afin d'estimer les valeurs de Tr et P(f) de manière plus robuste.





#### 3.2 Analyse de DRIR dans le domaine des harmoniques sphérique

Après avoir présenté les caractéristiques propres à chacune des sections temporelles et les différentes caractéristiques d'un champ réverbérant, nous pouvons nous concentrer sur l'analyse spatiale de ces réponses impulsionnelles. La méthode de décomposition en ondes planes combinée à la décomposition dans le domaine de Fourier spatial, offre des outils intéressants pour la visualisation des divers réflexions dans le dodomaine spatio-temporel.

#### 3.2.1 Décomposition en ondes planes

La décomposition en ondes planes (PWD<sup>13</sup>) facilite l'analyse de champs sonores mesurés par des microphones sphériques. Elle résulte de la fonction de densité d'amplitude spatiale des ondes planes composant le champ sonore, décrite comme suit [Raf04], [CSNW13]:

$$\Delta(k,\theta,\varphi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{1}{H_n(k,r,a)_{|r=a}} \left( \int_0^{2\pi} \int_0^{\pi} p(k,a,\theta,\varphi) Y_n^m(\theta,\varphi)^* \sin\theta \ d\theta \ d\varphi \right) Y_n^m(\theta,\varphi)$$
(7)

Avec  $H_n$  la fonction holographique (cf. explication détaillée chapitre 4) permettant de prendre en compte la diffraction des ondes au niveau de la sphère,  $Y_n^m$  l'harmonique sphérique d'ordre n et de degré m.  $p(k, a, \theta, \varphi)$  est la pression acoustique sur la sphère de rayon a.

En conséquence, l'échantillonnage spatial de la pression au niveau de la sphère composée de *C* capsules microphoniques, entraîne une approximation de l'intégrale par une somme discrète où l'on tronque l'ordre  $n \rightarrow \infty$  par un ordre fini  $n \leq N$ . Ceci limite la résolution spatiale angulaire de la décomposition, où la largeur du lobe de directivité de la direction considérée dépend de l'ordre N tel que [AAG<sup>+</sup>13] :

$$\Theta \approx \frac{\pi}{N} \left[ rad.s^{-1} \right] \tag{8}$$

Nous pouvons voir ci-dessous (Figure 7) la visualisation d'une réponse impulsionnelle de salle directionnelle obtenue par PWD. Il s'agit d'un enregistrement par microphone sphérique dans l'Espro, où la PWD est calculée sur des trames temporelles successives (avec recouvrement). Les données sont représentées sur le plan azimutal. La fréquence observée est 3500Hz. On peut clairement identifier les sections temporelles, à commencer par l'onde directe arrivant de l'angle 0° au temps t = 0. On peut également repérer les premières réflexions arrivant du sol au bout de quelques ms et des réflexions latérales importantes arrivant à  $\pm 50^{\circ}$  au bout de  $t \approx 20ms$ . Enfin, le champ sonore devient progressivement diffus au bout de 100ms.

<sup>&</sup>lt;sup>13</sup> Plane Wave Decomposition

<sup>©</sup>Ircam. EAC. - Caractérisation perceptive de champ sonore à résolution spatiale variable - Alexandre Panizzoli - ATIAM 2013









**Figure 7:** Exemple de représentation espace-temps obtenue à partir de décomposition en onde plane d'un enregistrement avec EigenMike®

[CSNW13]

#### 3.2.2 Phénomènes limitants la représentation exacte du champ sonore

#### **Recouvrement Spatial**

Similaire au recouvrement dans le domaine fréquentiel, l'aliasing ou recouvrement spatial peut se produire lorsqu'un champ sonore est capturé par un nombre fini de capsules microphoniques sur la sphère, c'est-à-dire en utilisant un échantillonnage discret pour les points de mesure de pression sur la sphère. La contrainte sur l'ordre N ambisonics maximum que l'on peut atteindre, impose un nombre fini de microphones suivant le type d'échantillonnage de la sphère. Nous serons donc limité par le nombre de microphones C tq.  $(N + 1)^2 \leq C$ .

#### Largeur de bande fréquentielle

La figure 8 nous montre que pour un ordre donné n, la fonction  $b_n(kr)$  a la forme d'un filtre passe-bande (en excluant l'ordre 0 qui se comporte comme un filtre passe-bas), avec un pic autour de la valeur  $N \approx kr$ .  $b_n$  (noté également  $H_n$ ) est la fonction holographique définissant la diffraction d'une onde sur une sphère tel que:

$$b_{n}(k,r,a) = H_{n}(k,r,a) = 4\pi i^{n} \begin{cases} j_{n}(kr) & \text{sphère ouverte} \\ \left(j_{n}(kr) - \frac{j_{n}^{'}(ka)}{h_{n}^{'}(ka)}h_{n}(kr)\right) & \text{sphère rigide} \end{cases}$$
(9)

où  $h_n$  et  $j_n$  représentent les fonctions de Hankel et de Bessel de premier type et leur dérivées (dans l'expression pour la sphère rigide), a le rayon de la sphère,  $r_s$  le rayon d'une source omnidirectionnelle, k le nombre d'onde et r le rayon de mesure entre la source et la sphère.



#### UPPC SORBONNE UNIVERSITÉS



Le comportement de cette fonction risque de créer un changement de timbre entre les différents ordres d'harmoniques sphériques et donc de changer la "couleur" des signaux synthétisés pour un ordre d'encodage différent. Lors des tests de pré-écoute, nous avons effectivement pu repérer ce changement de "couleur" qui pouvait varier suivant l'ordre d'encodage et entrainer un changement de timbre pour un même stimulus.









#### 3.3 Protocole de Mesure

Pour cette étude basée sur des réponses impulsionnelles enregistrées par microphone sphérique<sup>14</sup> le matériel utilisé était un microphone **Eigenmike (REM32)**. Il est composé de 32 capsules omnidirectionnelles, placées à la surface d'un icosahèdre tronqué (faisant partie des solides de platon autorisant un échantillonnage homogène de la sphère). Son rayon est de 4.2 cm est il est équipé de pré-amplificateurs et de convertisseurs analogique-numérique. Etant donné que le degré d'encodage ambisonics est dépendant du nombre de capsules, l'ordre maximum possible pour la restitution est l'ordre 4.

Lors des mesures, le microphone était placé à l'origine du système de coordonnée sphérique du dôme ambisonics de l'ESPRO. Les réponses ont été mesurées pour 3 configurations de salles: réfléchissant diffus (tous les murs sont positionnés en configuration avec des panneaux réfléchissants), mur sud réfléchissant (les autres panneaux en config. absorbante) et configuration avec des panneaux absorbants. Au lieu d'enregistrer des stimuli par impulsion, les réponses impulsionnelles spatiales provenant des 75 directions du système de haut-parleurs et ont été enregistrées à l'aide d'un sweep exponentiel glissant de manière croissante d'une longueur de 3 secondes à la fréquence d'échantillonnage de 44.1 kHz. Le sweep utilisé pour la déconvolution est par ailleurs incertain et ne correspond peut-être pas à l'original et il est possible qu'il soit la source du problème visible figure 9 dans la partie diffuse. Ce problème n'est finalement pas majeur car le débruitage intervient dans cette zone temporelle et a supprimé cet artefact.



**Figure 9:** Artefact rémanent sur la réponse impulsionnelle après déconvolution Il s'agit de la réponse du canal 1 du microphone en configuration diffuse réfléchissant. La mesure complète n'est pas présentée et on observe la réponse seulement sur une partie de la section R3.

<sup>&</sup>lt;sup>14</sup> SMA - Spherical Microphone Array





#### 3.4 Vérification des données expérimentales

Les réponses impulsionnelles de l'ESPRO fournies provenaient d'une base de données pré-existante et les mesures n'ont pas été effectuées au cours du stage. Elles proviennent d'une campagne de mesures prises au cours de l'année 2012 et elles étaient accompagnées de mesures effectuées à la Cité de la Musique (celles-ci n'ont pas été utilisées pour notre projet). Ces mesures avaient tout de même des problèmes liées au bruit de mesure de l'EigenMike fournis par Orange Labs (Lanion). Nous avons donc souhaité en premier lieu vérifier si elles étaient exploitables avant de se lancer dans des analyses plus poussées. La prochaine partie décrit cette étape d'étude de variance entre les capsules du microphone sphérique ainsi qu'entre les mesures sur les différentes configurations de salle.

#### 3.4.1 Variance des capsules

#### Sur une configuration de salle

Une des premières analyse des réponses consistait à étudier l'écart de différentes valeurs mesurées ou calculées à partir des 32 canaux du microphone sphérique pour une configuration précise d'enregistrement dans l'*Espro* (Configuration absorbante, Mur sud réfléchissant, Mode réverbérant). Ces valeurs, calculées à partir de l'analyse de l'*Energy Decay Relief* (EDR) correspondent au Po (le spectre de source initial sans perte d'énergie), au TR (temps de réverbération), ainsi qu'au T-Lim (temps à partir duquel le bruit de mesure devient prédominent par rapport à la mesure).

Le but de cette étude étant de voir s'il existe des différences systématiques pour une capsule donnée, sur les 75 réponses impulsionnelles correspondant aux différents haut-parleurs placés dans l'*Espro* et dans trois configurations de salle. Le débruitage conjoint de toutes ces réponses reposera donc sur cette analyse afin d'estimer au mieux la pente d'EDR et le champs diffus virtuel à synthétiser pour combler au mieux les défauts de mesures. On pourra également envisager différentes solutions applicables pour les prochaines mesures soit en recalibrant les microphones, soit en prenant en compte leur biais lors des futures analyses.

