

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS 6

École doctorale EDITE

Spécialité

Traitement du signal

présentée par

Terence Caulkins

terence.caulkins@ircam.fr

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE PARIS 6

Caractérisation et contrôle du rayonnement
d'un système de Wave Field Synthesis
pour la situation de concert.

soutenue le 18 décembre 2007

devant le jury composé de :

M. Jean-Dominique POLACK	Directeur de thèse
M. Diemer DE VRIES	Rapporteur
M. Alexandre GARCIA	Rapporteur
M. Yves GRENIER	Examineur
M. Emmanuel FRIOT	Examineur
M. Olivier WARUSFEL	Examineur

Remerciements

New York, 7 Novembre 2008

Je remercie tout d'abord mon directeur de thèse Prof. Jean Dominique Polack pour ces conseils avisés, sa confiance et sa sympathie tout au long de la thèse. Je suis ravi de constater que nous continuons à collaborer actuellement et j'espère que cette nouvelle collaboration sera fructueuse.

Je remercie mon encadrant de thèse Dr. Olivier Warusfel, qui a accepté de me guider de son excellence scientifique et de son expérience de chercheur pendant ces quatre années. Je reste admiratif des qualités humaines et éthiques d'Olivier, qui ont participé au fait que je retiendrais l'expérience de la thèse non seulement comme un enrichissement scientifique mais aussi comme un apprentissage humain.

Je remercie Prof. Yves Grenier et Dr. Emmanuel Friot pour avoir accepté de participer à mon jury de thèse, et plus particulièrement Prof. Diemer DeVries et Prof. Alexandre Garcia, les rapporteurs de ce travail, pour l'intérêt qu'ils ont porté à la lecture du document et leurs commentaires avisés.

Je remercie Dr. Etienne Corteel, collègue de bureau, mentor et ami. Je suis profondément reconnaissant de ses conseils avisés et de son aide qui ont été des facteurs déterminants dans la réussite de ma thèse, et qui m'ont permis de mûrir d'un point de vue scientifique.

Je remercie mes collègues de l'équipe acoustique des Salles qui ont su m'épauler dans des moments parfois difficiles - Olivier Delerue, Khoa-Van Nguyen, Emmanuel Rio, Gérard Bertrand, Alain Terrier, Stephanie Bertet, Alexis Baskind, Remy Muller, Joseph Sanson, Clara Suied, Isabelle Viaud-Delmon, et Brian Katz. Je suis enchanté d'avoir pu côtoyer des gens aussi brillants et sympathiques pendant cette aventure, et j'espère pouvoir collaborer avec eux à l'avenir.

Je remercie Hugues Vinet, pour son accueil au sein du secteur scientifique de l'Ircam et sa sympathie, Laurent Ghys, Arnaud Gomes-do-Vale, Youcef Bensaid, Olivier Labat, Fred Vandromme de l'équipe Informatique pour leur savoir-faire technique et leur assistance saupoudrée de bonne humeur, Sylvie Benoît et Carole Tan pour leur assistance et leur patience, l'ensemble de l'équipe de Production pour leur sympathie et leur compétence, Guillaume Boutard pour sa culture artistique et son sens de l'humour peu orthodoxe, Chungshin Yeh pour sa générosité et sa gentillesse, Sébastien Roux pour son esthétique musicale et sa sympathie, Damien Tardieu, Arshia Cont, Julien Bloit, Matthias Demoucron, Jean Bresson, et Nicolas Rasamamina pour avoir été mes joyeux compagnons de thèse. Plus généralement j'étends ma gratitude à tous les Ircamiens que j'ai eu la chance de côtoyer, et qui font de cet endroit un haut lieu de la culture et de la pensée.

Je remercie enfin du fond du coeur mes parents pour leur amour et leur soutien inconditionnel. Sans eux rien de tout cela n'aurait été possible.



Résumé

La Wave Field Synthesis (WFS) est une technique qui vise à reconstruire les caractéristiques physiques d'un champ sonore en zone étendue, à partir d'un réseau de haut-parleurs. Le cadre défini pour la thèse est celui de la situation de concert de musique mixte, dans lequel un système WFS placé en nez de scène est utilisé pour synthétiser des instruments virtuels partageant l'espace acoustique de la scène avec des instrumentistes réels. Etant donné ce cadre, on peut observer que si la WFS permettait de reconstruire à *l'identique* le son direct de l'instrument réel, il s'en suivrait naturellement que son effet de salle associé serait indissociable de celui produit par les instruments réels qui l'entourent sur la scène. Cependant, la technique repose sur un ensemble d'approximations des principes théoriques sous-jacents qui n'assurent pas l'exactitude du champ reproduit. Par conséquent, le son direct rayonné par l'instrument virtuel WFS n'engendre pas un effet de salle conforme à celui qui serait engendré par un instrument réel.

Le travail présenté consiste d'abord à analyser le champ effectivement rayonné par le banc de haut-parleurs, utilisant une méthode d'acoustique prédictive développée dans le cadre de la thèse. A partir de cette analyse, nous identifions la partie du champ rayonné par le banc de haut-parleurs dans la salle qui est naturellement cohérente avec celle d'un instrument réel. Nous proposons ensuite une méthode de compensation, permettant de rajouter la portion d'effet de salle manquante, basée sur une caractérisation préalable du système et l'injection d'énergie réverbérée via des sources WFS directives. Nous validons enfin la possibilité de contrôler l'effet de salle précoce en modifiant la directivité de rayonnement du banc WFS. Cette validation est basée sur des mesures réalisées *in situ* utilisant un microphone à haute résolution spatiale.

Mots clé : spatialisation, réseaux de haut-parleurs, Wave Field Synthesis, holophonie, compensation d'effet de salle, directivité de rayonnement, holographie sphérique, audio 3D, mesure de puissance acoustique, microphone haute résolution spatiale

Abstract

Wave Field Synthesis (WFS) is a sound reproduction technique that aims at reconstructing the physical characteristics of a sound field over an extended area, using a loudspeaker array. This thesis focuses on the "mixed concert", in which a WFS system is used to synthesize virtual instruments that share the acoustic space of the stage with real musicians. If the system were capable of replicating the exact acoustic field emitted by a real instrument, the ensuing room effect generated by the virtual instrument would be indistinguishable from that generated by real instruments. However, the WFS sound reproduction technique relies on a set of approximations of underlying physical principles, which introduce artifacts in the reproduced sound field. Thus, the direct sound field radiated by a virtual WFS instrument generates room effect which does not conform with room effect generated by real instruments.

The work presented here consists first of all in analyzing the sound field radiated by the loudspeaker array, by means of an acoustic radiation prediction method developed during the thesis. Working from the analysis, we identify the portion of the sound field radiated by the array that is naturally consistent with the one radiated by a real instrument. We then propose a compensation method, which allows to supply the missing portion of the room effect, based on a characterization of the system and the use of directional sources to inject additional reverberant energy. Finally, we validate the possibility of varying early room effect levels in the room by modifying the radiation directivity of the WFS loudspeaker array. The validation is based on measurements conducted *in situ* using a high spatial resolution microphone.

Key Words : spatialization, loudspeaker arrays, Wave Field Synthesis, holophony, room effect compensation, spherical holography, directivity, power measurement techniques, 3D audio, high spatial resolution microphones

Table de Matières

Notations	xiii
Introduction	1
1 Etat de l'art pour la synthèse d'une scène sonore par Wave Field Synthesis	7
1.1 Introduction	7
1.2 Bases théoriques de la Wave Field Synthesis	8
1.2.1 Le principe de Huygens	8
1.2.2 Intégrale de Kirchhoff-Helmholtz	9
1.2.3 Intégrale de Rayleigh I	10
1.3 Synthèse du son direct d'une source par Wave Field Synthesis	12
1.3.1 Introduction	12
1.3.2 Définition du concept de source virtuelle	13
1.3.3 Contraintes technologiques et pratiques	14
1.3.4 Dérivation des filtres WFS pour un banc linéaire continu infini	15
1.3.5 Cas de la source focalisée	20
1.3.6 Cas de l'onde plane	21
1.3.7 Conséquences de la troncature du banc de haut-parleurs	23
1.3.8 Conséquences de l'échantillonnage spatial du banc de haut-parleurs	25
1.3.9 Egalisation du son direct en conditions réelles	28
1.4 Description et synthèse d'un effet de salle pour la WFS	36

1.4.1	Modélisations de l'effet de salle	36
1.4.2	Le Spatialisateur	38
1.4.3	Système de rendu d'effet de salle WFS Ircam	41
1.5	Architecture globale du système WFS Ircam	44
1.6	Conclusion et positionnement de la thèse	50
2	Loudspeaker radiation prediction method based on a spherical harmonic expansion of the sound field	55
2.1	Theory	61
2.1.1	Spherical harmonic expansion of the sound field	61
2.1.2	Radiation prediction methods	63
2.1.3	The discrete spherical harmonic decomposition	66
2.2	Measurement protocol and apparatus	69
2.3	Simulations	72
2.3.1	Influence of spatial oversampling on measurement noise	72
2.3.2	Influence of source size on quality of the decomposition at a given order	73
2.3.3	Comparison of methods aimed at diminishing the influence of incoming reflections.	75
2.4	Practical Results	84
2.4.1	Sound field prediction results for a Multi-Actuator Panel loudspeaker .	85
2.4.2	Sound field prediction results for an electrodynamic loudspeaker	87
2.5	Conclusion	90
3	Contrôle de la directivité en Wave Field Synthesis	95
3.1	Introduction	95
3.2	Méthode modulaire pour la synthèse de directivité	98
3.2.1	Manipulation des figures de directivité	100
3.3	Formulation des filtres WFS pour une source virtuelle directive	102
3.4	Effet de la troncature du banc de haut-parleurs	105
3.5	Echantillonnage spatial du banc WFS et ordre maximal de reproduction d'une figure de directivité	106
3.5.1	Théorème d'échantillonnage dans l'espace des fonctions de directivité .	107
3.5.2	Ordre maximal de recombinaison d'une fonction de directivité par WFS	109

TABLE DE MATIÈRES

3.6	Evaluation de l'erreur de reconstruction du son direct associé à une source directive dans le plan horizontal	113
3.6.1	Evaluation de l'erreur complexe de reconstruction	114
3.6.2	Evaluation de l'erreur de reconstruction énergétique	116
3.6.3	Conclusions	119
3.7	Analyse du champ tridimensionnel effectivement rayonné par le banc WFS. . .	120
3.7.1	Analyse du champ en-dessous de la fréquence de repliement par décomposition sur la base des harmoniques sphériques	121
3.7.2	Rayonnement tridimensionnel du banc en hautes fréquences	126
3.7.3	Extension possible de l'analyse du rayonnement du banc	127
3.8	Discussion : De l'intérêt de pouvoir recréer une "Couleur Tonale Directionnelle"	130
3.9	Conclusion	132
4	Caractérisation et contrôle de l'effet de salle tardif pour la Wave Field Synthesis	135
4.1	Introduction	135
4.2	Proposition d'une méthode de compensation de l'effet de salle tardif pour la WFS	138
4.2.1	Définition d'une qualité acoustique "cible"	139
4.2.2	Définition d'une qualité acoustique "contexte"	144
4.2.3	Bilan énergétique et calcul de compensation	150
4.2.4	Conclusion	151
4.3	Elaboration d'une méthode de compensation de l'effet de salle tardif simplifiée	152
4.3.1	Introduction	152
4.3.2	Mesure de puissance en chambre réverbérante	153
4.3.3	Mesure de puissance par Intensimétrie	154
4.3.4	Méthode d'estimation par simulation d'un banc de haut-parleurs monopolaires	155
4.3.5	Méthode d'estimation de puissance par prédiction de rayonnement . . .	158
4.3.6	Méthode d'estimation de puissance in situ utilisant le relief de décroissance	160
4.3.7	Conclusion	164
4.3.8	Principe de mise en oeuvre de la compensation	164
4.3.9	Base de données pour la compensation	167
4.4	Conclusion	168

5	Use of a high spatial resolution microphone to characterize the early reflections generated by a WFS loudspeaker array	171
5.1	Introduction	173
5.2	Description of the High Spatial Resolution Microphone Array	175
5.2.1	Presentation	175
5.2.2	Principle	176
5.2.3	Spatial resolution of the microphone array	177
5.3	Characterizing the early room effect associated to a WFS array from simulated measurements made on the high resolution microphone	180
5.3.1	Description of the simulated measurement setup	180
5.3.2	Ideal source measurement	182
5.3.3	Measurement of an ideal WFS array	184
5.4	Measurements of a WFS loudspeaker array in the Espace de Projection concert hall	190
5.5	Conclusions	192
Conclusion		195
A Techniques de Mesure en Acoustique des Salles		201
A.1	Mesure d'une réponse impulsionnelle de salle	201
A.1.1	Principe général	201
A.1.2	Choix du signal de mesure	202
A.1.3	Déconvolution circulaire	203
A.1.4	Analyse de distorsions	203
A.1.5	Généralization de la méthode classique de mesure de réponse impulsionnelle	204
A.2	Mesure de réverbération	205
A.3	Mesure de la puissance émise par une source acoustique	208
A.3.1	Mesure de puissance en chambre réverbérante	209
A.3.2	Mesure de puissance basée sur le relief de décroissance énergétique de la réverbération.	210
A.3.3	Mesure de puissance basée sur l'intensité acoustique	216
A.3.4	Méthode basée sur la décomposition en harmoniques sphériques	225

TABLE DE MATIÈRES

B Les Harmoniques Sphériques	227
C Décomposition en harmoniques sphériques du champ associé à source virtuelle WFS	233
D Calcul de Compensation de Contexte	239
E Glossaire	243
Bibliographie	252
References	253

Les termes apparaissant en *italiques* dans le texte sont pour la plupart définis dans le Glossaire (Annexe E). Le Glossaire regroupe également la définition d'un ensemble d'indices objectifs communément utilisés dans la description d'un effet de salle.

Liste des notations employées	
* · <i>i</i> <i>Re</i> <i>Im</i>	opération de convolution opération de multiplication nombre imaginaire tel que $i^2 = -1$ partie réelle d'un nombre complexe partie imaginaire d'un nombre complexe
<i>t</i> <i>c</i> <i>T</i> <i>f</i> ω <i>k</i> \vec{k} λ <i>V</i> ρ <i>E</i> <i>W</i>	temps [<i>s</i>] célérité du son [$m \cdot s^{-1}$] période [<i>s</i>] fréquence [Hz] fréquence angulaire [radians] nombre d'onde [m^{-1}] vecteur d'onde longueur d'onde [<i>m</i>] volume [m^3] masse volumique de l'air [$kg \cdot m^{-3}$] densité d'énergie [$J \cdot m^{-3}$] \equiv [$kg \cdot m^{-1} \cdot s^{-2}$] puissance [<i>W</i>] \equiv [$J \cdot s^{-1}$]
log ln	logarithme à base 10 logarithme népérien
BF, MF, HF FFT IFFT ITD ILD RIF HRTF MAP QA TFCT TFTD WFS	basses, moyennes, hautes fréquences Fast Fourier Transform Inverse Fast Fourier Transform Interaural Time Difference Interaural Level Difference Réponse Impulsionnelle Finie Head Related Transfer Functions Multi Actuator Panel Qualité Acoustique Transformée de Fourier à Court Terme Transformée de Fourier à Temps Discret Wave Field Synthesis
$p_0 = 2 \cdot 10^{-5} Pa$ $I_0 = 10^{-12} W \cdot m^{-2}$ $W_0 = 10^{-12} W$	pression de référence (seuil d'audibilité) intensité de référence (seuil d'audibilité) puissance de référence (seuil d'audibilité)

Introduction

L'utilisation de technologies sonores de spatialisation connaît aujourd'hui un essor important dans les milieux artistiques et professionnels. Ces applications permettent de restituer une sensation d'espace par voie électroacoustique, et présentent un intérêt particulier dans le cadre de la réalité virtuelle ou augmentée, qui sont des applications visant à simuler la présence d'objets virtuels dans l'environnement avec lesquels l'individu peut interagir.

On distinguera trois catégories générales dans les technologies de spatialisation :

- Les technologies comme le dolby surround, 5.1, et plus généralement le N.1, que nous nommerons technologies *perceptives* du fait qu'elles se basent sur une illusion perceptive (le principe stéréophonique) qui n'est appréciable qu'au centre du dispositif.
- Les technologies de type *signal* basées sur des mesures de signaux acoustiques. On placera dans cette catégorie la technologie *binaurale*, qui utilise des mesures de fonctions de transfert faite au niveau des conduits auditifs (HRTF ou Head Related Transfer Function) de manière à restituer, sur casque, la sensation d'espace dans l'angle solide complet entourant la tête de l'auditeur.
- Les méthodes dites de *reconstruction physique du champ sonore*, qui visent à recréer l'impression d'espace sur une zone étendue en utilisant des réseaux de haut-parleurs. On compte aujourd'hui de multiples implémentations de ces techniques, dont les deux plus répandues sont, pour l'instant, la *Wave Field Synthesis* (ou WFS) qui vise une reproduction en 2D, et la technologie *Ambisonique*, qui vise une reproduction 3D.

La technologie binaurale (Larcher, 2001) restitue une impression d'espace qui est cohérente avec les déplacements et mouvements de tête de l'auditeur, à condition d'asservir le rendu à un système de suivi de position ("tracking") placé sur la tête de l'auditeur. La technique Ambisonique, introduite par (Gerzon, 1973; Gerzon, 1985), utilise une décomposition du champ

sonore, capté en un point de l'espace, sur un ensemble fonctions spatiales élémentaires (les harmoniques sphériques). A partir de cette décomposition, elle permet de reconstruire, par réciprocity de l'équation des ondes, le champ sonore tridimensionnel autour d'un point de la zone d'écoute, en utilisant un dispositif périphérique de haut-parleurs. Cette zone est réduite à une taille de "sweet spot" stéréophonique dans le cas de l'Ambisonique ordre 1, mais peut être étendue davantage grâce à l'utilisation de techniques microphoniques relativement récentes qui permettent de restituer un *Ambisonique d'Ordre Elevé* ou *HOA* (High Order Ambisonics) (Daniel *et al.*, 2003; Daniel, 2000). La WFS est quant à elle issue de l'université de Delft (Berkhout, 1988; Berkhout *et al.*, 1993) et inspirée de méthodes d'exploration géophysique mises au point par le professeur A.J. Berkhout pendant les années 1980. La technique se base sur un principe d'échantillonnage spatial des fronts d'ondes acoustiques émis par une source sonore virtuelle à travers la "fenêtre" formée par le banc de haut-parleurs dans le plan horizontal.

Ces technologies ont très vite été adoptés dans des milieux utilisant la réalité virtuelle, comme les laboratoires médicaux (binaural pour les traitements de réhabilitation en psychothérapie) et l'industrie (WFS pour les simulateurs de vol et de conduite). On commence à observer également des incursions de ces méthodes dans des domaines plus larges comme le divertissement, le cinéma, le musée. Ainsi, le rendu binaural est utilisé maintenant dans certains jeux vidéos, et la WFS fait son apparition dans quelques salles de cinéma comme moyen de d'adresser un contenu 5.1 sur l'étendue de la salle entière, ou encore dans des musées sous forme d'installations sonores (Desnoux & Caulkins, 2005; Grand *et al.*, 2005; Corsino & Corteel, 2006; Chiron *et al.*, 2007).

Dans le domaine du concert, on compte également un certain nombre de performances musicales publiques ayant fait usage des technologies audio citées ci-dessus. Parmi celles-ci, on peut citer des concerts de musique électroacoustique¹ utilisant le système de WFS du TU Berlin entre 2003 et 2007. On trouve également quelques événements musicaux sur casque pour public étendu, comme le "placard headphone festival"² (1998-2007), dans lequel certains compositeurs font usage de la technologie binaurale. Il convient aussi de mentionner une installation multicanale du Fraunhofer Institute IDMT³ dans le cadre du Bregenz Festival en 2005, où un système WFS a été utilisé pour générer un effet de salle virtuel dans un théâtre en plein air, en complément d'un système de delta stéréophonie pour le rendu du son direct. Enfin, on citera des esquisses musicales réalisées à l'Ircam entre 2004 et 2007 à partir de matériaux préexistants adaptés pour la diffusion sur WFS dans le cadre du festival Résonances et Agora (Verin *et al.*, 2003; Maresz & Corteel, 2004; Schoeller & T., 2004).

Malgré ces quelques événements, il est clair que les technologies de spatialisation comme

¹<http://gigant.kgw.tu-berlin.de/baalman/>

²www.leplacard.org

³<http://www.iosono-sound.com/>

Introduction

le binaural, l'Ambisonique, ou la WFS (et même plus généralement les techniques multicanales comme le dolby surround ou 5.1), sont loin de faire office de standard dans le domaine du concert, dont le panorama est encore dominé par la stéréophonie. Les raisons de cette absence sont multiples, et tiennent autant au problème du format de stockage multicanal⁴ qu'au coût de la technologie, voire même aux habitudes de travail des ingénieurs du son du domaine. Malgré tout, il est fort à parier que l'utilisation des technologies audio 3D s'étende davantage au domaine du concert, et plus généralement à celui du spectacle vivant, dans les années à venir.

Si on considère l'adaptabilité des trois méthodes décrites ci-dessus à la situation de concert, on se rend compte que la technique binaurale, qui requiert le port d'un casque audio, comporte a priori un intérêt réduit dans la situation de concert. Ceci est évident d'abord d'un point de vue matériel (nécessité d'équiper un public entier avec des casques), mais également du fait que le casque implique un masquage du contenu sonore qui serait éventuellement produit par des instruments non amplifiés sur la scène, ce qui pose un problème par exemple dans le cadre de la musique "mixte" (i.e. la musique instrumentale augmentée d'effets électroacoustique).

Contrairement à la technologie binaurale, rien n'empêche, dans le principe, d'utiliser les techniques dites "de reconstruction physique du champ sonore" dans le cadre du concert. En effet, la WFS et le HOA permettent tous les deux d'adresser un contenu spatialisé à un public étendu, sans avoir besoin de l'équiper d'un matériel spécifique (comme un casque ou un système de tracking). Les principes sous-jacents à ces méthodes impliquent cependant des artefacts de synthèse et des contraintes géométriques différentes.

La technique Ambisonique s'appuie sur une disposition périphérique des haut-parleurs permettant de reproduire un champ tridimensionnel. Pour résumer, on pourrait dire que plus l'ordre de description est élevé, plus on est capable d'étendre la zone de restitution privilégiée depuis le centre du dispositif vers la périphérie (Daniel, 2000). D'un autre côté, dans le cas de la WFS, la zone de restitution privilégiée n'est plus limitée en taille, et couvre l'ensemble du plan horizontal situé devant le banc de haut-parleurs. La contrainte d'une répartition périphérique est également relâchée ; on peut envisager une forme quelconque du banc de haut-parleurs utilisé pour réaliser la synthèse des fronts d'ondes, pour peu que celle-ci soit connexe. Par contre, le "compromis" physique entre la méthode WFS et la méthode HOA⁵ se situe au niveau de la fréquence ; en effet, la WFS possède une fréquence maximale de reproduction (la fréquence de repliement spatial), au-delà de laquelle le champ sonore produit n'est plus cohérent. Ceci implique de densifier la répartition des haut-parleurs si on souhaite recréer correctement les in-

⁴Des formats supportant les technologies multicanales ont aujourd'hui émergé (Blu-Ray et HD-DVD) mais pour l'instant représentent seulement 2.5% des ventes totales de DVD dans le monde pour les 6 premiers mois de 2007 (HDDigest, 2007).

⁵Nous employons ici le terme "compromis", car on peut montrer que les deux méthodes dérivent du même formalisme (i.e. l'intégrale de Kirchhoff-Helmholtz) et diffèrent simplement dans le choix des approximations faites au moment de l'implémentation (Nicol & Emerit, 1999).

dices de localisation, et il en découle une restriction au plan horizontal, étant donné le nombre important de canaux qu'impliquerait une distribution 2D de haut-parleurs.

Les avantages et inconvénients des deux méthodes que nous venons de citer impliquent qu'il vaut mieux adopter l'une ou l'autre suivant le contexte d'application. Dans la situation du concert, on est placé a priori dans un espace de restitution d'envergure importante (même pour des salles de taille moyenne) ce qui implique d'installer un grand nombre de haut-parleurs si on souhaite étendre la zone de restitution privilégiée dans le cas HOA, et encore plus élevée si on souhaite assurer une fréquence de repliement suffisante en WFS. Cependant, on peut remarquer que dans cette situation, l'attention du public est en général focalisée vers une seule zone de la salle : la scène.

Pour cette raison, nous pensons que la WFS présente un avantage vis-à-vis d'autres technologies immersives dans le cadre du concert, au sens où on peut adapter sa géométrie de manière à concentrer la qualité de restitution dans la zone frontale de la scène. On peut donc prétendre à la création d'une fenêtre acoustique ouvrant sur l'espace de la scène, permettant par la même occasion de recueillir les bénéfices de la réalité augmentée dans un contexte musical. La technique paraît particulièrement intéressante dans le cadre de la musique dite "mixte", dans laquelle des instrumentistes sont accompagnés par un dispositif électroacoustique qui diffuse des sons ou effets synthétisés par électronique.

Parmi ces bénéfices, on citera par exemple la possibilité de renforcer la présence des sources instrumentales sur scène tout en maintenant une localisation cohérente de celles-ci sur l'ensemble de la zone d'écoute. En effet, le renforcement sonore d'instruments par des systèmes standards stéréophoniques comporte le problème de la délocalisation du son produit vis-à-vis de l'instrument pour toutes les positions non centrées dans la salle. Un bénéfice supplémentaire peut être trouvé dans la possibilité d'introduire (et déplacer au besoin) des instruments virtuels sur la scène, de manière à compléter l'espace des timbres dans certaines sections de l'orchestre à des moments précis requis par la partition.

Etant donné cette situation de "concert augmenté", le public situé dans la salle va pouvoir cependant juger du réalisme de la reproduction par comparaison directe avec des instruments réels placés dans le même espace acoustique. La problématique de la thèse se situe autour de cette question : comment faut-il procéder pour assurer la synthèse d'une source instrumentale *virtuelle* sur la scène qui soit cohérente avec les instruments environnants ?

Dans cette situation, l'enjeu est double ; il faut d'une part s'assurer que la composante de son direct émise par l'instrument virtuel soit correctement restituée, et, d'autre part, il faut assurer que l'effet de salle associé à celui-ci demeure cohérent, quelle que soit sa position ou orientation sur la scène.

Introduction

La thèse s'articule autour de l'utilisation d'un réseau linéaire WFS en situation de concert. L'objectif est d'assurer, moyennant une caractérisation préalable du système, une reproduction cohérente à la fois du son direct et de l'effet de salle associé à un instrument virtuel situé sur scène. Nous visons un contrôle simultané de ces deux facettes du champ sonore, par le biais d'un jeu sur la directivité de rayonnement du banc de haut-parleurs utilisé.

Le travail présenté vient en continuité d'une thèse réalisée à l'Ircam par Etienne Corteel, qui a présenté un ensemble de méthodes permettant de prendre en compte les défauts d'un système WFS pour la synthèse du son direct, et qui a introduit le formalisme de synthèse de sources directives en WFS (Corteel, 2004). Il fait suite également à un travail de l'auteur sur la cohérence de l'effet de salle avec la salle de restitution, notamment pour la synthèse des sources situées à l'intérieur de l'espace d'écoute (Caulkins, 2003; Caulkins *et al.*, 2003).

Plan du document

Ce document peut être abordé de diverses manières, suivant l'intérêt et le niveau de familiarité avec les techniques utilisées.

Le **chapitre 1** présente un **tour d'horizon théorique, perceptif et pratique de l'état de l'art de Wave Field Synthesis**. Ce premier chapitre est scindé en trois sections : d'abord, nous introduisons la théorie sous-jacente à la technique, puis nous décrivons comment la synthèse du son direct et de l'effet de salle sont réalisées utilisant la WFS. En fonction de cet état de l'art, nous dégageons le positionnement de la thèse et les axes principaux de recherche suivant lesquels nous allons travailler de manière à adapter la technique de Wave Field Synthesis à la situation de concert. Le souhait de l'auteur est également que les pistes données dans ce premier chapitre puissent servir à des étudiants, ou des chercheurs peu familiers avec le domaine, comme un point d'entrée général sur la technique de Wave Field Synthesis.

Le **chapitre 2** est consacré à la description d'une **méthode de prédiction de rayonnement** permettant de caractériser le rayonnement d'un banc de haut-parleurs WFS. La méthode s'appuie sur une décomposition en harmoniques sphériques du rayonnement d'un ou plusieurs haut-parleurs constitutifs du banc, permettant de reconstruire par propagation le champ rayonné par le banc entier en champ libre. Trois approches théoriques possibles pour la décomposition du champ sont comparées, de manière à déterminer la méthode optimale étant donnée de l'installation de mesure dont nous disposons en chambre sourde. La méthode est utilisée ensuite pour prédire le rayonnement de deux types de haut-parleurs différents (MAP et électrodynamique), et nous comparons les résultats obtenus avec des mesures réelles des mêmes transducteurs faites en champ libre.

Le **chapitre 3** est consacré au **contrôle de la directivité** de rayonnement d'un banc WFS. On introduit d'abord les implications du contrôle de la directivité rayonnement pour la diffusion en salle. Suite à cette introduction, nous approfondissons le formalisme existant vis à vis de la directivité en WFS, en introduisant des outils théoriques permettant de contrôler celle-ci de manière simple et intuitive. Nous proposons ensuite une évaluation de l'erreur de restitution du son direct associé à une source virtuelle dans le plan horizontal par simulation d'un banc idéal, suivie d'une analyse du champ effectivement rayonné par le banc WFS dans les trois dimensions. Le chapitre se termine avec une discussion sur l'intérêt, pour le concert, de recréer une "couleur tonale directive", i.e. un rayonnement directif qui reproduise les caractéristiques modales d'un instrument réel dans une salle de concert.

Le **chapitre 4** part du constat que l'excitation de la salle de concert produite par le banc de haut-parleurs dépend à la fois de la position et de la directivité de la source virtuelle reproduite. Il s'agit donc d'introduire une **compensation de l'effet de salle** si on souhaite reproduire un effet de salle qui demeure cohérent avec les sources virtuelles reproduites au cours de la performance. Une première méthode de compensation est introduite, basée sur une méthode développée dans (Jot *et al.*, 1998), qui vise à reproduire un effet de salle cible à une position de référence dans la salle. Nous introduisons, dans le cadre de cette méthode, un **principe de rajout d'effet de salle tardif** basé sur le contrôle de la directivité d'émission du banc WFS. Nous introduisons ensuite une méthode simplifiée de compensation de l'effet de salle, basée sur une **caractérisation de la puissance émise par le banc WFS**. Différentes méthodes de mesure de la puissance du banc sont comparées dont une basée sur la méthode de prédiction développée au chapitre 2. Nous introduisons enfin un principe de compensation de l'effet de salle tardif utilisant les résultats de la caractérisation de puissance émise par le banc WFS, suivi de la description d'une architecture permettant d'intégrer la compensation au système de rendu WFS.

Le **chapitre 5** présente enfin une **validation objective du contrôle de la directivité**, par le biais d'une **mesure en salle** utilisant un **microphone à haute résolution spatiale**. Ce microphone, qui permet d'isoler les composantes provenant du plan horizontal, est utilisé conjointement avec un modèle de type source image de manière à analyser l'effet de salle précoce produit par le banc de haut-parleurs. Nous illustrons le fait que le banc de haut-parleurs est capable de focaliser le son direct vers le public en minimisant les premières réflexions, ou bien, au contraire, de stimuler les murs latéraux en minimisant le niveau du son direct.

Nous concluons enfin le document en résumant l'ensemble du travail effectué au cours de la thèse et en dégageant un certain nombre de travaux prospectifs inspirés de ce travail.

CHAPITRE 1

Etat de l'art pour la synthèse d'une scène sonore par Wave Field Synthesis

1.1 Introduction

Cette section a pour vocation d'introduire un ensemble de concepts et de méthodes utilisées couramment dans le cadre de la Wave Field Synthesis (WFS) pour la synthèse d'une scène sonore sur zone étendue. L'objectif est triple : d'une part, présenter l'état de l'art de la technique au début de la thèse ; d'autre part, présenter les outils théoriques et pratiques qui seront manipulés dans le reste du document ; enfin, dégager les voies de recherche proposées dans la thèse. Nous avons opté pour une méthode de présentation qui alterne des considérations théoriques ou mathématiques avec leurs implications pratiques pour l'implémentation.

La Wave Field Synthesis est une technique de reproduction sonore qui se base sur le principe de Huygens (sec. 1.2.1) et ses formalisations mathématiques (sec. 1.2.2 & 1.2.3), qui permettent théoriquement de recréer un champ sonore tridimensionnel quelconque à partir d'une distribution de sources sonores. D'un point de vue technologique et pratique il est impossible de mettre en oeuvre le principe de Huygens en respectant les hypothèses contenues dans ses formalisations (sec. 1.3.3). Il est donc question, dans un premier temps, de restreindre les objectifs de synthèse en se concentrant sur la reproduction du *son direct* émis par une source ponctuelle omnidirectionnelle située dans le plan horizontal. Ces hypothèses, qui sont les hypothèses "clas-

siques" de la WFS dans sa formulation initiale, permettent de dériver les filtres à appliquer à un banc de haut-parleurs monopolaires idéaux de manière à approcher au mieux un contrôle du champ sonore sur l'ensemble du plan horizontal (sec. 1.3.4). La méthode a par la suite été adaptée de manière à prendre en compte les défauts de directivité de haut-parleurs réels via un algorithme d'égalisation multicanal (sec. 1.3.9). La synthèse d'un effet de salle associé au son direct a également fait l'objet d'une implémentation (1.4), qui s'inscrit dans l'architecture globale du système WFS Ircam (sec. 1.5).

Le lecteur est invité à noter que certains aspects théoriques liés à la synthèse du son direct et de l'effet de salle sont parfois traités ici de manière synthétique, car ils ont déjà fait l'objet d'une description détaillée dans (Start, 1997), (Verheijen, 1997), (Nicol, 1999), (Corteel, 2004). Les références bibliographiques appropriées sont indiquées au cours du chapitre pour guider le lecteur qui souhaiterait approfondir l'étude des différents problèmes abordés.

1.2 Bases théoriques de la Wave Field Synthesis

1.2.1 Le principe de Huygens

La Wave field synthesis est basée sur le principe de Huygens, établi par le chercheur hollandais Christiaan Huygens en 1690 dans son *Traité de la Lumière* (Huygens, 1690). Le principe stipule que chaque élément d'un front d'onde peut être vu comme le centre d'une perturbation qui génère à son tour un front d'onde secondaire (Fig. 1.1).

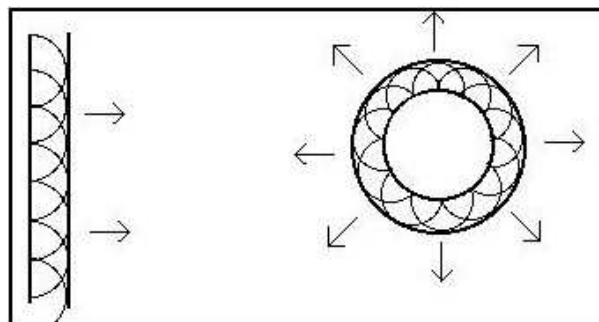


FIG. 1.1 – Principe de Huygens

Ainsi, en plaçant une distribution continue de sources dites "secondaires" le long d'un front d'onde émis par une source ponctuelle dite "primaire", on peut reproduire à l'identique le champ sonore situé en aval de la position du front d'onde *en l'absence de la source primaire*. On dira que les sources secondaires peuvent prendre le "relais" de la source primaire. Dans ces conditions, un auditeur situé en aval de la position du front d'onde ne sera pas capable de faire la distinction entre le champ original produit par la source primaire et un champ synthétisé par

la distribution de sources secondaires. L'intérêt potentiel de ce principe pour la reproduction de champs sonores est ici manifeste.

1.2.2 Intégrale de Kirchhoff-Helmholtz

Soit Ω un milieu 3D homogène et isotrope à la fois inviscide et non-conducteur de chaleur. On sépare Ω en deux sous-espaces Ω_R et Ω_Ψ par une surface $\partial\Omega$. Ω_Ψ , qui contient une distribution de source primaires, correspond au sous-espace des sources. Le sous-espace complémentaire Ω_R est appelé quant à lui le sous-espace de reproduction. Ces deux sous-espaces sont représentés sur la figure 1.2. L'intégrale de Kirchhoff-Helmholtz permet d'exprimer la va-

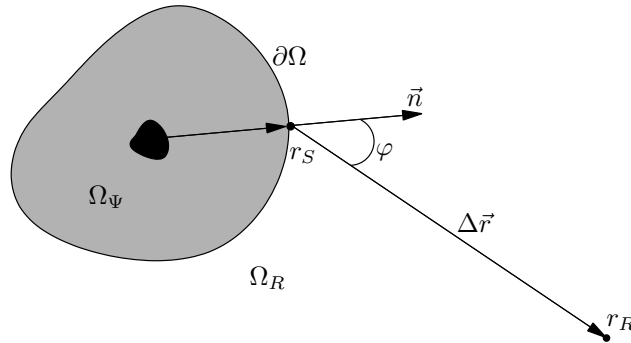


FIG. 1.2 – Géométrie pour l'intégrale de Kirchhoff-Helmholtz

leur de la pression $p(\vec{r}_R)$ dans le sous-espace de reproduction Ω_R en fonction de la pression et du gradient de pression produit par les sources primaires sur $\partial\Omega$:

$$\begin{cases} p(\vec{r}_R, \omega) = \int_{\partial\Omega} \left[p(\vec{r}_S, \omega) \vec{\nabla} G(\vec{r}_R | \vec{r}_S, \omega) - G(\vec{r}_R | \vec{r}_S, \omega) \vec{\nabla} p(\vec{r}_S, \omega) \right] \vec{n} dS & \text{pour } \vec{r}_R \in \Omega_R, \\ p(\vec{r}_R, \omega) = 0 & \text{pour } \vec{r}_R \in \Omega_\Psi, \end{cases} \quad (1.1)$$

avec ω la fréquence angulaire, R un point du sous-espace de reproduction Ω_R , $p(\vec{r}_S, \omega)$ le champ de pression capté sur $\partial\Omega$ et \vec{n} le vecteur normal à $\partial\Omega$ au point r_S orienté dans le sens entrant vers Ω_R . G est la *fonction de Green de l'espace libre* en 3 dimensions définie par :

$$G(\vec{r}_R | \vec{r}_S, \omega) = \frac{e^{-ik(\|\vec{r}_R - \vec{r}_S\|)}}{4\pi \|\vec{r}_R - \vec{r}_S\|}. \quad (1.2)$$

avec $k = \frac{\omega}{c}$ le nombre d'onde.