La Figure 10 nous montre les valeurs des trois caractéristiques Po, TLim et RT pour chacune des 32 capsules du microphone sphérique (EigenMike®) sur l'échelle des fréquences et pour une réponse impulsionnelle provenant du premier haut-parleur en configuration réfléchissante (pour le Po la valeur est en dB). On peut déjà repérer visuellement que l'écart de ces valeurs est relativement faible entre les différents microphones répartis sur la sphère.







(a) Po des 32 capsules pour le HP1



(b) RT des 32 capsules pour le HP1

(c) TLim des 32 capsules pour le HP1

Figure 10: Paramètres EDR pour l'analyse des RiR en configuration Réfléchissante (HP 1) Les valeurs pour les fréquences supérieures à 18kHz ne sont pas exploitables à cause du SNR trop important. Bien qu'elles influent sur l'affichage, une étude plus fine des différents indices présentés ici a montré que les écarts sur les fréquences inférieures n'étaient pas significatifs et n'étaient pas dommageables pour l'exploitation des données.





#### 3.4.2 Comportement global

On continue par la suite ce genre de représentation en effectuant un calcul de *ratio* entre une capsule spécifique du SMA et toutes les réponses impulsionnelles. On étend donc nos calculs sur les IR provenant des 75 positions des haut-parleurs du dôme de reproduction ambisonics et pour toutes les configurations de salles. Cela correspond donc à 3x75x32 réponses impulsionnelles à analyser et à représenter sur l'échelle des fréquences.

De fait, pour calculer le rapport d'un micro donné à la moyenne globale, on représente d'abord cette valeur moyenne pour vérifier qu'elle présente des valeurs cohérentes suivant la configuration de salle donnée. On observe donc figure 11 un accroissement significatif des valeurs de TLim et Tr en configuration diffuse par rapport aux configurations absorbante et mur sud réfléchissant. Les données affichées semblent par ailleurs correspondre aux temps de réverbération donnés pour l'Espro avec des valeurs inférieures à 1 seconde en configuration absorbantes et supérieures à 2 sec en mode réverbérant. Les valeurs de Po sont quant à elles à peu près égales (écarts inférieurs à 2 dB) puisqu'elles sont caractéristiques de la source. On retrouve ce faible écart pour toutes les configurations de salles, et les différences visibles seront donc plutôt liées au biais des haut-parleurs.

La figure 12 donne un comportement global d'une capsule pour les trois configurations et tous les hautparleurs. On se concentre sur l'étude du Po pour lequel on s'attend à ce que les valeurs calculées restent à peu près constantes, puisqu'il représente la puissance initiale de la source. On remarque que pour certaines positions de haut-parleur des valeurs deviennent assez extrêmes autour de 8kHz. La figure 29 (en Annexe section 8.3) nous confirme que certains enregistrements ont des valeurs extrêmes qui risquent de poser problème lors du débruitage de toutes les réponses.

En conclusion, on pourra donc envisager une correction basée sur les paramètres moyens de  $T_{lim}$ ,  $P_0$  et TR afin de réduire le niveau de variance pour éviter le biais des capsules et également de calibrer le niveau des haut-parleurs suivant leur position sur le dôme ambisonics.







(b) TR Moyen Global

(c) TLim Moyen Global









(a) Ratio du Po sur le Po Moyen Global; vue Fréquence-Haut-parleur pour la capsule 1; Configurations: Réfléchissante (valeurs de 1 à 75), Mur Sud Réfléchissant (de 75 à 150), Absorbante (de 151 à 225).



(b) Ratio du Po sur le Po Moyen Global; vue Fréquence-Valeur du Po pour la capsule 1

Figure 12: Ratio par rapport à la valeur moyenne globale du Po pour un microphone donné





#### 3.5 Débruitage des mesures

#### 3.5.1 Cas monophonique

Les réponses impulsionnelles mesurées à partir du microphone sphérique ne peuvent pas être utilisés directement pour une auralisation dans le domaine ambisonics. En effet, le niveau de bruit de mesure est trop important et effectuer une convolution avec ces signaux engendrerait des artefacts audibles notamment sur la queue de réverbération.

Une méthode de restauration pour supprimer le bruit de mesure des RIR a donc précédemment été proposée par JM. Jot [Jot92] au sein de l'équipe EAC. Il s'agissait de substituer le bruit indésirable par un bruit blanc gaussien filtré par une enveloppe spectro-temporelle. Le profil de cette enveloppe étant déterminé de sorte à prolonger le profil naturel de l'EDR. Le calcul des caractéristiques de l'enveloppe est basé sur l'estimation de la pente EDR présenté dans les chapitres précédents. On modélise donc l'enveloppe du signal en prolongeant les caractéristiques propres de chaque *bin* fréquentiel jusqu'au niveau de bruit, limité cette fois-ci par la résolution numérique de notre signal. Dans notre cas on fait une approximation de quantification uniforme à 6dB/bit pour une résolution de 24 bits soit un bruit de quantification situé à -144 dB par rapport au niveau du signal utile.



Figure 13: EDR avant (à gauche) et après (à droite) débruitage des RIR [CSNW13]





#### 3.5.2 Extension au débruitage de microphone sphérique

#### Calcul de Cohérence Spatiale

Le débruitage actuel repose sur le principe décrit section précédente. Il serait possible de l'appliquer tel quel sur chacune des capsules du microphone sphérique. Cependant ceci ne garantit pas de préserver certaines cohérences (spatio-fréquentielle) entre les capsules. A cet égard nous allons étendre la procédure de débruitage afin de vérifier un critère de cohérence spatiale du champ diffus. Nous utiliserons un bruit cohérent tel que le décrit [Kut00]. En effet, pour un champ de pression diffus idéal, la cohérence spatiale entre deux points  $p_1$  et  $p_2$  séparés d'une distance d s'écrit:

$$\Psi(p_1, p_2) = \frac{\sin(kd)}{kd} \tag{10}$$

Ainsi, nous voyons que les propriétés spatiales doivent mettre en évidence une corrélation entre les différentes capsules du microphone sphérique. Le signal de synthèse produit pour débruiter chacune de ces capsules devra donc respecter ces caractéristiques. Ce champ cohérent est donc reconstruit par une superposition d'ondes planes avec une projection dans le domaine des harmoniques sphériques. Afin de créer ce type de signal nous l'exprimons comme la superposition d'ondes planes provenant de *L* directions, d'amplitude constante unitaire et de phase  $\phi(k)$  aléatoire. La pression sur la sphère de rayon *a* s'exprime donc [Wil99], [CSNW13] :

$$p(k,r,\theta,\varphi)_{|r=a} = \sum_{l=1}^{L} \sum_{n=0}^{N} \sum_{m=-n}^{n} H_n(k,r,a)_{|r=a} Y_n^m(\theta,\varphi) Y_n^m(\theta_l,\varphi_l)^* e^{2i\pi\varphi_l(k)}$$
(11)

La fonction qui simule ce bruit requiert donc un nombre d'ondes planes en entrée dont on à pu déterminer empiriquement le nombre de 400 suffisant. Les directions d'arrivées doivent être uniformément réparties pour préserver l'énergie totale dans le volume considéré. La solution choisie est d'utiliser une grille d'hyperinterpolation d'ordre équivalent au nombre d'ondes planes désirées, soit l'expression suivante: ( $\sqrt{Nombre d'Ondes Planes} - 1$ ).





(a) Cohérence d'un champs diffus synthétisé, décroissant pour une Sphère Ouverte et d'un bruit blanc



(b) Cohérence d'un champs diffus pour une Sphère Rigide (Synthétique, Synthétique Décroissant & Réponse Impulsionnelle) - La courbe théorique est indiquée pour une sphère ouverte





On calcule ensuite la base des harmoniques sphériques correspondant à l'ordre de notre EigenMike® (soit l'ordre 4 correspondant à 25 degrés maximum pour les 32 capsules). En prenant en compte l'effet de diffraction de la sphère par la fonction holographique on peut donc retrouver la pression virtuelle au niveau du microphone par projection des ondes planes (d'amplitude unité et de phase aléatoire) sur chacun des degrés des harmoniques sphériques. Enfin, en effectuant une transformée de Fourier inverse classique, on obtient notre signal temporel de bruit cohérent sur la sphère que l'on peut normaliser par l'énergie de ce même signal. Une fois notre bruit synthétisé, le débruitage de toutes les réponses peut être effectué à partir des paramètres de P0 de TR et de TLim moyens, calculés dans la partie précédente (chap. 3.5.1).

#### Analyse de Cohérence Spatiale

En reprenant la formule (11), le champ de pression considéré est composé d'ondes planes d'amplitudes égales et de phase distribuées aléatoirement. Nous pouvons voir Figure 14 l'analyse du calcul de cohérence en champ libre et d'une sphère rigide. Cette étude va permettre de vérifier d'une part que le champ synthétisé pour le débruitage se comporte conformément à la théorie, et d'autre part que les réponses impulsionnelles enregistrées ont un comportement que l'on peut analyser par comparaison à nos simulations. La première (14(a)) nous montre bien que le champ synthétisé à partir de 400 ondes planes suit la forme théorique du Sinus Cardinal. Ce n'est pas le cas pour un bruit aléatoire centré. Ceci justifie donc l'extension de la méthode de débruitage au cas sphérique plutôt que l'usage d'un débruitage "monophonique" appliqué indépendamment sur chaque capsule. Il est par ailleurs intéressant de noter que l'application d'une enveloppe de décroissance temporelle (courbe verte marquée par le symbole \*) ne modifie pas l'allure des courbes de cohérence spatiale.

La Figure 14(b) représente le calcul de cohérence sur un bruit diffus synthétisé cette fois-ci pour une sphère fermée (ou rigide) qui correspond à notre cas réel de mesure par notre microphone sphérique. Cette simulation pour une sphère fermée est intéressante car la résolution analytique du calcul de cohérence est beaucoup plus complexe que pour une sphère ouverte où l'on prenait seulement 2 points dans l'espace libre mais sans obstacle. Ainsi, lorsque l'on analyse nos réponses impulsionnelles (partie décroissante, non débruitée, des RiR), on se rend compte que l'allure en sinus cardinal n'est pas valide pour une sphère rigide (superposition des courbes verte et rouge). A la surface d'une telle sphère, la cohérence spatiale décroit beaucoup rapidement. Le champ synthétique simulé sur une sphère rigide (courbe bleu) présente une allure semblable aux mesures (courbe rouge). Cela nous porte à croire que l'approche consistant à simuler un champ diffus par superposition d'un grand nombre d'ondes planes est légitime. Par la suite, cette approche sera retenue pour réaliser l'opération de débruitage des RIRs.







Figure 15: Cohérence partie diffuse d'une RiR (comparaison des microphones Adjacents et Opposés sur la sphère)

De plus, la figure 15 nous montre la comparaison de ce même calcul de cohérence sur des RiR mesurées par l'EigenMike, mais pour des capteurs positionnés différemment sur la sphère. Sur la courbe du haut est tracé le calcul de la fonction de cohérence pour un couple de capsules adjacentes (1 & 2), tandis que celle du bas marque le même calcul mais pour des capsules opposées sur la sphère (capsules 1 & 25). On superpose la courbe théorique pour les deux, où le maximum de la fonction est différent dans chaque cas, puisque la distance *d* est inférieure pour la courbe du haut (capsules adjacentes). On observe donc que la corrélation est plus forte pour deux capsules proches, ce qui semble suivre notre intuition pour le cas de points de mesure situés sur une sphère rigide, où des microphones éloignés sur la sphère sont certainement plus décorrélés.

Ces hypothèses restent tout de même à discuter avec la résolution analytique du calcul qui n'est - par ailleurs - pas un calcul trivial et n'a pas été trouvé dans la littérature parcouru pour ce travail. La courbe théorique de cohérence spatiale pour une sphère rigide n'étant pas connue, nous avons donc pu observer que la décroissance de la courbe était beaucoup plus rapide pour les simulations (et les mesures) avec une sphère rigide. Ceci confirme que le comportement spatial d'un champ diffus enregistré par microphone sphérique (rigide) et deux points de l'espace pris en champ libre n'ont pas le même comportement.





### 4 Analyse-Synthèse de réponses impulsionnelles

Le prochain chapitre décrit de manière détaillée les techniques employées pour le développement des méthodes de synthèse de signaux dans le domaine ambisonics suivant nos deux approches. Ceci nous permettra en amont du test perceptif, d'effectuer une évaluation objective des critères d'influence de l'ordre ambisonics ou du nombre de voies pour les différentes sections des Réponses Impulsionnelles mesurées dans *l'Espace de Projection* et la *Cité de la Musique*.

#### 4.1 Approche par transcodage HOA

Le transcodage est l'étape permettant d'effectuer les opérations partant du signal temporel débruité et de les passer dans le domaine des harmoniques sphériques selon l'ordre ambisonics désiré. La technologie ambisonics est basée sur la solution de l'équation d'onde en coordonnées sphériques [Col11]. En tout point de l'espace, la pression acoustique peut être décrite par une décomposition de Fourier-Bessel où apparaissent les fonctions harmoniques sphériques  $Y_n^m$  [DM04]. Le détail du calcul du passage dans le domaine des Harmoniques Sphériques est donné en Annexe (section 8.1), mais nous donnons ici l'expression de la décomposition du champ de pression sur la base des harmoniques sphériques pour des ondes propagatives:

$$p(r,\theta,\varphi,k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} (a_n^m(k)h_n^{(1)}(kr) + b_n^m(k)h_n^{(2)}(kr))Y_n^m(\theta,\varphi)$$
(12)

où *n* et *m* représentent l'ordre et le degré des harmoniques sphériques  $a_n^m$  et  $b_n^m$  les coefficients spatiaux de la décomposition du champ sonore, que l'on peut comparer aux coefficients des séries de Fourier et  $h_n^{(1),(2)}$  respectivement les fonctions sphériques de Hankel de première et seconde espèce; et  $Y_n^m$  sont les harmoniques sphériques définis tels que :

$$Y_n^m(\theta,\varphi) = \sqrt{\frac{(2n+1)}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta) e^{im\varphi}$$
(13)

avec  $P_n^m$  la fonction de Legendre associée,  $\theta$  l'angle d'élévation (ou site) et  $\varphi$  l'azimut.

Ainsi, en effectuant des opérations d'encodage spécifiques sur chacune des sections temporelles des RIR, nous pourrons par exemple intervenir indépendamment sur l'onde directe et les réflexions précoces en utilisant diverses méthodes d'égalisation pouvant être décrites et contrôlées par des paramètres de signaux, des paramètres physiques et acoustiques ou bien par des paramètres de haut niveau relatifs à la perception que l'on a de la salle.

Dans le cadre de cette étude nous avons utilisé une méthode d'encodage où nous agirons seulement sur des paramètres liés au signal; ces paramètres sont fixés dans un script matlab et sont donc seulement utiles pour préparer les stimuli avec un traitement optimisé. Un paramètre important du transcodage est celui





du *soft-limiting* à appliquer pour une compensation du filtre modal radial lié à la sphère de mesure tel que nous l'avons vu pour la fonction holographique.

Afin de compenser l'effet de diffraction sur la sphère nous devons donc calculer l'inverse de ces filtres holographiques en ajoutant une compensation selon un seuil  $\alpha = 10^{\alpha_{dB}/20}$  dans l'équation suivante [NC13]:

$$d_n(k,r) = \frac{1}{b_n(k,r)} = \frac{2\alpha}{\pi} \frac{|a_n(k,r)|}{b_n(k,r)} \arctan\left(\frac{\pi}{2\alpha |b_n(k,r)|}\right)$$
(14)

avec  $b_n(k,r) = H_n(k,r,a,r_s)$  la fonction holographique définie section 3.2.2. On remarque que le softlimiting est appliqué de manière proportionnelle suivant l'ordre de la décomposition et de manière plus importante pour les ordres élevés (cf Fig. 16). De plus, le début de nos réponses mesurées par le microphone sphérique restent bruitées. On utilise donc cette méthode de soft-limiting pour palier ce mauvais SNR.



**Figure 16:** Diagramme d'amplitude de filtres modal radial inverse Tracé pour une sphère fermée avec soft-limiting à 30 dB [NC13]

Ces filtres auront tendance à amplifier le bruit en basse fréquence. Ce paramètre a été fixé à 20dB en cherchant à obtenir une perception claire sur les stimuli écoutés, c'est-à-dire ayant un bon équilibre à l'écoute entre les hautes-fréquences et les basses fréquences. Cette opération vise à limiter l'effet de suramplification des basses fréquences pour des ordres élevés, ce qui peut être problématique en fonction du SNR des données. Elle entraîne en outre une perte de résolution spatiale en basses-fréquences.







Il faut noter qu'une fois les réponses transcodées et convoluées avec les stimuli il persiste un problème lié à la queue de réverbération, sur la partie débruitée avec un signal cohérent sur la sphère. On peut voir figure 17 qu'il reste un "gap" autour de 1,5 sec. et qui reste encore inexpliqué puisque lors de la partie débruitage les réponses impulsionnelles temporelles ne montraient pas ce défaut. Des travaux sont actuellement en cours pour rechercher la source de ce problème qui peut être variable (normalisation des IR insuffisante, bruit cohérent trop décorrélé de la réponse réelle, effet concentré sur certaine fréquences). Ce gap étant présent sur toutes les réponses (plus important dans les ordres HOA et Beams élevés) et situé aux alentours de -100dB il n'est toutefois pas audible et n'aura pas d'incidence sur le test perceptif.



Figure 17: Gap apparaissant sur la partie débruitée cohérente après transcodage HOA des IR (canal 25)





#### 4.2 Approche par formation de voies

La méthode de formation de voies dans le domaine des harmoniques sphériques permet d'améliorer le signal d'un champ sonore capturé pour une direction d'arrivée spécifique tout en atténuant le signal provenant des autres directions. Ce filtre spatial peut-être utile pour capter les directions d'arrivée des réflexions au niveau d'un microphone sphérique et permettre ainsi de déterminer les zones de transitions entre les différentes sections temporelles des réponses impulsionnelles.

La pression à la surface d'une sphère est pondérée par une fonction  $w(k, \theta, \varphi)$  produisant ainsi un signal de sortie pour k = w/c donné [ea10], [ME04], [SMKK12] :

$$y(k) = \int_{0}^{2\pi} \int_{0}^{\pi} w(k,\theta,\varphi) p(k,a,\theta,\varphi) \sin\theta d\theta d\varphi$$
(15)

Lorsque l'on exprime la pression  $p(k, a, \theta, \varphi)$  et sa fenêtre de pondération spatiale  $w(k, \theta, \varphi)$  dans le domaine des harmoniques sphériques, on obtient l'expression suivante :

$$y(k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} w_{nm}(k) p_{nm}(k)$$
(16)

Avec  $p_{nm}(k)$  la transformée de Fourier spatiale sphérique du champ de pression pour une onde plane arrivant de la direction ( $\theta_s$ ,  $\varphi_s$ ) exprimée tel que [ea10] [NC13] :

$$p_{nm}(k) = H_n(k, r, a)|_{r=a} Y_n^m(\theta_s, \varphi_s)^*$$
(17)

Où  $H_n(k, r, a)_{|r=a}$  la fonction holographique décrite équation (9) et  $w_{nm}(k)$  la fenêtre de pondération axisymétrique centrée autour de la direction de la voie ( $\theta_s, \varphi_s$ ) donnée par :

$$w_{nm}(k) = \frac{b_n Y_n^m(\theta_s, \varphi_s)}{H_n(k, r, a)_{|r=a}}$$
(18)

où  $b_n$  contrôle la fonction directivité du réseau microphonique. Par exemple,  $b_n = 1, \forall n \leq N$  est une pondération souvent utilisée comme beamformer de décomposition en onde plane. Cette pondération doit prendre en compte deux phénomènes: la divergence de la fonction holographique en basse fréquence et l'aliasing spatial en haute fréquence. Ce problème peut être évité en appliquant un filtrage passe-bande dépendant de l'ordre n à la fonction  $w_{nm}(k)$  en sélectionnant les bande fréquentielles proportionnelles à ka(cf. Fig. 18).