Par ailleurs, d'après l'équation d'Euler sur la conservation de la masse on a :

$$i\omega\rho v_n(r_S, \omega) = \vec{\nabla} p(\vec{r}_S, \omega) \cdot \vec{n} \quad (1.3)$$

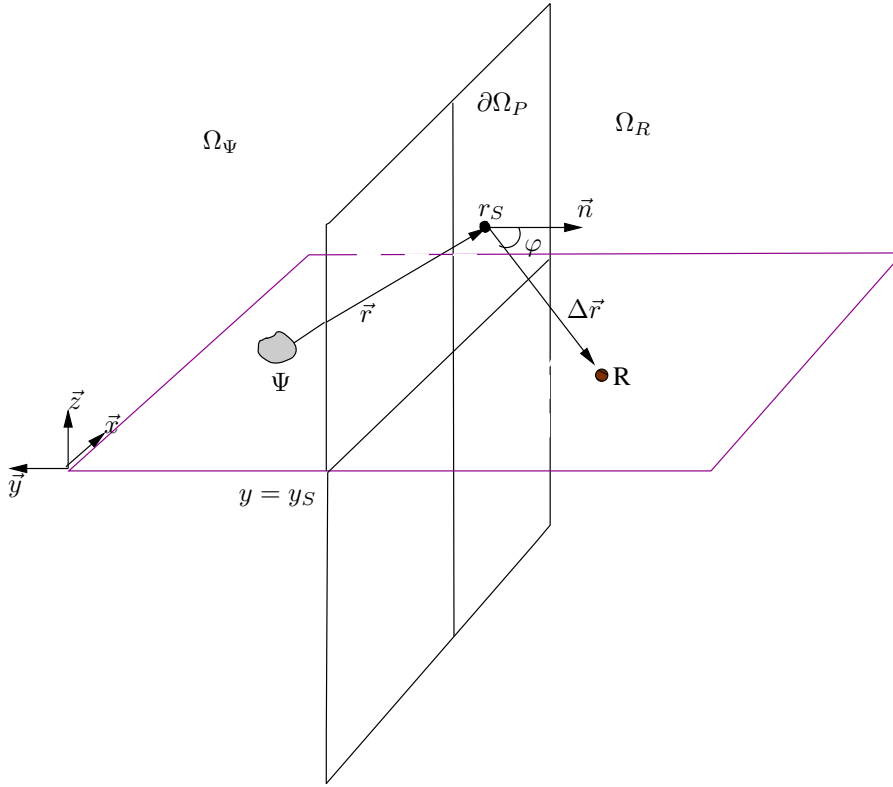


FIG. 1.3 – Géométrie pour l'intégrale de Rayleigh I

Cette condition de frontière (dite condition de Neumann) est réalisée dans le cas où $\partial\Omega$ est une paroi rigide parfaitement réfléchissante. On choisit la géométrie de la surface $\partial\Omega$ telle que celle-ci soit l'union d'une section de plan $\partial\Omega_P$ et d'une calotte sphérique $\partial\Omega_{CS}$ (Wapenaar & Berkhout, 1989). En faisant tendre le rayon de la calotte sphérique vers l'infini, on montre que l'intégrale de Kirchhoff-Helmholtz s'annule sur $\partial\Omega_{CS}$ (condition de Sommerfeld). Ceci permet de simplifier l'équation 1.4 et d'aboutir à l'expression suivante pour la pression au point R :

$$p(\vec{r}_R, \omega) = \frac{1}{2\pi} \int_{\partial\Omega_P} \left[i\omega\rho v_n(\vec{r}_S, \omega) \frac{e^{-ik\Delta r}}{\Delta r} \right] dS. \quad (1.6)$$

avec $\partial\Omega_P$ le plan infini qui sépare Ω_Ψ et Ω_R (Fig. 1.3).

L'équation 1.6 est connue sous le nom de l'Intégrale de Rayleigh I. Elle traduit le fait que tout champ produit par une distribution de sources primaires situées dans Ω_Ψ peut être reconstruit dans Ω_R en utilisant une distribution planaire de monopoles sur $\partial\Omega_P$. Pour ce faire, les monopoles doivent être alimentés par des signaux proportionnels à la composante normale de la vitesse particulière du champ incident sur $\partial\Omega_P$.

Les équations permettant de dériver les fonctions d'alimentation d'un banc WFS, qui sont explicitées à la section suivante, sont basées sur l'intégrale de Rayleigh I.

1.3 Synthèse du son direct d'une source par Wave Field Synthesis

1.3.1 Introduction

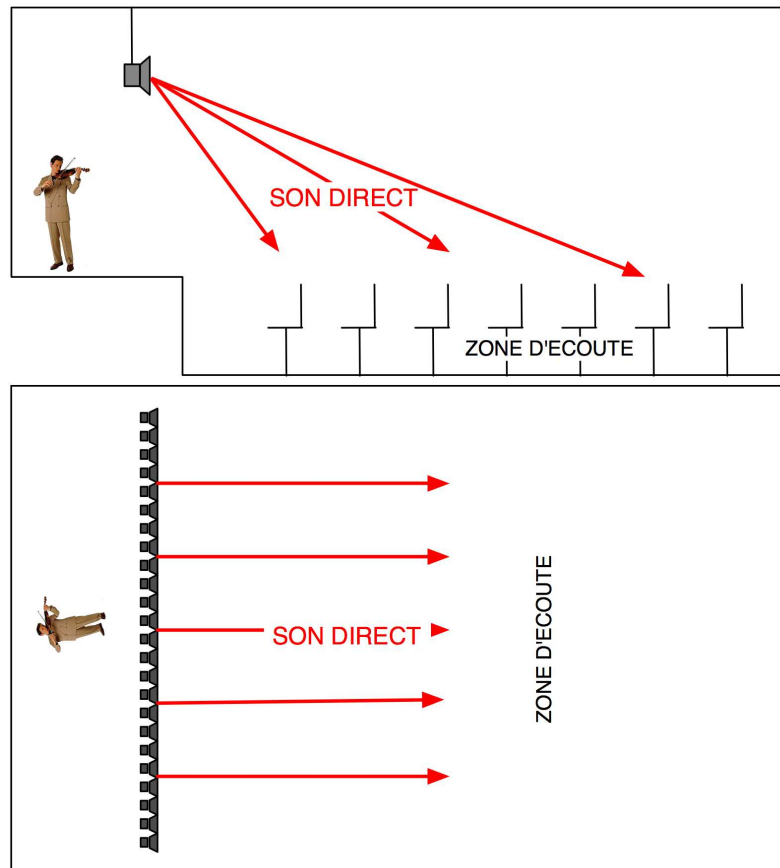


FIG. 1.4 – Vue schématique en coupe (haut) et de dessus (bas) d'une salle de concert équipée d'un banc de haut-parleurs WFS. Quelques composantes du son direct émis par le banc de haut parleurs sont indiquées par des flèches rouges.

Le *son direct* d'une source sonore désigne la partie du champ acoustique émise par la source et arrivant à l'auditeur suivant le trajet le plus court, sans subir l'influence des parois de la salle d'écoute (Fig. 1.4). En situation d'écoute, le son direct renseigne l'auditeur sur la localisation angulaire des différents événements sonores se déroulant dans une salle. Son niveau relatif à la réverbération constitue un indice absolu de distance. Le son direct revêt également une importance particulière dans les situations où l'intelligibilité du message véhiculé par la source est primordiale (cours magistral, conférence scientifique, pièce de théâtre, etc.). Les lieux destinés à ce genre de manifestation sont d'ailleurs souvent conçus de manière à optimiser le niveau du son direct à travers la zone d'écoute (Kuttruff, 1991). Le son direct constitue ainsi un élément

essentiel du champ sonore, dont il est indispensable de maîtriser la reproduction en situation de concert.

Cette section est dédiée à l'étude et la description de la synthèse du son direct dans le cadre de la Wave Field Synthesis. Le paradigme de *source virtuelle* ponctuelle est d'abord introduit de manière à pouvoir calculer une expression analytique pour les signaux d'alimentation des haut-parleurs, tout en conservant la possibilité de reconstruire des scènes sonores complexes par superposition de sources multiples. Les divers problèmes liés à l'implémentation pratique du principe de Huygens et de ses formalisations sont mis en évidence, et on montre comment on arrive à la formulation de fonctions d'alimentation permettant de synthétiser une source virtuelle à partir d'un banc linéaire de haut-parleurs. Une discussion sur les diverses limitations de la méthode est également abordée¹.

Notons que le choix d'envisager de manière séparée la synthèse du son direct et la synthèse de l'effet de salle a été fait dans le but de découpler les problèmes et de simplifier l'étude. Par ailleurs, pour isoler le son direct émis par le banc WFS il convient de se placer en conditions dites de *champ libre* (cf. glossaire), où l'on s'affranchit de toute signature d'un lieu d'écoute donné. Dans toute la suite de cette section nous supposons donc des conditions de champ libre.

1.3.2 Définition du concept de source virtuelle

La formulation des équations de la Wave Field Synthesis s'appuie sur une hypothèse simplificatrice vis-à-vis de la distribution $\partial\Omega$ de sources primaires, celle-ci étant modélisée sous la forme d'une source ponctuelle unique désignée par le terme **source virtuelle** ou source notionnelle². Cette description permet de donner une expression analytique aux fonctions d'alimentation du banc WFS, tout en offrant la possibilité de reconstruire par linéarité des scènes sonores complexes comportant des sources multiples.

Le champ de pression produit en $r_S^{\vec{}}$ par une source virtuelle monopolaire située en $r_{\Psi}^{\vec{}}$ s'exprime de la manière suivante :

$$p(r_S^{\vec{}}, \omega) = S(\omega) \frac{e^{-ikr}}{4\pi r}. \quad (1.7)$$

où $r = \|\vec{r}\| = \|\vec{r}_{\Psi} - \vec{r}_S\|$ et $S(\omega)$ représente le signal d'alimentation de la source virtuelle (Fig. 1.3). En pratique, ce signal d'alimentation correspond à un flux audio monophonique obtenu soit par enregistrement d'une source réelle, soit par un procédé de synthèse.

¹Pour plus de détails sur la synthèse du son direct en WFS le lecteur est invité à consulter (Start, 1997), (Nicol, 1999) et (Corteel, 2004).

²L'appellation "source virtuelle" sera retenue dans la suite de ce document.

1.3.3 Contraintes technologiques et pratiques

Les fonctions d'alimentation d'une distribution de sources planaire, continue et de taille infinie permettant de synthétiser à l'identique le champ émis par une source virtuelle pourraient être déduites directement de l'intégrale de Rayleigh I (éq. 1.6). Cependant, des facteurs technologiques et pratiques empêchent l'implémentation directe de cette intégrale dans le monde réel :

- a. L'état de l'art en matière d'ingénierie de transducteurs ne permet pas, à la connaissance de l'auteur, de fabriquer des distributions continues de sources acoustiques contrôlables en tout point. Il faut donc *échantillonner la distribution* de sources secondaires lors de l'implémentation de manière à pouvoir utiliser les transducteurs électroacoustiques existant.
- b. L'implémentation pratique requiert la *troncature de la distribution linéaire infinie* de manière à la ramener à une taille finie.
- c. Le coût de la technologie nécessaire à la réalisation d'un canal WFS (transducteur, amplificateur, convertisseur numérique/analogique, traitement du signal numérique) est prohibitif du point de vue de la réalisation de systèmes comportant de très grands nombres de haut-parleurs. Par conséquent on doit limiter le nombre de canaux de diffusion en minimisant la géométrie du dispositif. Le choix retenu consiste à *concentrer les haut-parleurs dans le plan horizontal* situé à la hauteur de la tête des auditeurs. Ce choix se justifie par le fait que dans beaucoup d'applications les têtes des auditeurs demeurent voisines du plan contenant les haut-parleurs (installations sonores muséographiques, cinéma, simulateurs, etc.³).

En réduisant ainsi la nature⁴ et l'étendue⁵ de la distribution de sources secondaires on introduit des erreurs dans le champ reconstruit. Ces erreurs peuvent être compensées en partie par une modification des fonctions d'alimentation appliquées à la distribution de sources secondaires. Ces modifications, qui sont explicitées en détail dans (Start, 1997) et (Verheijen, 1997), seront traitées dans leurs grandes lignes au cours des sections suivantes. On notera par ailleurs que, dans le cadre de ce travail, on s'intéresse uniquement à des bancs de haut-parleurs linéaires. Le lecteur peut cependant se référer à (Daniel *et al.*, 2003; Nicol, 1999) pour la formulation des fonctions d'alimentation d'un banc WFS à géométrie circulaire.

³On notera que cette configuration n'est pas toujours envisageable, comme par exemple dans des applications de renforcement sonore de sources instrumentales en situation de concert. Dans ce cas, le banc de haut-parleurs doit être sur-élevé pour ne pas gêner la visibilité de la scène (cf. Fig. 1.4). Les effets perceptifs de ce décalage spatial seront discutés plus loin dans cette section.

⁴continu → discret

⁵plan infini → segment fini

1.3.4 Dérivation des filtres WFS pour un banc linéaire continu infini

La première restriction géométrique imposée à la distribution de Rayleigh est une réduction du plan infini de sources secondaires $\partial\Omega_P$ en une droite L située dans le plan horizontal. On suppose que la source virtuelle Ψ et le point de réception R se situent dans ce même plan horizontal d'équation $z = 0$ (Fig. 1.5). Dans un premier temps on modélise le plan de Rayleigh

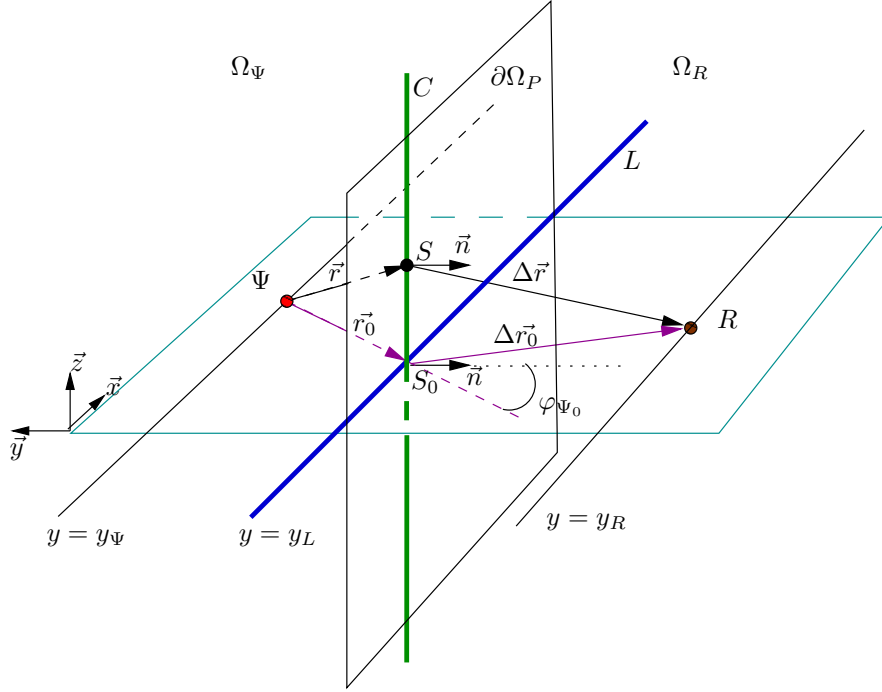


FIG. 1.5 – Géométrie pour le passage d'un plan infini de sources secondaires à une droite.

$\partial\Omega_P$ par un agencement continu de colonnes parallèles C . L'intégrale de Rayleigh I (eq. 1.6) s'exprime alors de la manière suivante :

$$p(\vec{r}_R, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P_C(\vec{r}_R, \omega) dx_C \quad \text{avec} \quad (1.8a)$$

$$P_C(\vec{r}_R, \omega) = \int_{-\infty}^{\infty} \left[i\omega \rho v_n(\vec{r}_S, \omega) \frac{e^{-ik\Delta r}}{\Delta r} \right] dz_C \quad (1.8b)$$

où x_C représente l'abscisse de la colonne C et z_C représente la hauteur d'un point de la colonne C . L'intégrale 1.8b quantifie la contribution en R d'une colonne C de la surface $\partial\Omega_P$.

En supposant que la source virtuelle Ψ est *monopolaire*, la composante normale de vitesse particulière $v_n(\vec{r}_S)$ du champ émis par la source virtuelle le long d'une colonne C peut s'écrire (eqs. 1.3 et 1.7) :

$$v_n(\vec{r}_S, \omega) = \frac{S(\omega)}{\rho c} \left(\frac{1 + ikr}{ikr} \right) \cos \varphi_\Psi \frac{e^{-ikr}}{r} \quad (1.9)$$

où φ_Ψ représente l'angle entre le vecteur normal \vec{n} à $\partial\Omega_P$ et \vec{r} au point S.

En supposant que l'on se place en champ lointain (i.e. $kr \gg 1$, source virtuelle lointaine ou en hautes fréquences), équation 1.9 devient :

$$v_n(\vec{r}_S, \omega) = \frac{S(\omega)}{\rho c} \cos \varphi_\Psi \frac{e^{-ikr}}{r} \quad (1.10)$$

La contribution d'une colonne C au point de réception R peut donc s'écrire :

$$P_C(\vec{r}_R, \omega) = \int_{-\infty}^{\infty} \left[\frac{ikS(\omega) \cos \varphi_\Psi}{r \Delta r} \right] e^{-ik(r+\Delta r)} dz_C \quad (1.11)$$

L'approximation dite de *phase stationnaire* permet d'approcher la valeur de cette intégrale, ce qui donne :

$$P_C(\vec{r}_R, \omega) = \sqrt{2\pi ik} \cdot S(\omega) \cos \varphi_{\Psi_0} \sqrt{\frac{\Delta r_0}{r_0 + \Delta r_0}} \frac{e^{-ik(r_0 + \Delta r_0)}}{\sqrt{r_0 \Delta r_0}} \quad (1.12)$$

avec $r_0 = \|\vec{r}_0\| = \|\vec{r}_\Psi - \vec{r}_{S_0}\|$, $\Delta r_0 = \|\Delta \vec{r}_0\| = \|\vec{r}_R - \vec{r}_{S_0}\|$ et φ_{Ψ_0} l'angle entre \vec{n} et \vec{r}_0 . Ici on observe que la valeur de $P_C(\vec{r}_R, \omega)$ ne dépend plus que du point d'intersection S_0 de la colonne C avec le plan horizontal contenant la source Ψ et le point de réception R . En d'autres termes, la contribution de la colonne C au point R est approchée par la contribution d'un seul monopole placé au *point de phase stationnaire* S_0 (Fig. 1.5). Les conditions de validité de l'approximation de phase stationnaire, qui dépassent le cadre de cette étude, sont décrites dans (Guillemin & Sternberg, 1977). On peut cependant noter de manière générale que cette approximation se dégrade à mesure que kr tend vers des valeurs faibles (Corteel, 2004).

En remplaçant $P_C(\vec{r}_R, \omega)$ par sa valeur approchée dans l'équation 1.8a on forme l'expression du champ de pression $P_L(\vec{r}_R, \omega)$ produit dans Ω_R par la ligne L de sources secondaires :

$$P_L(\vec{r}_R, \omega) = \int_{-\infty}^{\infty} S(\omega) \mathbf{u}_\Psi(x_L, \omega) \frac{e^{-ik\Delta r_0}}{\Delta r_0} dx_L \quad (1.13a)$$

$$\text{avec } \mathbf{u}_\Psi(x_L, \omega) = \left(\frac{\cos \varphi_{\Psi_0}}{\sqrt{r_0}} e^{-ikr_0} \right) \sqrt{\frac{\Delta r_0}{r_0 + \Delta r_0}} \sqrt{\frac{ik}{2\pi}}. \quad (1.13b)$$

L'équation 1.13a traduit le fait que le champ de pression cible $p(\vec{r}_R, \omega)$ émis par une source virtuelle Ψ dans le plan horizontal de Ω_R peut être approché par un champ $p(\vec{r}_R, \omega)$ synthétisé par une ligne infinie de monopoles alimentée par une fonction d'alimentation

$$F_{alim}(x_L, \omega) = S(\omega) \mathbf{u}_\Psi(x_L, \omega), \quad (1.14)$$

où \mathbf{u}_Ψ désigne le filtre WFS associé à la source virtuelle Ψ et $S(\omega)$ représente son signal d'alimentation. Cette synthèse demeure le résultat d'une approximation qui n'est pas sans conséquences sur la validité du champ effectivement reproduit. Une décomposition terme à terme du filtre $\mathbf{u}_\Psi(x_L, \omega)$ est proposée pour permettre de mieux cerner la signification physique des différents termes de l'équation 1.13b et anticiper les caractéristiques du champ reproduit dans l'ensemble de Ω_R :

- Le terme $\left(\frac{\cos \varphi_{\Psi_0}}{\sqrt{r_0}} e^{-ikr_0}\right)$ est lié aux propriétés de la source virtuelle. Le facteur $\cos \varphi_{\Psi_0}$ dénote une dépendance vis à vis de la composante normale du gradient de pression à la position de la source secondaire.
- Le terme $\sqrt{\frac{ik}{2\pi}}$ entraîne une modification globale du champ reproduit dans Ω_R (i.e. indépendante de la position de la source virtuelle et de la position de réception). Le terme \sqrt{k} implique un filtrage passe-haut de pente $+3dB/octave$ du signal émis par les sources secondaires. Le terme \sqrt{i} (qui peut également s'écrire $e^{i(\frac{\pi}{8})\omega}$) indique quant à lui une avance d'un huitième de période pour chaque fréquence. Ceci traduit une *avance* et une *atténuation* des basses fréquences vis-à-vis des hautes fréquences au moment de l'émission (Berkhout, 1988).
- Le terme $\sqrt{\frac{\Delta r_0}{r_0 + \Delta r_0}}$ est potentiellement problématique pour la restitution du champ sur zone étendue du fait qu'il comporte une dépendance vis-à-vis du point d'écoute R. En manipulant ce terme, on obtient :

$$\sqrt{\frac{\Delta r_0}{r_0 + \Delta r_0}} = \sqrt{\frac{|y_R - y_L|}{|y_R - y_\Psi|}}, \quad (1.15)$$

ce qui montre que la dépendance dans la fonction d'alimentation se situe plus précisément dans la profondeur d'écoute y_R . On définit une fonction d'atténuation $g_\Psi(y_R)$ par

$$g_\Psi(y_R) = \sqrt{\frac{|y_R - y_L|}{|y_R - y_\Psi|}} \quad (1.16)$$

Un calcul rapide montre que le facteur $g_\Psi(y_R)$ est faible pour des positions d'écoute très proches de la ligne L et tend asymptotiquement vers 1 lorsqu'on s'éloigne de la ligne. La solution retenue pour lever la dépendance consiste à définir une profondeur d'écoute moyenne $y_{R_{moy}}$ pour laquelle l'atténuation de la source virtuelle est parfaitement restituée (Start, 1997). De ce fait, on explicitera le filtre WFS associé à la synthèse d'une source

ponctuelle monopolaire de la manière suivante :

$$\mathbf{u}_\Psi(x_L, \omega) = \cos \varphi_{\Psi_0} \sqrt{\frac{k}{2\pi}} g_\Psi(y_{R_{\text{moy}}}) \frac{e^{-i[kr_0 - (\frac{T}{8})\omega]}}{\sqrt{r_0}}. \quad (1.17)$$

Il s'en suit une sous-estimation de l'amplitude du champ dans la zone située derrière la ligne $y = y_{R_{av}}$ et une surestimation du gain global dans la zone située devant cette ligne (à proximité de la ligne de sources L). Dans (Corteele, 2004), l'auteur montre comment dériver une fonction d'atténuation de la source virtuelle Ψ à partir de l'équation 1.17, qui s'écrit :

$$Att_\Psi(r_R^{\vec{r}}) = \sqrt{\frac{|y_{R_{av}}|}{|y_R|}} \sqrt{\frac{|y_R| + |y_\Psi|}{|y_{R_{av}}| + |y_\Psi|}} \frac{1}{4\pi d_\Psi^R} \quad (1.18)$$

où d_Ψ^R représente la distance entre la source virtuelle et le point de réception R. Cette expression fait apparaître le fait que l'atténuation d'une source virtuelle WFS peut s'exprimer comme une combinaison de l'atténuation de la source ponctuelle (en $\frac{1}{d_R^R}$) et de la source linéaire (en $\frac{1}{\sqrt{y_R}}$), avec un facteur correctif permettant de définir une profondeur d'écoute privilégiée.

Parallaxe de mouvement

D'un point de vue perceptif, la courbure locale du front d'onde associé à une source virtuelle monopolaire renseigne l'auditeur sur la direction de la source virtuelle. Cependant, en absence d'effet de salle additionnel, et si la distance auditeur/source est maintenue constante, la courbure à elle seule ne constitue pas un estimateur viable de distance absolu de la source (Witek, 2004). Pour pouvoir estimer correctement la position d'une source en absence d'effet de salle, l'auditeur doit pouvoir déambuler dans la zone de reproduction. Dans ce cas, les indices directionnels fournis par la courbure du front d'onde permettent d'estimer indirectement la position de la source, alors que les variations dynamiques de niveau soulignent l'impression de distance⁶. Dans des scènes sonores composées de sources situées à différentes profondeurs, un auditeur pourra ressentir une *parallaxe de mouvement* (cf. glossaire) cohérente avec la position des différentes sources lors de ses déplacements dans la zone d'écoute (Fig. 1.6). Les variations de perspective perçues lors des déplacements de l'auditeur vont alors contribuer à renforcer sa sensation d'immersion au sein d'une scène sonore.

L'existence de différents points de vue possibles sur une scène sonore implique par ailleurs un changement de paradigme pour le mixage audio. En effet, il ne s'agit plus de se placer au

⁶Bien que la loi d'atténuation d'une sources virtuelle ne soit pas parfaitement conforme à celle d'une source ponctuelle idéale, et ce surtout lorsque la source virtuelle est située loin du banc de haut-parleurs, elle demeure tout de même cohérente avec la position de la source [ou "quasi-naturelle" (Witek, 2004)] au sens ou un rapprochement de la source engendre une augmentation du niveau perçu et un éloignement engendre une chute du niveau perçu.

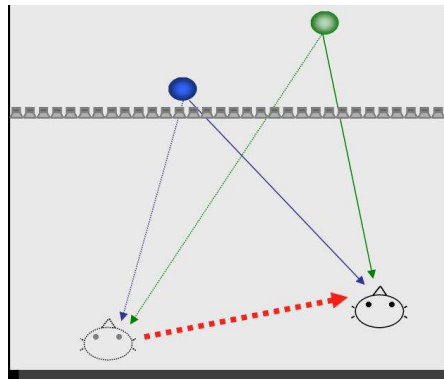


FIG. 1.6 – Mise en évidence de la parallaxe de mouvement. Les points colorés situés derrière le banc de haut-parleurs correspondent à des sources virtuelles monopolaires fixes.

centre du lieu d'écoute et de mixer de manière "égocentrée" comme c'est le cas pour les techniques basées sur la stéréophonie ; le mixage de scènes sonores en WFS implique un mixage "exocentré", où le compositeur doit se déplacer dans l'espace de manière à ajuster les paramètres ne pouvant être appréciés que dans le cadre d'une navigation (Nogues *et al.*, 2003; Warusfel *et al.*, 2004a).

Symétrie axiale du banc - Implications pour la localisation

De par la géométrie linéaire du banc WFS et l'omnidirectivité des sources qui le composent, le champ effectivement rayonné par le banc de haut-parleurs présente une symétrie de révolution autour de son axe. Cette symétrie implique en particulier que la source virtuelle est toujours perçue comme étant située dans le plan contenant le banc de haut-parleurs et la tête de l'auditeur. Ainsi, si l'auditeur se situe en dehors du plan horizontal, la position apparente de la source virtuelle se décale verticalement (Fig. 1.7) et entraîne une variation du site perçu, accompagnée éventuellement d'un très léger décalage en azimuth⁷.

L'utilisation de la WFS pour le renforcement sonore de sources instrumentales sur scène requiert typiquement de placer le banc de haut-parleurs en dehors du plan horizontal pour ne pas gêner la visibilité des sources (cf. Fig. 1.4, Fig. 1.8). Dans (Start, 1997), l'auteur montre que dans la plupart des cas où le banc de haut-parleurs est surélevé (Fig. 1.8, haut), la source virtuelle peut-être assimilée à une réflexion provenant du plafond, dont l'azimuth se confond avec la source à amplifier. De ce fait, et d'après les travaux de Hartmann (Hartmann, 1983), la source virtuelle aura donc tendance à renforcer la capacité de localisation azimuthale de l'auditeur. Par contre, un glissement de localisation vers le haut risque de s'opérer si le niveau d'intensité de la source virtuelle excède largement le niveau d'intensité de la source instrumentale. Cet

⁷Le décalage en azimuth, qui apparaît si l'auditeur et la source ne sont pas alignés selon une droite perpendiculaire au banc, peut-être négligé dans la plupart des cas pratiques (Start, 1997)

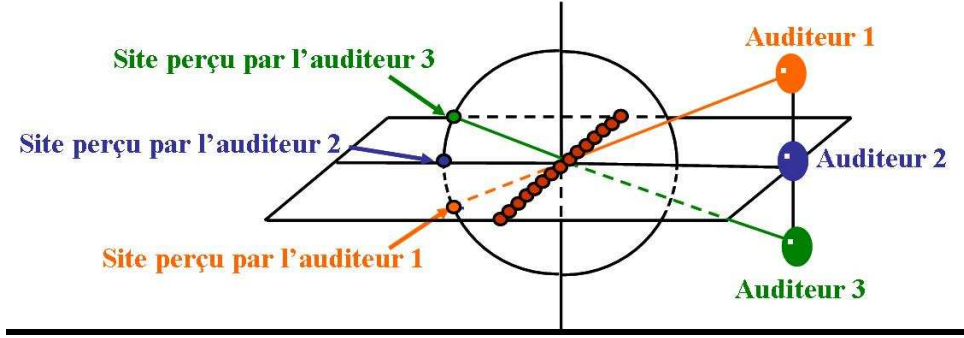


FIG. 1.7 – Mise en évidence de la variation de site apparent de la source virtuelle lorsque l'auditeur se situe en dehors du plan horizontal.

effet de délocalisation pourra toutefois être diminué en appliquant un retard à la source virtuelle (2-20ms) en vertu de l'effet de précedence⁸.

La symétrie de révolution du rayonnement émis par le banc influe également sur l'effet de salle précoce associé aux sources virtuelles, ce qui est mis en évidence au chapitre 5.

1.3.5 Cas de la source focalisée

Une extension du formalisme exposé ci-dessus permet de synthétiser le champ sonore associé à des sources virtuelles situées à l'intérieur du sous-espace de reproduction et désignées par le terme *sources focalisées* (cf. glossaire). Ces sources, qui sortent du cadre prévu par le principe de Huygens⁹, sont synthétisées en opérant un retournement temporel du front d'onde émis par la source et capté par les haut-parleurs. Ceci revient à faire converger des fronts d'ondes au point correspondant à la position de la source virtuelle (Fig 1.9, droite). En aval de ce point, les fronts d'onde propagés sont identiques, aux artefacts de reconstruction près, à ceux qui seraient émis par une source se situant au point de focalisation. En d'autres termes, la position apparente de la source virtuelle perçue par un auditeur situé en aval est cohérente avec le point de focalisation.

Le filtre WFS associé à la synthèse d'une source focalisée peut s'exprimer de la manière suivante :

$$\mathbf{u}_{\Psi}(x_L, \omega) = \cos \varphi_{\Psi_0} \sqrt{\frac{k}{2\pi}} g_{\Psi}(y_{R_{moy}}) \frac{e^{-i[(\frac{T}{8})\omega - kr_0]}}{\sqrt{r_0}}. \quad (1.19)$$

La différence principale entre une sources focalisée et une source non-focalisée, pour une distance égale au banc de haut-parleurs, se situe dans l'inversion de phase (retards opposés appliqués au haut-parleurs). Ainsi, le terme $e^{-i(\frac{T}{8})\omega}$ indique cette fois un *retard* des basses fréquences

⁸Le fonctionnement de l'effet de précedence dans le plan vertical est mis en évidence dans (Somerville *et al.*, 1966)

⁹du fait qu'elles se situent en dehors du sous-espace des sources Ω_{Ψ} , cf. sec. 1.2.2

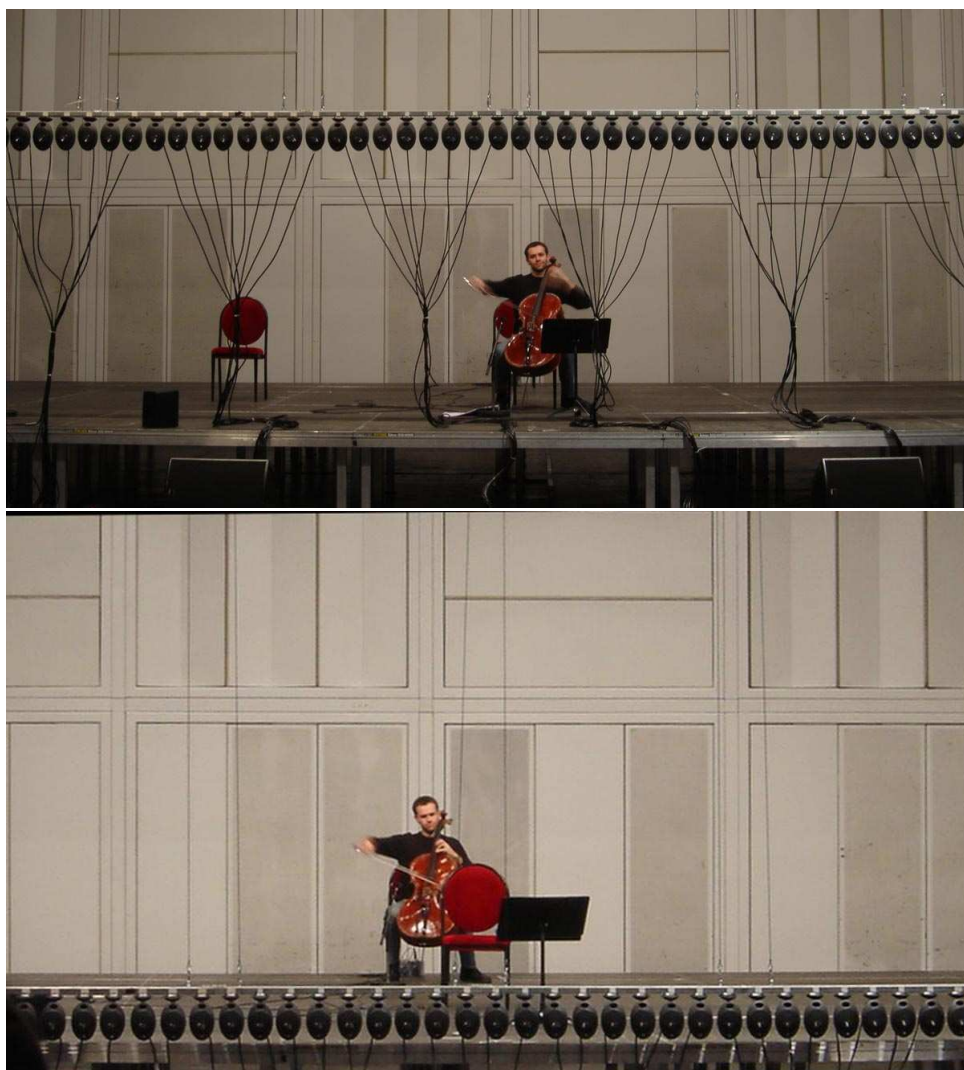


FIG. 1.8 – Session de travail dans l’Espace de Projection avec le violoncelliste Alexis Descharmes. Un système WFS 48 canaux avec haut-parleurs KEF est placé à différentes hauteurs de manière à tester, entre autres, les limites du niveau d’amplification au delà desquels surgissent des problèmes de feedback (effet Larsen), ainsi que la cohérence entre le musicien et la source virtuelle en situation de renforcement sonore.

par rapport aux hautes fréquences.