Une fois la fonction de pondération calculée pour un type de beam pattern<sup>15</sup>, et les directions de faisceaux calculées dans le domaine modal (cf. Fig 20), les filtres de formation de voies peuvent être dérivés dans le domaine spatial. Ils sont évalués en calculant la transformée de Fourier spatiale inverse exprimée par :

<sup>&</sup>lt;sup>15</sup> *beam pattern* - forme des lobes de directivité (cardioïde, hypercardioïde) changeante suivant l'ordre *N* désiré et dépendante de la fréquence





**Figure 18:** Exemple de lobes spatiaux hypercardioïde d'une formation de voie Pour une sphère de rayon a = 0.042m et N = 4, l'aliasing spatial apparaît au-dessus de 5kHz.

$$w(k,\theta,\varphi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} w_{nm}(k) Y_n^m(\theta,\varphi)^*$$
 [CSNW13], [ea10], [ME04] (19)





#### 4.3 Transcodage binaural

Nous allons présenter deux méthodes de transcodage à partir d'un signal transcodé dans le domaine des harmoniques sphériques. Il s'agit de deux approches différentes utilisant les fonctions de directivité de la tête (HRTF) propres à chaque oreille.

La première présentée par [Dan01] est une méthode de reproduction sur haut-parleur virtuel. Le principe est visible figure 19. Cette technique met en jeu un procédé de convolution entre les signaux ambisonics décodés et les HRTF prisent dans les différentes positions virtuelles des hauts-parleurs.





L'inconvénient de cette approche est qu'elle nécessite une discrétisation spatiale limitée par le nombre de hauts-parleurs. Ceci peut entraîner un coût de calcul très important puisque si l'on souhaite reproduire un champ sonore avec de nombreuses directions (discrètes), nous devrons effectuer un nombre proportionnel de convolutions.

La méthode choisie pour notre test perceptif en binaural repose sur un autre principe. Il s'agit d'une resynthèse dans le domaine modal. Elle repose sur une décomposition des HRTF sur la base des harmoniques sphériques. Nous aurons donc pour chaque oreille une fonction de directivité spatiale prise à partir des mesures de HRTF et décomposée dans le domaine ambisonics. Il s'agit en faite d'un "filtrage" du champ ambisonics par les HRTF. Cette méthode est donc intéressante puisqu'elle part d'une description du champ spatial pour simuler un champ sonore à reproduire sur 2 canaux (propre à chaque oreille).





## 5 Evaluation perceptive des approches

La prochaine partie est consacrée à la comparaison des différentes approches et ordres d'encodage des réponses impulsionnnelles de salle. Ces réponses seront ensuite exploitées pour convoluer différents stimuli sonores. Pour notre test perceptif, nous nous concentrons sur la partie diffuse de l'effet de salle. Ce test vise à déterminer quel ordre d'encodage minimum est requis pour la partie tardive de l'effet de salle. Ceci permettra à plus long terme de trouver le meilleur rapport entre rapidité d'encodage et qualité d'écoute afin de donner les spécifications techniques suffisantes pour manipuler *Spat* en temps-réel à l'aide du logiciel Max-MSP.



#### 5.1 Enjeux du test ABX

**Figure 20:** Visualisation Géométrique des ordres HOA et du nombre de Voies Correspond aux divers ordres utilisés pour le test perceptif Source de la figure du haut : [Dan03]



De nombreuses études ont déjà été menées pour l'étude du champ sonore capturé par microphone sphérique. Les articles faisant références dans le domaine [MP05], [PM06], [Pul07] se sont concentrés sur les méthodes d'analyses de signaux tels que le recouvrement spatial, la largeur de bande et les méthodes de transcodage comme nous l'avons présenté dans les chapitres précédents. Peu d'études ont cependant été présentées sur les caractéristiques perceptives de signaux convolués en ambisonics à partir de ces réponses impulsionnelles [Avn10], [AAG<sup>+</sup>13], [PF12].

L'expérience présentée ici est un test ABX visant à déterminer si des auditeurs sont capables de repérer des différences entre des signaux encodés soit dans le format ambisonics soit par formation de voies. Dans le premier cas il s'agit de déterminer l'ordre minimum nécessaire à une bonne reconstruction du champ diffus, dans le second cas il s'agit de déterminer le nombre de voies minimum. La figure 20 montre un tableau récapitulatif des deux approches utilisées dans notre test Perceptif. On peut y voir les ordres ambisonicss testées avec la visualisation 3D correspondante dans le domaine des harmoniques sphérique. La partie du bas représente la méthode par formation de voies, on peut y voir un lobe de faisceau hypercardioïde à l'ordre 3 et les directions de tir avec le nombre correspondant.

Dans un test aveugle ABX on présente trois stimuli sonores à l'auditeur que l'on nomme successivement A, B et X. Le stimulus A est un échantillon pris aléatoirement dans la base de test, de même pour le stimulus B qui est censé être différent. La tâche pour l'auditeur est d'identifier si le stimulus X correspond au son A (i.e. X = A) ou s'il s'agit du son B (i.e. X = B) [BL09].

Lors de l'analyse des résultats, s'il n'existe aucune différence audible entre A et B, la réponse de l'auditeur devrait être équiprobablement distribué entre "X = A" et "X = B". Ainsi si nous trouvons un résultat de 50% de bonnes (ou mauvaises) réponses pour une paire de son particulière, nous pourrons conclure qu'il n'y a pas de différence "perceptible" entre ces deux échantillons. Il faut cependant se poser la question sur le degré de confiance que nous accordons à ces résultats, à savoir s'il ne s'agit pas de réponses dues au hasard mais bien au fait qu'il est impossible de distinguer deux stimuli particuliers. Nous verrons qu'il existe un indice capable de nous donner cet intervalle de confiance nommé *p-value*.

Dans de nombreuses études, les sujets expérimentaux sont recrutés en tant qu'auditeurs experts et qualifiés. Il s'agit d'auditeurs qui ont été sélectionnés pour leur compétence d'écoute et leur sensibilité aux qualités qui nous intéressent d'étudier dans ce test. Ce choix a été fait pour assurer une cohérence dans les réponses et une constance dans l'appréciation et l'évaluation des stimuli proposés [BL09].

En amont du test et de la création de l'interface, de nombreux stimuli ont été générés correspondant à divers niveaux d'encodage et différents types de stimuli. Une des questions les plus importante était de savoir avec quelle référence nous devions effectuer nos tests et quels types de paires devraient être générés. Après de nombreuses comparaisons et modifications pour obtenir des stimuli normalisés et au timbre sensiblement identique, nous avons choisi de sélectionner le HOA4 comme référence. On rappelle que l'ordre 4 est l'ordre maximum qu'il est possible d'atteindre en utilisant les RIR de l'EigenMike, donc supposé être avec le plus haut niveau de définition et de résolution spatiale.





#### 5.2 Procédure

#### 5.2.1 Préparation des stimuli

#### Normalisation

Les techniques de transcodage n'étant pas optimisées pour la préservation de l'énergie du signal original avant décodage, les différents ordres ambisonicss et différentes formations de voies relatifs à chacun des ordres testés ont dû être normalisés. Le problème de répartition énergétique non équilibré entre les ordres vient probablement du recouvrement temporel en haute fréquence et des lobes secondaires qui ajoutent à l'énergie totale du signal. De plus, le soft-limiting utile en basse-fréquence n'a pas été effectué proportionnellement à l'ordre d'encodage mais est resté le même pour tous les stimuli générés, ce qui a certainement rajouté au phénomène entrainant des différences importante de niveau. La méthode retenue à été de normaliser par l'énergie totale du signal encodé à l'ordre 4, pris comme référence.

D'un point de vue pratique, pour la partie HOA, en amont de la normalisation par l'énergie du signal, une première normalisation sur chacune des sections temporelles avec son ordre d'encodage correspondant est effectuée avec la relation suivante:

$$Section\_HOA(canal_i) = Section\_HOA(canal_i) * \sqrt{\left(\frac{(Ordre\_HOA\_Max+1)^2}{(Ordre\_HOA\_SectionCourante+1)^2}\right)}$$
(20)

Cette normalisation vise à compenser le nombre de composantes intervenant dans la décomposition. De même pour l'approche par formation de voies, comme on utilise toujours des formations de voies d'ordre 4 (donc de même largeur de lobes) on normalise par le nombre de formation de voies.

En pratique cette normalisation théorique ne s'est pas montrée parfaite. Il a donc été décidé de mener une égalisation déduite de la mesure de l'énergie effective des signaux générés par les différentes méthodes d'encodage. Pour cela le bilan énergétique est effectué à partir du décodage binaural permettant ainsi de mesurer ce qui est effectivement transmis à l'auditeur. Enfin, une normalisation par le maximum de tous les signaux est également effectuée ( $Max_All_Signal_HOA$  dans la formule (21)) afin d'éviter la saturation des fichiers. Les réponses sont donc normalisées par la relation suivante :

$$Signal\_HOA_i = Signal\_HOA_i / \left[ \sqrt{\left(\frac{Energie\_Signal\_HOA_i}{Energie\_Signal\_HOA\_ordre4} \right) * Max\_All\_Signal\_HOA} \right]$$
(21)

avec l'indice i = 1, ..., 4 correspondant à l'ordre d'encodage de la section considérée. La référence étant toujours l'ordre 4 pour la formation de voies, cette formule est également valable.

Cette solution nous permettra d'avoir des niveaux équivalents pour les différents stimuli présentés lors du test. En effet, la distinction entre les différents ordres se faisait facilement à cause d'un grande différence de niveau sonore et rendait le test biaisé par des différences sans lien avec la résolution spatiale proposée.

La Figure 21 montre l'énergie présente dans chaque section ambisonics après la première normalisation (équation 20), pour un signal transcodé en ambisonics puis codé en binaural.











En abscisse les 4 premiers correspondent aux signaux HOA, puis les numéros 5 6 et 7 correspondent aux formations de voies avec 4, 9 et 16 directions.

On peut voir sur cette figure que le signal ayant le maximum d'énergie est celui de la section R3 encodée à l'ordre 1 (bar rouge numéro 1). L'onde direct et les premières réflexions ayant une durée très courte, la première normalisation relative à l'ordre d'encodage à été suffisante pour obtenir des valeurs constantes entre les différentes sections. Une fois normalisé, ces différences entre sections (i.e. différence de niveau entre les 3 couleurs de la figure 21) ne sont plus que de quelques dB, tandis que le niveau d'énergie pour R3 devient constant pour tous les types d'encodage.



#### Stimuli proposés

Pour le test perceptif, 28 stimuli différents ont été générés correspondant à deux types de configuration de salle (absorbante et diffuse), deux sources sonores différentes (guitare jouant des accords rythmiques avec des transitoires importants et une voix d'acteur assez claire). Ces sons provenaient d'une base de données interne à l'équipe testée et approuvée pour ce type d'expérience. L'idée initiale était de tester des stimuli avec OD R1 et R2 encodés à l'ordre 4 et différents ordres pour R3, mais trop difficilement discernables entres eux nous avons opté pour un test sur la section R3 seulement car elle contient plus d'information relative aux caractéristiques de la salle et non directe de la source.

Section		Ordre HOA				Nombre de Voies	
R3 Only	1	2	3	4	4	9	16
OD R1 R2 + R3	1	2	3	4	4	9	16

#### Table 1: Récapitulatif des ordres des stimuli de test ABX

Découpage des sections temporelles valables pour les échantillons de guitare de voix en configuration absorbante et réfléchissante

#### 5.2.2 Dispositif expérimental

L'expérience n°1 à été menée dans le studio 1 de l'IRCAM qui possède un dôme de restitution ambisonics avec 24 Haut-Parleurs. Les participants incluaient 5 hommes et une femmes ayant déjà des expériences d'écoute sur ce type d'expérimentation. La seconde expérience en binaural à été menée dans le studio 4 de l'IRCAM à l'aide d'un casque *Sennheiser HD650*. Pour ce deuxième test, deux femmes et 3 hommes ont participé avec les mêmes compétences que précédemment. Au début du test, le texte explicatif suivant leur à été fourni :

Dans cette expérience nous allons vous demander d'écouter successivement 3 stimuli sonores et de les appairer entre eux. Une interface vous est présentée mettant en évidence le stimulus actuellement joué. Les deux premiers stimuli joués seront notés A et B et auront toujours une différence (audible ou non). Le dernier noté X correspondra aléatoirement au son B ou au son A. A vous de noter sur l'interface lequel du stimulus A ou B ressemble au son X. Votre temps de réaction est également noté mais vous pouvez faire une pause une fois votre choix indiqué. Lorsque votre choix est fait vous pouvez passer à l'essai suivant et écouter attentivement les 3 nouveaux stimuli en cliquant sur "Next Trial".

L'expérience a une durée approximative de 30min. Vous devrez écouter 48 exemples comprenant 3 stimuli de 8 secondes chacun. Ces stimuli seront aléatoirement un son de Guitare ou une Voix parlée. Une fois toutes les paires notées l'interface se fermera automatiquement.





ircam
Z Centre
Pompidou

😝 🔿 😁 Bli	indTestABX
Playback-	Choice
Play X	X is A
Play A Play B	X is B
Playing Status	-
A Playing	Valid Choice

Figure 22: Interface du Test ABX communiquant à Spat via OSC et MaxMsp

Avec les besoins de présenter les triplets de stimuli de façon aléatoire, la génération de la liste de lecture était faite à partir d'une interface Matlab qui permettait également de sauvegarder les réponses utiles pour l'étape de l'analyse. Les auditeurs avaient seulement accès à l'interface visible Fig. 22. cette interface communiquait en OSC vers un patch Max utilisant les objets spécifiques à *Spat* offrant la possibilité de décoder les signaux ambisonics et de régler la balance des niveaux vers le système de diffusion en temps réel.





#### 5.3 Analyse des résultats

#### 5.3.1 ANOVA

L'analyse statistique est une étape importante pour les tests psychoacoustiques. Pour une tâche de discrimination comme le test ABX il s'agit d'une épreuve de Bernouilli comportant deux issues: Succès (bonne réponse de l'auditeur sur X) ou échec (mauvaise assignation du stimulus X proposé) [BL09]. Etant donné que nous avons effectué 2 répétitions des mêmes paires pour chaque auditeur et qu'il y a eu 6 participants pour le test en Studio, nous avons moyenné les réponses des auditeurs afin de réduire la dimension de la variable à tester et d'obtenir un score de performance moyen. Du coup ce n'est plus une variable binaire qu'on observe mais une variable continue.

La question principale qui se pose est de savoir si les conditions d'encodage affectent la performance (i.e. la capacité de l'auditeur à reconnaître si X vaut A ou bien B). Mais on souhaite également observer l'influence d'autres facteurs tels que la configuration de la salle (absorbant, réfléchissant), et le choix du type de son (guitare, voix). On veut aussi étudier l'effet de l'interaction des facteurs entre eux, par exemple si le type d'encodage lié à la salle ont une influence sur les réponses.

L'analyse de la variance<sup>16</sup> est un test statistique permettant de vérifier l'effet de plusieurs facteurs une variable continue [How02] [Gel08]. On a donc conduit une ANOVA à 3 facteurs sur notre variable "score de performance moyen". Les trois facteurs sont les suivant : Le Facteur Type d'encodage (6 niveaux: HOA-1,2 et 3 et Formations à 4, 9 et 16 voies), le stimulus à deux niveaux (guitare et voix), le facteur Configuration de salle à deux niveaux (réfléchissant diffus et absorbant).

Dans notre cas, on va chercher à déterminer s'il existe une dépendance entre la méthode et le niveau d'encodage ambisonics proposé et les réponses des auditeurs. Cette analyse va donc nous permettre de vérifier si la distinction est faite entre la référence et les autres ordres et éventuellement de vérifier si celleci varie de manière corrélative aux différences objectives entre les différents stimuli. L'analyse ANOVA suppose deux conditions fondamentales, à savoir que la forme générale de l'analyse de variance repose sur la normalité des distributions et l'indépendance des échantillons.

Pour un test ANOVA il convient d'expliciter les hypothèses en jeu. L'hypothèse nulle correspond au cas où les distributions de la variable sous les différentes conditions ne sont pas différentes. Dans notre expérience elle correspond au cas où nos différents facteurs n'ont pas d'influence sur la performance de l'auditeur. L'hypothèse alternative, est qu'il existe au moins une distribution dont la moyenne s'écarte des autres et qui correspondrait au cas où au moins un facteur a une influence sur les réponses des auditeurs.

Un intervalle de confiance nommé *p-value* est indispensable pour savoir si notre résultat est fiable. Cette valeur indique la probabilité d'obtenir la même valeur (ou une valeur encore plus extrême) du test si l'hypothèse nulle était vraie. Plus simplement elle représente la probabilité d'obtenir une réponse correcte de l'auditeur par chance. On admet généralement qu'un résultat dont la *p-value* est inférieur à 0.05 (1 chance sur 20 d'avoir répondu correctement de manière aléatoire) est admissible et que 0.01 donne un résultat extrêmement fiable. [Goo99]

<sup>&</sup>lt;sup>16</sup> Terme souvent abrégé par le terme anglais ANOVA : ANalysis Of VAriance - appelé N-Way ANOVA dans Matlab<sup>®</sup>



Le tableau ci-dessous (23) présente les résultats du test ANOVA sur les réponses moyennées [TM]. Lorsque "Prob > F" est inférieur à 0.05 pour un facteur donné, cela signifie que la probabilité pour que ce facteur n'ai pas d'influence sur la variable est inférieure à 1/20. On considère alors que la différence entre les moyennes des différents niveau du facteur n'est pas due au hasard et que ce facteur a un effet significatif sur le score de performance (les réponses moyennées) :

\varTheta 🔿 🔿	Figure	1: N-W	ay ANOVA/			
	Anal	ysis of	Variance			
Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F	
Encoding Type Stimulus Room Encoding Type*Stimulus Encoding Type*Room Stimulus*Room Error Total	0.93981 0.01042 0 0.02778 0.13542 0.00463 0.01968 1.13773	5 1 5 5 1 5 23	0.18796 0.01042 0 0.00556 0.02708 0.00463 0.00394	47.76 2.65 0 1.41 6.88 1.18	0.0003 0.1647 1 0.3572 0.027 0.3276	

Constrained (Type III) sums of squares.