1.3.6 Cas de l’onde plane

A partir du banc WFS on peut également synthétiser une approximation d’onde plane dans le plan horizontal, en formant des fronts d’ondes de rayon de courbure infini. Dans le cas le plus simple, qui est celui de l’onde plane se propageant perpendiculairement au banc de haut-parleurs, il suffit d’appliquer un gain et un retard constant à tous les haut-parleurs. Le banc

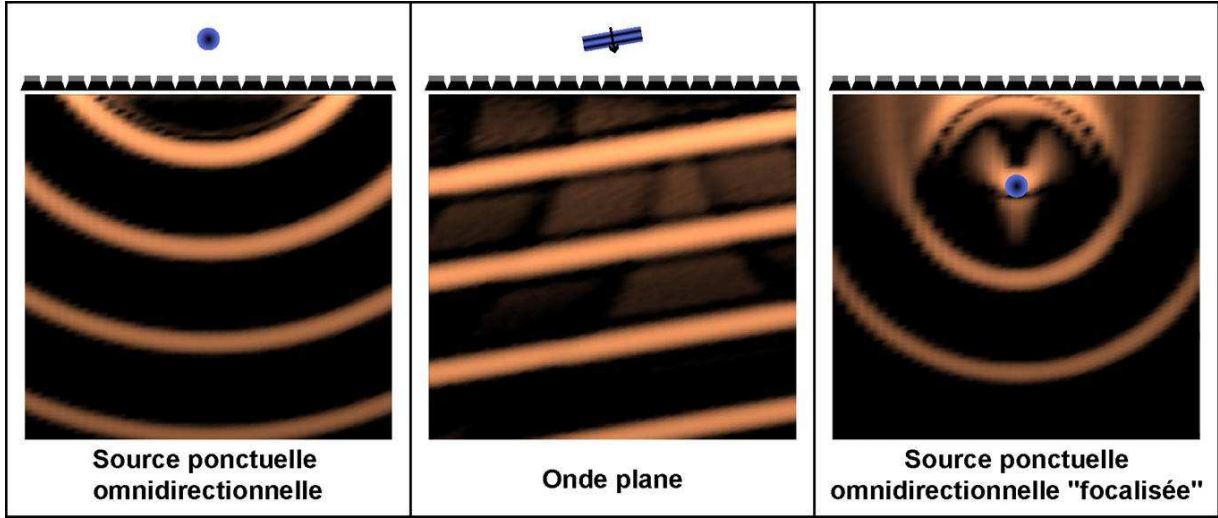


FIG. 1.9 – Evolution spatiale des fronts d'ondes émis par un banc de haut-parleurs WFS pour trois types de sources virtuelles.

revêt alors un comportement de source linéaire, dont les propriétés sont bien décrites dans la littérature. Dans ce cas, au lieu de générer le champ associé à l'onde plane idéale, qui serait de la forme :

$$p(y_R, k) = Ae^{-iky_R}, \quad (1.20)$$

avec A réel et y_R représentant la profondeur d'écoute, on génère un champ de la forme :

$$p(y_R, k) = AH_0^1(ky_R), \quad (1.21)$$

où H_0^1 représente la fonction de Hankel cylindrique de première espèce d'ordre 0. En champ lointain ($ky_R \gg 1$), on peut donc approximer le champ associé à l'onde plane perpendiculaire au banc par l'expression suivante (Williams, 1999) :

$$p(y_R, k) = A\sqrt{\frac{2}{\piiky_R}}e^{i(ky_R)}. \quad (1.22)$$

Dans (Caulkins *et al.*, 2004), l'auteur montre que le banc linéaire alimenté ainsi fournit une bonne approximation de l'onde plane pour une position de réception lointaine, au facteur d'atténuation en $\frac{1}{\sqrt{y_R}}$ près. En grande proximité du banc ($y_R \leq 10\text{cm}$) on observe cependant une chute de niveau de l'ordre de -10dB pour les fréquences entre 20Hz et 100Hz. Néanmoins, ce problème reste cantonné aux très basses fréquences dès que l'on s'éloigne du banc, et demeure largement en dessous du seuil de perception dans le reste de la zone d'écoute (Corteel, 2004).

Contrairement aux sources virtuelles ponctuelles, qui sont perçues comme étant localisées à une **position** fixe, les ondes planes sont perçues comme provenant d'une **direction** fixe. Lors

de ses déplacements dans la zone d'écoute, l'auditeur a donc l'impression que le signal diffusé par onde plane le "suit". Ce type d'effet revêt un intérêt particulier pour les installations sonores muséographiques dans lesquelles l'auditeur est invité à naviguer à l'intérieur l'espace d'écoute.

1.3.7 Conséquences de la troncature du banc de haut-parleurs

En pratique, la distribution de sources secondaires utilisée dans un système de reproduction réel n'est pas une ligne infinie, mais un segment. Cette réduction géométrique par rapport à la situation idéale entraîne principalement deux conséquences sur le champ effectivement rayonné dans la zone d'écoute :

- D'une part, la troncature de la ligne de sources secondaires entraîne une réduction de la "visibilité" de la source virtuelle à travers la fenêtre acoustique que constitue le banc de haut-parleurs, illustrée à la figure 1.10. Sur cette figure, les zones sombres correspondent à des "zones d'ombre" dans lesquelles les sources virtuelles ne sont pas correctement reproduites. Dans le cas de la source focalisée (Fig. 1.10, droite), la zone de visibilité est d'autant plus réduite que l'espace situé entre le banc de haut-parleurs et la source correspond à la zone de focalisation, où les fronts d'ondes synthétisés ont une courbure inversée.

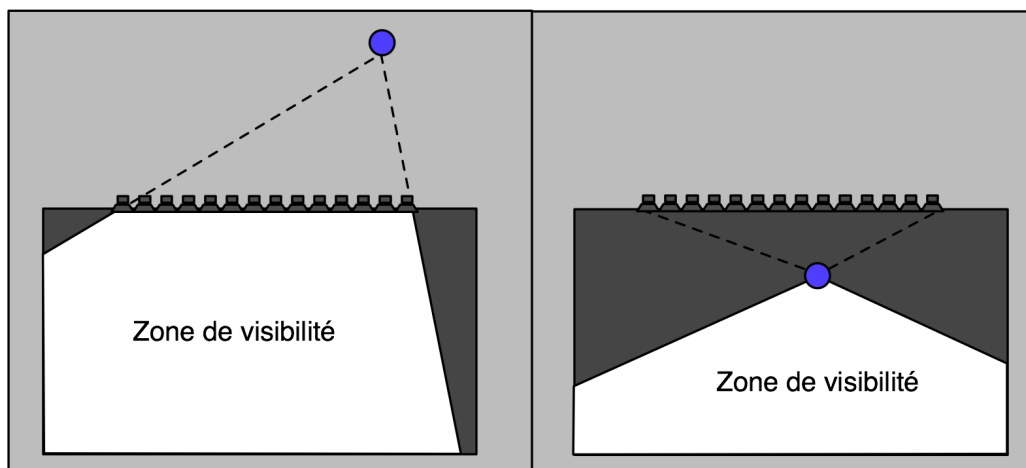


FIG. 1.10 – Mise en évidence de la zone de visibilité d'une source virtuelle focalisée (droite) et non-focalisée (gauche). La source virtuelle est indiquée par un cercle bleu. La zone sombre correspond à la portion de de l'espace de reproduction dans laquelle la source n'est pas reproduite correctement.

- D'autre part, la troncature de la ligne de sources secondaires entraîne l'apparition d'ondes de diffraction qui suivent le front d'onde principal. Pour visualiser le phénomène, on simule la synthèse d'une onde plane en basses fréquences [20-1000Hz] par un banc de

haut-parleurs monopolaires de taille finie (Fig. 1.11, gauche). Le résultat est mesuré sur une ligne de microphones virtuels situés dans la zone de restitution. Le graphique de

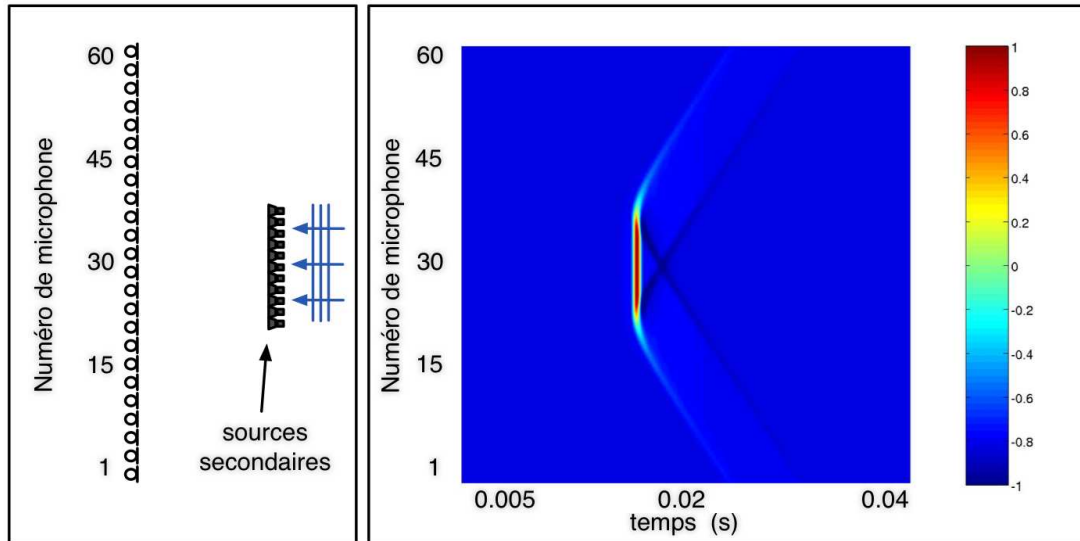


FIG. 1.11 – Mise en évidence de l'effet de diffraction pour un banc de haut-parleurs de taille finie. La figure de gauche représente la configuration de simulation, dans laquelle un banc de haut-parleurs WFS synthétise une onde plane d'incidence perpendiculaire au banc. L'évolution temporelle de l'amplitude du champ capté par un ensemble de microphones situés dans la zone d'écoute est représentée sur le graphique de droite selon un échelle linéaire.

droite de la figure 1.11 montre l'évolution temporelle du champ capté par la ligne de microphones. On observe l'apparition de deux fronts d'ondes parasites qui suivent le front d'onde principal et paraissent provenir de sources secondaires qui seraient situées à chaque extrémité du banc de haut-parleurs. Cet effet de diffraction a été étudié en détail dans (Nicol, 1999; Start, 1997), et les auteurs montrent que les artefacts de diffraction peuvent effectivement être modélisés formellement par les contributions de deux sources de directivité complexe situées aux extrémités du banc de haut-parleurs. La méthode classiquement employée pour réduire ces artefacts consiste à atténuer les filtres WFS associés aux haut-parleurs situés aux extrémités. Cette méthode présente toutefois le désavantage de réduire le niveau des basses fréquences pour des positions d'écoute excentrées (Cor-teeel, 2004).

La troncature de la ligne de sources secondaires provoque par ailleurs une modification de la loi d'atténuation associée aux sources WFS, qui passe graduellement d'un comportement de source linéaire (atténuation d'amplitude en $1/\sqrt{r}$) en champ proche à un comportement de source ponctuelle (atténuation d'amplitude en $1/r$) en champ lointain.

1.3.8 Conséquences de l'échantillonnage spatial du banc de haut-parleurs

Outre la troncature de la ligne de sources secondaires, une implémentation pratique de la WFS exige un échantillonnage spatial du continuum de sources secondaires. En effet, l'état de l'art actuel ne permet pas de fabriquer une ligne continue de monopoles permettant de synthétiser un champ sonore directement suivant l'équation 1.13a. Il faut par conséquent opérer une discrétisation de cette équation, impliquant un repliement de spectre spatial dans le champ effectivement reproduit dans la zone d'écoute (Start, 1997; Nicol, 1999). Ces auteurs montrent que la fréquence de repliement spatial¹⁰ dépend à la fois de la position de la source et de l'espacement des haut-parleurs.

La figure 1.12 met en évidence le phénomène de repliement de spectre spatial pour un banc de haut-parleurs linéaire synthétisant une source virtuelle décentrée, trois mètres derrière le banc. En observant les réponses impulsionnelles du banc, on voit que le front d'onde émis par le banc de haut-parleur demeure compact en BF, en-dessous de 1.5kHz (Fig. 1.12.c). Cependant, en HF (Fig. 1.12.d) on observe un étalement temporel du champ, qui traduit une incohérence entre les ondes émises par les haut-parleurs individuels. Malgré tout, ces artefacts de repliement succèdent au front d'onde basses fréquences dans la réponse, ce qui implique que le premier front d'onde qui arrive aux oreilles de l'auditeur demeure bien cohérent avec la position de la source virtuelle.

Dans le domaine fréquentiel (Fig. 1.12.b), le champ cible est correctement reproduit jusqu'à la fréquence de repliement, au-delà de laquelle apparaissent des artefacts de repliement de spectre. En dehors de la zone de visibilité (microphones [41-60]) et entre 500Hz et 1500Hz, le champ reproduit s'annule. Cette zone espace-fréquence correspond à la zone où la fréquence est suffisamment haute pour éviter les phénomènes de diffraction en BF, et suffisamment basse pour éviter les phénomènes de repliement du spectre en HF. On note par ailleurs que la fréquence de repliement varie en fonction de la position d'écoute. Cette variation est liée au fait que l'échantillonnage **spatial régulier** de la distribution de sources secondaires se traduit par un échantillonnage **temporel irrégulier** de la forme d'onde associée à la source virtuelle (Fig. 1.13), dont les caractéristiques varient selon la position d'écoute. Dans (Corteel, 2004), l'auteur montre que l'on peut exprimer la fréquence de repliement à une position donnée pour une source virtuelle Ψ donnée par la formule suivante :

$$f_{al}(\vec{r}, \Psi) = \frac{1}{\max_{l=1, \dots, L-1} (|t_{l+1}(\vec{r}, \Psi) - t_l(\vec{r}, \Psi)|)} \quad (1.23)$$

où $\max_{l=1, \dots, L-1} (|t_{l+1}(\vec{r}, \Psi) - t_l(\vec{r}, \Psi)|)$ représente la différence de temps maximale entre deux contributions, en considérant l'ensemble des contributions successives émises par les haut-

¹⁰i.e. limite fréquentielle à partir de laquelle des phénomènes de repliement de spectre spatial se produisent

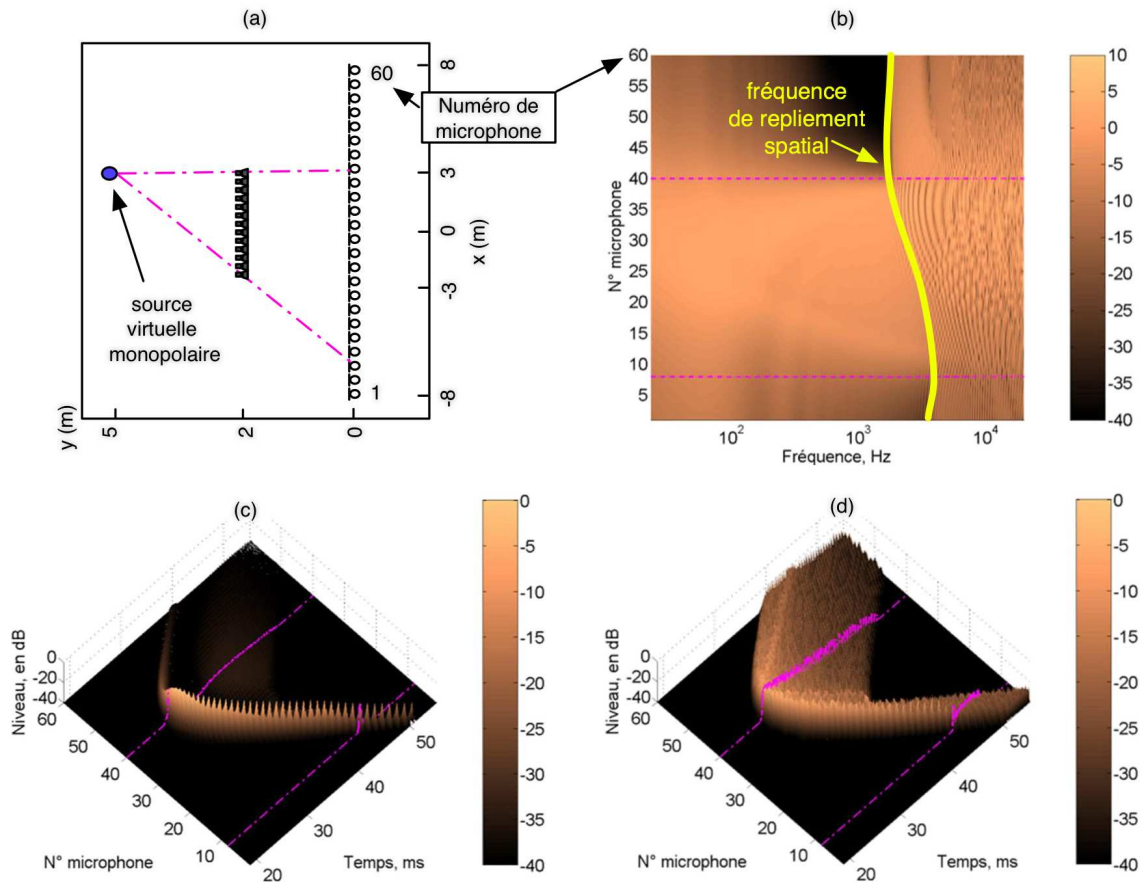


FIG. 1.12 – Mise en évidence de l'effet de repliement spatial pour un banc linéaire. **(a)** : Configuration utilisée. Le banc de 60 haut-parleurs mesure 6m de long, la distance entre chaque haut-parleur est de 10cm, et on mesure sa réponse sur un banc de 60 microphones espacés de 15cm chacun. La source virtuelle synthétisée est un monopole situé 3m derrière le banc de haut-parleurs, 3m sur la droite, et sa zone de visibilité est délimitée par des traits pointillés mauves sur l'ensemble des figures. **(b)** : Magnitude (dB) de la réponse fréquentielle mesurée au niveau des microphones. La fréquence de repliement est indiquée par une courbe jaune. **(c)** : Niveau (dB) de la réponse impulsionnelle mesurée au niveau des microphones, filtré passe-bas en dessous de 1.5kHz. **(d)** : Niveau (dB) de la réponse impulsionnelle mesurée au niveau des microphones et filtrée passe-haut au dessus de 1.5kHz.

parleurs du banc WFS à la position d'écoute \vec{r} . Il est intéressant de noter que la fréquence de repliement spatial tend à augmenter avec la compacité du support temporel du front d'onde. Ainsi, la fréquence de repliement associée à des sources monopolaires et des ondes planes est plus élevée pour des positions d'écoutes lointaines, car les contributions du banc tendent à être "vues" comme émanant d'un seul point, d'où une cohérence accrue entre les contributions individuelles.

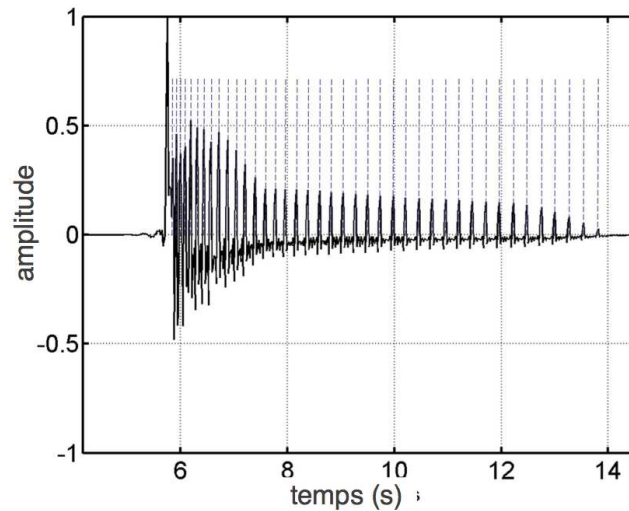


FIG. 1.13 – Echantillonnage irrégulier de la forme d'onde temporelle associée à une source virtuelle synthétisée en WFS.

Impact perceptif du repliement spatial

D'un point de vue perceptif, les performances mesurées pour la localisation de sources virtuelles large bande synthétisés par WFS sont équivalentes à celles mesurées pour des sources réelles, à condition que la fréquence de repliement excède 1.5kHz. De même, dans ces conditions, la résolution angulaire de localisation dans des champs synthétisés en WFS est équivalente à celle mesurée dans des champs réels (Vogel, 1993; Start, 1997). On attribue cette qualité en termes de performances de localisation au fait que le système WFS restitue correctement l'ITD¹¹ en dessous de la fréquence de repliement. Or, on montre que pour des signaux large bande, l'ITD domine à la fois l'ILD¹² et les indices monauraux liés au filtrage effectué par l'oreille externe (Wightman & Kistler, 1992).

Les artefacts de repliement au-dessus de la fréquence de repliement engendrent une augmentation de la taille apparente de la source¹³ (Start, 1997) ainsi qu'une *coloration* (cf. glossaire) du signal original, qui s'apparente à un filtrage en peigne. Ce phénomène a d'autant plus d'impact que la coloration du signal varie avec la position, ce qui implique une gêne supplémentaire lors de déplacements dans la zone d'écoute. Cette gêne est objectivée au moyen d'un indice permettant de caractériser la variation de coloration sur une zone étendue, le SCVI (Spatial Coloration Variation Index) (DeBruijn, 2004). On a montré récemment la possibilité de réduire ces varia-

¹¹Interaural Time Difference, i.e. différence interaurale de temps. On regroupe dans l'ITD deux indices différents : l'IPD (Interaural Phase Difference - différence interaurale de phase) qui est perceptible jusqu'à 1500Hz environ, et l'IED (Interaural Envelope Difference - différence interaurale d'enveloppe) qui permet de détecter une différence de temps d'arrivée aux deux oreilles pour des signaux d'attaque brève au-delà de 1500Hz.

¹²Interaural Level Difference, i.e. différence interaurale de niveau

¹³ASW - Apparent Source Width (voir Glossaire)

tions spatiales de coloration en utilisant des filtres diffus en hautes fréquences, au prix d'une augmentation de la coloration à position fixe [mesurée par l'indice D décrit dans (Moore & Tan, 2004)] et d'un élargissement perceptif de la source (Corteel *et al.*, 2007). D'autres études sont actuellement en cours pour tenter de mieux cerner l'impact des erreurs de reproduction liées au repliement de spectre spatial sur la qualité de restitution des indices binauraux de localisation (ITD, ILD) en zone étendue (Sanson, 2007).

1.3.9 Egalisation du son direct en conditions réelles

Les deux sections précédentes (1.3.7 et 1.3.8) ont mis en évidence des artefacts inhérents à la synthèse d'un champ sonore en WFS, i.e. les effets de champ proche, la diffraction et de repliement spatial. Dans des conditions réelles, la directivité non-idéale des transducteurs employés vient se rajouter à ces artefacts et implique d'envisager une égalisation du système. Cette étape d'égalisation est d'autant plus nécessaire que les haut-parleurs utilisés présentent des caractéristiques de directivité complexes, ou encore une grande variabilité inter-transducteur, ce qui est typiquement le cas pour les haut-parleurs MAP (Multi-Actuator Panels).

Les haut-parleurs MAP sont utilisés dans le cadre de la WFS pour leur qualités de *transparence acoustique* (cf. glossaire) (Corteel *et al.*, 2007), ainsi que pour des qualités esthétiques évidentes (notamment le fait que leur forme plate (Fig. 1.14) permet de les dissimuler facilement dans l'environnement existant¹⁴). Ces haut-parleurs, qui sont basés sur la technologie DML

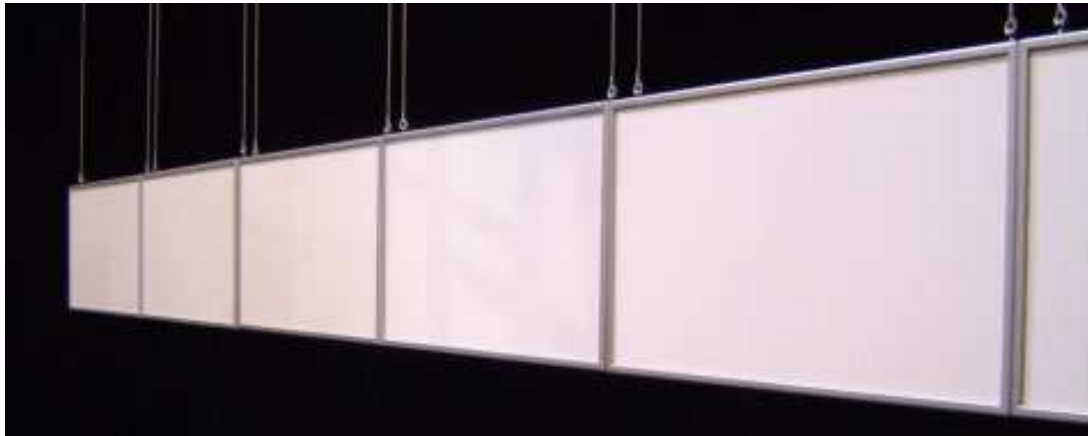


FIG. 1.14 – Haut-parleurs MAP.

(Distributed Mode Loudspeaker) sont constitués d'une plaque rigide derrière laquelle on fixe entre 4 et 8 transducteurs. Une étude menée par STUDER (Corteel *et al.*, 2002) a montré que

¹⁴La possibilité de d'intégration dans l'environnement existant ("seamless integration") offerte par les MAP constitue un argument fort pour leur utilisation dans le cadre de d'installations sonores muséographiques (Grand *et al.*, 2005; Chiron *et al.*, 2007).

le niveau de distorsion mesuré pour un excitateur MAP est suffisamment faible (moins de 1%) dans la bande passante pour pouvoir assimiler ces haut-parleurs à des systèmes multilinéaires. Les haut-parleurs MAP présentent toutefois des caractéristiques de directivité complexes et variant d'un excitateur à un autre. Ceci apparaît clairement sur la figure 1.15, qui présente les réponses fréquentielles de deux excitateurs adjacents d'un MAP mesurées le long d'une ligne située 2m devant les haut-parleurs.

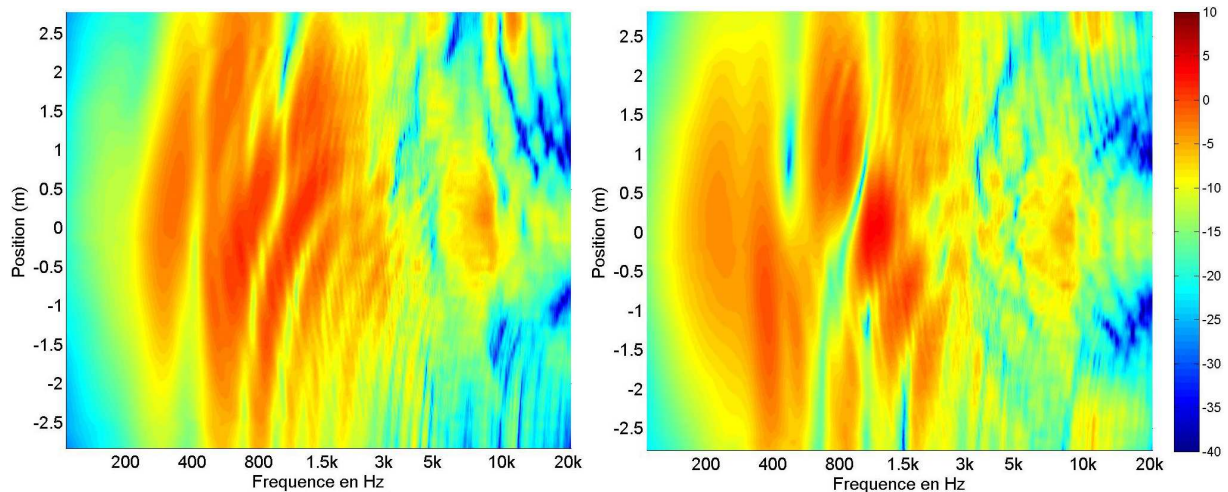


FIG. 1.15 – Réponses fréquentielles de deux excitateurs adjacents d'un haut-parleur MAP mesurées en champ libre le long d'une ligne de microphones espacés de 10cm située 2m devant les haut-parleurs.

Etant donné ces caractéristiques de directivité complexes, un processus de traitement de signal permettant de contrôler et réduire la coloration du champ effectivement reproduit par le système WFS a été mis en place (Corteel *et al.*, 2002) puis perfectionné par la suite (Corteel, 2004). Des méthodes plus simples, ne nécessitant pas une procédure d'inversion multicanale, ont également été mises au point pour traiter les situations moins critiques où le banc WFS est constitué de haut-parleurs présentant des caractéristiques de directivité plus régulières. Trois méthodes d'égalisation de complexité croissante seront donc brièvement décrites ici ; pour plus de détails sur ces méthodes le lecteur est invité à consulter (Corteel, 2004; Corteel, 2006a; Corteel, 2007). Notons que ces méthodes visent uniquement à égaliser la composante de son direct émise par le banc de haut-parleurs ; en d'autres termes, on s'intéresse uniquement au rayonnement du système WFS en champ libre. D'autres méthodes qui tentent de prendre en compte et d'annuler les artefacts liés à la salle d'écoute (éventuellement de manière active) sont décrites dans (Spors *et al.*, 2003; Corteel & Nicol, 2003; Spors *et al.*, 2004a; Spors *et al.*, 2004b; Spors *et al.*, 2007; Gauthier & Berry, 2005; Gauthier & Berry, 2006). Ces méthodes s'appuient cependant sur des configurations 2D de haut-parleurs ce qui implique l'impossibilité de réaliser une compensation parfaite dans l'environnement 3D d'une salle d'écoute, notamment à cause du

repliement de la composante verticale liée aux réflexions sur le sol/plafond (Corteel, 2004). Dans le cadre de ce travail, on se restreint volontairement à une description de techniques d'égalisation permettant d'ajuster le son direct de manière privilégiée. La caractérisation/compensation de l'effet de salle est abordée de manière dé耦plée au problème de rendu du son direct aux chapitres 4 et 5.

Principe général

On considère un système WFS constitué de I haut-parleurs repérés par \vec{r}_i destinés à synthétiser un champ sonore dans un espace d'écoute dans lequel on a placé J microphones repérés par \vec{r}^j (Fig. 1.16). On cherche à synthétiser une source virtuelle Ψ alimentée par un signal $S(\omega)$. Une cible \mathbf{a}_Ψ est définie à chaque position de microphone par :

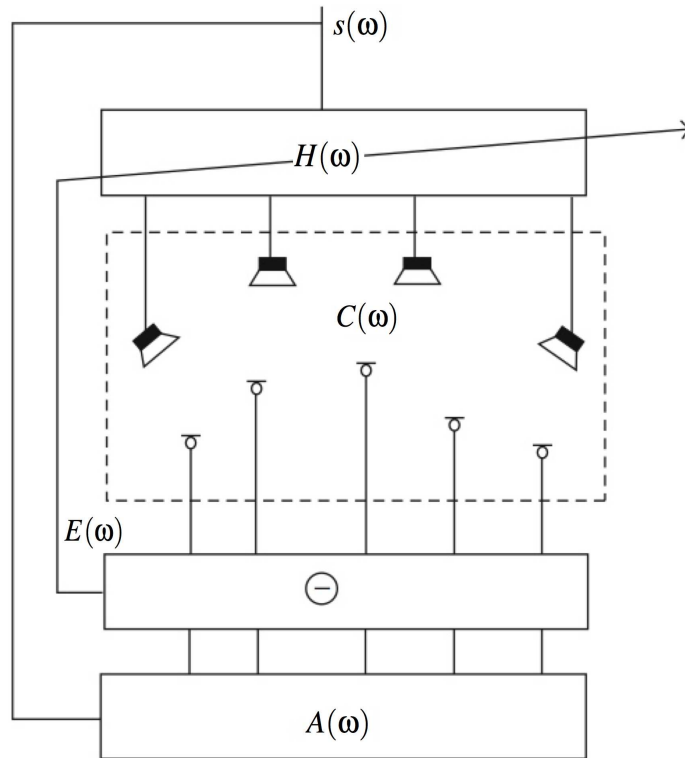


FIG. 1.16 – Principe de l'égalisation multicanale appliquée à la synthèse de champs sonores.

$$\mathbf{a}_\Psi(\vec{r}^j, \omega) = Att_\Psi(\vec{r}^j) e^{-jk\|\vec{r}^j - \vec{r}_\Psi\|} \quad (1.24)$$

où $Att_\Psi(\vec{r}^j)$ représente l'atténuation d'une source virtuelle synthétisée par un banc linéaire continu de longueur infinie (eq. 1.18) et $\|\vec{r}^j - \vec{r}_\Psi\|$ représente la distance entre le microphone

considéré et la source virtuelle. \mathbf{a}_Ψ correspond donc à la TFTD¹⁵ de la réponse impulsionnelle "idéale" qu'aurait un banc linéaire continu de longueur infinie aux positions des microphones. Cette cible est valable pour une source virtuelle monopolaire ; dans le cas où l'on cherche à reproduire une source de directivité plus complexe, on rajoutera simplement une fonction de directivité dans son expression. Les J cibles idéales sont répertoriées dans une matrice A .

On définit une matrice C qui regroupe l'ensemble des fonctions de transfert $c_i^j(\omega)$ entre le haut-parleur i et le microphone j . La matrice C , qui décrit entièrement les canaux de transmission du système haut-parleurs/microphones, peut être considérée comme la base d'un système MIMO (Multiple Input Multiple Output) dont les entrées sont les I signaux fournis aux haut-parleurs et les sorties sont les J signaux captés par les microphones.

Le principe de l'égalisation consiste alors à trouver un ensemble de I filtres $\mathbf{h}_i(\omega)$ regroupés dans une matrice H tels que l'erreur

$$E(\omega) = A(\omega) - C(\omega)H(\omega) \quad (1.25)$$

entre le champ synthétisé et le champ cible soit minimisée suivant un critère de distance convenable (Fig. 1.16).

Egalisation individuelle

Dans cette approche du filtrage, on cherche uniquement à corriger la réponse individuelle des haut-parleurs avant l'étape de synthèse du champ sonore ; aucune correction tenant compte de l'objectif global de la reproduction n'est appliquée. Chaque haut-parleur du banc est mesuré en champ libre selon une ligne de microphones parallèle au banc. On sélectionne les J microphones contenus dans un angle plan de 60 degrés autour de l'axe principal du haut-parleur (Fig. 1.17). Le filtre d'égalisation individuel \mathbf{h}_i^{Ind} du haut-parleur i se calcule alors en moyennant les

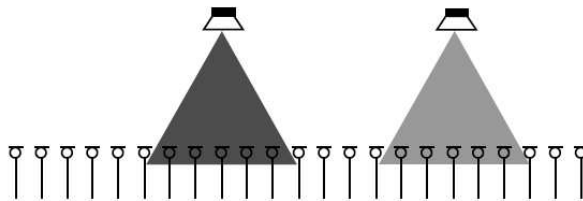


FIG. 1.17 – Sélection de microphones pour la procédure d'égalisation individuelle de microphones.

¹⁵Dans toute la suite du document, l'utilisation d'une fonte en gras indique que l'on manipule la Transformée de Fourier à Temps Discret de la réponse impulsionnelle.

J réponses (Corteeel, 2007) :

$$\mathbf{h}_i^{Ind}(\omega) = J \times \sum_{j=1}^J \frac{\|\vec{r}_i - \vec{r}^j\|}{\mathbf{c}_i^j(\omega)}. \quad (1.26)$$

Le facteur $\|\vec{r}_i - \vec{r}^j\|$ est introduit de manière à compenser l'effet de la propagation entre le haut-parleur et le microphone dans chaque réponse mesurée. Notons qu'avant ce calcul de réponses moyennées, on lisse les réponses fréquentielles $\mathbf{c}_i^j(\omega)$ suivant un procédé non-linéaire décrit dans (Corteeel *et al.*, 2002) de manière à garder les pics mais à éliminer les creux qui peuvent être problématiques du point de vue de l'égalisation. Les filtres \mathbf{h}_i^{Ind} finaux sont typiquement des filtres RIF (Réponse Impulsionnelle Finie) à minimum de phase de longueur 800 échantillons pour une fréquence d'échantillonnage à 48kHz.

Egalisation individuelle avec compensation de l'erreur moyenne de synthèse

Dans la méthode d'égalisation individuelle avec compensation d'erreur moyenne, on rajoute une étape de correction globale du champ synthétisé de manière à corriger, en plus des imperfections liées à la directivité des haut-parleurs, les erreurs inhérentes à la technique WFS décrites dans les sections précédentes. Pour ce faire, on définit une fonction de qualité \mathbf{q}_{Ψ}^{Ind} décrivant les erreurs de synthèse à chaque position de microphone :

$$\mathbf{q}_{\Psi}^{Ind}(\vec{r}^j, \omega) = \frac{\sum_{i=1}^J \mathbf{c}_i^j(\omega) \times \mathbf{h}_i^{Ind}(\omega) \times \mathbf{u}_{\Psi}(x_i, \omega)}{\mathbf{a}_{\Psi}(\vec{r}^j, \omega)} \quad (1.27)$$

où \mathbf{u}_{Ψ} représente le filtre WFS appliqué au haut-parleur situé en $\vec{r}_i = x_i\vec{x} + y_i\vec{y} + z_i\vec{z}$. On modifie alors les filtres individuels pour tenir compte de l'erreur de synthèse globale, ce qui permet d'obtenir les filtres suivants :

$$\mathbf{h}_{i,\Psi}^{comp}(\omega) = \frac{J \times \mathbf{h}_i^{Ind}(\omega)}{\sum_{i=1}^J \mathbf{q}_{\Psi}^{Ind}(\vec{r}^j, \omega)} \quad (1.28)$$

Un lissage non-linéaire est appliqué au facteur de qualité avant l'étape de déconvolution. Les filtres $\mathbf{h}_{i,\Psi}^{comp}$ finaux sont typiquement des filtres RIF (Réponse Impulsionnelle Finie) à minimum de phase de longueur 800 échantillons pour une fréquence d'échantillonnage à 48kHz.

Egalisation basée sur une inversion multicanale

Le principe de l'inversion multicanale est de trouver la matrice de filtrage H qui minimise la valeur de l'erreur E définie à l'équation 1.25. Si on se place dans le domaine fréquentiel et

on calcule les filtres optimaux $H_0^{Meq}(\omega)$ au sens des moindres carrés on obtient

$$H_0^{Meq}(\omega) = [C^{*T}(\omega)C(\omega)]^{-1}C^{*T}(\omega)A(\omega) \quad (1.29)$$

où $[C^{*T}(\omega)C(\omega)]^{-1}C^{*T}(\omega)$ représente la pseudo-inverse de la matrice C . Si la matrice $C^{*T}(\omega)C(\omega)$ admet des singularités, le problème sera mal conditionné, d'où l'introduction d'une matrice de régularisation dépendante de la fréquence $B(\omega)$ et d'un gain de régularisation γ tels que :

$$H_{0,reg}^{Meq}(\omega) = [C^{*T}(\omega)C(\omega) + \gamma B^{*T}(\omega)B(\omega)]^{-1}C^{*T}(\omega)A(\omega). \quad (1.30)$$

La matrice B permet d'éviter des pics de facteur de qualité élevés dans les filtres calculés (pouvant causer des surcharges des haut-parleurs), au prix d'une introduction d'erreurs dans le résultat final. Le facteur γ permet quant à lui de régler l'équilibre entre l'inversibilité de la matrice et la qualité du résultat final.

Le problème des méthodes fréquentielles réside dans la circularité de la FFT, qui peut introduire des artefacts dans les filtres calculés (Norcross *et al.*, 2004). Pour cette raison, on fera appel à une méthode temporelle plus coûteuse d'un point de vue calcul mais permettant de circonvier les problèmes inhérents aux méthodes fréquentielles. Le fait que les temps de calcul soient augmentés demeure un problème secondaire, car l'implémentation visée n'est pas une solution de type "active" ou "closed-loop". En effet, le calcul des filtres s'effectue dans un processus qui n'est pas soumis aux contraintes du rendu en temps-réel ; ils sont ensuite stockés dans une base de données et rappelés au moment de la restitution¹⁶. On tolère donc des temps de calcul assez importants, ce qui permet d'envisager des méthodes d'inversion temporelles.

Les réponses impulsionnelles du système MIMO $c_j^i(t)$ sont modifiées avant l'étape d'inversion multicanale en leur affectant les composantes de retard pur $k_{i,\Psi}(t)$ extraites des filtres WFS décrits aux equations 1.17 et 1.19. Cette affectation se fait par une opération de convolution :

$$\widetilde{c}_{i,\Psi}^j(t) = k_{i,\Psi}(t) * c_j^i(t) \quad (1.31)$$

De cette manière, on sépare l'opération de formation des fronts d'ondes (qui dépend en grande partie des retards WFS) de l'opération de filtrage destinée à corriger les artefacts de coloration. On peut alors choisir la taille des filtres contenus dans H indépendamment de la longueur des réponses contenues dans \widetilde{C}_Ψ et donc indépendamment de la source virtuelle. Le calcul des filtres \widetilde{H}_Ψ se fait en utilisant une version multicanale d'un algorithme MFAP (Modified Fast Affine Projection). Pour plus de détails à ce sujet, le lecteur est invité à consulter (Corteel, 2004; Corteel, 2006a). Dans le cadre des travaux présentés dans le reste du document, on utilise des filtres $\mathbf{h}_{i,\Psi}^{Meq}$ de longueur 800 échantillons à une fréquence d'échantillonnage de 48kHz.

¹⁶la structure générale du système de rendu est détaillée à la section 1.5

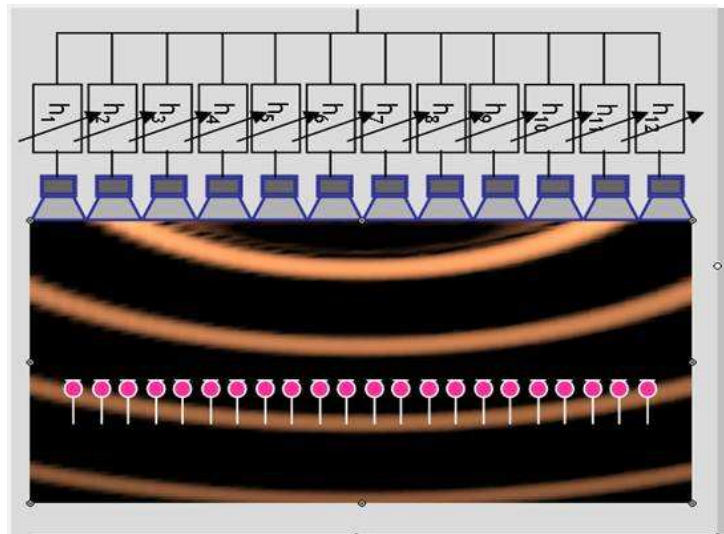


FIG. 1.18 – Configuration géométrique de haut-parleurs et microphones pour le calcul des filtres par inversion multicanale.

La particularité principale de cette méthode réside dans le fait que l'on choisit les points de contrôle le long d'une ligne parallèle au banc de haut-parleurs (Fig. 1.18), ce qui permet de caractériser les contributions principales émises par le banc de haut-parleurs. Par ailleurs, on sélectionne uniquement les microphones utiles étant donné la visibilité de la source virtuelle (Corteel, 2006a), et on effectue l'inversion multicanale uniquement en-dessous de la fréquence de repliement. Ces précautions permettent d'assurer que la solution numérique obtenue par le processus d'inversion multicanale soit cohérente d'un point de vue physique avec une solution d'onde propagative qui soit valable en dehors des points de contrôle, sur une zone étendue. Dans (Corteel, 2006a), l'auteur vérifie effectivement que la solution fournie par l'algorithme d'inversion multicanale se propage bien au-delà de la ligne de contrôle, en regardant l'évolution de la coloration du champ à différentes distances.

Au-dessus de la fréquence de repliement, les filtres sont calculés suivant la méthode d'égalisation individuelle avec compensation de l'erreur moyenne (eq. 1.28). On constate que par cette technique le temps d'arrivée du front d'onde principal est cohérent avec la position de la source virtuelle, après quoi se succèdent des artefacts liés notamment au repliement spatial du champ sonore (sauf dans le cas des sources focalisées où ces artefacts *précèdent* le pic principal). En vertu des phénomènes de masquage temporel (forward and backward masking) et de l'effet de préférence (Gardner, 1968) on est assuré de la bonne localisation des sources virtuelles alimentées par des signaux large-bande.

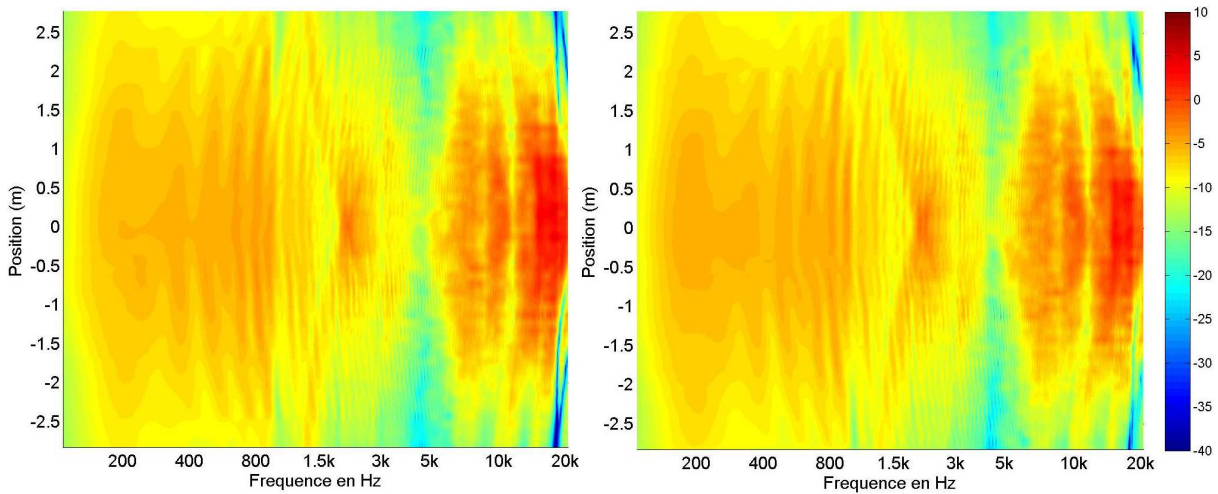


FIG. 1.19 – Réponses fréquentielles de deux haut-parleurs adjacents dans un banc de haut-parleurs électrodynamiques de type KEF mesurées en champ libre le long d'une ligne de microphones espacés de 10cm située 2m devant le banc de haut-parleurs.

Discussion : Choix de la méthode d'égalisation en fonction du type de haut-parleur

La méthode de calcul des filtres $\mathbf{h}_{i,\Psi}^{comp}$ (eq. 1.28) permet d'assurer une correction des artefacts de reproduction dépendant de la source virtuelle reproduite (position, directivité). Cependant, ces filtres ne permettent pas de compenser certains artefacts qui dépendent de la position d'écoute (Corteel, 2007). De ce fait, on montre que pour des haut-parleurs MAP, les filtres $\mathbf{h}_{i,\Psi}^{Meq}$ issus de l'inversion multicanale produisent des résultats sensiblement meilleurs que les filtres $\mathbf{h}_{i,\Psi}^{comp}$ (Corteel, 2006a). Malgré tout, des tests d'écoute informels montrent que des filtres de type $\mathbf{h}_{i,\Psi}^{comp}$ produisent des résultats très satisfaisants pour ce qui est de l'égalisation de systèmes WFS utilisant des haut-parleurs électrodynamiques de type KEF. Ces observations s'expliquent en partie par le fait que les haut-parleurs électrodynamiques utilisés présentent des caractéristiques de directivité plus régulières que les MAPs dans le plan frontal (Fig. 1.19, à comparer avec Fig. 1.15). De plus, la variabilité d'un haut-parleur à un autre est plus faible qu'entre deux excitateurs consécutifs d'un même panneau MAP. On utilisera donc des filtres de type $\mathbf{h}_{i,\Psi}^{comp}$ par la suite pour la synthèse de sources virtuelles par banc de haut-parleurs KEF.

1.4 Description et synthèse d'un effet de salle pour la WFS

Ayant décrit comment on réalise la synthèse du son direct d'une source virtuelle en WFS, on s'intéresse dans cette section à l'état de l'art pour la synthèse d'un effet de salle associé. Pour ce faire, on expose dans un premier temps trois modélisations (architecturale, physique et perceptive) couramment employées pour décrire un effet de salle, en donnant des exemples de leur utilisation dans le cadre de la WFS. Cette description est suivie d'une présentation du *Spatialisateur*, qui est le processeur d'acoustique virtuelle développé par l'équipe Acoustique des Salles de l'Ircam. Ce processeur est employé dans le système utilisé à l'Ircam pour synthétiser l'effet de salle associé aux sources virtuelles, suivant un schéma qui est détaillé à la fin de cette section.

1.4.1 Modélisations de l'effet de salle

On distinguera trois approches principales dans la modélisation de l'effet de salle : l'approche physique, l'approche perceptive, et l'approche signal.

Approche Physique

L'approche "physique" de modélisation de l'effet de salle repose sur un modèle architectural de la salle, qui décrit la position/directivité des sources sonores ainsi que les caractéristiques matérielles des différentes parois (coefficients d'absorption et diffusion). Ce type d'approche, qui s'intéresse aux *causes* du champ sonore au point d'écoute, permet par la suite de calculer le champ émis par une source sonore en tout point de la salle. On distingue deux classes de méthodes dans cette approche : les méthodes dites *physiques* et les méthodes dites *géométriques*. Les méthodes *physiques* sont basées sur des modèles mathématiques dérivés directement des lois de l'acoustique. Les méthodes physiques les plus connues sont les Méthodes d'Eléments de Frontière (Boundary Element Methods - BEM) dans lesquelles les parois de la salle sont subdivisées en éléments de surface discrets. Chaque élément de surface est considéré comme étant une source acoustique individuelle, qui contribue à la réponse impulsionnelle totale. Les Méthodes Eléments Finis (Finite Element Methods - FEM) ajoutent un degré de liberté au modèle précédent, et tiennent compte du comportement des éléments de volume de l'espace modélisé. Les méthodes BEM et FEM produisent des résultats proches de la réalité, mais requièrent des temps de calcul importants, qui peuvent être prohibitifs dans beaucoup de cas (Kleiner *et al.*, 1993).

Les méthodes *géométriques* fournissent une manière moins coûteuse de description de l'effet de salle. Parmi ces méthodes, on trouve la méthode de lancer de rayons (ray tracing) et la

méthode de sources images (mirror image source model - MISM). Ces méthodes permettent de prédire rapidement la distribution des réflexions spéculaires précoces en un point d'une salle. Elles sont cependant peu adaptées à la description de l'effet de salle tardif et la diffusion. Pour décrire les composantes diffuses de l'effet de salle, on utilise plutôt des méthodes de radiosité, qui permettent de modéliser les échanges énergétiques entre les différentes parois en supposant qu'elles ont un comportement parfaitement diffusant. La méthode WRW (Berkhout *et al.*, 1998) permet quant à elle de prendre en compte efficacement les effets de diffraction, qui ne sont pas pris en compte dans les méthodes de lancer de rayons et sources images.

Approche Signal

L'approche "signal" de modélisation de l'effet de salle se base classiquement sur la réponse d'un ou plusieurs haut-parleurs à une position d'écoute de référence. L'idée est de stocker l'image du canal acoustique entre la source et l'auditeur sous forme d'une réponse impulsionnelle, avec en plus (éventuellement) la direction de provenance de la source de manière à permettre des effets de panoramique au moment du rendu. La réponse impulsionnelle peut être obtenue par un calcul basé sur un modèle physique ou perceptif ou bien par une mesure dans l'espace réel.

Dans le cadre des méthodes de synthèse de champs sonores (WFS, Ambisonique), le paradigme du point d'écoute de référence n'est plus valable, car le système de reproduction fournit une réponse sur zone étendue. Il est préférable dans ce cas de procéder à une décomposition du champ sur une base de fonctions spatiales (ondes planes, harmoniques cylindriques, transformée de Radon). La décomposition du champ nécessite de mesurer l'effet de salle sur un réseau multi-microphonique. Les techniques utilisant une mesure multi-microphonique associée à une décomposition du champ sonore sont regroupées sous l'appellation Wave Field Analysis (WFA), par analogie avec la méthode de synthèse multi haut-parleurs. Elles permettent d'obtenir une représentation de l'effet de salle sur un nombre fini de composantes spatiales. Ce nombre est souvent très élevé, d'où la nécessité de réduire le nombre de composantes via une paramétrisation temps/fréquence/espace du champ sonore (Hulsebos & de Vries, 2002). Au moment du rendu, la WFS permet ensuite de recréer efficacement l'effet de salle ainsi caractérisé, en le diffusant sur un ensemble d'ondes planes équi-réparties angulairement.

En ce qui concerne l'utilisation de la WFA comme moyen de mesure associé à une *auralisation* (cf. glossaire) de salles de concert en WFS, on trouve dans la littérature des exemples utilisant des bancs de microphones linéaires (Sonke, 2000) ou circulaires (Hulsebos *et al.*, 2001). Ces techniques sont cependant basées sur des réseaux bidimensionnels, et souffrent du problème de repliement de la composante vertical du champ sonore au moment de la mesure. Ce problème est résoluble par l'utilisation de réseaux microphoniques tridimensionnels (Laborie

et al., 2003; Laborie *et al.*, 2004b), dont l'intérêt est mis en évidence au chapitre 5.

Approche Perceptive

L'approche "perceptive" de modélisation d'effet de salle est une approche qui s'intéresse à l'impression spatiale perçue par l'auditeur lors de l'écoute en salle. Dans le cadre du projet Européen Carrouso (Vaananen *et al.*, 2002), deux approches perceptives ont été abordées pour la description de scènes sonores, l'une basée sur les travaux de Pellegrini (Pellegrini, 2001) et l'autre basée sur le Spatialisateur de l'Ircam (Jot & Warusfel, 1995; Jot, 1992). Ces descriptions donnent toutes les deux la possibilité au créateur du contenu de manipuler la qualité acoustique de la salle sous la forme de paramètres perceptifs.

La prochaine section aborde une description du Spatialisateur, qui est le processeur d'acoustique virtuelle employé pour synthétiser l'effet de salle dans le système de diffusion WFS à l'Ircam.

1.4.2 Le Spatialisateur

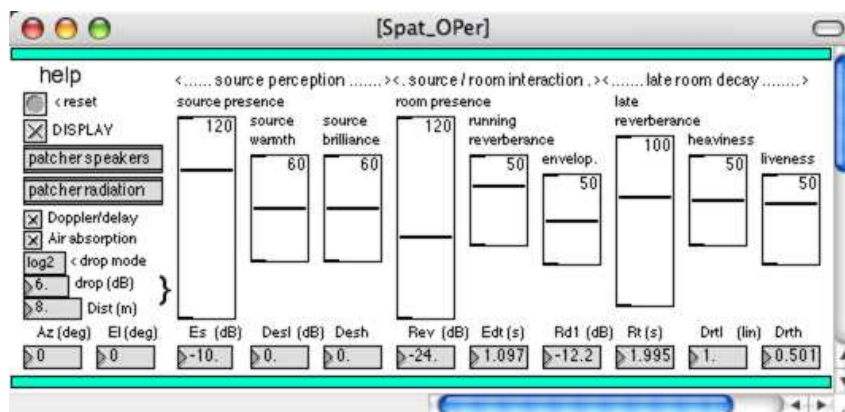


FIG. 1.20 – Implémentation du Spatialisateur dans l'environnement Max/MSP - Interface de réglage des paramètres perceptifs.

Le *Spatialisateur* (ou Spat~) est un processeur d'acoustique virtuelle conçu originalement pour l'environnement de programmation graphique Max/MSP. Il permet de contrôler en temps réel la spatialisation d'une source (position, orientation, directivité) ainsi que la qualité acoustique de la salle virtuelle dans laquelle la source évolue (Jot *et al.*, 1998). Le Spat~ intègre divers formats de reproduction multicanale permettant de s'adapter à des dispositifs électroacoustiques variables (casque, haut-parleurs stéréo, système périphonique,...). Les formats de reproduction supportés incluent le Binaural (Larcher, 2001), le Transaural (Gardner, 1995), le VBAP (Vector Based Amplitude Panning) (Pulkki, 1997), ou bien encore Ambisonics (Gerzon, 1985). Une

1.4. Description et synthèse d'un effet de salle pour la WFS

adaptation de l'architecture du Spatialisateur pour la synthèse d'effet de salle en WFS a également fait l'objet d'une implémentation, que nous décrivons dans cette section.

Le Spat-permet de manipuler la qualité acoustique de la salle sur deux niveaux : un niveau perceptif ou "haut-niveau", et un niveau objectif ou "bas-niveau". La manipulation du niveau perceptif se fait par le biais d'une interface regroupant 9 paramètres perceptifs indépendants (Fig. 1.20) qui sont issus de recherches psychoacoustiques menées à l'Ircam entre 1989 et 1995 sur l'objectivation de l'impression spatiale (Lavandier, 1989; Jullien *et al.*, 1992; Jullien, 1995; Kahle, 1995) :

- 3 paramètres pour décrire la source (*source presence, source warmth, source brilliance*).
- 3 paramètres pour décrire l'interaction entre la source et la salle (*room presence, running reverberance, envelopment*).
- 3 paramètres pour décrire la salle (*late reverberance, heaviness, liveness*).

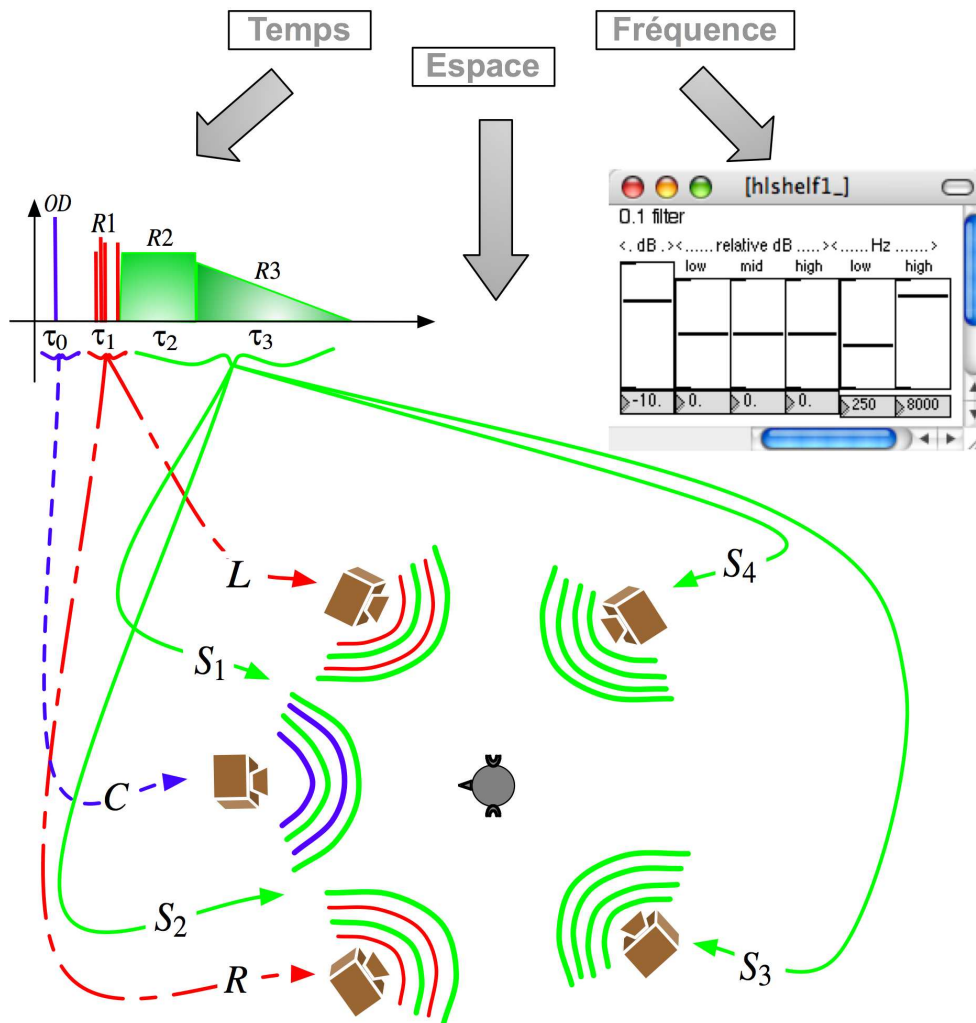


FIG. 1.21 – Représentation schématique de la distribution énergétique temps/espace/fréquence d'une réponse impulsionnelle par le Spatialisateur.

Ces paramètres perceptifs sont traduits en paramètres objectifs “bas-niveau” au cours de la synthèse de l'effet de salle. Les paramètres bas-niveau déterminent la répartition temps/espace/fréquence de l'énergie acoustique dans la salle, ce qui est représenté de manière schématique à la figure 1.21.

En haut à gauche de la figure 1.21 se trouve une représentation temporelle de la réponse impulsionnelle de la salle, découpée en 4 sections : OD (son direct), $R1$ (réflexions précoces), $R2$ (réflexions tardives ou “cluster”) et $R3$ (réverbération tardive). Les durées τ_0 , τ_1 , τ_2 , et τ_3 de chaque section peuvent être réglées indépendamment. Le réglage “par défaut” considère que $R1$ couvre la plage temporelle [20-40]ms, $R2$ couvre la plage [40-100]ms et $R3$ couvre la plage supérieure à 100ms. Notons que l'on emploie également les termes OD , $R1$, $R2$ et $R3$ pour désigner l'énergie dans ces sections temporelles, dont la valeur par défaut vaut donc :

$$OD = \int_{t=0ms}^{t=20ms} p^2 dt, \quad (1.32)$$

$$R1 = \int_{t=20ms}^{t=40ms} p^2 dt, \quad (1.33)$$

$$R2 = \int_{t=40ms}^{t=100ms} p^2 dt, \quad (1.34)$$

$$R3 = \int_{t=100ms}^{t=\infty} p^2 dt. \quad (1.35)$$

En haut à droite de la figure 1.21 se trouve une représentation fréquentielle de la réponse impulsionnelle de la salle, dont la manipulation se fait par le biais d'un filtre récursif du second ordre, permettant de régler le niveau des différentes sections temporelles dans 3 bandes “low” (BF), “mid” (MF), et “high” (HF).

En bas et au centre de la figure 1.21 se trouve enfin une représentation de la répartition spatiale de l'énergie au moment de la diffusion. En pratique, la répartition spatiale des sections temporelles se fait suivant des groupes de haut-parleurs (Jot *et al.*, 1998) :

- Un groupe “central” constitué d'un canal virtuel C employé pour diffuser le son direct OD .
- Un groupe “latéral” constitué de deux canaux virtuels L et R situés de part et d'autre du son direct, employés pour diffuser les réflexions précoces $R1$.
- Un groupe “diffus” constitué de 4 canaux S_1 , S_2 , S_3 , et S_4 répartis autour du point central via des panoramiques d'intensité sur l'ensemble des haut-parleurs disponibles¹⁷, employés pour diffuser les réflexions tardives $R2$ et la réverbération tardive $R3$. Les canaux décorrélés de réverbération tardive $R3$ sont calculés en utilisant des matrices de bouclage unitaires, suivant un procédé décrit dans (Jot, 1992). Le nombre de canaux diffus a été

¹⁷de manière à produire la meilleure approximation spatiale de l'isotropie d'un champ diffus

fixé initialement à 4 en raison de limitations technologiques liées au coût de ce calcul. La dernière version du Spatialisateur permet cependant à priori de synthétiser un nombre illimité de canaux décorrélés pour l'alimentation de l'effet de salle.

Dans les systèmes de diffusion périphoniques, les groupes "central" et "latéral" sont assignés dynamiquement moyennant des lois de panoramique ou matricages au système de diffusion suivant les déplacements de la source, de manière à maintenir une direction de provenance cohérente avec la position de la source virtuelle. Par contre, les composantes tardives de l'effet de salle sont distribuées de manière figée, puisqu'on vise une sensation de champ diffus quel que soit la position de la source virtuelle.

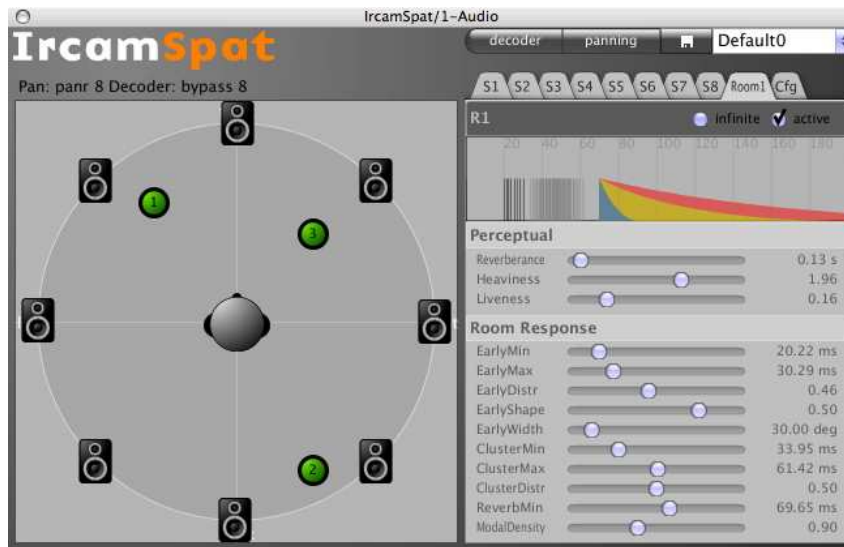


FIG. 1.22 – Implémentation plug-in du Spatialisateur v4.0 - fenêtre de contrôle de la qualité acoustique (droite) et interface de positionnement des sources (gauche).

L'ensemble des paramètres objectifs bas-niveau sont modifiables directement, sans passer par le haut-niveau. Dans la version plug-in du Spatialisateur implémentée en 2007, on accède aux paramètres perceptifs et objectifs dans la même fenêtre de contrôle (Fig. 1.22).

1.4.3 Système de rendu d'effet de salle WFS Ircam

Haut-parleurs virtuels

Le système de rendu d'effet de salle WFS utilisé à l'Ircam fait intervenir un certain nombre de "haut-parleurs virtuels" Ψ_{HP} , qui sont des sources virtuelles WFS fixes employées pour reproduire les canaux L , R , S_1 , S_2 , S_3 , S_4 du Spat~. Idéalement, on dispose d'un système WFS périphonique, dans quel cas les haut-parleurs virtuels sont équirépartis autour de la salle d'écoute. Cependant, il se peut que pour des raisons technologiques, le système WFS soit complet uniquement dans la zone frontale, avec un nombre limité de canaux pour alimenter le reste

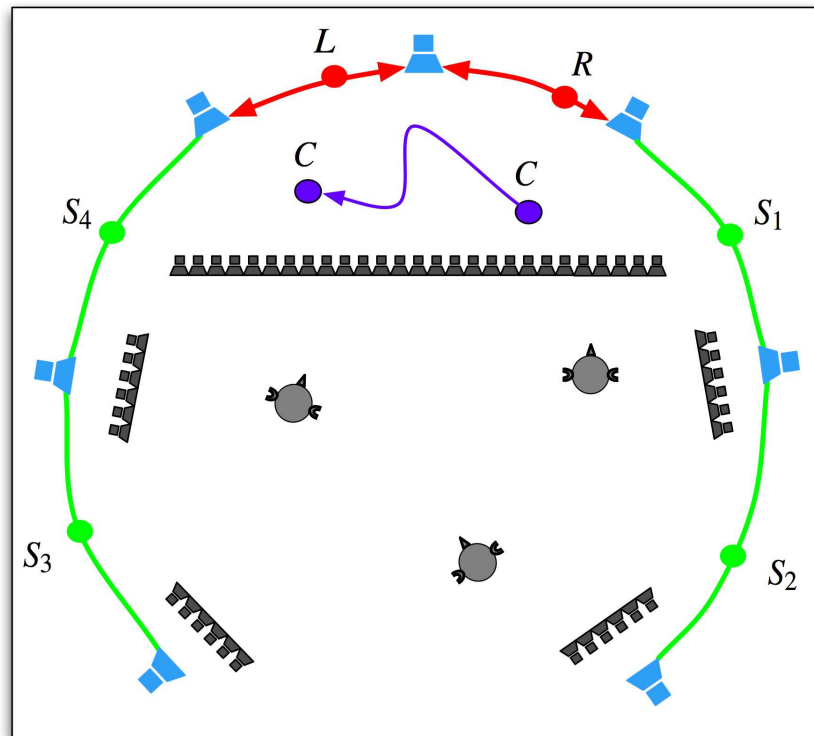


FIG. 1.23 – Vision schématique d'un système de rendu d'effet de salle pour la WFS. On représente en bleu clair les haut-parleurs virtuels employés pour diffuser l'effet de salle. La source virtuelle Ψ est représentée en violet, les sources panoramiques mobiles employées pour diffuser les réflexions précoces sont représentées par des points rouges, et les sources fixes employées pour diffuser les réflexions tardives sont représentées par des points verts clairs.

du pourtour de la zone d'écoute. Dans ce cas, les haut-parleurs virtuels latéraux et arrières sont diffusés via des bancs WFS réduits, placés en périphérie, comme il a été représenté à la figure 1.23 (les haut-parleurs virtuels y sont représentés en bleu).

Répartition spatiale des sections temporelles de l'effet de salle

La répartition des sections temporelles de l'effet de salle se décompose de la manière suivante :

- Le canal C , qui transmet le son direct OD , est diffusé via une source virtuelle WFS mobile, représentée par un point violet sur la figure 1.23.
- Les canaux L et R transmettent les réflexions précoces $R1$. Ces canaux sont synthétisés par panoramique d'intensité, et leur position varie dynamiquement avec la position de la source virtuelle, de manière à entourer celle-ci à tout moment. Ces canaux sont représentés par des points rouges sur la figure 1.23.
- Les canaux S_1, S_2, S_3, S_4 , qui transmettent les réflexions tardives $R2$, sont distribués sur

1.4. Description et synthèse d'un effet de salle pour la WFS

l'ensemble des haut-parleurs virtuels.

- On synthétise autant de signaux décorrélés de réverbération tardive $R3$ que de haut-parleurs virtuels, et on diffuse directement ces signaux par le biais de ces haut-parleurs.

Un système WFS a été monté suivant ce schéma lors d'une série d'ateliers destinés aux compositeurs et ingénieurs du son, organisés conjointement par l'Ircam et sonic emotion ag. en Mai 2005 au Forum Neues Musiktheatre de Stuttgart (Fig. 1.24).



FIG. 1.24 – Vue d'ensemble d'un démonstrateur WFS installé lors d'une série d'ateliers montés conjointement par l'Ircam et sonic emotion ag. au Forum Neues Musiktheatre de Stuttgart en Mai 2005.

Discussion : validité du rendu d'effet de salle tardif

Le synthèse de l'effet de salle en WFS a fait l'objet d'une étude perceptive dans le cadre d'une thèse à l'université de Delft (Sonke & de Vries, 1997; Sonke, 2000), dans laquelle l'auteur s'est penché en particulier sur la synthèse de l'effet de salle tardif. L'effet de salle tardif désigne par définition la section temporelle de la réponse impulsionnelle de la salle qui arrive à partir du moment où il règne un régime de *champ diffus* dans la salle (cf. Annexe A.2). La définition courante du champ diffus fait intervenir un "ensemble d'ondes planes décorrélées provenant d'incidences uniformément distribuées dans l'espace". Sonke a tenté de déterminer le nombre minimal d'ondes planes à synthétiser par un système WFS périphonique de manière à assurer une isotropie perceptive du champ dans le plan horizontal. Les résultats d'une campagne de tests perceptifs ont montré que 11 ondes planes décorrélées étaient suffisantes pour assurer l'isotropie

perceptive du champ réverbéré dans le plan horizontal, même pour des sujets entraînés et/ou experts.

Dans le modus operandi de synthèse d'effet de salle présenté dans cette section, les canaux d'effet de salle tardif sont au nombre de 8 au lieu de 11, et sont diffusés via des sources ponctuelles éloignées (et non des ondes planes). Comme le fait remarquer (Corteel, 2004), cette situation diverge quelque peu par rapport aux recommandations de Sonke, bien que l'éloignement important des sources virtuelles fasse que leur champ sonore associé soit relativement similaire à celui produit par une onde plane. La validité perceptive de ce champ resterait cependant à être vérifiée par des tests perceptifs, ce qui n'a pas encore été fait à ce jour.

1.5 Architecture globale du système WFS Ircam

Une vision d'ensemble de la chaîne de production WFS telle qu'elle est implémentée à l'Ircam est donnée dans cette section. Pour une description détaillée le lecteur est invité à consulter (Corteel, 2004).

Répartition des canaux audio

Si on suit la formulation originale de la WFS, l'idée est de faire correspondre un flux audio monophonique donné à une source virtuelle donnée. Ce flux audio peut être issu de diverses provenances, que ce soit un microphone placé devant un instrumentiste dans une situation "live" (cf. Fig. 1.8) ou bien un signal de sortie d'un séquenceur audio dans une situation de composition en studio. On peut également récupérer des flux audio multicanaux et les diffuser par des techniques panoramiques stéréophoniques classiques (intensité ou retard) sur un nombre fini de sources virtuelles dédiées, par la méthode des "Virtual Panning Spots" ou VPS (Theile *et al.*, 2003). Ceci permet de réduire le nombre de sources virtuelles à synthétiser dans des scènes sonores comportant un grand nombre d'éléments.

Outils auteur

Les paramètres de description de la scène sonore sont ajustables via des plug-ins qui communiquent par réseau avec le processeur d'acoustique virtuelle et le système de rendu WFS, permettant d'ajuster les paramètres de source (position, orientation, directivité) ainsi que la qualité acoustique de la salle (Fig. 1.25 - droite et centre). Les paramètres sont entièrement automatisables depuis l'interface native du séquenceur, ce qui permet de régler des trajectoires précises en concordance avec les événements audio diffusés.

Les paramètres de spatialisation sont également manipulables via l'outil auteur Listenspace

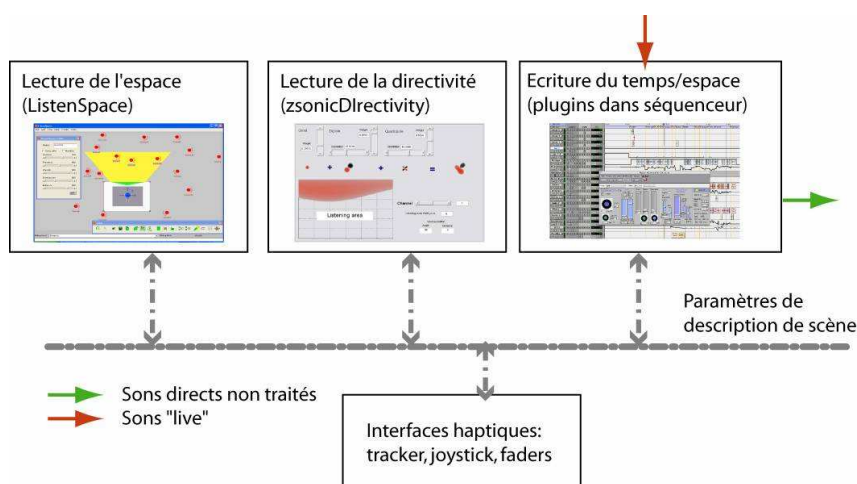


FIG. 1.25 – Outils auteurs pour la création de scène sonore en WFS.

(Fig. 1.25 gauche), qui donne une vue de haut de la disposition des sources dans la scène sonore. Cet outil auteur peut être placé sur une interface graphique portable, comme un Tablet PC, de manière à permettre au créateur de contenu de déambuler dans la scène en écoutant le rendu de la scène sous différents points de vue pendant la phase compositionnelle.

Chaîne de production complète

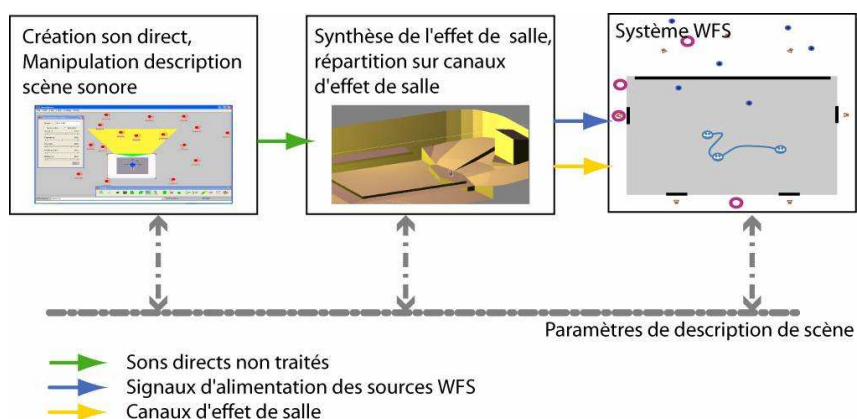


FIG. 1.26 – Vision complète de la chaîne de production WFS.