Figure 23: Analyse N-Way ANOVA sur la moyenne des résultats du test en studio

On observe que seulement deux valeurs de p sont significatives : Celle du type d'encodage (F(5,5) = 47.76; p = 0.0003 < 0.05) et celle du type d'encodage lié à la configuration de la salle (F(5,5) = 6.88; p = 0.027 < 0.05). Où F représente le facteur total de déviation utile pour calculer la *p-value* et qui dépend du nombre de degrés de liberté dont la valeur est consultable dans la table de Fisher. On peut donc conclure que le type d'encodage a un effet sur la perception que les auditeurs ont de la salle. De même l'interaction avec le facteur *Room* nous permet de dire que suivant la longueur de la queue de réverbération, la perception que l'on peut avoir du stimulus est différente.

Le tableau ci-dessous (23) présente les résultats du test ANOVA sur les réponses moyennées mais dans le cas du test mené en binaural. On observe cette fois-ci que les deux valeurs de p significatives sont toujours celle du type d'encodage (F(5,5) = 95.83; p = 0.0001 < 0.05) mais que l'interaction significative est passée sur le type d'encodage lié au type de stimulus convolué (F(5,5) = 12; p = 0.0082 < 0.05). Il est assez difficile d'émettre une hypothèse sur ce changement dans l'analyse de facteurs multiple. Ceci est peut-être dû au fait que les niveaux sonores ne correspondaient pas entre le test en studio et le test binaural. Les gens auront donc été plus sensible au timbre de l'instrument et moins à la queue de réverbération plus longue dans le cas diffusant et donc plus facilement différentiable.



0 0	Figure 1: N-Way ANOVA					
	Anal	ysis of \	Variance			
Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F	
Encoding Type Stimulus Room Encoding Type*Stimulus Encoding Type*Room Stimulus*Room Error Total	0.95833 0.06 0.00167 0.12 0.02833 0 0.01 1.17833	5 1 5 5 1 5 23	0.19167 0.06 0.00167 0.024 0.00567 0 0.002	95.83 30 0.83 12 2.83 0	0.0001 0.0028 0.4032 0.0082 0.1388 1	

Constrained (Type III) sums of squares.

Figure 24: Analyse N-Way ANOVA sur la moyenne des résultats du test en binaural

#### 5.3.2 Multiple Comparison Test

L'analyse de variance permet simplement de répondre à la question de savoir si tous les échantillons suivent une même loi normale. Dans le cas où l'on rejette l'hypothèse nulle, cette analyse ne permet pas de savoir quels sont les échantillons qui s'écartent de cette loi. Pour identifier les échantillons correspondant, on utilise différents tests «post-hoc», ou tests de comparaisons multiples [HT87], [MJ92], [ea80]. L'analyse  $MCP^{17}$  sous Matlab permet de fixer le seuil de confiance sur la matrice de sortie de l'analyse ANOVA et de spécifier sur quelle dimension nous voulons effectuer notre comparaison. Etant donné le résultat de l'analyse ANOVA sur les valeurs de p, nous observerons donc en premier lieu les groupes liés au type d'encodage (HOA et formation de voies).

La figure 25 présente les résultats de l'analyse MCP. On peut donc voir de manière détaillée que pour l'encodage de Type HOA3, la moyenne des réponses est centrée autour de 0.5, soit 50% de bonnes ou mauvaises réponses. Ceci veut dire qu'il n'existe pas de différence audible entre l'ordre HOA 3 et l'ordre 4. De plus la couleur du graphe nous montre que 4 groupes ont des différences significatives avec HOA 3, il s'agit des trois types d'encodage en formation de voies ainsi que l'ambisonics d'ordre 1.

Nous pouvons également observer la moyenne des réponses pour l'ordre HOA 2. Cette moyenne tend aussi vers une valeur de 0.5 et nous montre que ce type d'encodage n'a pas de différence audible avec l'ordre 4 dans ces conditions de test. De plus, on remarque que ce groupe de réponses représente une population significativement différente des autres niveaux d'encodage (avec toujours un intervalle de confiance de 95%).

Etant donné que l'analyse de facteur multiple sur les résultats du Studio ont montré que le facteur salle était perçu différemment (Rappel: F(5,5) = 6,88; p = 0.027 < 0.05), nous pouvons voir figure 26 l'analyse détaillée de ces deux facteurs. Les résultats confirment que c'est la configuration "réflec. diffus" qui est le plus difficilement discernable. En effet on remarque que les réponses des candidats pour "Encoding Type = HOA-2" et "Room = D" (pour diffusant), le groupe a une moyenne significativement différente des

<sup>&</sup>lt;sup>17</sup> *MCP* - Multiple Comparison Test





Figure 25: Analyse de Comparaison Multiple sur le type d'encodage pour les résultats du test en Studio

groupes "HOA1" et "Beam" dans toutes les configurations (cf. les 3 points rouge à distinguer des groupes en gris et bleu). De plus les réponses pour HOA-2 (config. reflec. diffus) et HOA-3 ( config. reflec. diffus et absorbant) montrent des moyennes situées autour de 0.5, ce qui confirme une nouvelle fois que ce type d'encodage ne montre pas de différence perceptible par rapport à l'encodage d'ordre 4.

Le test en Binaural (fig. fig:MultCmpBinaural) confirme les résultats pour l'encodage ambisonics d'ordre 3. Cependant pour l'écoute au casque, le groupe HOA 2 devient indéterminé dans le sens ou les résultats ne permettent pas d'affirmer que ce niveau d'encodage à un effet sur la performance. Le groupe HOA2 fait désormais partie de la population significativement différente du groupe HOA 3.

Pour l'analyse statistique sur l'encodage de type formation de voies (représenté par les 3 points rouges les plus bas de la figure 27(a)), on peut dire qu'ils sont tous discernables de HOA3 et donc de HOA4.

De la même façon que le test en studio a montré des différences entre les facteurs de configuration de salle et d'encodage, nous avons remarqué pour le test binaural que c'était le facteur multiple lié au "stimulus" qui montrait des résultats significativement différents (Rappel: F(5,5) = 12; p = 0.0082 < 0.05). Nous pouvons observer figure 27(b) que la guitare est l'exemple sonore qui permet le plus facilement de juger des différences entre les divers ordres d'encodage (point bleu le plus à gauche de la figure 27(b) pour la formation avec 16 voies). Toutefois, les moyennes des réponses ne montrent pas une différence aussi









Figure 26: Analyse de Comparaison Multiple à deux facteurs (type d'encodage et "room") pour les résultats du test en Studio

marquante et les moyennes liées seulement au facteur du stimulus sont plus significatives que lorsqu'on fait l'analyse double facteurs.







(a) Analyse de Comparaison Multiple sur le type d'encodage pour les résultats du test en Binaural



(b) Analyse de Comparaison Multiple à deux facteurs (type d'encodage et stimulus) pour les résultats du test en Binaural

Figure 27: Analyse MCP du test Binaural)





## 6 Critiques et perspectives

### 6.1 Développements possibles du test perceptif

L'étude à montré que des différences audibles existaient entre les différents ordres d'encodage, mais de nombreuses questions d'ordre psychoacoustique restent à déterminer et résoudre. En effet, l'analyse ANOVA donne seulement une confirmation que l'encodage a un effet sur la perception. L'analyse Post hoc comme indiqué dans le chapitre 5.3.2 permet tout de même de comprendre quels ordres ne sont pas distingués de la référence mais elle n'explique pas de quelle manière elle influe et sur quels critères perceptifs (enveloppement, clarté, …). La réponse du comment pourrait être donné par un test ABC ou MUSHRA.

Avni *et. al* [Avn10] ont utilisé une méthode intéressante<sup>18</sup> [BR99] pour déterminer de manière plus fine comment les différents ordres étaient perçus et surtout quels paramètres objectifs liés aux attributs de la salle étaient mis en avant lors de la distinction entre plusieurs stimuli. Leur méthode préconisait une phase de distinction des stimuli où les utilisateurs fixaient eux-même les attributs perceptifs. Dans une deuxième phase de nouvelles paires étaient générées et offraient la possibilité de noter sur une échelle une dizaine de stimuli en les réécoutant plusieurs fois. Cette méthode leur à permis de classifier de nombreux groupes et surtout de déterminer en quoi les différences étaient perceptibles.

En effet, dans notre cas, certain stimuli étaient différentiables non pas à cause d'une différence de résolution spatiale, mais par exemple à cause d'une différence de position de la source, ou bien d'une différence de timbre. Donner la possibilité de noter sur une échelle perceptive des critères avec un attribut perceptif aurait pu donner à notre étude la possibilité de comprendre l'origine des lacunes existantes dans le système d'encodage et de décodage actuel.

De plus, lors du test ABX, plusieurs candidats ont remarqué qu'ils n'effectuaient pas la distinction entre les divers stimuli de manière globale, mais attendaient simplement l'écoute entre le stimulus B et le stimulus X pour effectuer leur notation. Ainsi, laisser la possibilité de rejouer des sons plusieurs fois dans une première phase aurait donné la possibilité pour la deuxième phase de ne marquer que des critères propres au stimulus noté et de détailler plus encore les spécificités de chaque ordre et des deux approches.

#### 6.2 Modèle de convolution hybride

Comme nous l'avons évoqué en introduction, ce stage s'inscrit dans un projet à plus long terme visant à utiliser une approche hybride entre l'implémentation actuelle du *Spat* basé des réseaux de retards bouclés et l'approche convolutive. Le fait d'aborder la convolution dans le domaine ambisonics a montré que cette technique offrait de nombreuses possibilités quant à la paramétrisation des RIR. On pourra par exemple utiliser la même approche que dans notre test en ne codant que l'onde direct simulée à l'aide d'une onde plane dans la direction correspondante à l'analyse de la réponse et envoyer cette onde ainsi que les réflexions précoces dans un réseau de retard fonctionnant dans le domaine des harmoniques sphériques.