Une vision complète de la chaîne de production se trouve à la figure 1.26. On voit sur cette figure que les paramètres de description de la scène sont partagés et accessibles à tous les éléments de la chaîne. Ceci se fait par le biais d'une base de données distribuée sur réseau ethernet (développé par sonic emotion). Le système est dénué de serveur principal et entièrement modulaire ; le fait de rajouter des canaux physiques supplémentaires se fait très simplement du point de vue de la configuration du réseau (mais demande cependant à priori de recalculer les filtres

d'égalisation).

Le réseau exhibe une latence faible (<10 ms), et permet une mise à jour de tous les paramètres depuis n'importe quel noeud de la chaîne en temps réel. Il permet de transmettre les données audio et les paramètres de contrôle de manière synchrone à toutes les machines chargées de faire les calculs de convolution WFS en temps réel.

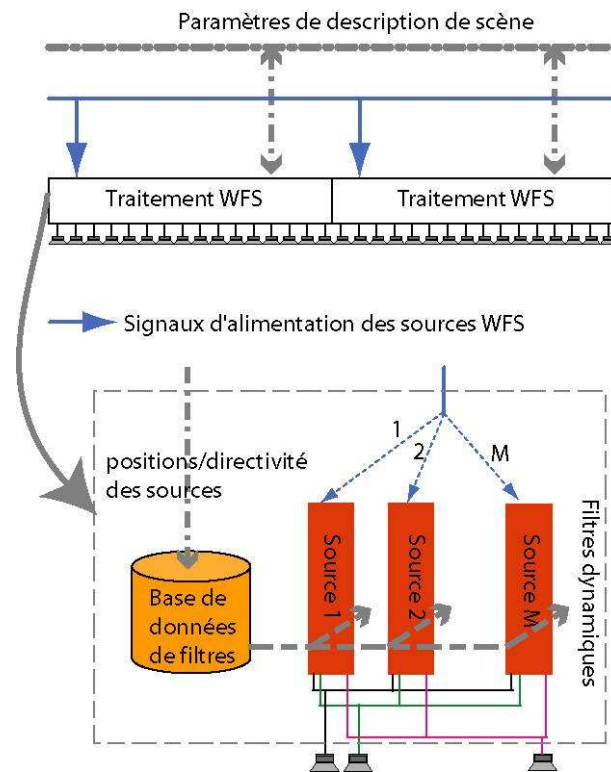


FIG. 1.27 – Schéma fonctionnel du traitement WFS. **Haut** : Les paramètres de description de la scène sont communiqués via le réseau zsonicnet aux machines de traitement convolutif. **Bas** : Chaque machine de traitement individuelle reçoit les ordres de changement de position de la source et charge un nouvel ensemble de filtres permettant d'alimenter les 16 canaux dont elle a la charge.

Détail du traitement convolutif temps réel

La figure 1.27 montre le détail de l'opération de traitement convolutif réalisée pendant la synthèse de la scène sonore. On fait correspondre 16 canaux physiques à chaque machine de traitement, qui gère au plus M sources simultanément¹⁸. A chaque fois qu'une des M sources change de position ou de directivité, chaque machine charge un nouvel ensemble de 16 filtres

¹⁸On peut estimer qu'un PC à l'état de l'art permet a priori d'effectuer le filtrage simultané d'environ 24 sources virtuelles WFS en temps réel permettant d'alimenter 16 canaux audio. Pour une latence réduite (utilisation en situation de concert "mixte" avec instrumentistes), le nombre de sources simultanées est réduit à 16. Etant donné la loi de Moore, ce nombre est cependant en augmentation constante.

FIR à appliquer à ses haut-parleurs nominaux. Les filtres sont chargés à partir d'une base de données contenant des filtres WFS correspondant à un ensemble de positions qui définissent un échantillonnage de l'espace situé derrière et devant le banc de haut-parleurs (Fig. 1.28). Cet échantillonnage est fait de telle manière à assurer un angle maximum de $\pm 2.5^\circ$ entre deux positions successives de sources virtuelles, vu d'un point situé dans la zone d'écoute, ce qui est une optimisation de la distribution spatiale des sources tenant compte du flou de localisation dans le plan azimutal (Corteel, 2006a). A chaque position de l'échantillonnage, on stocke un ensemble de filtres de directivité élémentaire, permettant de synthétiser n'importe quelle figure de directivité dans le plan horizontal par combinaison linéaire (cf. chapitre 3).

Les changements de filtres à la volée se font par des fondus d'amplitude (crossfade), avec une interpolation sur les composantes de retards $k_{i,\psi}(t)$ extraites des filtres WFS, de manière à réduire la latence (cf. éq. 1.31).

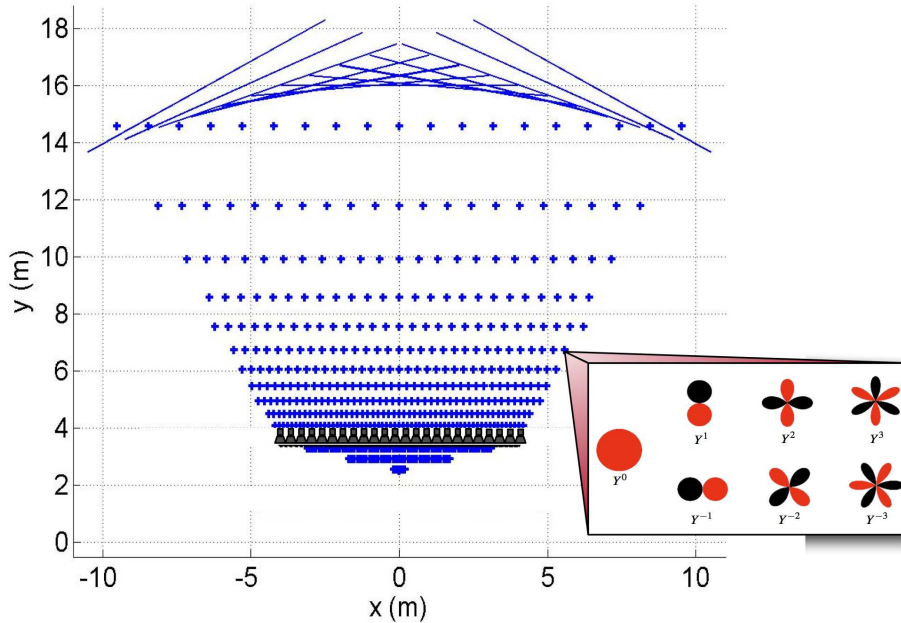


FIG. 1.28 – La base de données de filtres WFS définit un quadrillage de l'espace autour du banc de haut-parleurs. A chaque position de ce quadrillage sont stockés un ensemble de filtres de directivité élémentaire.

Manifestations artistiques et scientifiques récentes utilisant la WFS

Un certain nombre de manifestations artistiques ont été produites au sein de l'Ircam utilisant un démonstrateur WFS issu du projet Carrouso, dont une création dans le cadre de l'exposition DADA en juin 2005 au Centre Georges Pompidou (Grand *et al.*, 2005) (Fig. 1.30), une exposition audio-vidéo installée de manière permanente à la Cinémathèque Française en octobre 2005

(Desnoux & Caulkins, 2005) (Fig. 1.29), une installation interactive au Centre Georges Pompidou en juin 2006 (Corsino & Corteel, 2006) ainsi qu'une installation sonore pour l'exposition Samuel Beckett au Centre Pompidou en 2007 (Chiron *et al.*, 2007) (Fig. 1.31).



FIG. 1.29 – Installation sonore audio-vidéo utilisant la WFS dans l'exposition permanente de la Cinémathèque Française. Les haut-parleurs MAP sont situés derrière les écrans de projection en toile acoustiquement transparents.

1.5. Architecture globale du système WFS Ircam



FIG. 1.30 – Installation sonore WFS au Centre Georges Pompidou dans le cadre de l'exposition DADA (juin 2005). Les haut-parleurs MAP utilisés sont ici dissimulés derrière une toile blanche acoustiquement transparente.

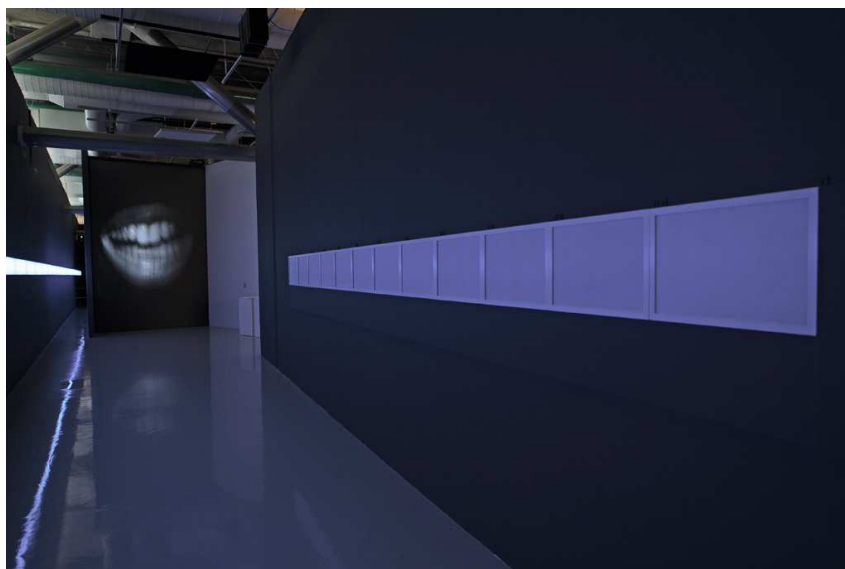


FIG. 1.31 – Installation sonore WFS au Centre Georges Pompidou dans le cadre de l'exposition Samuel Beckett (Mars-Juin 2007). Les haut-parleurs MAP de l'installation sont visibles sur le mur de droite.

1.6 Conclusion et positionnement de la thèse

Ce chapitre a été l'occasion de dresser un état de l'art pour la synthèse d'une scène sonore sur zone étendue par WFS. On pourrait dire, pour simplifier, que ce chapitre aborde un ensemble de techniques et de concepts qui avaient déjà été mis au point au moment du début de la thèse. Ces techniques et concepts ont aussi continué à évoluer en parallèle du travail, et nous avons essayé de mentionner aussi souvent que possible ces évolutions lorsqu'il y avait lieu de le faire.

L'objectif du chapitre n'a donc pas été de résumer un travail personnel, mais plutôt d'introduire les notions indispensables à la compréhension, voire à la construction, des concepts qui vont leur succéder dans les sections suivantes.

Dans la première section ce chapitre, nous avons introduit le principe de Huygens, qui sous-tend la WFS, et décrit ses formalisations sous forme d'équations intégrales. Partant de ces équations, nous avons montré dans la seconde section comment on dérivait les fonctions d'alimentation permettant *synthétiser le son direct* associé à une *source virtuelle monopolaire*, en utilisant un banc linéaire et continu de sources secondaires. Nous avons ensuite montré comment la réduction des hypothèses de base des formalisations (i.e. limitation au plan horizontal, troncature et échantillonnage de la distribution de sources secondaires) se répercutait sur le champ effectivement rayonné par le système. Enfin, nous avons exposé un ensemble de techniques d'égalisation du son direct développées dans le cadre de la WFS, permettant de prendre en compte les défauts des transducteurs dans des conditions réelles d'utilisation.

La section finale de ce chapitre a été consacrée à une description de la *synthèse d'un effet de salle* en Wave Field Synthesis. Nous avons montré comment, en utilisant le processeur d'acoustique virtuelle de l'Ircam, il est possible de synthétiser une qualité acoustique associée à une source virtuelle.

Le lecteur aura pu remarqué qu'aucune mention n'est faite au cours du chapitre de la *prise en compte de l'acoustique de la salle*, que ce soit dans la section consacrée au rendu du son direct ou celle consacrée au rendu de l'effet de salle. En effet, bien que l'équation intégrale de Kirchhoff-Helmholtz permette en théorie de s'affranchir de l'effet de salle, la WFS est, dans sa formulation initiale, une méthode de type "open loop" qui présuppose un environnement de restitution anéchoïque.

En réalité, le contexte de diffusion introduit des artefacts dans la construction des fronts d'ondes, ce qui a poussé de nombreux auteurs à proposer des méthodes de compensation de contexte dans les dernières années (Corteel & Nicol, 2003; Spors *et al.*, 2003; van Zon, 2003; Lopez *et al.*, 2005). Actuellement ces recherches se poursuivent, et certains adoptent des solutions "actives" ou "closed-loop", permettant de tenir compte de l'évolution des conditions dans le contexte de diffusion (présence d'un public, changements de température ou de niveau d'humidité). Pour réduire les coûts de calcul importants inhérents à l'inversion multicanale, certains auteurs proposent de travailler dans un domaine transformé, basé sur une description physique du champ sonore (Spors *et al.*, 2007). D'autres proposent d'utiliser la solution "open loop" WFS comme solution *a priori* de l'algorithme de compensation de la salle, et d'ajuster par un facteur de pondération le poids de la solution "open loop" et "closed loop" au cours de la diffusion (Gauthier & Berry, 2005; Gauthier & Berry, 2006).

On pourra dire, de manière caricaturale, que les applications de compensation de contexte citées visent à *rendre la salle de diffusion anéchoïque*, en tirant partie des possibilités offertes par le réseau de haut-parleurs utilisé pour le rendu du son direct. De cette manière, l'information de la scène sonore qui est restituée dans la zone d'écoute provient uniquement des haut-parleurs eux-mêmes, et se substitue à la salle proprement dite.

L'approche que nous proposons dans ce travail provient d'une réflexion dans un contexte musical, et la problématique à laquelle elle tente de répondre est vraisemblablement restreinte par rapport aux démarches que nous venons de décrire dans les deux paragraphes précédents. En effet, nous nous plaçons dans le cadre de musique "mixte", où le système WFS est employé pour *synthétiser des sources virtuelles qui partagent l'espace acoustique de la scène avec des sources instrumentales*. L'idée n'est donc plus de "s'affranchir" de la salle, car sa qualité acoustique est a priori nécessaire à l'appréciation de la performance par le public. On peut d'ailleurs remarquer que si la WFS permettait de reconstruire à *l'identique* le son direct de la source virtuelle cible, il s'en suivrait naturellement que son effet de salle associé serait conforme à celui produit par les instruments réels qui l'entourent sur la scène.

Cependant, les approximations faites par rapport au cadre idéal décrit dans les équations intégrales impliquent que *le champ effectivement rayonné par le système WFS n'est pas conforme à celui des sources idéales* que nous cherchons à reproduire, ce qui implique une différence, *in fine*, entre l'effet de salle idéal et l'effet de salle qui est effectivement produit par le banc WFS.

Etant donné ce cadre, notre objectif est d'*identifier la partie de l'effet de salle produite par le système qui est bien cohérente avec celui d'une source réelle, et rajouter l'information nécessaire* en s'appuyant sur la salle. Nous avons déjà adopté ce type d'approche précédemment, en proposant un schéma permettant de compenser le manque de réflexions précoces lors de la synthèse des sources focalisées en salle (Caulkins, 2003; Caulkins *et al.*, 2003). Dans le travail que nous présentons ici, l'accent est mis cette fois sur une compensation de l'énergie émise dans la salle par le système lors du rendu du son direct. Notons toutefois que dans cette approche, on ne peut pas enlever de l'énergie s'il y en a déjà trop dans la salle.

L'idée est, d'une part, de *développer des outils permettant de caractériser la puissance associée à la source virtuelle effectivement émise par le banc de haut-parleurs* dans la salle. En effet nous savons, en vertu de l'équation A.8, que la densité d'énergie présente dans la salle, qui détermine le niveau perçu du champ réverbéré, est liée à la puissance émise par la source et aux caractéristiques de la salle (volume V et temps de réverbération Tr_{60}). Ainsi, si on souhaite assurer que le niveau du champ réverbéré perçu par les auditeurs demeure constant lors des déplacements et rotations de sources sur scène, il faut d'abord caractériser puis compenser la puissance émise par le banc de haut-parleurs. La caractérisation de puissance émise peut être conduite de diverses manières, qui sont décrites au chapitre 4 et en Annexe A.3. Il apparaît cependant que le banc dont nous cherchons à caractériser la puissance possède des propriétés tout à fait particulières : il est d'une grande étendue, il émet un rayonnement cohérent en dessous de 1500Hz, et sa géométrie peut éventuellement varier suivant la situation de diffusion (répartition spatiale variable des haut-parleurs suivant le contexte, permettant de s'adapter au contenu visé par l'application (Corteel, 2006b)). Par conséquent, nous proposons d'adopter une méthode modulaire de caractérisation de la puissance émise par le banc, basée sur une *méthode de prédiction du rayonnement* qui vise à reconstruire le champ émis par le banc WFS complet à partir de la décomposition du rayonnement d'un (ou plusieurs) de ses haut-parleurs constitutifs. Cette méthode fait l'objet d'une description détaillée dans le Chapitre 2.

D'autre part, au chapitre 3 nous proposons d'*approfondir des formalismes existants de contrôle de directivité en WFS* (Corteel, 2007; Corteel, 2004), en adaptant des outils de manipulation des sources virtuelles dérivées du formalisme Ambisonique (rotation, synthèse de sources hyper-directives) pour l'utilisation dans le cadre de la WFS. Ces outils seront ensuite utilisés pour *compenser l'effet de salle* (chapitre 4), en injectant

de l'énergie réverbérée par le biais de sources virtuelles hyper-directives synthétisées par le système WFS. Le chapitre 5 est enfin consacré à la *vérification de la pertinence des effets de contrôle de directivité* des sources virtuelles WFS dans une salle de concert. Nous développons pour cela une méthode basée sur l'utilisation d'un microphone hyper-directif, en combinaison avec un modèle géométrique de la salle permettant d'identifier les différents régimes de réflexions précoces captés par le microphone.

CHAPITRE 2

Loudspeaker radiation prediction method based on a spherical harmonic expansion of the sound field

Terence Caulkins¹, André Girier², and Olivier Warusfel¹

Introduction et contexte

Ce chapitre, qui reprend un article écrit en collaboration avec André Girier et Olivier Warusfel, vise à introduire une méthode systématique et robuste d'estimation du champ sonore émis par un banc de haut-parleurs. La motivation première de l'étude découle de la nécessité de développer un outil de travail permettant, dans le cadre plus général de la thèse, de réaliser la caractérisation de la puissance d'un banc de haut-parleurs WFS.

¹UMR9912 IRCAM-CNRS, 1 Pl. Igor Stravinsky 75004 Paris, France

²École Nationale Supérieure des Télécommunications, 46 rue Barrault 75013 Paris, France

En effet, comme le montre l'étude conduite au chapitre 4, section 4.3, la caractérisation de puissance d'un banc WFS n'est a priori pas viable par des méthodes classiques, étant donné la taille du banc et la cohérence des ondes qu'il émet en dessous de la fréquence de repliement. De plus, les caractéristiques d'un banc WFS peuvent être amenées à varier suivant le contenu que l'on souhaite diffuser. En effet, (Corteel, 2006b) montre que si l'on souhaite synthétiser uniquement des sources non-focalisées, l'espacement inter haut-parleurs optimal est de type logarithmique (avec une concentration supérieur des haut-parleurs vers les extrémités du banc) ; au contraire, si l'on souhaite synthétiser des sources à toutes les positions possibles, l'idée est d'adopter un espacement régulier les haut-parleurs.

En vue de ces contraintes, la nécessité de développer une méthode modulaire pour la mesure de la puissance émise par le banc de haut-parleurs WFS est apparue. Cette nécessité fait écho par ailleurs à la nécessité de trouver une manière simplifiée de conduire la phase de mesure en champ libre du banc de haut-parleurs WFS, qui est indispensable au processus d'égalisation du son direct que nous avons décrit à la section 1.3.9. La méthode de prédiction du rayonnement du banc de haut-parleurs que nous proposons adresse donc à la fois la caractérisation du son direct, et la caractérisation de la puissance émise pour la diffusion en situation de concert. Nous montrons, par ailleurs, qu'elle est applicable dans un contexte plus large englobant d'autres méthodes de reproduction sonore, ainsi que la modélisation en acoustique des salles.

La méthode décrite s'appuie sur une décomposition du champ rayonné par un ou plusieurs haut-parleurs constitutifs du banc WFS. Des facteurs tels que l'étendue de la source mesurée, la densité d'échantillonnage, la position du centre acoustique du haut-parleur, et l'interférence due à la présence de haut-parleurs adjacents dans le banc sont isolés et analysés de manière à déterminer leur influence sur la qualité de la prédiction effectuée. Trois approches, basées alternativement sur la mesure suivant sur une ou deux sphères concentriques, sont détaillées théoriquement et comparées dans des simulations numériques, de manière à trouver la méthode la plus efficace pour la tâche d'extrapolation du champ. La méthode retenue est alors utilisée pour prédire le champ émis par un haut-parleur de type Multi-Actuator Panel (MAP) et un satellite électrodynamique, et les résultats sont comparés avec des mesures réelles faites en champ libre.

ABSTRACT

This article presents an exploratory study aimed at developing a systematic and robust method for predicting the 3D free field radiation characteristics of a loudspeaker array. The method relies on the spherical harmonic expansion of the sound field emitted by the individual loudspeakers of the array. Factors such as source extent, spatial sampling density, position of acoustic origin, and interference from adjacent array speakers are isolated and analyzed so as to underline their influence on the quality of the radiation prediction. Three approaches, based either on the measurement of the sound field over a single sphere of microphones or over two concentric spheres of microphones, are detailed theoretically and compared in numerical simulations so as to find the most effective one in terms of sound field extrapolation. The retained method is used to predict the sound field of a Multi-Actuator Panel and an electrodynamic satellite speaker, and results are compared with actual free field measurements. Developed initially as a means of simplifying the sound field equalization process used in Wave Field Synthesis, as well as characterizing the power emitted by the loudspeaker array, the proposed method is shown to be applicable in a broader context encompassing other sound reproduction techniques and room acoustics simulations.

INTRODUCTION

Wave Field Synthesis (WFS) is a loudspeaker array-based sound reproduction technique relying on Huyghens' principle, which potentially allows for the exact reconstruction of any given sound field over an extended area (Berkhout, 1988; Berkhout *et al.*, 1993). However, practical issues —notably irregularities in loudspeaker radiation characteristics— can be shown to introduce audible artifacts in the effectively reconstructed sound field. Multichannel equalization algorithms have been developed to limit the scope of these artifacts (Corteel, 2004; Corteel, 2006b). The effectiveness of these algorithms relies on the measurement of the free-field response of each individual loudspeaker along a control line in the horizontal plane. Since most anechoic chambers are too small to fit massive multi-loudspeaker arrays, measurements must be conducted in non-anechoic environments, which often requires a lengthy and complex measurement process. The array must then be placed sufficiently far away from all room boundaries to avoid unwanted reflections in the measurements (an example measurement setup can be seen on Figure 2.10, section 2.4).

The technical difficulties and errors inherent to this task have underlined the need for a systematic and robust method of predicting the sound field emitted by the array of loudspeakers. The need for such a method has further been reasserted in the context of ongoing research on the subject of characterizing the interaction of a WFS array with the listening room (Caulkins &

Warusfel, 2006; Caulkins *et al.*, 2006) which will gain from knowledge of the full-space radiation properties of an array. In addition, a loudspeaker radiation prediction method will provide a convenient means of simulating other potentially difficult and time-consuming laboratory studies involving multi-loudspeaker arrays, such as :

- Designing or enhancing listening environments equipped with public address or sound reinforcement loudspeaker arrays : In these types of environments the directivity characteristics of the loudspeaker array influence both sound level and intelligibility throughout the listening area. Knowledge of array directivity characteristics should therefore be incorporated into the room acoustics simulations in order to enable the proper prediction of acoustic quality and serve as a guide in the architectural design process. In the case of listening rooms that have already been built but require some form of sound reinforcement, a precise knowledge of array directivity can serve as an aid for determining optimal speaker placement.
- Evaluating the free field loudspeaker array response at different distances : This type of evaluation can be necessary in the context of WFS to verify the quality of the synthesized sound field (Corteel, 2006b). Indeed, this sound field synthesis process involves calculating loudspeaker filters that ensure the minimization of the synthesis error in a least squares sense on a set of control points, but there is no *a priori* guarantee that outside of these points the sound field will be conform to the target sound field. The verification process involves measurements of the sound field at different distances so as to ensure that the sound field propagates correctly throughout the listening area. It requires the same type of measurements as the equalization process, but at multiple distances, hence the accrued benefit of a proper prediction method.
- Equalizing an array that has undergone a modification in terms of layout : This situation can arise when a given system is meant to be transported and used in different environments (concert halls, theaters, museums, etc.). The architecture of the different environments may impose changes in the layout of the system and require to repeat the measurement/equalization process for each new location. An automatic equalization procedure taking into account the loudspeaker array's radiation characteristics can be of benefit in this type of situation. The need for a feasible prediction method will be all the more important for very long arrays used in walkthrough sound installations, where the large dimensions of the array may preclude the possibility to isolate the array for a free-field characterization in a measurement room.

Having established the relevance of loudspeaker radiation prediction techniques in the context of loudspeaker array technology, it is important to note that the demand for these techniques existed long before the advent of multi-loudspeaker arrays in research and sound arts. Indeed, a review of scientific literature related to sound radiation prediction for loudspeakers reveals numerous methods, that can generally be classified into one of three categories :

- **Closed form equations** : In these methods, a closed form equation gives access to the radiated amplitude and/or phase characteristics of the sound field radiated by the sound source at any point in space. The most commonly spread model for electrodynamic loudspeakers is that of the ideal monopole, omnidirectional at all frequencies and presenting a phase response that results strictly from propagation delay. This model is widely used in the context of loudspeaker array-based sound field synthesis techniques such as WFS and Ambisonics [see for instance (Berkhout, 1988; Berkhout *et al.*, 1993; Corteel, 2006b; Betlehem & Abhayapala, 2005; Daniel, 2000)]. More sophisticated models of loudspeaker radiation, such as that of the piston set in an infinite baffle or in a sphere, produce more precise results than the simple monopole model (Morse & Ingard, 1968). However, these models remain unadapted for the description of the rear radiation region of loudspeakers. They are equally unadapted to describe loudspeakers presenting characteristics that differ fundamentally from those of the classical electrodynamic loudspeaker, such as horns or MAPs (Multi-Actuator Panels).
- **Numerical Methods** : Work on boundary element (BEM) and Finite Difference Time Domain (FDTD) methods has yielded good results in the context of sound source radiation prediction and more particularly loudspeaker radiation prediction [see for instance (Shindo *et al.*, 1990; Johansen, 1994)]. These models can be combined with analytical models in order to reduce calculation load, and obtain good results in terms of agreement with measurement (Karjalainen *et al.*, 2001). Finite element methods (FEM), which require more complex modelling and heavier calculations, have also been employed successfully for predicting diverse loudspeakers' radiation patterns (Tsuchiya *et al.*, 2003; Pellerin *et al.*, 2003). Generally speaking, FEM and BEM are well adapted to the parametric study of shape and materials during the design phase of the loudspeaker. They are however less suited to the proposed task of providing a simple and systematic characterization method of a wide selection of finished transducers, since in this case the experimenter does not necessarily possess knowledge of materials and/or manner of assembly.
- **Measurement-based methods** : Simple techniques for predicting the radiation of loudspeaker arrays are reviewed in (Jacob & Birkle, 1990). In so-called *power sum* techniques,

a non-omnidirectional magnitude response is supposed—and measured—for each loudspeaker, but phase interactions between different elements of the array are neglected. In so-called *phasor sum* techniques, a full-space magnitude response is also measured for each loudspeaker and a phase response due solely to propagation delay is factored into the calculation of the array response. None of these methods are suitable for the laboratory studies enumerated above since these studies require a precise full-space characterization of individual loudspeakers' phase responses. However, more precise methods based on solutions to the Helmholtz integral equation measurements have been employed as early as the 1960's to project acoustic fields in the underwater and airborne domain (Horton & Innis, 1961; Schenck, 1968; Copley, 1968). These methods generally rely on the measurement of the pressure or normal velocity either on a planar surface or on a surface enclosing the source, and can be used to predict the sound field along a more distant position (forward projection) or at a position which is closer to the source (backward projection). Among these methods, Nearfield Acoustic Holography (NAH) has proven to be useful for the study of the radiation of harmonic sound fields (Maynard *et al.*, 1985; Veronesi & Maynard, 1987; Williams, 1999)). Other methods have since been developed to deal with transient, time varying sound fields (Forbes *et al.*, 1991; Clement *et al.*, 1997; Rochefoucauld *et al.*, 2004). Among all of these methods, work by (Weinreich & Arnold, 1980) on the measurement of violin radiation patterns using a spherical harmonic expansion of the sound field appears particularly well adapted to the task of characterizing loudspeaker radiation patterns. Their method employs measurements of the source on two concentric spheres in order to separate incoming and outgoing components. The technique can be seen to fall into the category of Generalized Spherical Holography (Maynard *et al.*, 1985), and constitutes the inspiration for the following study.

The goal of this article is to develop a systematic and robust method for predicting the radiation of an array of loudspeakers. To do so, the authors propose to develop a measurement-based description of individual component loudspeakers belonging to the array, allowing the prediction of their complex response at any position in space. Determining the full array response at any position is then easily achieved using the principle of superposition.

Section 2.1 is dedicated to establishing the theoretical formulation of three radiation prediction methods based on a spherical decomposition of the sound field. These methods rely on the measurement of loudspeaker radiation patterns over either a single sphere or a double sphere of microphones. Section 2.2 gives a description of the measurement apparatus and protocol. In section 2.3, computer simulations are carried out in order to determine the influence of parameters such as source size, number of sampling points, and measurement noise. Performances of the three prediction methods are compared using a energy-based and complex criteria in order

to determine the best possible method to use in the context of this study. Finally, section 2.4 presents practical results of the prediction method applied to electrodynamic satellite speakers and Multi-Actuator Panel (MAP) loudspeakers mounted in an array configuration.

2.1 Theory

This section is dedicated to formalizing the proposed radiation prediction method.

2.1.1 Spherical harmonic expansion of the sound field

The proposed radiation prediction method derives from a resolution of the wave equation in the spherical coordinate system (Fig. 2.1). This coordinate system is particularly well adapted to the description of a sound field emitted by a centered sound source radiating outwards, such as a loudspeaker.

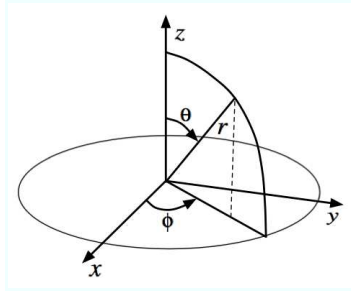


FIG. 2.1 – Definition of spherical coordinates relative to the Cartesian coordinate system. ϕ represents the azimuthal angle, measured from the x axis within the (x,y) plane ; θ represents the elevation angle, measured from the z axis.

Supposing separability of the three variables $\{r, \theta, \phi\}$, the resolution of the wave equation using spherical coordinates yields the following expansion, with e^{-ikct} implicit, for the acoustic pressure field p :

$$p(r, \theta, \phi, k) = \sum_{l=0}^{+\infty} \sum_{m=-l}^l \left(b_{lm}^{(1)}(k) h_l^{(1)}(kr) + b_{lm}^{(2)}(k) h_l^{(2)}(kr) \right) Y_l^m(\theta, \phi) \quad (2.1)$$

This expansion derives from the homogeneous wave equation, and is therefore valid only within a region of space devoid of any sources or scatterers.

For the sake of interpretation, equation (2.1) is split into two parts :

- a **propagation** term $(b_{lm}^{(1)}(k) h_l^{(1)}(kr) + b_{lm}^{(2)}(k) h_l^{(2)}(kr))$

– a **directivity** term $Y_l^m(\theta, \phi)$

These terms are explicated in the following subsections.

Propagation term of the expansion

The term $(b_{lm}^{(1)}(k)h_l^{(1)}(kr) + b_{lm}^{(2)}(k)h_l^{(2)}(kr))$ in equation (2.1) is a weighted sum of spherical Hankel functions of first and second kind describing the propagation of the sound field. Spherical Hankel functions of the first kind $h_l^{(1)}(kr)$ are associated to the outgoing component of the sound field. They are composed of a sum of a spherical Bessel functions, such that :

$$h_l^{(1)}(kr) = j_l(kr) + in_l(kr) \quad (2.2)$$

where $j_l(kr)$ represents the first order spherical Bessel function and $n_l(kr)$ represents the second order spherical Bessel function, which is divergent at the origin. Spherical Hankel functions of the first kind are therefore divergent at the origin, as shown by their nearfield ($kr \ll l$) approximation

$$h_l^{(1)}(kr) \approx -i \frac{(2l-1)!!}{(kr)^{l+1}}, \quad (2.3)$$

and asymptotically equivalent to a simple monopole source in the farfield ($kr \gg l$), with characteristic phase shifts depending on their order l :

$$h_l^{(1)}(kr) \approx (-i)^{l+1} \frac{e^{ikr}}{kr}. \quad (2.4)$$

Spherical Hankel functions of the second kind $h_l^{(2)}(kr)$ are the complex conjugates of the spherical Hankel functions of the first kind. They are associated with the incoming component of the sound field.

Directivity term of the expansion

The term $Y_l^m(\theta, \phi)$ in equation (2.1) describes the angular dependence or directivity of the sound field. It represents a family of functions $Y_l^m(\theta, \phi)_{l>0, |m|<l}$ called spherical harmonics, which follow increasingly complex 3D spatial patterns as their order m and degree l increase. To put it simply, the order m determines the rate of variation of Y_l^m following the azimuth angle ϕ , and the degree l determines the rate of variation of Y_l^m following the elevation angle θ . A more detailed description of spherical harmonics is provided¹, for instance, in (Williams, 1999). For the purpose of this work we follow the common practice of using the term "order" to designate both "order" and "degree" (Daniel, 2000).

¹See also Annex B

2.1. Theory

The spherical harmonics can be chosen so as to form an orthonormal basis of the solutions of equation 2.1, i.e. :

$$\langle Y_l^m(\theta, \phi) Y_{l'}^{m'}(\theta, \phi) \rangle_{4\pi} = \int_{\theta=0}^{\pi} \int_{\phi=-\pi}^{\pi} Y_l^m(\theta, \phi) Y_{l'}^{m'}(\theta, \phi)^* \sin\theta d\theta d\phi = \delta_{ll'} \delta_{mm'}, \quad (2.5)$$

where δ_{ij} represents the Kronecker delta, equal to one when $i = j$, and zero otherwise. It follows that any given directivity function f is decomposable upon the spherical harmonic basis using the so-called *inverse spherical fourier transform* :

$$\forall f(r, \theta, \phi, k), \quad \exists! \{a_{lm}(r, k)\}_{l>0, |m|<l}, \quad f(r, \theta, \phi, k) = \sum_{l=0}^{+\infty} \sum_{m=-l}^l a_{lm}(r, k) Y_l^m(\theta, \phi), \quad (2.6)$$

such that the knowledge of the series of a_{lm} coefficients is equivalent to the knowledge of values of f upon the surface of a sphere. In light of equation (2.5), the series of a_{lm} coefficients can be determined by projecting the directivity function f onto each individual spherical harmonic basis function :

$$\begin{aligned} \langle f(r, \theta, \phi, k), Y_l^m(\theta, \phi) \rangle_{4\pi} &= \int_{\theta=0}^{\pi} \int_{\phi=-\pi}^{\pi} \left(\sum_{l=0}^{+\infty} \sum_{m=-l}^l a_{lm}(r, k) Y_l^m(\theta, \phi) \right) Y_l^m(\theta, \phi)^* d\theta d\phi \\ &= a_{lm}(r, k) \end{aligned} \quad (2.7)$$

This projection-based analysis² forms the ground of the proposed radiation prediction method. When used to analyze the measurements of a directivity pattern over a sphere, it yields a description of the sound field as a sum of radiation modes. For example, the Y_0^0 mode can be assimilated to an acoustic monopole, just as the Y_1^1 , Y_1^{-1} , and Y_1^0 modes can be assimilated to acoustic dipoles along the x , y and z axes. These radiation modes each possess their own propagation function, and by summing their contributions one can effectively extrapolate the value of the measured sound field to any point in 3D space.

2.1.2 Radiation prediction methods

In this section the theoretical elements of sound field analysis presented in section 2.1.1 are used as a basis for establishing a full-space radiation prediction from the measurement of a loudspeaker on one or two concentric spheres.

First off, one should note that equation (2.1) is equivalent to stating that for a fixed wave number k the full-space pressure field is entirely determined by the spherical harmonics expansion coefficients $\{b_{lm}^{(1)}(k), b_{lm}^{(2)}(k)\}_{l>0, |m|<l}$. In other words, to be able to estimate the pressure

²referred to as the Continuous Spherical Harmonic Transform in certain publications (Laborie, 2000).

field at any point in space one need only determine the $\{b_{lm}^{(1)}(k), b_{lm}^{(2)}(k)\}_{l>0, |m|<l}$ coefficients. Moreover, in light of equation (2.7), the measurement of the sound field on a sphere of radius r_i and its subsequent projection onto the spherical harmonics basis is sufficient to determine the value of the $a_{lm}(r_i, k)_{l>0, |m|<l}$ coefficients. They are directly linked to the coefficients of the spherical harmonic expansion by the following relationship [using eq. (2.6) and (2.1)] :

$$a_{lm}(r_i, k) = b_{lm}^{(1)}(k)h_l^{(1)}(kr_i) + b_{lm}^{(2)}(k)h_l^{(2)}(kr_i) \quad (2.8)$$

Sections 2.1.1 and 2.1.2 present methods that use this relationship to determine the spherical harmonic expansion of the sound field associated to the loudspeaker —i.e. its full-space radiation pattern— from either single or double sphere measurements.

Radiation prediction method based on single sphere measurements

The first method is based on the hypothesis that all sources and scatterers are constrained to the zone situated inside the measurement sphere. This is a valid supposition if true free field conditions exist outside of the measurement sphere. In this case, the incoming component of the sound field is null, which implies that $b_{lm}^{(2)}(k) = 0$ in equation (2.1).

Projecting the measurements conducted along the sphere of radius $r = r_0$ onto the spherical harmonics basis yields $a_{lm}(r_0, k)$ [eq. (2.7)]. It follows from equation (2.8) that :

$$\forall(l, m), \quad b_{lm}^{(1)}(k) = \frac{a_{lm}(r_0, k)}{h_l^{(1)}(kr_0)}. \quad (2.9)$$

The extrapolation \tilde{p} of the sound field to another position (r_1, θ_1, ϕ_1) is then obtained by replacing $b_{lm}^{(1)}(k)$ in expression (2.1) :

$$\tilde{p}(r_1, \theta_1, \phi_1, k) = \sum_{l=0}^{+\infty} \sum_{m=-l}^l \left(b_{lm}^{(1)}(k) h_l^{(1)}(kr_1) Y_l^m(\theta_1, \phi_1) \right). \quad (2.10)$$

Radiation prediction method based on double sphere measurements

If no suppositions are made about the sound field, the determination of the coefficients of the spherical harmonic expansion requires the measurement of the sound field on two concentric spheres of radius r_1 and r_2 (Weinreich & Arnold, 1980). The projection of these measurements onto the spherical harmonics basis [eq. (2.7)] yields the value of $a_{lm}(r_1, k)$ and $a_{lm}(r_2, k)$, after which the resolution of a linear equation system leads to the determination of $\{b_{lm}^{(1)}(k), b_{lm}^{(2)}(k)\}_{l>0, |m|<l}$ from equation (2.8).

The extrapolation of the sound field to any given position (r, θ, ϕ) is then obtained directly

2.1. Theory

from the expansion of eq. (2.1). This expansion is valid in a region of space $r_0 < r < r_1$ such that all sources and scatterers are either situated within the sphere $r < r_0$ or outside the sphere $r < r_1$.

In cases where the incoming sound field consists entirely of unwanted reflections off of the measurement structure, it can be interesting to try to distinguish the outgoing (useful) sound field from the incoming one. In such cases, the sound field can be decomposed into the sum of an outgoing sound field produced by the centered loudspeaker and the incoming sound field produced by external sources that, as such, do not create a singularity at the center of the sphere. This implies that the a_{lm} coefficients obtained from the projection of the measured sound field onto the spherical harmonics can alternatively be expressed in a manner that separates outgoing and incoming contributions :

$$a_{lm}(r, k) = c_{lm}(k)h_l^{(1)}(kr) + d_{lm}(k)j_l(kr) \quad (2.11)$$

Since this decomposition and that of equation (2.8) are valid for all r values in a given domain, the following relations are equally valid :

$$\begin{cases} d_{lm}(k) &= 2b_{lm}^{(2)}(k) \\ c_{lm}(k) &= b_{lm}^{(1)}(k) - b_{lm}^{(2)}(k) \end{cases} \quad (2.12)$$

Incoming and outgoing contributions are thus separable from the measurement of the sound field on two spheres. This means in particular that the sound field emitted by the loudspeaker being measured can be isolated so as to simulate a situation resembling a single sphere measurement in which the incoming sound field is null. The prediction of the sound field radiated by the loudspeaker at any position in space is then made using the $c_{lm}(k)$ coefficients.

Maximum order for the spherical harmonic decomposition

Let r_{min} represent the radius of the smallest concentric sphere encompassing the loudspeaker. The $a_{lm}(r_{min}, k)$ series necessarily decreases for high values of l for reasons of convergence of the spherical harmonic series and continuity of the loudspeaker's directivity pattern. Supposing the absence of sources on the outside of the sphere encompassing the loudspeaker, the sound field emitted by the loudspeaker can then be predicted correctly throughout the domain defined by $r > r_{min}$.

Let us consider a measurement sphere of radius $r_0 > r_{min}$. In this case the $a_{lm}(r_0, k)$ series is such that $a_{lm}(r_0, k) = a_{lm}(r_{min}, k) \left(\frac{kr_{min}}{kr_0}\right)^{l+1}$ in the near field (eq. (2.3) & (2.9)) and $a_{lm}(r_0, k) = a_{lm}(r_{min}, k) \left(\frac{kr_{min}}{kr_0}\right)$ in the far field (eq. (2.4) & (2.9)). These relations point to the fact that high l components exert a much greater influence near the origin than in the far

field, which translates the fact that the loudspeaker effectively transmits very little energy to the measurement sphere for l -order components such that $l > kr_{min}$. If the measurements are being used to extrapolate the sound field towards the vicinity of the loudspeaker, this can be a problem since in that case the high- l noise caused by random measurement errors will entail large errors in the estimation of the radiation pattern (Weinreich & Arnold, 1980). In the present work, predictions only concern the outward-going radiation, so this is not an issue.

As a consequence, the components of interest for the prediction method can be restricted to an order l such that :

$$l \leq l_{max} \quad \text{where} \quad l_{max} = kr_{min} \quad (2.13)$$

2.1.3 The discrete spherical harmonic decomposition

The model established thus far for full-space radiation prediction requires the knowledge of the pressure field $p(r, \theta, \phi, \omega)$ for all (θ, ϕ) on a continuous spherical surface. In practice, the microphones used to measure the pressure field only allow the estimation of the sound field one point at a time, whereby surface data must necessarily be sampled at a discrete number of points. The notion of *discrete spherical harmonics* must therefore be introduced.

In the discrete case, directivities are defined by a set of indexes I of cardinal N and a data vector $\{f(\theta_i, \phi_i)\}_{i \in I}$. The discrete scalar product can thus be defined as :

$$\langle f, g \rangle_{4\pi d} = \frac{1}{4\pi} \sum_{i \in I} f(\theta_i, \phi_i) g(\theta_i, \phi_i)^* \Omega_i \quad (2.14)$$

where Ω_i is a weighting factor equal to the solid angle subtended by the spherical surface surrounding the point i over which the value of (θ_i, ϕ_i) needs to be extrapolated. For instance, if the sphere is uniformly sampled, then for all i , $\Omega_i = \frac{4\pi}{N}$. If the sampling scheme is irregular, the weighting factor becomes indispensable in order to adjust the influence of points situated in dense or sparse zones.

The computation of Ω_i is conducted in the $\phi \times \eta :] - \pi : \pi] \times [-\frac{\pi}{2} : \frac{\pi}{2}]$ plane, where η is defined as $\eta = \theta - \frac{\pi}{2}$. Using the right bisectors between every point of I , a Voronoi diagram representing *zones of influence* for each sampling point is calculated. (Fig. 2.2, top). Once the Voronoi polygon corresponding to point (θ_i, ϕ_i) has been calculated, the value of Ω_i is then calculated using the following integral :

$$\Omega_i = \iint_{P_i} \sin \theta d\theta d\phi \quad (2.15)$$

The numerical resolution of this integral is facilitated by the observation that given a sufficiently dense sampling of the sphere one can approximate $\sin \theta$ by its mean value (or its value

2.1. Theory

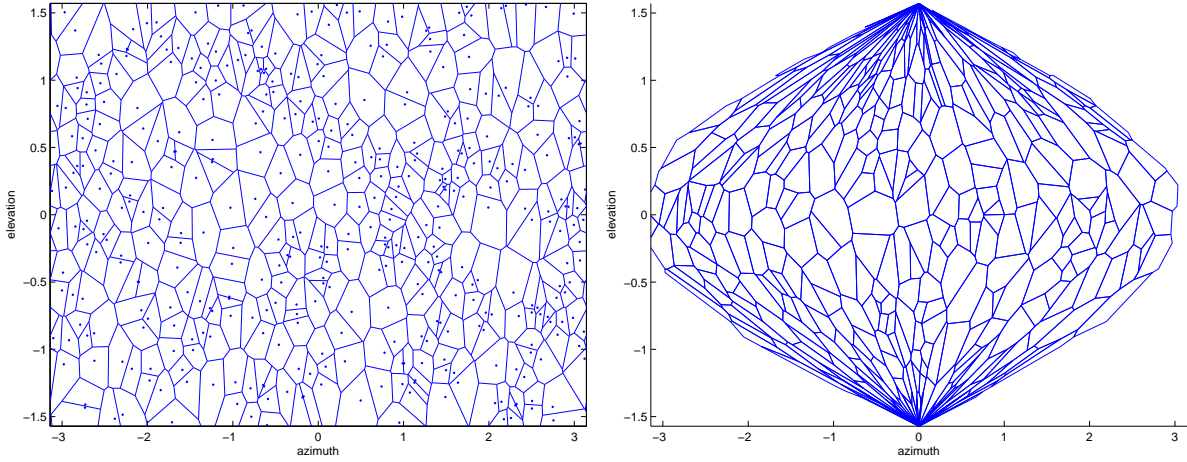


FIG. 2.2 – **Left** : Voronoi diagram for a random sampling of the $\phi \times \eta$ plane. **Right** : Voronoi diagram in which the polygons are distorted by a factor of $\cos \eta = \sin \theta$ so that their area may better represent the solid angle surrounding the corresponding point on the sphere.

at (θ_i, ϕ_i)). Another solution consists in distorting the polygons by a factor of $\sin \theta$ before evaluating the area of the resulting polygons (Fig. 2.2, bottom).

Orthonormality of the discrete spherical harmonics

In light of the previous section, it is clear that, in practice, loudspeaker directivity functions are manipulated as discrete size- N vectors, thus imposing an N -dimensional vector space as the environment for the analysis. Similarly, the continuous spherical harmonics Y_l^m are replaced by a discrete $N \times 1$ sized vector representation y_s , in which harmonics of increasing degree and order are indexed by an integer $s \in \mathbb{N}$.

Thus $y_0 = \{Y_0^0(\theta_i, \phi_i)\}_{i \in I}$, $y_1 = \{Y_{-1}^1(\theta_i, \phi_i)\}_{i \in I}$, $y_2 = \{Y_0^1(\theta_i, \phi_i)\}_{i \in I}$, etc.

Given the finite dimension of the vector space, the spherical harmonics no longer form an orthonormal basis. We define their scalar product as :

$$\langle Y_l^m, Y_{l'}^{m'} \rangle_{4\pi d} = \frac{1}{4\pi} y_s^T \Omega y_{s'} \quad (2.16)$$

where $\Omega(N \times N)$ is a matrix of weighting factors specific to each measurement point, defined in eq. (2.15).

To obtain the best possible description of the directivity function from its projection upon the discrete spherical harmonic basis, $\langle Y_l^m, Y_{l'}^{m'} \rangle_{4\pi d}$ must approach the Kronecker delta $\delta_{ss'}$. In other words, if the first harmonics are grouped into a matrix $Y = [y_1 y_2 \cdots y_P]$ it should ideally respect the following relation :

$$Y^T \Omega Y = I_P, \quad (2.17)$$

where I_P represents the order P identity matrix.

The total number of harmonics up to an order l_{max} is $\sum_{l=0}^{l_{max}} (2l + 1) = (l_{max} + 1)^2$. This signifies that a minimum of $N_{harm} = (l_{max} + 1)^2$ measurement points are needed to ensure that $Y = [y_1 y_2 \dots y_P]$ represents a linearly-independent set, which is a necessary condition for the fulfillment of equation (2.17). Y is henceforth defined as an $N \times N_{harm}$ matrix grouping all spherical harmonics up to order l_{max} .

Decomposition using the pseudo-inverse

Despite the fact that the discrete spherical harmonics do not form an ideal orthonormal basis, it is nonetheless possible to decompose a directivity function onto the vector space that they generate, by means of an orthogonal projection. The basic supposition here is that the measured directivity function $f(r, \theta_i, \phi_i, k)$ is *well-described* by the spherical harmonic components contained in Y , meaning that its contributions to the higher orders are negligible.

The relation between the continuous directivity function f and the coefficients of the spherical harmonic decomposition is given by the inverse spherical fourier transform [eq. (2.6)]. In the discrete case presented here, the fact that $N > N_{harm}$ implies that determining the series of a_{lm} coefficients from equation (2.6) is equivalent to the resolution of an overdetermined equation system. This implies in turn that contrarily to the continuous spherical harmonics transform, the discrete spherical harmonics transform is non-reversible. This absence of reversibility of the discrete transform can be modelled as the inclusion of a transformation noise. In the discrete case, eq. (2.6) is therefore represented by the following matrix equation :

$$f = Ya + \epsilon \quad (2.18)$$

where a is a $N_{harm} \times N_\omega$ matrix containing the value of the a_{lm} coefficients of the decomposition for each pulsation ω (N_ω represents the total number of pulsations), f is a $N \times N_\omega$ matrix containing the measured directivity function at each position for each pulsation ω , and ϵ is a matrix orthogonal to the basis functions in Y representing errors due to the non-reversibility of the system. The solution to this problem that minimizes the error vector ϵ can be obtained by use of the pseudoinverse matrix :

$$a = (Y^T Y)^{-1} Y^T f \quad (2.19)$$

This solution can be modified taking into account the relative weight of each sampling point, using a diagonal weighting matrix W such that :

$$a = (Y^T W Y)^{-1} Y^T W f \quad (2.20)$$

It is indispensable to ensure that the rank of the y_s family is at least N_{harm} , otherwise $Y^T W Y$ is non-invertible. The quality of the sound field description relies on the proper conditioning of this matrix ; the better it is conditioned, the less ambiguous the directivity will be. This conditioning can be optimized by modifying the spatial distribution of sampling points and the weighting matrix. In practice, the choice $W = \Omega$ improves the orthonormality of the discrete spherical harmonics basis.

Sampling scheme and maximum order of decomposition

Periodicities in the distribution of measurement points have been shown to cause the aliasing of higher order components (see for instance (Laborie *et al.*, 2003)). To put it simply, when spatial variations of the directivity function occur more frequently than the distance between measurement points they are incorrectly interpreted as lower order variations and introduce errors in the subsequent radiation prediction. This aliasing phenomenon can be reduced by choosing an irregular or random distribution of measurement points. In the context of the present study however, a sampling scheme in which measurement points are regularly distributed over the $\theta \times \phi$ grid is nonetheless retained for practical reasons linked to the geometry of the measurement apparatus (section 2.2).

In theory, the minimum amount of sampling points necessary to ensure linear independence between the spherical harmonics is $l_{max} + 1$, where l_{max} represents the maximum order of decomposition. Given the periodicities of the $Y_l^m(\theta, \phi)$ function, for an order L decomposition on a regular grid, at least $l_{max} + 1$ points are needed in θ and $(2l_{max} + 1)$ points are needed in ϕ . However, to obtain better conditioning of the $Y^T W Y$ matrix it is preferable in practice to increase the number of sampling points, especially if the sound field is poorly described for the retained order of decomposition l_{max} (i.e. if the measured field is either noisy or if it contains non-negligible contributions from higher orders).

2.2 Measurement protocol and apparatus

In this section, the apparatus and experimental protocol involved in acquiring the loudspeaker directivity measurements used for the radiation prediction method are described.

Loudspeaker directivity measurements are conducted in a fully anechoic chamber located inside IRCAM. The chamber is equipped with a MIDI-controlled mechanical arm that supports a half-circle of microphones (Fig. 2.3, left & center). The arm can pivot on its axis from a position where the half-circle is contained within the horizontal plane to a fully vertical position, allowing the coverage of a quarter sphere at a time. Microphone capsules are mounted so as to allow the measurement of the sound field on two concentric spheres (Fig. 2.3, center). The inner

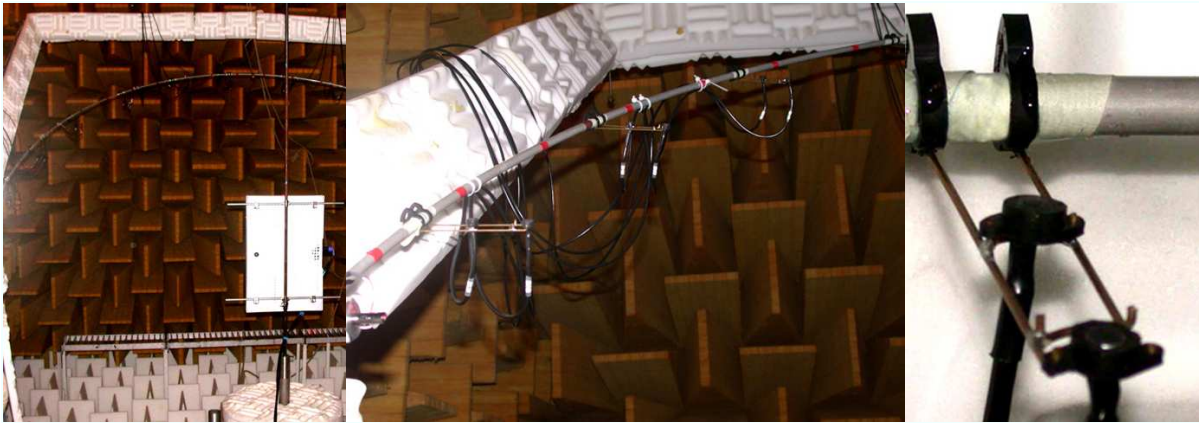


FIG. 2.3 – Anechoic chamber measurement apparatus. **Left** : Ensemble view of the measurement apparatus during the measurement of a Multi-Actuator Panel loudspeaker. The loudspeaker is mounted on an axis fixed onto a turntable, visible at the bottom right. **Center** : Partial view of the half-circle of microphones used for single and double sphere measurements. **Right** : Close-up shot of two adjacent microphone capsules.

sphere has a 1.43m radius, and the radius of the outer sphere is set to 1.48m (5cm spacing).

The acoustic origin of the loudspeaker³ is positioned at the center of the quarter sphere using a set of three coincident laser beams. The loudspeaker is fixed to a metal rod that follows the axis of a turntable, which allows variation of its orientation within the horizontal plane. The axis of the turntable coincides with a line extending through the acoustic origin of the loudspeaker, therefore minimizing displacements of the acoustic origin during rotations of the turntable.

Measurements are conducted with the loudspeaker positioned sideways, so that vertical movements of the mechanical arm measure azimuthal (ϕ) directivity variations, and the successive microphone positions along the arm measure directivity variations along the elevation angle θ . The data acquisition system used in this setup allows for 24 simultaneous measurements ; 12 microphone pairs are therefore installed along the half-circle. These pairs are arranged with constant angular spacing along the half circle, covering 12 different elevations. For practical reasons, the microphones remain in a fixed position along the half-circle during the measurement sessions. This sets the maximum sampling resolution for elevation to 15° . Azimuthal sampling resolution can however be varied at will by adjusting the rotation angle increment of the mechanical arm, so that 0.3° is the maximum possible resolution in terms of azimuth angle.

The system is entirely automated via a software interface that allows sequencing of movements of the mechanical arm, movements of the turntable and acquisition of loudspeaker measurements on a 24 channel soundcard. The loudspeaker responses are measured using a logarithmically swept-sine technique (Farina, 2000).

³See section 2.4.1 for a discussion regarding the acoustic origin of MAP loudspeakers.

2.2. Measurement protocol and apparatus

The microphones used for the measurements are Panasonic MCE2000 capsules. They offer the advantage of presenting small dimensions (6mm diameter, 5mm height) in comparison to the wavelengths of interest, as well as relatively omnidirectional characteristics within the audible frequency range.

The microphones are set up so as to measure the incoming sound field with a lateral incidence along the membrane, as shown on the center and right hand photos in Figure 2.3. This choice is made to diminish any masking or diffraction effects caused by the presence of the inner-circle microphones on the sound field measured by the outer-circle microphones. A comparison between frontal and lateral incidences reveals a loss for lateral incidences above 2kHz that progressively increases to become equal to 1dB above 4kHz and superior to 4.5dB above 15kHz (Fig. 2.4). This is taken into consideration in the interpretation of results presented in section 2.4, where predictions based on anechoic chamber measurements are compared with free-field measurements conducted using the same capsules with a frontal incidence.

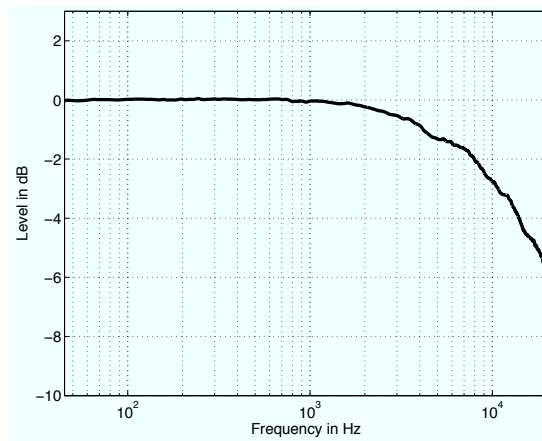


FIG. 2.4 – Smoothed magnitude response (dB) of a microphone capsule at lateral incidence deconvolved by a measurement of the same capsule at frontal incidence.

Microphone calibration measurements are carried out in the anechoic chamber prior to the installation setup. Each microphone is measured individually in identical conditions, i.e. along the frontal axis of a studio monitor loudspeaker at a 1.3m distance, the capsule set up to be measured with lateral incidence. Results show that for the mid-frequency range each measured frequency response differs only by a constant gain factor. An individual correction gain allowing compensation for these errors is calculated using the following formula :

$$G_{corr} = \frac{f_{max} - f_{min}}{\int_{f_{min}}^{f_{max}} H(f)df} \quad (2.21)$$

where $H(f)$ represents the frequency response of the capsule being measured, $f_{min} = 100Hz$

and $f_{max} = 3000Hz$. After gain corrections using G_{corr} and deconvolution by a reference measurement made with a *B&K* type 4149 microphone at the same distance, the capsules' magnitude responses present a standard deviation in amplitude inferior to $0.4dB$ over the $[20 - 9000]Hz$ frequency band. Phase comparisons between the different capsules after deconvolution by the *B&K* response and removal of the linear component introduced by propagation reveal that the standard deviation between the phase response of each capsule is inferior to 0.06 radians over the $[20 - 9000]Hz$ frequency band. At higher frequencies, the effect of placement errors coupled with the individual differences between microphones cause the standard deviation between individual magnitude responses to increase up to $2dB$ and the standard phase deviation to increase up to 0.3 radians.

2.3 Simulations

Numerous factors are likely to introduce errors in the anechoic measurements, therefore affecting the subsequent radiation predictions. For the most part, these errors can be isolated and analyzed by means of computer simulations. This section describes a few of the simulations conducted by the authors to try to shed light on the influence of factors such as microphone and loudspeaker placement errors, measurement noise, or unwanted reflections off of the measurement structure.

The modelling environment used here allows the simulation of the response of an ideal monopole point source responses ($\frac{1}{r}$ amplitude attenuation and phase response consisting solely of propagation delay) on a distribution of ideal omnidirectional microphones.

2.3.1 Influence of spatial oversampling on measurement noise

In section 2.1.3, the relation between the sampling scheme and the overall quality of the sound field description is underlined : in order to reduce the influence of poorly described phenomena in the discrete spherical harmonic at a given order, it is preferable to increase the density of sampling points. Indeed, increasing the number of sampling points on the measurement sphere can be seen to cause noise to spread out onto higher order harmonics, so that for a fixed order its influence on the discrete spherical harmonic decomposition is diminished.

To verify this, simulation measurements perturbed by white noise are carried out. They involve an ideal loudspeaker at a centered position measured on a set of regularly sampled measurement spheres, for which the number of measurement points is progressively increased. In each case, the measured sound field is decomposed into spherical harmonic partials up to order 11 and subsequently recomposed at a distant position, at which point the signal to noise

ratio is calculated. This ratio is represented as a function of the number of sampling points on Figure 2.5.

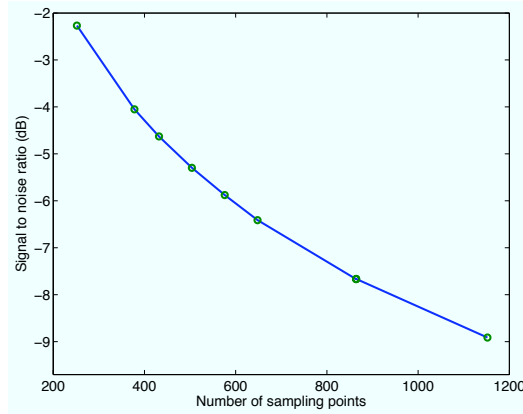


FIG. 2.5 – Signal to noise ratio as a function of the number of sampling points, considering an 11th order spherical harmonic description of the sound field.

Strictly speaking, the benefit of increasing the density of sampling points does not lie in reducing background noise or electric noise due to the measurement chain, since the logarithmic sweep method used to measure the signals typically yields signal to noise ratios equal to 60dB. However, random amplitude and phase errors due to imprecise microphone and loudspeaker positioning, as well as individual differences in microphone responses incur uniformly distributed contributions to all partial waves of the spherical harmonic expansion (Weinreich & Arnold, 1980) that remain unmitigated by the measurement method. These can effectively be reduced by increasing the number of sampling points.

The practical measurements described in section 2.4 of this article are carried out over 12 elevations and 48 azimuths, making a total of 576 measurement points, for an expansion up to order 11. This amounts to two times the minimum number of points necessary for a valid decomposition (see sec. 2.1.3), thus ensuring a significant reduction of noise due to individual microphone differences and positioning errors.

2.3.2 Influence of source size on quality of the decomposition at a given order

A theoretical formula for determining the minimum order of decomposition required to ensure a proper description of the sound field radiated by the source being measured according to its radius is recalled in section 2.1.3. This formula is verified here in a series of simulations involving a virtual microphone setup identical to the setup used during actual measurement sessions.

The extended source is modelled as a distribution of three point sources, randomly distributed around the center of the sphere. This formalism, referred to as “monopole synthesis”, was used by (Giron, 1996) to describe the radiation of extended sources. The radius of the extended source modelled this way is defined as the distance of the outermost monopole from the center.

The goal of the simulation is to determine the maximum frequency at which an order L decomposition remains valid for a given source size, i.e. the maximum frequency for which the harmonics contained in the expansion are sufficient to describe the sound field emitted by the measured sound source.

Let M be the largest modulus out of all of the a_{lm} coefficients of the spherical harmonic expansion at the retained order L . The criterion chosen to determine the maximum frequency of validity is as follows : if the coefficient of any mode belonging to a higher order $L_h > L$ has a modulus M_h such that $M_h \geq M/100$, then the decomposition at order L is considered insufficient to describe the measured sound field at that frequency.

A series of simulations is carried out in which the positions of the three monopoles used for the “monopole synthesis” are varied randomly inside a sphere of fixed radius. In each case, the maximum frequency of validity of the decomposition is calculated using the criterion enunciated above. Results are averaged over the 20 simulations to obtain an estimation of the maximum frequency for a given radius. This operation is repeated for a series of radii varying between 2cm and 20cm.

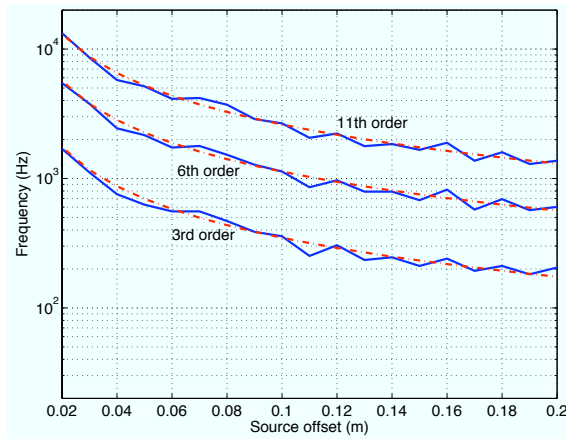


FIG. 2.6 – (color online) **Solid lines** : Estimation of the maximum frequency of proper sound field description as a function of source radius for a 3rd, 6th and 11th order decomposition. **Dash-dotted lines** : Maximum theoretical frequency of proper sound field description as a function of source radius for a 3rd, 6th and 11th order decomposition

The maximum frequency for proper sound field description according to source radius is represented for 3 different orders of decomposition in Figure (2.6). The maximum theoretical frequency for proper decomposition according to the $kr_{min} < l_{max}$ criteria given in section 2.1.3

is indicated as a reference on the same figure. The estimations derived from the simulation can be seen to agree closely with the theoretical curve.

2.3.3 Comparison of methods aimed at diminishing the influence of incoming reflections.

The measurement setup described in section 2.2 presents the drawback of generating a number of unwanted reflections that can interfere with the characterization of the loudspeaker. These reflections arise from different sources :

- **The door to the anechoic chamber.** The door to the chamber is a clearly identified source of broadband reflections. This door is situated approximately 4m from the center of the measurement sphere, which means that an acoustic wave takes approximately 0.023s to go back and forth between the loudspeaker and the closest measurement microphone when it is reflected off of the door. It can however be observed that the majority of the impulse response's energy is contained in the first hundredth of a second, after which only very weak low frequency contributions are emitted. A satisfactory elimination of this reflection is ensured by truncating responses to a length of 0.02s.
- **The mechanical arm supporting the half-circle of microphones.** Even though its surface is acoustically treated using absorbant foam material (see Figure 2.3, center), the arm remains a source of mid and high frequency reflections arriving too quickly to be truncated effectively. Contrarily to the door of the chamber, each time the half circle of microphones is moved to measure a new azimuth, the mechanical arm moves with it. The arm therefore generates an identical echo pattern on each microphone pair of a given elevation.
- **The metallic grating and turntable supporting the loudspeaker.** These portions of the measurement apparatus are the most problematic sources of reflections, since they are situated close to the loudspeaker and inside the measurement sphere itself. When the loudspeaker is rotated to measure a new quarter sphere they remain at a fixed position, which means that when the four quarter spheres are reassembled in the post-processing phase so as to recreate the measurement of a full-sphere system, the reflections created by these sources cannot be assimilated to a real sound field. The only way to deal with these reflections is to minimize them as much as possible by reducing the supporting surface to a minimum and adding acoustically absorbant material on the surface of the turntable, as shown on the bottom of the leftmost picture of Figure 2.3.

Faced with the impossibility of entirely suppressing incoming reflections, it can be interesting to investigate techniques aimed at diminishing their overall influence. This section exposes three radiation prediction methods based on the formalism developed in section 2.1. Each method is designed to minimize the influence of a given type of interference (measurement noise, reflections from a fixed external source, or reflections emanating from a “moving” source like the measurement arm). The methods are tested using computer simulations that model the different types of interference. Advantages and disadvantages of each method are weighed according to the existing constraints in the measurement installation.

Double sphere method

The radiation prediction method based on double sphere measurements presented in section 2.1.2 is referred to as the “double sphere method” in this section. This method is tested for the purpose of reducing the influence of fixed ⁴ sources of reflection situated outside of the measurement sphere.

In section 2.1.2 it was shown that for continuous directivity measurements and an infinite order spherical harmonic expansion, the double sphere measurement allows the separation of the incoming portion of the sound field from its outgoing portion. In the discrete case presented here, a perfect suppression of the incoming sound field is possible only to the extent that the incoming wave field is decomposable into a *finite* series of spherical harmonics. For an outgoing wave produced by an acoustic source of limited extension situated inside the measurement sphere it is shown in section 2.1.2 that this is automatically the case. However, for sources situated outside of the measurement sphere this is no longer necessarily true, as shown by the limit case of the incoming plane wave which incurs equal contributions on the entire spherical harmonic series.

In terms of the analysis of the incoming sound field, the measurement sphere can be assimilated to a large-radius, high-order soundfield microphone (Abhayapala & Ward, 2002; Meyer & Elko, 2002)). In (Rafaely, 2005), it is shown that for a spherical microphone of radius R_{arr} the analysis error of an incoming plane-wave sound field at low frequencies ($kR_{arr} \leq l_{max}$) can mainly be attributed to noise and positioning errors. However, at high frequencies, the dominant cause for error becomes spatial aliasing. In the setup presented here, the sound field is decomposed up to order 11 on a microphone array such that $R_{arr} = 1.48m$; in other words, above 400Hz it is expected that spatial aliasing will become the dominant source of error for the incoming sound field estimation.

⁴The term “fixed” is used to imply that the position of these sources does not shift when the measurement arm is moved to a measure a new azimuth.

Single sphere cardioid microphone method

This method is inspired from the single sphere method described in section 2.1.2. The idea is to use the available pairs of omnidirectional microphones to form cardioid patterns oriented towards the center of the circle in order to cancel out reflections caused by the presence of the measurement arm. The equations allowing the formation of the response of a cardioid microphone (combination of an omni and bidirectional microphone, weighted by $(1 + \cos \alpha)$ where α is the favored direction) around a point \vec{r}_0 are written as follows, considering that the microphones are situated in $\vec{r}_0 + \frac{\Delta r}{2}$ and $\vec{r}_0 - \frac{\Delta r}{2}$:

- The omnidirectional component $\tilde{p}_{omni}(\vec{r}_0) = p(\vec{r}_0)$ is estimated by averaging the two microphone signals :

$$\hat{p}_1(\vec{r}_0) = \frac{p(\vec{r}_0 + \frac{\Delta r}{2}) + p(\vec{r}_0 - \frac{\Delta r}{2})}{2} \quad (2.22)$$

- The bidirectional component $\tilde{p}_{bi}(\vec{r}_0) = \cos \alpha p(\vec{r}_0)$ can be estimated by a pressure gradient. A first order approximation in $\|k\| \|\Delta r\|$ of this gradient is written :

$$\hat{p}_2(\vec{r}_0) = p(\vec{r}_0 + \frac{\Delta r}{2}) - p(\vec{r}_0 - \frac{\Delta r}{2}) \quad (2.23)$$

A combination of these two contributions yields the following approximation \hat{p}_3 for the response of a cardioid microphone to an incoming plane wave :

$$\hat{p}_3(\vec{r}_0) = \frac{1}{4j\|k\|\|\Delta r\|} \cdot \hat{p}_2(\vec{r}_0) + \frac{1}{2} \cdot \hat{p}_1(\vec{r}_0) \quad (2.24)$$

This virtual directivity pattern is the result of a first order approximation in $\|k\| \|\Delta r\|$, meaning that it loses validity when $\|k\|^2 \|\Delta r\|^2$ becomes non-negligible. Moreover, it involves a division by k , which implies increased errors in the low frequency range and sets limits on the virtual cardioid microphone's valid frequency range.

Single spheres averaged method

The two previous methods attempt a suppression of external reflections incurring limitations in terms of frequency. The idea in this third method is simply to ignore any structural reflections and apply the single sphere method described in section 2.1.2 directly.

Since two measurement spheres are available, two separate sets of coefficients $\{\hat{a}_{lm}(r_1, k), \hat{a}_{lm}(r_2, k)\}$ can be extracted and averaged so as to limit the influence of measurement noise and placement errors. The following estimator proposed from the expansion coefficients measured on each sphere :

$$\hat{c}_{lm}(k) = \frac{\hat{a}_{lm}(r_1, k)h_l^{(2)}(kr_1) + \hat{a}_{lm}(r_2, k)h_l^{(2)}(kr_2)}{\|h_l(kr_1)\|^2 + \|h_l(kr_2)\|^2} \quad (2.25)$$

Since $h_l^{(2)}(kr) = h_l^{(1)*}(kr)$ if the measured sound field is that of a strictly outgoing wave, and supposing the absence of measurement errors (i.e. $\hat{a}_{lm}(r, k) = a_{lm}(r, k) = c_{lm}^{(1)}(k)h_l^{(1)}(kr)$), then :

$$\begin{aligned}\hat{c}_{lm}(k) &= \frac{a_{lm}(r_1, k)h_l^{(2)}(kr_1) + a_{lm}(r_2, k)h_l^{(2)}(kr_2)}{\|h_l(kr_1)\|^2 + \|h_l(kr_2)\|^2} \\ &= c_{lm}^{(1)}(k)\end{aligned}\quad (2.26)$$

On the other hand if the incoming sound field is non-null because of the presence of reflected components, and supposing the absence of measurement errors (i.e. $\hat{a}_{lm}(r, k) = a_{lm}(r, k) = c_{lm}^{(1)}(k)h_l^{(1)}(kr) + c_{lm}^{(2)}(k)h_l^{(2)}(kr)$), then :

$$\begin{aligned}\hat{c}_{lm}(k) &= c_{lm}^{(1)}(k) + c_{lm}^{(2)}(k) \frac{h_l^{(2)}(kr_1)^2 + h_l^{(2)}(kr_2)^2}{\|h_l(kr_1)\|^2 + \|h_l(kr_2)\|^2} \\ &= c_{lm}^{(1)}(k) + c_{lm}^{(2)}(k)Q\end{aligned}\quad (2.27)$$

The estimated quantity $c_{lm}^{(1)}(k)$ depends on the weighting factor $Q \leq 1$. In virtue of eq. (2.3) if $kr_{i=1,2} \ll l$ then Q reaches its maximum value ($Q \approx 1$) and the estimation of the outgoing field loses in accuracy.

Simulations : Choice of error criteria

In order to compare the characteristics of the three methods described in section 2.3.3, 2.3.3 and 2.3.3, a series of simulations is proposed. In each simulation, a virtual microphone setup identical to the one described in section 2.2 is used to measure a given sound field. The sound field is decomposed and forward projected according to each method onto a reference sphere of radius $R_{ref} = 5m$ (each projected point conserves the same azimuth and elevation). The result of this projection constitutes the so-called ‘‘predicted’’ sound field. This ‘‘predicted’’ sound field is compared to the ‘‘target’’ sound field produced by the measured source at the same positions on the reference sphere, using two criteria described below.

The first criterion C_d is designed to place equal importance on phase and amplitude in the evaluation of the predicted sound field. Phase is given particular attention here, notably because the predicted sound fields are to be used in applications requiring precise knowledge of the phase response. C_d therefore based on the complex vector \vec{X}_d , defined as follows :

$$\vec{X}_d(nMic, f) = \vec{X}_p(nMic, f) - \vec{X}_t(nMic, f), \quad (2.28)$$

with $nMic$ the index of the microphone being considered, f the frequency, \vec{X}_t the Fourier transform of the target impulse response and \vec{X}_p the Fourier transform of the impulse response

2.3. Simulations

effectively predicted by the method. Let $\langle \rangle_{mics}$ design the average over all microphone positions and $\|\cdot\|$ the L2-norm of a vector. The complex criterion is defined as :

$$C_d = 10 \cdot \log_{10} \left(\frac{\langle \|\vec{X}_d(nMic, f)\|^2 \rangle_{mics}}{\langle E_t(nMic, f) \rangle_{mics}} \right) \quad (2.29)$$

with $E_t(nMic, f)$ the energy received at microphone position $nMic$ at the frequency f .