<sup>&</sup>lt;sup>18</sup> *RGT* - Repertory Grid Technique





## 7 Conclusion

Ce compte-rendu présente une expérience perceptive menée dans le domaine de l'analyse et la synthèse de réponses impulsionnelles de salles en ambisonics. Ces réponses ont été mesurées à l'aide d'un microphone sphérique dans l'Espace de Projection de l'IRCAM. Une analyse de la variance des capsules à été menée pour déterminer si les mesures étaient exploitables et n'avaient pas de différences systématiques dans les différentes configurations de salle enregistrées (absorbant, mur sud réfléchissant, réfléchissant diffus). L'étude de la cohérence spatiale d'un champ diffus a permis de déterminer une méthode de débruitage des capsules par synthèse de bruit reconstruit à l'aide de la superposition d'ondes planes. Deux approches d'encodage dans le domaine des harmoniques sphériques ont été proposées. La première est la méthode ambisonics d'ordre élevé, la seconde est une approche par formation de voies équivalente à un filtrage spatial.

Après la simulation et le traitement des enregistrements par microphone sphérique, des stimuli sonores adaptés à un test d'écoute ont été générés dans le domaine des harmoniques sphériques. Des choix pour les différents stimuli ont été faits et ont permis la sélection sur des critères d'ordre perceptifs (contenu timbral, clarté et longueur des échantillons). Le choix s'est orienté vers un son de voix et d'une guitare dans deux configurations de salle (absorbant et réfléchissant). En amont de l'expérimentation avec des participants nous avons effectué une normalisation des stimuli afin de présenter des échantillons sonores équilibrés en niveau pour le test aveugle ABX en prenant comme référence l'ordre HOA-4.

Le développement de l'interface de test et de la procédure ont été effectués dans le but d'une reproduction ambisonics en studio. Un second test à été mené par diffusion sur casque, en utilisant une synthèse modale de ce studio en binaural en utilisant des HRTF. L'analyse ANOVA ainsi que l'analyse "post hoc" sur les réponses moyennées des participants, ont permis de repérer qu'il existait des différences significatives pour le type d'encodage, le type de stimulus ainsi qu'un effet sur le type de salle choisi pour la convolution.

L'étude perceptive a montré que les participants n'arrivaient pas à effectuer une distinction entre l'ordre HOA-3 et l'ordre HOA-4. Des erreurs de réponses indiquaient même qu'il était parfois difficile de distinguer l'ordre HOA-2 avec cette référence à l'ordre 4. Nous avons également pu remarqué que l'approche par Formation de Voies était distinguée de l'ordre HOA-4 par les participants. Ces résultats nous font donc pencher en faveur d'une solution d'encodage dans le domaine HOA plutôt que la technique de formation de voies, qui donnait de plus des grandes différences au niveau du timbre de l'échantillon sonore proposé. Les différences sont donc pour l'instant d'ordre technique et dépendent de critères objectifs liés à l'implémentation de la méthode. Mais cette approche par filtrage spatial reste tout de même intéressante et de futurs travaux de développement pour résoudre ces problèmes sont sans doute possible.

En conclusion, le résultat le plus important est que cette étude a montré qu'il est possible de réduire de manière significative le nombre de canaux tout en gardant une qualité d'encodage élevée et donc suffisante pour une restitution à haute résolution spatiale.









## 8 ANNEXES

#### 8.1 Décomposition en Harmoniques Sphériques

Cette partie résume brièvement la solution de l'équation des ondes en coordonnées sphériques dont une explication plus complète est consultable dans [Wil99] Une bibliothèque d'outils *Matlab* à été développée par l'équipe afin de manipuler rapidement les équations qui sont présentées ici.

Solution de l'équation d'onde en coordonnées sphériques :

L'équation des ondes en coordonnées sphériques est donnée par :

$$\frac{1}{r^2}\frac{\partial}{\partial r}(r^2\frac{\partial p}{\partial r}) + \frac{1}{r^2sin\theta}\frac{\partial}{\partial \theta}(sin\theta\frac{\partial p}{\partial \theta}) + \frac{1}{r^2sin^2\theta}\frac{\partial^2 p}{\partial \varphi^2} - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = 0$$
(22)

Dont la résolution nous donne l'expression du champ de pression donnée par l'équation [DM04] :

$$p(r,\theta,\varphi,k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} (a_n^m(k)y_n(kr) + b_n^m(k)j_n(kr))Y_n^m(\theta,\varphi) + \sum_{n=0}^{\infty} \sum_{m=-n}^{n} (c_n^m(k)h_n^{(1)}(kr) + b_n^m(k)h_n^{(2)}(kr))Y_n^m(\theta,\varphi)$$
(23)

où  $Y_n^m$  sont les harmoniques sphériques et  $a_n^m$  et  $b_n^m$  les coefficients spatiaux de la décomposition du champ sonore, que l'on peut comparer aux coefficients des séries de Fourier.  $y_n$  et  $j_n$  sont les fonctions de Bessel sphériques de première et deuxième espèce décrivant les ondes progressives convergentes;  $h_n^{(1)}$  et  $h_n^{(2)}$  sont les fonction de Hankel sphériques de première et deuxième espèce et décrivent l'expansion du champ radial des ondes progressives divergentes.

Les fonctions angulaires solutions des équations différentielles en coordonnées sphériques peuvent être combinées en une fonction  $Y_n^m$  avec la fonction de Legendre associée de première espèce  $P_n^m$  et donner l'expression suivante:

$$Y_n^m(\theta,\varphi) = \sqrt{\frac{(2n+1)}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta) e^{im\varphi}$$
(24)







**Figure 28:** Représentation des Harmoniques Sphériques Réelles pour n = 0, 1, 2 et m = -n, ..., n source: [Col11]

#### 8.2 Discrétisation de la Sphère

Afin d'obtenir une transformée en harmonique discrète avec un nombre minimal de points de mesures, la sphère doit être discrétisée de la manière la plus uniforme possible pour préserver la répartition énergétique entre les différents points de mesure ou lors de la synthèse sur une sphère virtuelle. Les solides de Platon discrétisent la sphère uniformément, cependant ils ne sont que 5 et ont un nombre maximum de 20 sommets pour le dodécaèdre.

Ainsi, si nous souhaitons effectuer des mesures avec des ordres d'harmoniques sphériques élevés, il est nécessaire de proposer des méthodes alternatives [Col11]:

- Extremal points for hyperinterpolation [SW03]
- Spiral points [RSZ94]
- Equal-area partition [SK97]
- Gauss-Legendre scheme [Sne94]
- Equi-distant cylindrical grid [RSZ94]
- Equi-angle grid [DH94]

Tous ces schémas de discrétisation ont la possibilité de mettre en jeu un grand nombre de sommets, cependant ils ne sont pas strictement uniformes. Aucune de ces méthodes n'est parfaite, il faudra donc choisir la solution optimale en fonction du type d'interpolation choisi.

La grille d'hyperinterpolation est une méthode que nous avons utilisé pour la reconstruction du champ sonore lors du transcodage des Formation de voies. Ce schéma de discrétisation sur la sphère est liée aux travaux se basant sur la théorie des systèmes extrêmes. Le lecteur peut se reporter à l'ouvrage [SW03] pour plus de détails théoriques.





#### 8.3 Min et Max des Variances des Capsules

La figure 29 du haut représente les valeurs extrêmes des variances des capsules du microphone sphérique pour toutes les configurations ainsi que tous les haut-parleurs (correspond au numéro de la réponse impulsionnelle mesurée). On remarque que certaines valeurs extrêmes sont liées à des directions précises de haut-parleurs. Par exemple entre 700Hz et 3kHz les valeurs min proviennent de la même direction (segment bleu).

Ceci nous permet de confirmer que le débruitage des capsules doit être effectué en fonction des paramètres moyens de TR, de  $P_0$  et de  $T_{lim}$ .



**Figure 29:** Affichage des occurrences des valeurs extrêmes (min et max) pour le calcul de la variance des capsules en fonction de la fréquence











## References

- [AAG<sup>+</sup>13] Amir Avni, Jens Ahrens, Matthias Geier, Sascha Spors, Hagen Wierstorf, and Boaz Rafaely. Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution. J Acoust Soc Am, 133(5):2711–2721, 2013.
  - [AB04] J.S Abel and D.P. Berners. Reverberation: acoustics, analysis and synthesis. *AES 117th Convention*, October 2004.
  - [AH06] J.S. Abel and P. Huang. A simple, robust measure of reverberation echo density. *121st AES Convention*, 2006.
  - [AP09] J. Ahonen and V. Pulkki. Diffuseness estimation using temporal variation of intensity vectors. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, October 2009.
  - [Avn10] Amir Avni. Spaciousness of sound fields captured by spherical microphone arrays. Master's thesis, Electrical and Computer Engineering, Ben Gurion University of the Negev, 2010.
- [BDPW13] S. Bertet, J Daniel, E. Parizet, and O. Warusfel. Investigation on localisation accuracy for first and higher order ambisonics reproduced sound sources. *Ambisonic Symposium*, Volume 99(4):pp. 642–657(16), July/August 2013.
  - [Bea07] Bertet and Warusfel et al. Investigation of the perceived spatial resolution of higher order ambisonics sound fields: a subjective evaluation involving virtual and real 3d microphones. AES 30th International Conference, Saariselkä, Finland, March 15-17 2007.
  - [BF12] S. Braun and S. Frank. Localization of 3d ambisonic recordings and ambisonic virtual sources. Master Thesis Article, 2012.
  - [BL09] Boley and Lester. Statistical analysis of abx results using signal detection theory. *AES* 127th *Convention*, October 2009.
  - [BR99] Berg and Rumsey. Spatial attribute identification and scaling by repertory grid technique and other methods. *AES 16th Conference*, April 1999.
  - [Bra10] J.S. Bradley. Review of objective room acoustics measures and future need. *In Proceedings of the International Symposium on Room Acoustics, ISRA Melbourne, Australia,* pages 1–8, August 2010.
  - [Col11] J Colafrancesco. Caractérisation du champ sonore dans les salles, application à la mesure et au transcodage de réponses impulsionnelles multi-canaux. Master's thesis, Mémoire de Stage ATIAM - IRCAM - UPMC - Telecom Paris, 2011.
- [CSNW13] Thibaut Carpentier, Tal Szpruch, Markus Noisternig, and Olivier Warusfel. Parametric control of convolution based room simulators. In *ISRA-Toronto*, 2013.
  - [Dan01] J. Daniel. *Représentation de ChampsAcoustiques, Application à la Transmission et à la Reproduction de Scènes Sonores Complexes dans un Contexte Multimédia*. PhD thesis, Université de Paris VI, 2001.