Above the aliasing frequency since proper control of the phase is impossible the complex criterion becomes ill-adapted for the evaluation of the prediction error. A second criterion E_d , based on the energy measured over all the microphone positions, is proposed to complete the study. Let $E_t(nMic, f)$ represent the energy of the target sound at the microphone position $nMic$ and $E_p(nMic, f)$ represent the energy predicted by the method at the same position ; the energy criterion is defined as :

$$E_d = 10 \cdot \log_{10} \left(\frac{\langle E_p(nMic, f) \rangle_{mics}}{\langle E_t(nMic, f) \rangle_{mics}} \right) \quad (2.30)$$

Simulation results : Source situated inside the measurement sphere

In the first simulation, the sound field emitted by a fixed source situated inside the measurement sphere, 0.2m from the center, is measured on the virtual microphone setup, decomposed up to order 11, and predicted along the reference sphere according to the three methods described in sections 2.3.3, 2.3.3 and 2.3.3. The predictions are made with a 60dB signal to noise ratio so as to reflect typical measurement conditions. An additional “single sphere 24 elevations” curve is indicated, corresponding to a measurement of the same sound field on a single sphere (1.43m radius) using twice as many elevations as were used in this study (24 instead of 12). A 23rd order single sphere decomposition (sec. 2.1.2) made from this measurement is used to predict the sound field along the reference sphere. This curve is added to show the performances of the prediction method when the 24 available channels are used on a single sphere instead of applying the double sphere techniques.

The following observations can be made from Figure 2.7 :

- The energy difference E_d criterion (which, in this case, can be assimilated to a difference in emitted power, since the sound field radiates through the measurement sphere) shows that the “Single spheres averaged” and the “Double sphere” methods give a correct estimation of the emitted power up to 3kHz. This is the expected value at which the order 11 decomposition is no longer sufficient to describe the outgoing sound field ($kr \leq l_{max}$ with $r = 0.2m$ and $l_{max} = 11$ implies a proper estimation for all $f \leq 3kHz$). The complex estimator C_d shows the same tendencies ; below 3kHz, the complex difference between the predicted sound field and the target sound field remains equal to 10dB ; however, above

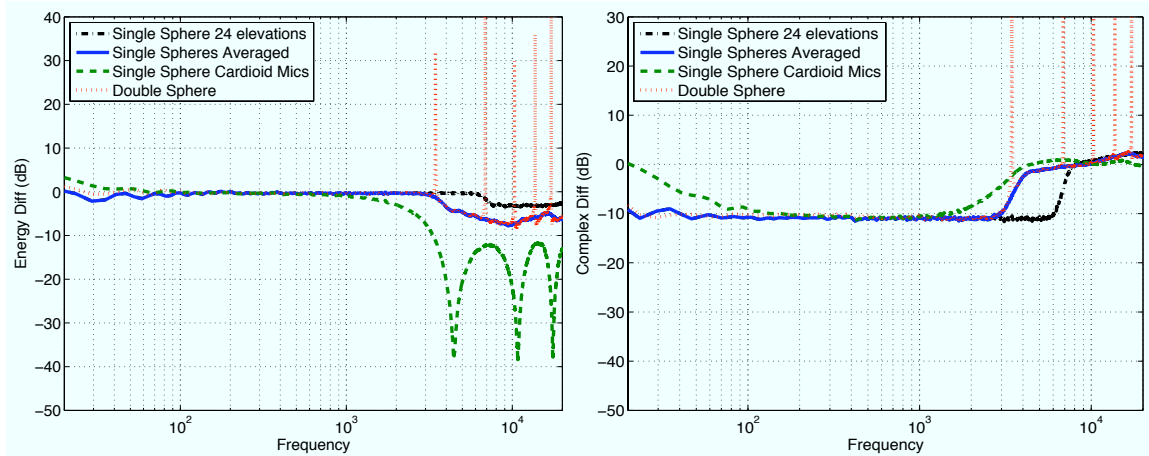


FIG. 2.7 – (color online) **Left** : Average energy difference E_d (dB) between the predicted sound field and the target sound field along the reference sphere. **Right** : Average complex difference C_d (dB) between the predicted sound field and the target sound field along the reference sphere. In both cases the sound field is produced by a fixed source situated inside the measurement sphere at a 0.2m distance from the center of the sphere, with a signal to noise ratio of 60dB on each microphone.

3kHz, the norm of the complex vector X_d tends to become equal to the norm of the target sound field, which reflects the incoherence between the prediction and the target.

- The “Double sphere” estimation diverges in both plots at certain frequencies because it involves a division by a Hankel function that tends to zero for certain values of kr . The presence of low level measurement noise helps reduce the extent of these divergences.
- The “Single sphere cardioid mics” estimation works correctly up to approximately 1kHz. Above this frequency, the second order term $\|k\|^2\|\Delta r\|^2$ is no longer negligible compared to $\|k\|\|\Delta r\|$ and the pressure gradient approximation begins to fail. Moreover, the presence of very strong divergences in the E_d criterion can be observed above 4kHz as a result of comb filtering. The cardioid microphone estimation is also affected at low frequencies because of the division by k that enters into the gradient calculation (see section 2.3.3).
- The “Single sphere 24 elevations” estimation allows an increase in the maximum frequency of correct estimation of the sound field from 3kHz to approximately 6.4kHz.

Simulation results : Source situated inside the sphere + Measurement arm reflections

In the second simulation, a simple model of the reflection off of the measurement arm is added to the measured sound field used for the prediction. The effect of the arm is modelled separately on each microphone as a simple reflection off of a flat surface situated 0.5m behind the outer sphere. The flat surface is tangential to the sphere at the angle defined by the

2.3. Simulations

microphone position. It creates an echo that reaches the outer microphone $\frac{(0.5+0.5)}{343} = 2.9ms$ after the direct sound and the inner microphone $\frac{(0.55+0.55)}{343} = 3.2ms$ after the direct sound when the source is centered. The magnitude of the reflected component is attenuated by an arbitrarily chosen factor of 0.5 over the entire frequency band. No diffraction is taken into account.

The sound field produced by source situated 0.2m from the center of the sphere is simulated on each microphone of the measurement sphere using the described reflection model. The result is decomposed up to order 11, and propagated to the reference sphere according to the three methods. This prediction is compared with the target sound field, which consists in the response of the measured source at the reference sphere positions devoid of any reflections. For the sake of interpretation, the signal to noise ratio is increased to 400dB, and an additional “single sphere” curve is added (Figure 2.8). This curve represents a situation in which the prediction is made from a simple order 11 single sphere decomposition made using the inner sphere.

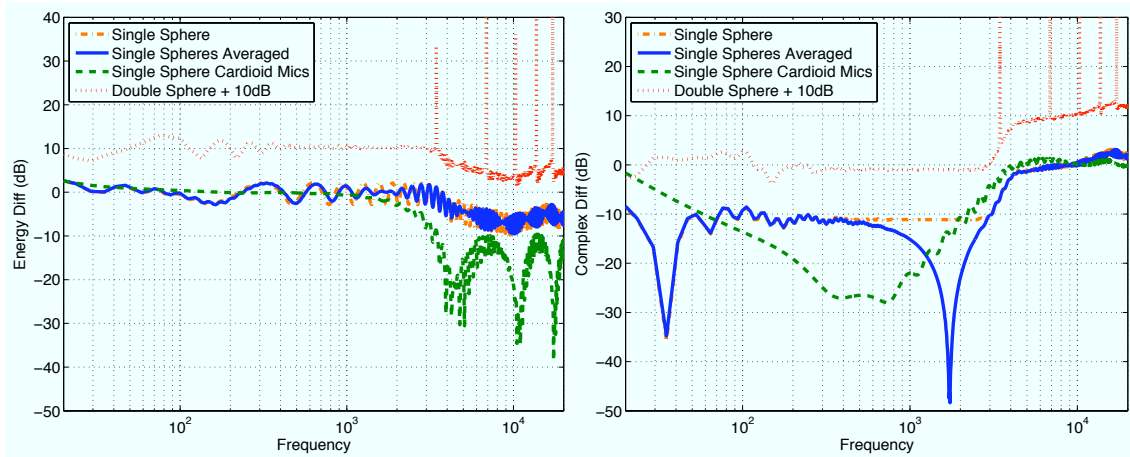


FIG. 2.8 – (color online) **Left** : Average level difference E_d (dB) between the predicted sound field and the target sound field along the reference sphere. **Right** : Average complex difference C_d (dB) between the predicted sound field and the target sound field along the reference sphere. In this simulation, the prediction is based on the sound field emitted by a fixed source situated inside the measurement sphere at an 0.2m distance from the center perturbed by echo patterns that model the presence of a measurement arm. The target sound field consists in the response of the same source along reference sphere, devoid of any reflections. These simulations are made supposing a signal to noise ratio of 400dB on each microphone. The “Double Sphere” curve is represented 10dB above the others to facilitate the legibility of the two figures.

The following observations can be made from Figure 2.8 :

- The simulated reflection creates a beating pattern in the energy difference criterion for the “single spheres averaged” method (Fig. 2.8, left), due to a difference in arrival times between the reflections at the microphone positions. This also causes a notch in the complex difference (Fig. 2.8, right). This notch should not be interpreted as a good performance of the prediction algorithm, but rather as the result of a comb filtering produced

by the addition of coherent waves along the measurement sphere. In the “single sphere” prediction, a beating pattern also appears in energy difference ; it is however no longer modulated as is the case for the “single spheres averaged” prediction. The notch observed at 1.7kHz for the “single spheres averaged” method also disappears in the “single sphere” prediction. The reader may note that these interference phenomena are reduced when the measurement noise is increased to -60dB.

- Modulations set aside, the “Double Sphere” and “Single Spheres Averaged” methods similar display characteristics to those exhibited in the previous simulation (see Fig. 2.7). The “Single sphere cardioid mics” method is however unaffected by the beating patterns observed in the energy criterion for the other methods (Fig. 2.8, left). Moreover, this method shows a lower complex difference in the [100-1000]Hz band. These improvements reflect the capacity of the cardioid microphones to cancel out reflections emanating from behind the microphone pairs. The “Double Sphere” method does not manage to cancel out these reflections since they can in no way be attributed to a stable and physically coherent source given that the image sources created by the measurement arm “move” during the session.

Simulation results : Source situated inside the sphere + Reflection from fixed external source

In the third and final simulation, the reflection off of a fixed obstacle in the anechoic chamber perturbs the measured sound field. The obstacle is modelled to be a perfectly reflecting surface situated 25cm outside of the measurement sphere. When a centered source emits a sound wave, the obstacle generates a sound field corresponding to an image source situated 3.5m from the center of the sphere. No diffraction effects are taken into account.

Using this model, the sound field produced by source situated 0.2m from the center of the sphere, along with the perturbing image source created by the obstacle, is simulated on each microphone of the two measurement spheres. The result is decomposed up to order 11, and propagated to the reference sphere according to the three methods. This prediction is compared with the target sound field, which consists in the response of the measured source at the reference sphere positions, in absence of the perturbing image source. The simulations are made supposing a signal to noise ratio of 60dB on each microphone. Results of this simulation can be seen on Figure 2.9.

The observation of average level difference curves E_d (Fig. 2.9, left) does not show a significant difference with the unperturbed situation (Fig. 2.7, left), aside from a slight decrease in the “Single spheres averaged” curve at low frequencies (-0.5dB to -1dB below 300Hz). However, the complex criterion C_d represented on the right hand graph of Fig. 2.9 undergoes more impact

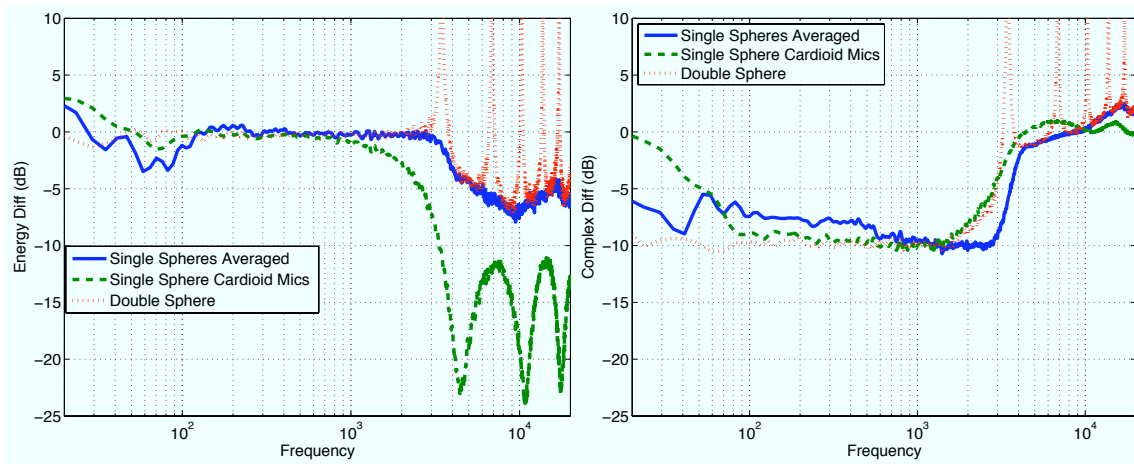


FIG. 2.9 – (color online) **Left** : Average level difference E_d (dB) between the predicted sound field and the target sound field along the reference sphere. **Right** : Average complex difference C_d (dB) between the predicted sound field and the target sound field along the reference sphere. In this simulation, the prediction is based on the sound field emitted by a fixed source situated inside the measurement sphere at an 0.2m distance from the center, perturbed by an image source situated 3.5m from the center modelling the effect of a fixed obstacle in the anechoic chamber. The target sound field consists in the ideal response of the same source along the reference sphere, in absence of any obstacles. The simulations are made supposing a signal to noise ratio of 60dB on each microphone.

from the perturbing reflection. For the ‘Single spheres averaged’ technique, one can observe that C_d increases from -10dB in the [100-500]Hz frequency band in the non-perturbed case (Fig. 2.7, right) to -7dB in the [100-500]Hz frequency band in this case (Fig. 2.9, right). This increase in the complex difference does not occur for the ‘Double sphere’ technique, indicating that the method manages to cancel out the perturbing reflection. Similarly, the ‘Single sphere cardioid mics’ method exhibits the same complex difference as it did in the first simulation (Fig. 2.9, right, and Fig. 2.7, right), indicating that it also cancels out the perturbing reflection in the [100-1200]Hz band. However, below 100Hz it continues to display errors, since the method relies on the estimation of the phase difference between the two capsules and becomes perturbed by the measurement noise at low frequencies.

Above 1.8kHz, the perturbations caused by the external obstacle begin exciting the higher orders of the spherical harmonic decomposition and introducing aliasing errors in the estimations. Given that the ‘Double sphere’ technique relies on a difference between the estimations on two spheres (eq. 2.12) it suffers from accrued sensitivity to estimation errors, which explains the increase in complex error for the ‘Double sphere’ technique above 2kHz as compared to the ‘Single spheres averaged’ method (Fig. 2.7, right).

Simulations : Conclusions

Generally speaking, the simulations presented above show that the choice of a prediction method should be guided by an analysis of the practical conditions in the measurement room. If the mobile part of the measurement apparatus presents a strong influence, the “Single sphere cardioid” method should be employed. If fixed obstacles are present in the measurement environment, the “Double sphere” method should be preferred. In truly free-field conditions, however, the “Single spheres averaged” method provides the same spatial resolution as the “Single sphere cardioid” and “Double sphere” methods, with a wider frequency bandwidth. Moreover, since it involves averaging two estimations, it is naturally more robust to measurement and placement noise than the decomposition on either the inner or the outer sphere considered individually. In the framework of the present study, it was observed that the fixed obstacle constituted by the door to the chamber was sufficiently distant so that it could be removed by simple truncation of the response, thereby limiting the relevance of the “Double sphere” method in this case. The influence of the measurement arm was also seen to be strongly attenuated by absorbant material, which tended to demote the idea of employing the “Single sphere cardioid” method. The “Single spheres averaged” method therefore appeared as the optimal prediction method for the double sphere setup built in the context of this study. Informal practical tests, involving comparisons with free-field measurements, confirmed that the “Single spheres averaged” method was indeed the most precise method for predicting loudspeaker radiation patterns at different distances for this configuration. In retrospect, however, it was observed that a low radius, 24-elevation single sphere setup would undoubtedly produce the better results than the double-sphere 12-elevation setup used here to test the different methods. This point is further discussed in the conclusion of the article as a basis for improving the installation.

2.4 Practical Results

In this section practical results of the loudspeaker radiation prediction technique are presented. Two different types of loudspeakers are considered for this study : electrodynamic satellite loudspeakers and Multi-Actuator Panel (MAP) loudspeakers. The satellite speakers used for the study are KEF model HTS-2001 2-way loudspeakers with coaxial (1/2” and 4”) drivers. The MAP loudspeakers are 40cm×60cm panels with 4 exciters per panel, manufactured by SonicEmotion.

Loudspeaker measurements are conducted in the anechoic chamber using the setup described in section 2.2 and decomposed onto the spherical harmonic basis following the “Single spheres averaged” method described in section 2.3.3. The decompositions are then employed to predict the sound field radiated by each loudspeaker along a line situated 2 or 3m away in

2.4. Practical Results

the horizontal plane. Quality assessment for the method is carried out by comparing the obtained predictions with reference measurements conducted in the Espace de Projection concert hall. The reference measurements involve a linear array of 48 MAP or satellite speakers with 15cm spacing between each speaker, measured in the free field along a line of omnidirectional microphones situated in front of the array (Fig. 2.10).

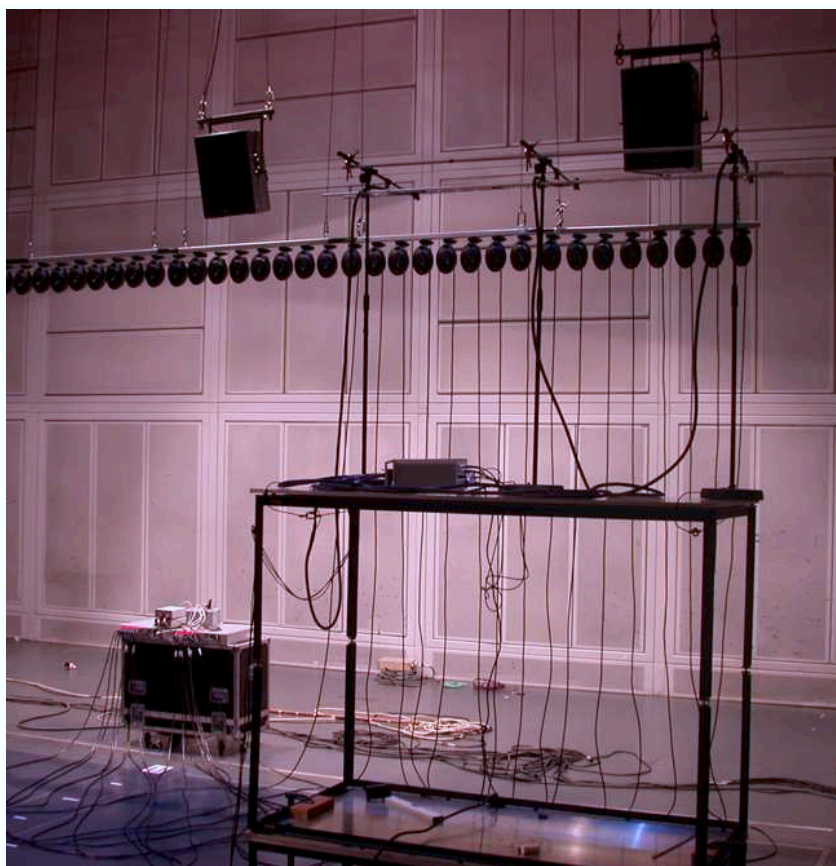


FIG. 2.10 – Free field measurement setup in the Espace de Projection concert hall. A 48 channel loudspeaker array (15cm spacing) is hung from the ceiling and measured on a 24 channel linear microphone array (10cm spacing) propped on a moveable base. The microphone array is laterally repositioned 4 times in order to cover the entire control line (96 positions).

2.4.1 Sound field prediction results for a Multi-Actuator Panel loudspeaker

A MAP loudspeaker can essentially be described as a stiff, light panel undergoing bending wave excitation by means of individually driven exciters (in this case, 4) affixed to its rear surface. This results in a modal directivity pattern : in a given direction, the MAP panel alternatively amplifies or attenuates certain frequencies, depending notably on the size of the panel,

the way it is clamped to its frame, and the position of the exciter on the panel. MAP panels also possess a quality of spatial diffuseness that has been shown to be perceptually desirable in the context of Wave Field Synthesis applications, (Corteel *et al.*, 2007), hence the interest of applying the radiation prediction technique to the study of these loudspeakers.

Choice of a measurement origin

Radiation predictions for the MAP loudspeakers is preceded by the choice of an acoustic origin on which to center the measurement. In (Jacob & Birkle, 1990), it is shown that the acoustic origin of a loudspeaker (notably horns) can incur spatial variations as a function of frequency. In the case of MAP loudspeakers, none of the exciters are situated at the center of the panel. A first possible choice would be to center the measurement on the panel, therefore limiting the radius r_{min} of the smallest sphere surrounding the source and, in virtue of section 2.3.2, increasing the maximum frequency for which the order 11 decomposition remains valid. On the other hand, various structural and vibro-acoustic studies have shown that, though the entire MAP panel radiates acoustic energy, most of the energy that it produces arises from the vicinity of the exciter itself [see for instance (Beer *et al.*, 2005)]. This tends to indicate the necessity to center the measurement on the position of the exciter.

In order to decide an optimal placement for the MAP loudspeaker, a comparative measurement session is carried out in which the loudspeaker is either centered on the exciter or on the panel. The reconstruction of the sound field at the exact points of measurement is made in each case and compared to the original measurement. The results, shown in Figure 2.11, indicate that the average prediction error greatly increases at higher frequencies when the measurement origin is centered on the panel. The choice of an acoustic origin coincident with the exciter itself should thus be made for any predictions involving MAP panels.

Results

Having resolved the question of choosing a proper measurement origin, the measurement and radiation prediction of the sound field emitted by the leftmost exciter on a MAP panel is carried out and compared with its measurement in the free field measurements at a 2m distance over a 110° angular width. Results of the prediction and the free field measurements are represented in Figure 2.12. Observation of the two graphs reveals that the general aspect of the modal radiation patterns for the MAP panel is correctly estimated by the prediction. One can observe that the sound field prediction (Fig. 2.12, left) displays more regular patterns than the free field measurements (Fig. 2.12, right) in which small striations appear indicating interferences caused by parasite reflections off of the measurement setup. In a direct comparison between free field and anechoic measurements, errors manifest as shifted or attenuated zeros that tend to follow

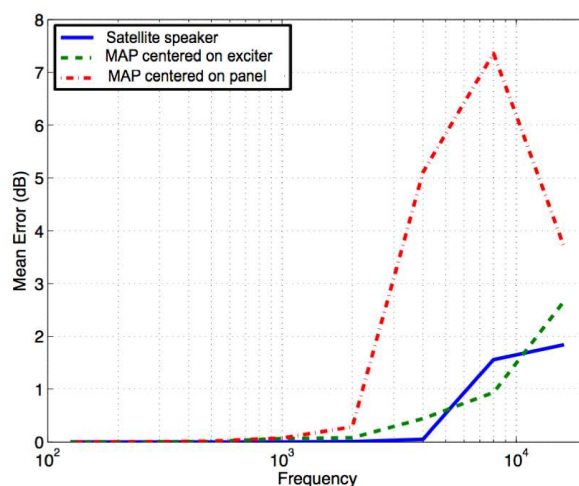


FIG. 2.11 – (color online) Error in magnitude (dB) averaged over all microphone positions per frequency band for a sound field prediction based on the order 11 spherical harmonic expansion of a satellite speaker (**solid line**), a MAP loudspeaker centered on the exciter being measured (**green dashed line**) and a MAP loudspeaker centered on the panel itself (**red dash-dotted line**).

the modal patterns. These errors are greatly increased by the fact that the reference free field measurements suffer from a slight lack of microphone positioning (up to 1cm) which in the case of MAP panels can greatly affect results, due to the dependency of their fine structure of radiation on frequency (Gontcharov *et al.*, 1999). Indeed, fluctuations over a range of 10dB can be observed over small angular sectors on the graphs Figure 2.12.

Errors begin to increase above 1.5kHz, since above this frequency the sound field is poorly described by the order 11 decomposition. Errors also increase along the sides, where the energy emitted by the panel is lower and therefore more difficult to estimate properly. An overall level underestimation is also apparent in the prediction, above 2kHz. This can be partly attributed to the difference between capsule incidences during the two measurements (see section 2.2).

2.4.2 Sound field prediction results for an electrodynamic loudspeaker

The measurement and radiation prediction of the sound field emitted by an electrodynamic satellite speaker at a 3m distance, over a 110° angular width, is carried out so as to test its accuracy in comparison with free field measurements. Results for the prediction and the free field measurements of the same speaker are represented on Figure 2.13.

The observation of the two graphs reveals that below 1.5kHz, the prediction fits the measurements quite well with the exception of two zones presenting visible artifacts. These error zones are visible in the left-hand plot of Figure 2.15, which represents the magnitude of the

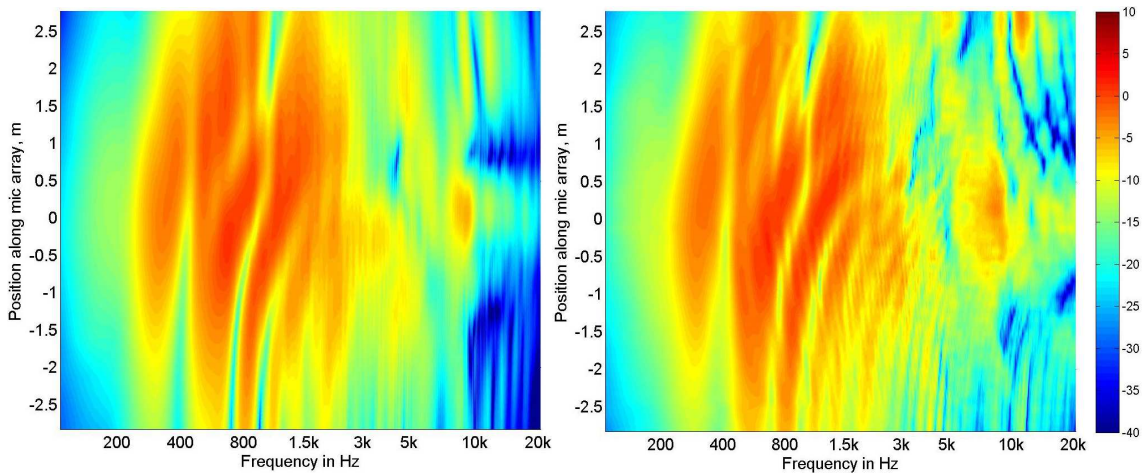


FIG. 2.12 – **Left** : Magnitude (dB) of the sound field emitted by a MAP exciter along a line situated 2m in front of it in the horizontal plane, as indicated by the sound field prediction. **Right** : Magnitude (dB) obtained from free field measurements of the MAP exciter along the same line.

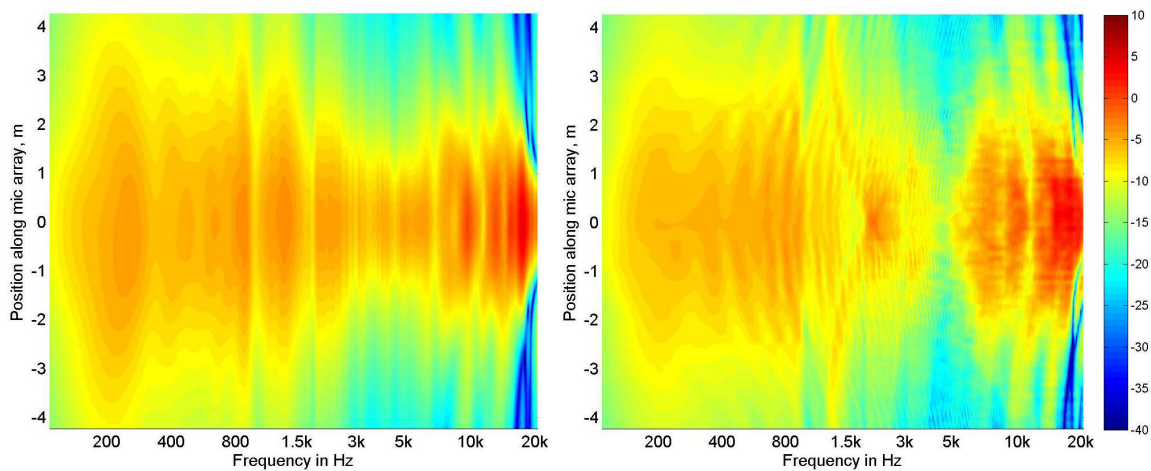


FIG. 2.13 – **Left** : Magnitude (dB) of the sound field emitted by a satellite speaker along a line situated 3m in front of it in the horizontal plane, as indicated by the sound field prediction. **Right** : Magnitude (dB) obtained from free field measurements of the satellite speaker along the same line.

ratio (prediction)/(free field measurement) in dB. The first zone, situated around 300 Hz and lower (especially along the sides) can be attributed to the lack of coherence between the four quarter spheres used to recombine the full measurement sphere in the anechoic chamber setup. Errors, notably due to the presence of the floor gratings (see section 2.3.3), are introduced in the spherical harmonic expansion, and manifest as low frequency spatial variations, that skew the subsequent sound field prediction at low frequencies. Similar discrepancies between the MAP predictions and measurements can also be observed below 300Hz (Fig. 2.12).

2.4. Practical Results

The second error zone appears as a triangular interference pattern that begins along the 1kHz crossover frequency of the loudspeaker⁵ and ends around 2.5kHz. This error zone, which has a variable width as a function of frequency, represents the interference effect due to the presence of adjacent loudspeakers of the array. Indeed, given the dimensions of the loudspeakers [198mm(l)×130mm(w)×150mm(d)], the passive presence of adjacent speakers at a 15cm distance is enough to create an interference pattern in the sound field for frequencies of similar wavelengths, i.e. between 1kHz-3kHz. Above 3kHz, the loudspeaker becomes sufficiently directional along its axis that the interference no longer takes place. This appears quite clearly in the right hand plot of Figure 2.13. On the other hand, it does not appear for the prediction, which is based on an anechoic measurement of a single loudspeaker isolated from the array.

To verify this assertion, the measurement is replicated using a loudspeaker mounted on a bar with two loudspeakers to each side, so as to recreate the passive presence of the array during the anechoic measurement session (figure 2.14).



FIG. 2.14 – Measurement of an electrodynamic KEF loudspeaker mounted between four other loudspeakers (two to each side) to recreate the presence of the adjacent loudspeakers in a WFS array.

⁵The crossover frequency is a notch at 1kHz, visible on the both graphs of Fig. 2.13

The resulting prediction is compared to the free field measurement of the same loudspeaker using the magnitude (dB) of ratio (Prediction)/(Free field measurement) on the right hand plot of Figure 2.15, alongside the same ratio (Prediction)/(Free field measurement) for an isolated loudspeaker (left hand plot). The presence of the array during the measurement improves the prediction : one can observe that the triangular interference pattern is attenuated by up to 5 or 6dB. This supports the hypothesis that the effect of neighboring loudspeakers creates a non negligible interference on the sound field, which should be taken into account when the prediction method is applied to loudspeaker arrays. If this interference is taken into account during the measurement, it is observed that the prediction remains within 2dB of the actual measurements at all positions below 1.5kHz, with the exception of a small zone at the 1kHz crossover frequency along the sides. Similar experiments conducted on MAP loudspeakers reveal that interference due to the presence of adjacent loudspeakers is negligible. Above 1.5kHz,

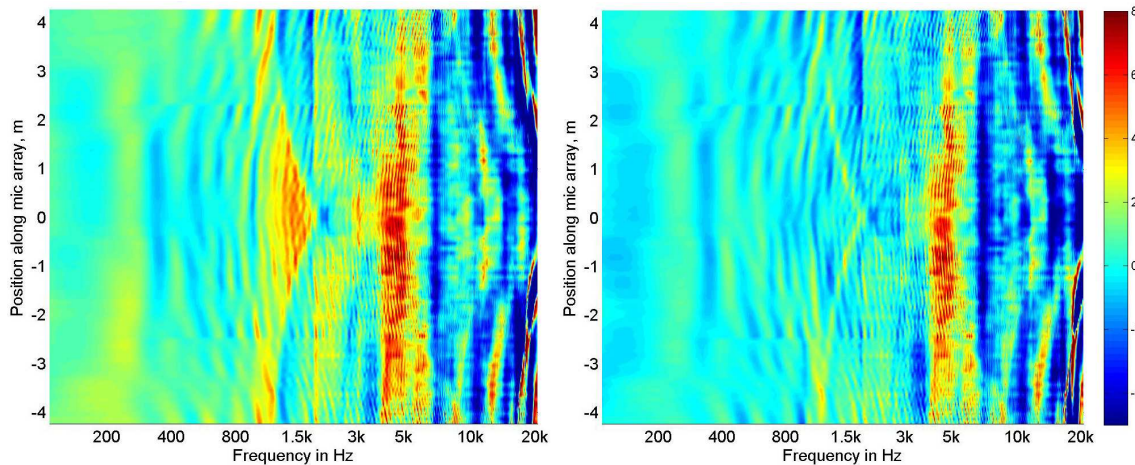


FIG. 2.15 – **Left** : Magnitude (dB) of the ratio (Prediction)/(Free field measurement) in which the prediction is based on measurements of a single loudspeaker. **Right** : Magnitude (dB) of the ratio (Prediction)/(Free field measurement) in which the prediction is based on the measurement of the same loudspeaker placed in an array.

the magnitude of the sound field prediction contains considerable errors for the same reasons as listed in the previous section (inadequate order of decomposition and influence of microphone incidence).

2.5 Conclusion

This article presents an exploratory study aimed at developing a simple and systematic laboratory method for predicting the sound field emitted by an array of loudspeakers. To do so, the means for extrapolating the complex sound field radiated by a single loudspeaker in three

dimensions are established, from which the sound field emitted by a composite array can be deduced in light of the principle of superposition.

The study begins with a theoretical description of radiation prediction methods based on the spherical harmonic expansion in the continuous domain (section 2.1). Two situations are distinguished : one in which the loudspeaker is measured with a single sphere of microphones, and the other in which the loudspeaker is measured over two concentric spheres of microphones. In both cases, the maximum order of decomposition is shown to theoretically depend on the extent of the source : the smaller the source, the lower the order needed to describe its radiation.

Implementation of the discrete spherical harmonic decomposition is described in section 2.1.3. A simple means of determining weighting factors associated with sampling points for non-uniform sampling schemes is expounded, based on the Voronoi diagram of the sampling grid. These weighting factors allow in practice to optimize the conditioning of the decomposition problem and improve upon the orthonormality of the discrete spherical harmonic basis.

Simulations are carried out in section 2.3 to isolate and analyze the influence of various factors on the quality of the prediction. Spatial oversampling is shown to reduce the effect of random amplitude and phase errors due to either imprecise positioning of the microphone and loudspeaker or dispersion in microphone characteristics. The previously established formula relating maximum valid frequency of decomposition to source extent at a given order is verified. Finally, three different methods of sound field decomposition are explored based on either a single or double sphere measurement. In the single sphere measurement, the outgoing sound field is acquired and decomposed directly onto the spherical harmonics, supposing perfect anechoicity of the environment surrounding the measurement microphones. Conversely, the double sphere measurement allows to acknowledge the existence of an incoming field caused by unwanted reflections, and effectively reduce its influence on the spherical harmonic expansion. This can be achieved in one of two ways : either by separating the outgoing field from the incoming field [following a method described in (Weinreich & Arnold, 1980)], or by weighting the concentric microphone pairs so as to form a sphere of virtual cardioid microphones pointing inwards.

The different approaches are compared in a set of simulations in section 2.3.3. The simulations indicate that the “single sphere cardioid” method is well adapted to the reduction of reflections caused by the measurement apparatus itself. It suffers however from a limited bandwidth linked to the estimation of the gradient component using two pressure capsules, which implies a tradeoff between a proper phase estimation at low frequencies and a higher spatial aliasing frequency. The tradeoff is linked to the choice of capsule spacing (5cm in this work). Better results could be obtained by combining multiple measurement sessions using different microphone spacings. The “double sphere” method is shown to allow the reduction of incoming reflections created by fixed obstacles situated outside of the measurement sphere ; on the other

hand it lacks in robustness to measurement noise and displays artifacts in the high frequency estimation, due to a division by a function that tends to zero for certain values of kr . Conversely, the “single spheres averaged” method does not allow the reduction of the effect of unwanted reflections but remains efficient in the estimation of the outgoing sound field up to relatively high frequencies, depending on the radius of the measured source. For the same number of elevations it also provides a better robustness to measurement noise than a single sphere decomposition. This method was thus adopted for the analysis of the practical measurements of section 2.4.

In section 2.4 practical measurement aspects are explored, and results of the prediction method are presented for Multi Actuator Panel and electrodynamic satellite speakers. It is shown that centering the MAP on the exciter during the measurement phase improves high frequency predictions as compared to a measurement in which the MAP is centered on the panel itself. In the case of satellite speakers, the passive presence of neighboring loudspeakers is shown to affect the response of the speaker being measured by up to 5-6 dB. When this is taken into account, the results obtained by the prediction method are found to be equal to actual free field measurements within a range of 2dB up to 1.5kHz. In the case of MAP loudspeakers, the verification process is made more difficult by the sensitivity of their fine frequency structure to slight angular variations. Nonetheless, the general aspect of the modal patterns are properly restituted up to 1.5kHz. Given that 1.5 kHz is a typical aliasing frequency for WFS systems, the radiation prediction method is therefore applicable in the initially prescribed context of WFS equalization. For other applications such as room acoustics simulations, further accuracy may be required in the high frequency domain.