- [Dan03] J. Daniel. Spatial sound encoding including near field effect: Introducing distance coding filters and a viable new ambisonic format. In Audio Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction, Copenhagen, Denmark, May 2003.
- [DDP08] G. Defrance, L Daudet, and J.-D. Polack. Detecting arrivals within room impulse responses using matching pursuit. *Proc. of the 11th Int. Conference on Digital Audio Effects*, 2008.
- [DDP09] G. Defrance, L Daudet, and J.-D. Polack. Using matching pursuit for estimating mixing time within room impulse responses. *Acta Acustica united with Acustica*, 2009.
- [DH94] J.R. Driscoll and D.M. Healy. Computing fourier transforms and convolutions on the 2-sphere. *Adv. Appl. Math.*, 15:202–250, 1994.
- [DM04] J. Daniel and S. Moreau. Further study of sound field coding with higher order ambisonics. In *Audio Engineering Society Convention 116*, 5 2004.
  - [ea80] Milliken et al. *Population marginal means in the linear model: an alternative to least-squares means.* American Statistician, 1980.
  - [ea10] B. Rafaely et al. *Speech Processing in Modern Communications,* chapter Spherical microphones array beamforming Chap. 11. Springer-Verlag, Berlin, 2010.
- [Gel08] A. Gelman. *Analysis of Variance The new Palgrave dictionary of economics*. Basingstoke, Hampshire New York: Palgrave Macmillan, 2008.
- [Ger75] Michael Gerzon. Ambisonics. part two: Studio techniques. Studio Sound, 17(8):24-26, oct 1975.
- [Ger85] M. Gerzon. Ambisonics in multichannel broadcasting and video, *J. Audio Eng. Soc.*, 33(11):859–871, November 1985.
- [Goo99] SN. Goodman. Toward evidence-based medical statistics. 1: The p value fallacy. *Annals of Internal Medicine*, 130:995–1004, 1999.
- [How02] David C. Howell. *Statistical methods for psychology*. Pacific Grove, CA: Duxbury/Thomson Learning., 2002.
  - [HT87] Hochberg and Tamhane. Multiple Comparison Procedures. Hoboken John Wiley and Sons, 1987.
  - [IR10] ITU-R. Multichannel stereophonic sound system with and without accompanying picture. Technical report, ITU-R BS.775-2, http://www.itu.int/dms\_pubrec/itu-r/rec/bs/R-REC-BS.775-2-200607-S!!PDF-E.pdf, 2010.
  - [JC91] J.M Jot and A. Chaigne. Digital delay networks for designing artificial reverberators. *Proceed*ings of the 90th AES Convention - Paris, 1991.
  - [Jot92] J.M. Jot. *Etude et réalisation d'un spatialisateur de sons par modèles physiques et perceptifs*. PhD thesis, Telecom Paris, 1992.



- [Jot99] J.M Jot. Real-time spatial processing of sounds for music, multimedia and interactive humancomputer interfaces. *ACM Multimedia Systems Journal*, 7(1), 1999.
- [Kah95] Eckhard Kahle. Validation d'un modèle objectif de la perception de la qualité acoustique dans un ensemble de salles de concerts. PhD thesis, Laboratoire d'Acoustique, Université du Maine, 1995.
- [Kut00] H. Kutruff. Room Acoustics. Spon Press, New York, 2000.
- [Lav89] C. Lavandier. *Validation perceptive d'un modèle objectif de caractérisation de la qualité acoustique des salles*. PhD thesis, Université du Main, Le Mans, 1989.
- [ME04] J. Meyer and G. W. Elko. Spherical microphones array for 3D sound recording, chapter Audio Signal Processing For Next-Generation Multimedia Communication Systems. Kluwer Academic, 2004.
- [MJ92] Milliken and D. E. Johnson. *Analysis of Messy Data, Volume 1: Designed Experiments*. Boca Raton, FL: Chapman Hall/CRC Press, 1992.
- [MJN07] M. Morimoto, M. Jinya, and K. Nakagawa. Effects of frequency charac- teristics of reverberation time on listener envelopment. Effects of frequency charac- teristics of reverberation time on listener envelopment," J. Acoust. Soc. Am., 122:1611–1615, 2007.
- [MNL08] M. Morimoto, K. Nakagawa, and K. Lida. The relation between spatial impression and the law of the first wavefront. *Appl. Acoust.*, 69:132–140, 2008.
  - [MP05] J. Merimaa and V. Pulkki. Spatial impulse response rendering i: Analysis and synthesis. *AES*, 2005.
- [MPL05] J. Merimaa, T. Peltonen, and T. Lokki. Concert hall impulse responses, 2005.
- [NC13] M. Noisternig and T. Carpentier. Technical report on spherical acoustics mathematical basis and implementation in spat,à<sup>o</sup>. Technical report, IRCAM EAC, March 18 2013.
- [Nic10] J. Nicolle. Analyse spatio-temporelle de réponses impulsionnelles. Master's thesis, Master ATIAM IRCAM UPMC Telecom, 2010.
- [PF12] J.D. Polack and F. Leão Figueiredo. Room acoustics auralization with ambisonics. *proceedings* of Acoustics 2012, 2012.
- [Ple09] P. Plessas. Rigid sphere microphone arrays for spatial recording and holography. Master's thesis, Graz University of Technology, Austria, Nov 2009.
- [PM06] V. Pulkki and J. Merimaa. Spatial impulse response rendering ii: Reproduction of diffuse sound and listening tests. *AES*, 2006.
- [Pol93] J.D. Polack. laying billiards in the concert hall: the mathematical foundations of geometrical room acoustics. *Applied Acoustics (special issue on Computer Modelling and Auralisation of Sound Fields in Rooms)*, 38:235–244, 1993.



- [Pul97] V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45(6):456–466, June 1997.
- [Pul07] V. Pulkki. Spatial sound reproduction with directional audio coding. AES, 2007.
- [Raf04] B. Rafaely. Plane-wave decomposition of the sound field on a sphere by spherical convolution. JASA, 116(4), October 2004.
- [Raf05] Boaz Rafaely. Analysis and design of spherical microphone arrays. *IEEE Transactions on Speech and Audio Processing*, 13(1):135–143, 2005.
- [RSZ94] E. A. Rakhmanov, E.B. Saffand, and Y.M. Zhou. Minimal discrete energy on the sphere, 1994.
- [RT13] B. Rafaely and V. Tourbabin. The relation between the information delivered by head-related transfer function and human spatial hearing. *J. Acoust. Soc. Am.*, 2013.
- [Sch54] M.R. Schroeder. Statistical parameters of the frequency response curves of large rooms. *J. Audio Eng. Soc. 1987,* 1954.
- [SJHL99] L. Savioja, J. J. Huopaniemi, and T. Lokki. Creating interactive virtual acoustic environments. *J. Audio Eng. Soc.*, 47:675–705, 1999.
  - [SK97] E. B. Saff and A. B. J. Kuijlaars. Distributing many points on a sphere. *The Mathematical Intelli*gencer, 19:5–11, 1997.
- [SMKK12] Haohai Sun, Edwin Mabande, Konrad Kowalczyk, and Walter Kellermann. Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing. *Journal* of the Acoustical Society of America, 131(4), April 2012.
  - [Sne94] N. Sneeuw. Global spherical harmonic analysis by least-squares and numerical quadrature methods in historical perspective, 1994.
  - [SP82] John Stautner and Miller Puckette. Designing multi-channel reverberators. *Computer Music Journal*, 6(1), 1982.
  - [SS07] R. Stewart and M. Sandler. Statistical measures of early reflections of room impulse responses. *Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07),* 2007.
  - [Ste06] R. Stewart. Hybrid convolution and filterbank artificial reverberation algorithm using statistical analysis and synthesis. Master's thesis, Msc Music Technology, University of York, England, 2006.
  - [SW03] I. H. Sloan and R. S. Womersley. Extremal systems of points and numerical integration on the sphere. *Adv Comput Math*, 21:107–125, April 2003.
    - [TM] Inc The MathWorks. N-way analysis of variance.
  - [Wil99] Earl G. Williams. *Fourier acoustics : sound radiation and nearfield acoustical holography*. Academic Press, June 1999.





- [WJC97] O. Warusfel, JM. Jot, and L. Cerveau. Analysis and synthesis of room reverberation based on a statistical time-frequency model. *AES Convention*, 1997.
- [ZPN12] Franz Zotter, Hannes Pomberger, and Markus Noisternig. Energy-preserving ambisonic decoding. *Acta Acust United Ac*, 98(31):1, 37-47 2012.