In practical terms, the different experiments that were conducted show that overall accuracy of the predictions will accrue given the following enhancements in the measurement structure :

- Use of an irregular distribution of the sampling points so as increase the aliasing of higher order components of the spherical harmonic expansion (see section 2.1.3)
- Concentration of measurement points around the axis of maximum energy output for the measured loudspeaker and ensure a better conditioning of the decomposition matrix.
- Minimization of the support structure (turntable, metallic grating) and reduction of the diameter of the microphone circle. This will offer the benefit of allowing the elimination of structural reflections by simple truncation. Given the limited influence of obstacles in the available anechoic chamber, this will then allow to avoid using the bandwidth limited double sphere techniques and exploit the 24 available channels to conduct a single sphere 24 elevation measurement. In the case of a regular distribution this increases the order of decomposition from 11 to 23, therefore doubling the bandwidth of the description.

Future work will be devoted to implementing the suggested improvements of the measurement apparatus and furthering the evaluation of the prediction technique so as to include an assessment of its performances when applied to Wave Field Synthesis filter calculations and

2.5. Conclusion

room acoustics simulations.

Contrôle de la directivité en Wave Field Synthesis

3.1 Introduction

Une source sonore instrumentale est caractérisée d'une part par les propriétés spectrales et temporelles du son qu'elle rayonne, que l'on désigne communément par le terme *timbre* (cf. glossaire). On la caractérise d'autre part par sa *directivité*, qui détermine comment la source rayonne dans l'espace et, par extension, la manière dont celle-ci excite le lieu d'écoute lorsque celui-ci est un espace clos. Ces deux caractéristiques participent à l'identité de l'instrument, et doivent être répliquées de la manière la plus fidèle possible par le système de diffusion si l'on souhaite recréer une présence convaincante de l'instrument par voie électroacoustique.

Ceci devient d'autant plus délicat dans le cas où l'instrument virtuel est synthétisé sur scène, de manière contiguë à l'instrument réel, voire à la même position que celui-ci (situation de renforcement sonore). Dans ce cas, les auditeurs situés dans la salle de concert peuvent juger du réalisme de la scène sonore suivant deux indices acoustiques (Fig. 3.1) :

- d'une part, le son direct émis par l'instrument virtuel
- d'autre part l'effet de salle associé à l'instrument virtuel

Au chapitre 1, nous avons dressé un état de l'art de l'utilisation de la WFS pour la synthèse du son direct de sources à rayonnement omnidirectionnel. La zone spatiale visée était le plan horizontal contenant la tête des auditeurs, et nous avons supposé, par souci de simplification, des conditions de champ libre. La problématique plus complexe d'une diffusion en salle est illustrée par la figure 3.1. Dans ce cas, la directivité de la source virtuelle influe directement sur les propriétés spectrales et temporelles du son qui arrive aux oreilles de l'auditeur via son

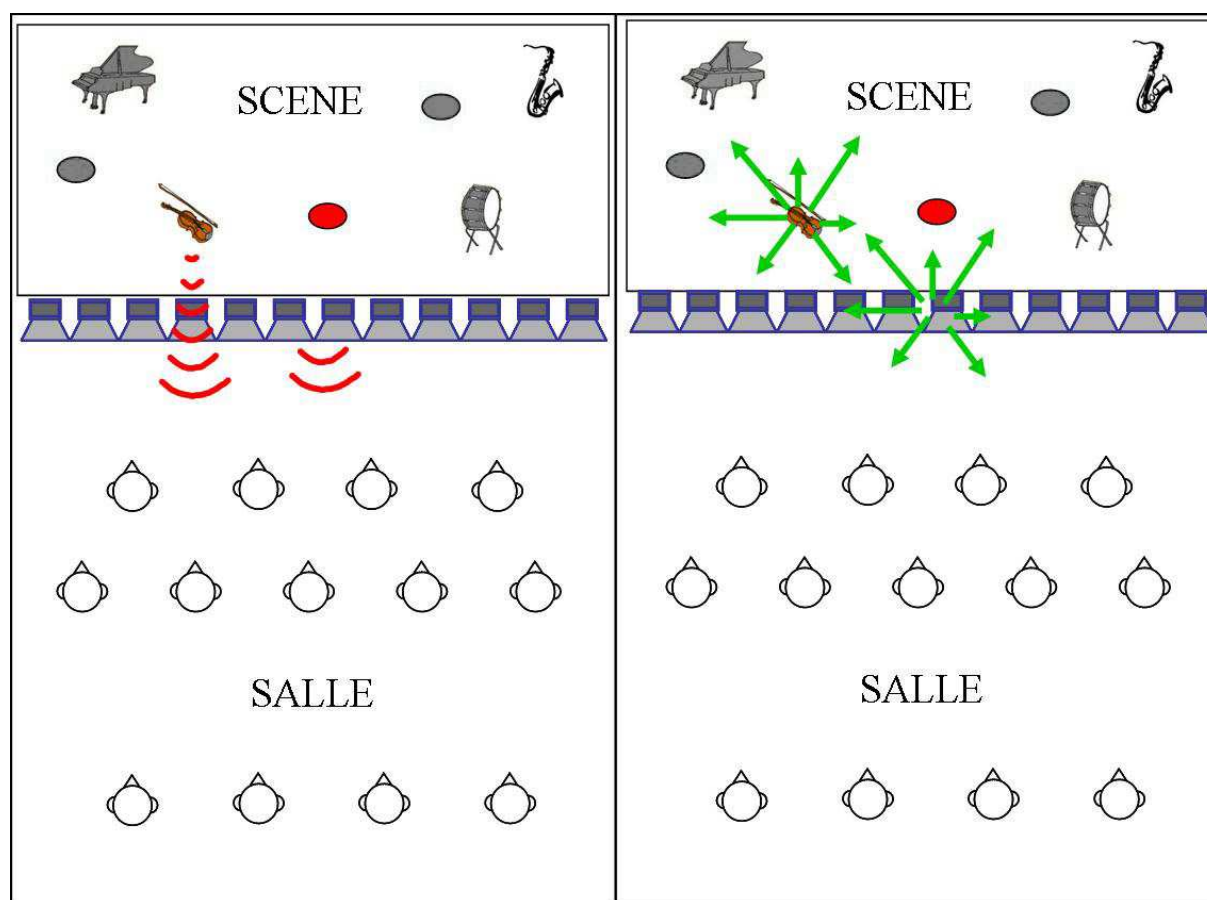


FIG. 3.1 – Schéma de la synthèse d’un instrument virtuel par un système WFS placé en nez de scène. Les sources virtuelles synthétisées par le système sont indiquées par des ronds. La figure de gauche illustre le rendu du son direct par le système WFS, indiqué par des ondes rouges. La figure de droite illustre le son émis par la source instrumentale et le système WFS qui va ensuite alimenter la salle, et générer l’effet de salle associé aux deux sources.

excitation de la salle.

Ce problème est d’ailleurs bien connu dans le cadre de la musique orchestrale. On a ainsi montré, par exemple, que les instruments d’orchestre comme les violons, violes, et violoncelles ont des caractéristiques de directivité distinctes dans lesquelles on distingue des régions de rayonnement favorisées. La direction et la largeur angulaire de ces régions varient selon la fréquence, ce qui a motivé l’idée d’adapter le positionnement de ces différents groupes d’instruments dans l’orchestre en fonction de la géométrie de la salle de concert et du style de composition (Meyer, 1972). Pour d’autres instruments comme la flûte ou le trombone, l’orientation influe de manière très sensible sur le timbre perçu à la position d’écoute. Le fait d’être en mesure de reproduire, ou tout du moins contrôler, la manière dont la source virtuelle rayonne dans l’espace tridimensionnel apparaît donc comme un facteur déterminant pour ce qui est du réalisme de l’avatar instrumental que l’on souhaite créer sur la scène.

3.1. Introduction

Cette observation a motivé le développement d'une source à directivité contrôlée : "la Timée" (Warusfel & Misdariis, 2001; Warusfel *et al.*, 2004b). Cette source est un assemblage de haut-parleurs filtrés indépendamment (Fig. 3.2, gauche) de manière à recréer les 4 premières figures de directivité de la base des harmoniques sphériques, i.e. le monopole Y_0^0 et trois dipôles Y_1^1 selon les axes \vec{x} , \vec{y} , et \vec{z} (voir Annexe B pour une description détaillée des harmoniques sphériques). Par conséquent, il devient possible de contrôler l'indice de directivité DI (cf. glossaire) du champ émis par la source, en temps réel. En effet, définissant un axe principal p_{ax} pour la Timée fixe et pris dans la direction de l'auditeur (Fig. 3.2), on peut faire varier continuellement l'indice de directivité DI dans l'intervalle $] -\infty; 4.8]$ par un jeu sur les poids respectifs du dipôle latéral Y_1^{-1} ($DI_{Y_1^{-1}} = -\infty$), du monopole Y_0^0 ($DI_{Y_0^0} = 0$), et du dipôle axial Y_1^1 ($DI_{Y_1^1} \approx 4.8dB$).

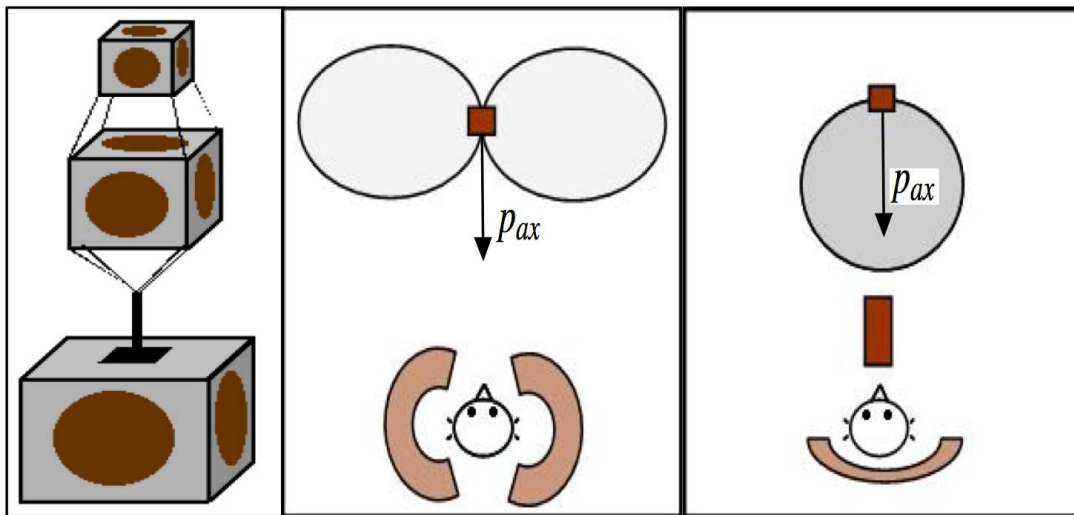


FIG. 3.2 – Gauche : Vue schématique de la Timée, assemblage de haut-parleurs permettant de contrôler la directivité du champ sonore rayonné en un point. Centre : Figure de directivité synthétisée par la Timée qui génère un champ sonore d'indice de directivité $DI = -\infty$ suivant l'axe acoustique p_{ax} , ce qui tend à réduire le niveau de son direct à la position de l'auditeur, tout en conservant des réflexions latérales. Droite : Figure de directivité synthétisée par la Timée qui génère un champ sonore d'indice de directivité $DI = 4.8dB$ suivant l'axe acoustique p_{ax} , ce qui tend à conserver un niveau de son direct et réflexions arrières élevé, tout en réduisant le niveau des réflexions latérales à la position de l'auditeur.

Par extension de ce principe, on peut également modifier l'excitation de la salle par la source. En formant une figure de directivité cardioïde orientée vers le public (Fig. 3.2, droite), on assure un niveau de son direct élevé et réflexions arrières élevé ; d'un point de vue perceptif, l'auditeur localise donc mieux la source. Par contre, en formant une figure dipolaire latérale dont le zéro est orienté vers le public (Fig. 3.2, centre), on réduit largement le niveau du son direct et on favorise les réflexions latérales, ce qui tend à augmenter la largeur apparente de la source (*Apparent Source Width*) via les réflexions précoces latérales, ainsi que l'enveloppement de l'auditeur (*Lis-*

tener Envelopment) via les réflexions tardives, jusqu'à aboutir à l'impossibilité de localisation de la source. On voit donc que la Timée fournit un moyen de modifier l'impression d'espace perçue par les auditeurs au cours d'une pièce par le biais d'un jeu sur sa directivité d'émission. Ceci illustre l'idée de "s'appuyer" sur la salle pour produire un effet. Ceci est particulièrement clair dans le cas de la figure dipolaire latérale, qui peut être utilisée pour créer un champ diffus *à la seule condition que la salle existe*.

En vertu de la pertinence perceptive des effets offerts par la Timée, l'idée est apparue d'introduire un contrôle sur la directivité des sources virtuelles en WFS. Verheijen avait déjà montré la possibilité d'inclure un opérateur de directivité dans la synthèse d'une source WFS (Verheijen, 1997), et des développements relativement récents (Caulkins, 2003; Caulkins *et al.*, 2003; Caulkins *et al.*, 2004; Warusfel *et al.*, 2004b; Corteel, 2007) ont vu l'implémentation de sources virtuelles à directivité de rayonnement variable dans le cadre de l'utilisation de la WFS à l'Ircam. Ces développements sont détaillés dans la section suivante.

3.2 Méthode modulaire pour la synthèse de directivité

L'introduction d'une directivité dans l'opérateur de synthèse WFS fut introduit initialement par Verheijen (Verheijen, 1997). Il proposa d'incorporer un terme caractéristique de la directivité dans la définition de source virtuelle monopolaire (eq. 1.7), qui avait été adoptée dans (Vogel, 1993; Start, 1997). En se plaçant en coordonnées sphériques¹, il exprime la pression rayonnée par la source virtuelle Ψ au point de réception $R(r, \theta, \phi)$ en introduisant une fonction de directivité Γ dans le calcul du champ rayonné en R :

$$p(r, \theta, \phi, \omega) = \frac{S(\omega)}{r} \Gamma(\theta, \phi, \omega) e^{-ikr}. \quad (3.1)$$

Cette expression correspond à une approximation champ lointain² du diagramme de directivité de la source virtuelle. La dépendance de Γ en ω indique la possibilité de pouvoir faire varier la directivité de la source en fonction de la fréquence, ce qui montre que le modèle incorpore la possibilité de décrire le rayonnement modal d'instruments réels. Cette description peut cependant poser certains problèmes dans le cadre d'une implémentation temps réel, du fait qu'il faille calculer et stocker un nombre de filtres proportionnel au nombre d'instruments virtuels. L'ajout d'un nouvel instrument au répertoire implique forcément une phase de calcul d'un nouvel ensemble de filtres WFS et un accroissement de la taille de la base de données utilisée pour le rendu temps réel.

¹Les conventions prises pour le site θ et l'azimut ϕ sont illustrées à la figure 2.1.

²(Kinsler *et al.*, 2000) suggèrent une représentation similaire pour les champs lointains de sources sonores, rappelé en équation A.56, annexe A.3.3.

3.2. Méthode modulaire pour la synthèse de directivité

Pour éviter ce problème, nous proposons une architecture plus modulaire pour la synthèse de directivité, dans laquelle la directivité de rayonnement de l'instrument virtuel est recomposée à partir d'une somme pondérée des fonctions de directivité de base (Caulkins *et al.*, 2003; Warusfel *et al.*, 2004b). Ces fonctions de directivité Y^m , que nous nommerons "harmoniques circulaires", sont issues d'une restriction planaire des harmoniques cylindriques réelles³, et s'expriment de la manière suivante :

$$Y^m(\phi) = \begin{cases} \cos(m\phi) & \text{pour } m \geq 0 \\ \sin(m\phi) & \text{pour } m < 0 \end{cases} \quad (3.2)$$

Les harmoniques circulaires possèdent des périodicités spatiales de complexité croissante à

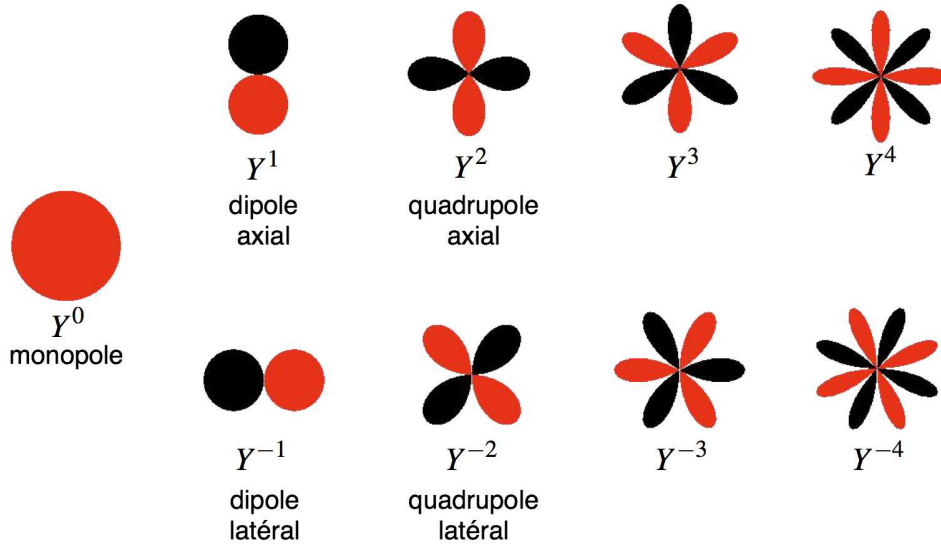


FIG. 3.3 – Harmoniques circulaires réelles d'ordre 0,1,2,3 et 4.

mesure que l'on monte dans les ordres (cf. Fig. 3.2), ce qui confère une caractère intuitif à leur manipulation. Elles permettent à priori de recomposer n'importe quelle figure de directivité dans le plan horizontal, à condition de monter suffisamment haut dans les ordres.

En adoptant ce paradigme pour la représentation de la directivité d'une source, la fonction de directivité Γ de l'équation 3.1 devient :

$$\Gamma(\theta = \frac{\pi}{2}, \phi, \omega) = \Gamma(\phi, \omega) = \sum_{m=-M}^M a_m(\omega) Y^m(\phi) \quad (3.3)$$

où a_m est un nombre réel représentant la pondération de l'harmonique d'ordre m et M représente l'ordre maximal de décomposition. Les coefficients a_m correspondent au spectre de

³On montre en annexe B que l'on peut également dériver les harmoniques circulaires à partir des harmoniques sphériques en fixant $\theta = \frac{\pi}{2}$ dans les harmoniques sectorielles réelles.

Fourier circulaire de la fonction de directivité Γ .

Ce paradigme offre une grande flexibilité pour la manipulation de la directivité. En effet, il existe des relations analytiques simples permettant de faire varier l'orientation ou l'ouverture angulaire des figures de directivité synthétisées, à partir de combinaisons des harmoniques de base. Ces relations sont explicitées dans la section suivante.

3.2.1 Manipulation des figures de directivité

Rotation d'une fonction de directivité

La rotation d'une fonction de directivité élémentaire Y^m se fait très simplement, en vertu des propriétés de base des fonctions trigonométriques. En effet, on peut écrire que :

$$\begin{aligned}\cos(m(\phi - \phi')) &= \cos m\phi \cos m\phi' + \sin m\phi \sin m\phi' \\ \sin(m(\phi - \phi')) &= \sin m\phi \cos m\phi' - \cos m\phi \sin m\phi',\end{aligned}$$

ce qui implique les relations suivantes pour la rotation d'une harmonique élémentaire :

$$Y^m(\phi - \phi') = \begin{cases} Y^m(\phi)Y^m(\phi') + Y^{-m}(\phi)Y^{-m}(\phi') & \text{pour } m > 0 \\ Y^m(\phi) & \text{pour } m = 0 \\ Y^m(\phi)Y^{-m}(\phi') - Y^{-m}(\phi)Y^m(\phi') & \text{pour } m < 0 \end{cases} \quad (3.4)$$

La rotation d'une figure de directivité $\Gamma(\phi, \omega)$ ⁴ d'un angle ϕ' quelconque se fait en effectuant la rotation de chacune de ses composantes élémentaires. En manipulant l'expression des harmoniques $Y^m(\phi - \phi')$ donnée en 3.4, on obtient la formule de recombinaison suivante :

$$\Gamma(\phi - \phi', \omega) = \sum_{m=-M}^M a_m(\omega)Y^m(\phi - \phi') = \sum_{m=-M}^M a'_m(\omega)Y^m(\phi) \quad (3.5)$$

avec

$$a'_m(\omega) = \begin{cases} a_m(\omega)Y^m(\phi') - a_{-m}(\omega)Y^{-m}(\phi') & \text{pour } m > 0 \\ a_m(\omega) & \text{pour } m = 0 \\ a_m(\omega)Y^{-m}(\phi') + a_{-m}(\omega)Y^m(\phi') & \text{pour } m < 0 \end{cases} \quad (3.6)$$

Le Dirac directif

Outre le fait de permettre la reproduction de figures de directivité existantes (comme celle de la voix ou d'un instrument de musique), la combinaison des composantes de la base des harmoniques circulaires peut donner lieu à la synthèse de figures artificielles, destinées à faire

⁴décomposée en harmoniques circulaires selon l'équation 3.3

varier l'excitation de la salle par le système d'une manière contrôlée. Le Dirac directif est un exemple d'une telle figure, du fait qu'il permet de concentrer l'énergie acoustique émise par le banc de haut-parleurs selon un faisceau de largeur variable, dépendant de l'ordre auquel il est décrit.

On définit le Dirac directif $\delta^{2\pi}$ sur un cercle unitaire de manière analogue au Dirac directif sur une sphère unitaire $\delta^{4\pi}$ défini dans (Laborie, 2000), c'est à dire :

$$\delta^{2\pi}(\phi) = \frac{1}{2\pi}\delta(\phi) \quad (3.7)$$

où $\delta(\phi)$ représente la distribution de Dirac unidimensionnelle (lobe infiniment étroit localisé en $\phi = 0$ et d'aire unitaire).

Cette définition de $\delta^{2\pi}$ permet d'assurer la propriété de longueur unitaire suivante :

$$\int_0^{2\pi} \delta^{2\pi}(\phi)d\phi = \int_0^{2\pi} \delta(\phi)d\phi = 1. \quad (3.8)$$

La distribution de Dirac directive possède un certain nombre de propriétés intéressantes du point de vue de la manipulation de la directivité, qui découlent des propriétés du Dirac unidimensionnel. Parmi ces propriétés, la propriété d'échantillonnage d'une fonction de directivité $f(\phi)$ nous intéresse plus particulièrement :

$$\delta^{2\pi}(\phi - \phi_0)f(\phi) = \delta^{2\pi}(\phi - \phi_0)f(\phi_0). \quad (3.9)$$

Grâce à cette propriété, le spectre de Fourier circulaire a_m^δ du Dirac directif se calcule aisément. En effet, la transformée de Fourier circulaire du Dirac directif s'énonce de la manière suivante :

$$a_m^\delta = \int_0^{2\pi} \delta^{2\pi}(\phi)Y^m(\phi)d\phi \quad (3.10)$$

En vertu de la propriété d'échantillonnage 3.9, on peut donc écrire immédiatement que :

$$a_m^\delta = Y^m(0) \int_0^{2\pi} \delta^{2\pi}(\phi)d\phi = Y^m(0) \quad (3.11)$$

En d'autres termes, le spectre de la distribution de Dirac directive n'est autre que l'ensemble des fonctions de base Y^m prises en $\phi = 0$.

Le spectre $a_m^{\delta'}$ d'une fonction de Dirac $\delta^{2\pi}(\phi - \phi')$ orientée selon un angle quelconque ϕ' se calcule alors aisément :

$$a_m^{\delta'} = \int_0^{2\pi} \delta^{2\pi}(\phi - \phi')Y^m(\phi)d\phi = Y^m(\phi - \phi'). \quad (3.12)$$

Le fait qu'un signal $s(t)$ convolué par un Dirac $\delta(t - \tau)$ retardé d'un temps τ produise le signal retardé $s(t - \tau)$ dans le cas unidimensionnel se retrouve également dans le cas bidimensionnel. Dans ce cas, on considère la fonction de directivité $\Gamma(\phi)$ définie en 3.3, que l'on convolue par un Dirac directif $\delta^{2\pi}(\phi - \phi')$ tourné (ou "retardé angulairement") d'un angle ϕ' . Etant donné que la convolution est transformée en produit simple dans l'espace de Fourier, le spectre $X_m(\phi)$ résultant de cette convolution s'écrit :

$$\begin{aligned} X_m(\phi, \omega) &= \sum_{m=-M}^M a_m(\omega) a'_m{}^\delta \\ &= \sum_{m=-M}^M a_m(\omega) Y^m(\phi - \phi') \\ &= \Gamma(\phi - \phi', \omega) \quad (\text{d'après 3.5}). \end{aligned}$$

On retrouve bien que la convolution d'une figure de directivité par un Dirac directif tourné d'un angle ϕ' résulte en la rotation de la figure de directivité d'un angle ϕ' .

Des distributions de Dirac directives discrètes d'ordre maximal l_{max} croissant sont représentées à la figure 3.4. Elles sont reproduites par sommation d'harmoniques Y^m sur un cercle échantillonné régulièrement en $N = 100$ points.

Une distribution de Dirac d'ordre l_{max} pourra être correctement synthétisée par le système WFS à une position donnée à condition que la distribution spatiale des haut-parleurs le permette. Une condition sur l'ordre maximal l_{max} recomposable par un banc WFS est énoncée à la section 3.5.2.

3.3 Formulation des filtres WFS pour une source virtuelle directive

Dans cette section, on propose de recalculer les filtres WFS en supposant que la source virtuelle possède une directivité de rayonnement quelconque⁵. On caractérisera ensuite l'effet des différentes approximations inhérentes à la technique WFS dans les sections suivantes, de manière à mieux cerner le cadre offert par un système WFS pour la synthèse d'instruments virtuels à directivité complexe en salle.

On se place en coordonnées sphériques et on considère une source virtuelle située en un point Ψ du plan horizontal contenant le banc de haut-parleurs (Fig. 3.5). En supposant des

⁵Au chapitre 1 on a fait l'hypothèse simplificatrice d'une source virtuelle omnidirective pour le calcul des filtres WFS.

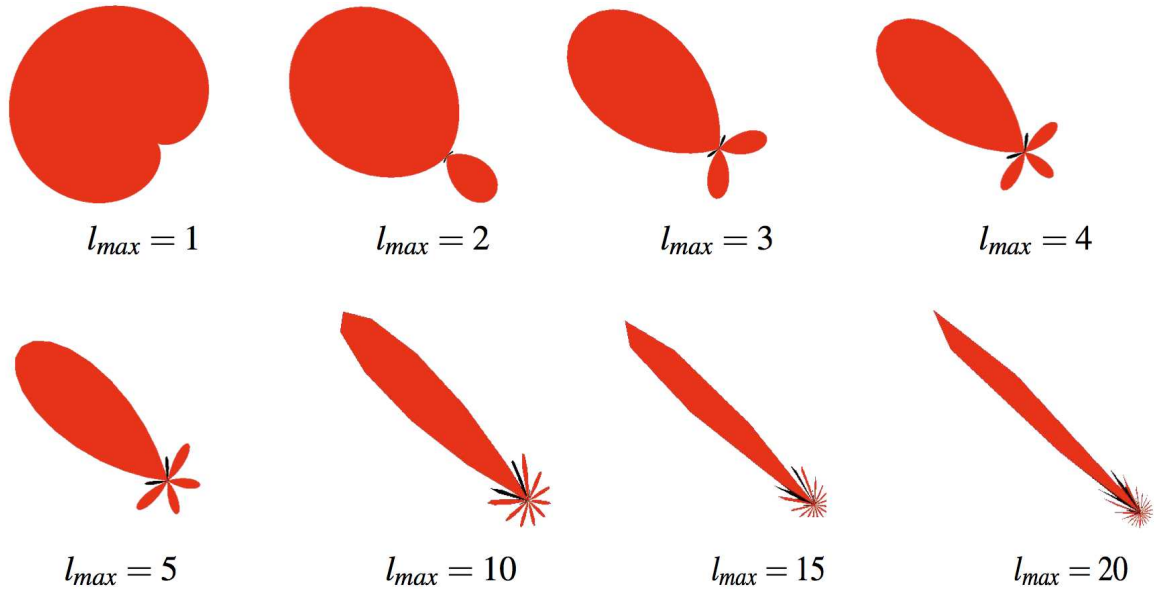


FIG. 3.4 – Distribution de dirac directive $\delta^{2\pi}(\psi)$ reproduite dans un plan d'échantillonnage de taille $N = 100$ à un ordre l_{max} croissant.

conditions de champ libre, on peut écrire⁶ que le champ de pression produit par cette source en un point $R(r, \theta, \phi)$ de la zone d'écoute s'exprime de la manière suivante :

$$p_{\Psi}(r, \theta, \phi, k) = \sum_{l=0}^{+\infty} \sum_{m=-l}^l \left(b_{lm}(k) h_l^{(1)}(kr) Y_l^m(\theta, \phi) \right). \quad (3.13)$$

où $h_l^{(1)}$ représente une fonction de Hankel sphérique divergente⁷ d'ordre l , Y_l^m représente l'harmonique sphérique de degré l et d'ordre m ⁸, et b_{lm} représente un coefficient de pondération réel.

En vertu de l'intégrale de Rayleigh I (eq. 1.5), on peut synthétiser un champ identique à celui produit par la source primaire Ψ dans la zone de reproduction Ω_R par le biais d'une distribution planaire $\partial\Omega$ de monopoles. Soit $P_C(\vec{r}_R, \omega)$ la contribution d'une ligne de sources verticale C appartenant au plan de Rayleigh $\partial\Omega_P$ (Fig. 3.5) ; en utilisant les équations 1.3, 1.8b et 3.13 on peut écrire que

$$p(\vec{r}_R, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P_C(\vec{r}_R, \omega) dx_C \quad \text{avec} \quad (3.14a)$$

$$P_C(\vec{r}_R, \omega) = \int_{-\infty}^{\infty} \left[\vec{\nabla}(h_l^{(1)}(kr) Y_l^m(\theta, \phi)) \cdot \vec{n} \frac{e^{-ik\Delta r}}{\Delta r} \right] dz_C \quad (3.14b)$$

⁶cf. section 2.1.2

⁷Les fonctions de Hankel sont décrites en sec. 2.1.1

⁸Les harmoniques sphériques sont décrites en Annexe B

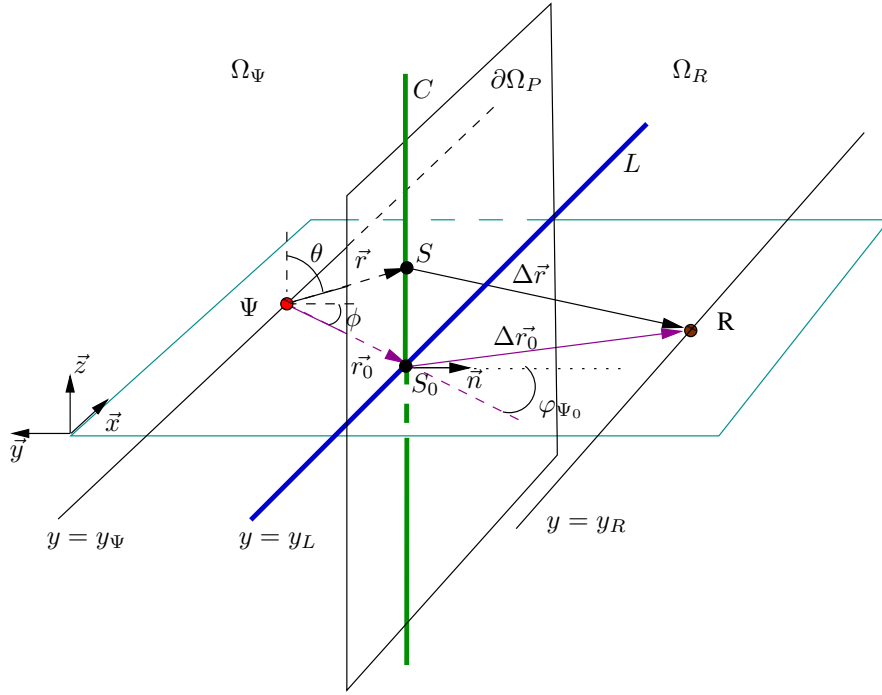


FIG. 3.5 – Géométrie pour la formulation des opérateurs de synthèse de sources directives

où x_C représente l'abscisse de la colonne C et z_C représente la hauteur d'un point de la colonne C . En introduisant l'expression de l'opérateur gradient en coordonnées sphériques, on a :

$$\begin{aligned} \vec{\nabla}(h_l^{(1)}(kr)Y_l^m(\theta, \phi)) \cdot \vec{n} &= \frac{\partial h_l^{(1)}(kr)}{\partial r} Y_l^m(\theta, \phi) \vec{e}_r \cdot \vec{n} \\ &+ \left(\frac{1}{r} \frac{\partial Y_l^m}{\partial \theta}(\theta, \phi) \vec{e}_\theta \cdot \vec{n} + \frac{1}{r \sin \theta} \frac{\partial Y_l^m}{\partial \phi}(\theta, \phi) \vec{e}_\phi \cdot \vec{n} \right) h_l^{(1)}(kr) \end{aligned}$$

Cette expression se simplifie du fait que l'on se restreint au plan horizontal ($\theta = \frac{\pi}{2}$) ; on a alors $\vec{e}_\phi \cdot \vec{n} = \cos \varphi_{\Psi_0}$ et $\vec{e}_\theta \cdot \vec{n} = 0$, ce qui permet d'annuler le terme de dérivée partielle par rapport à θ . D'autre part, étant donné l'expression de Y_l^m détaillée en Annexe , on peut écrire :

$$\frac{1}{r \sin \theta} \frac{\partial Y_l^m}{\partial \phi}(\theta, \phi) = \frac{1}{r} \times \begin{cases} \frac{-m}{\sqrt{\pi}} \tilde{P}_l^{|m|}(0) \sin(m\phi) & \text{pour } m > 0 \\ 0 & \text{pour } m = 0 \\ \frac{m}{\sqrt{\pi}} \tilde{P}_l^{|m|}(0) \cos(m\phi) & \text{pour } m < 0 \end{cases} \quad (3.15)$$

Ce terme peut être alors être négligé du fait de sa dépendance en $\frac{1}{r}$ si l'on suppose que la source est en champ lointain ($kr \gg 1$). Dans ces conditions, et en vertu de l'équation 2.4, on a également :

$$\frac{\partial h_l^{(1)}(kr)}{\partial r} \approx (-i)^l k \frac{e^{ikr}}{kr} \quad (3.16)$$

3.4. Effet de la troncature du banc de haut-parleurs

Par ailleurs, la restriction géométrique au plan horizontal implique que seul les harmoniques sphériques sectorielles Y_l^m telles que $l = |m|$ sont utiles d'un point de vue de la description du champ. En effet, bien que certaines harmoniques zonales et tesserales aient une influence dans le plan horizontal, leur influence demeure redondante avec celle harmoniques sectorielles. Ceci se voit en observant la figure B.3, où on peut observer par exemple que l'harmonique zonale Y_2^0 est en opposition de phase au monopole Y_0^0 dans le plan horizontal ; les harmoniques tesserales Y_3^{-1} et Y_3^1 sont également en opposition de phase avec les harmoniques sectorielles Y_1^{-1} et Y_1^1 . On retiendra donc uniquement les harmoniques sectorielles pour la description du champ rayonné dans le plan horizontal.

In fine, compte tenu de ces simplifications, l'application de l'approximation de phase stationnaire à l'équation 3.14b donne l'expression suivante pour la fonction de filtrage WFS associée à la synthèse d'une source de directivité élémentaire d'ordre m par la ligne continue de monopoles :

$$\mathbf{u}_{\Psi_m}(x_L, \omega) = \cos \varphi_{\Psi_0} \sqrt{\frac{k}{2\pi}} g_{\Psi}(y_{R_{\text{moy}}}) \frac{e^{-i[kr_0 - (\frac{T}{8})\omega]}}{\sqrt{r_0}} Y^m(\phi). \quad (3.17)$$

où $Y^m(\phi)$ représente l'harmonique circulaire d'ordre m décrite à l'équation 3.2. On choisit ici de retirer le terme $P_m^{|m|}(0)$ qui est constant à un ordre donné ; il faut cependant tenir compte de son influence si on souhaite transcoder une directivité de source décomposée selon les harmoniques sphériques pour la diffusion via la WFS.

Pour des ordres élevés (m grand) et des distances proches du banc, le terme négligé dans le calcul du filtre WFS (détaillé à équation 3.15) devient non-négligeable et l'expression 3.17 n'est plus valable. L'utilisation du processus d'inversion multicanale pour le calcul des filtres d'égalisation permet cependant de compenser son absence (Corteel, 2007) ; cependant, si on souhaite égaliser le champ suivant la technique d'égalisation individuelle (sec. 1.3.9) ou la technique d'égalisation individuelle avec compensation de l'erreur moyenne de synthèse (sec. 1.3.9), il faut réintroduire ce terme dans les filtres associés aux sources virtuelles proches du banc de haut-parleurs.

3.4 Effet de la troncature du banc de haut-parleurs

Le problème de la troncature du banc de haut-parleurs a été mis en évidence au chapitre 1. On a montré que cette troncature implique une réduction de la zone de visibilité de la source virtuelle ainsi que l'introduction d'artefacts de diffraction dans le champ effectivement reproduit. Si la source est directive, cela se traduit par une réduction du secteur angulaire de sa figure de directivité effectivement rayonnée par le banc de haut-parleurs dans la zone d'écoute (cf. Fig. 3.6). Bien qu'elle n'affecte pas la composante de son direct qui arrive aux oreilles de l'auditeur,

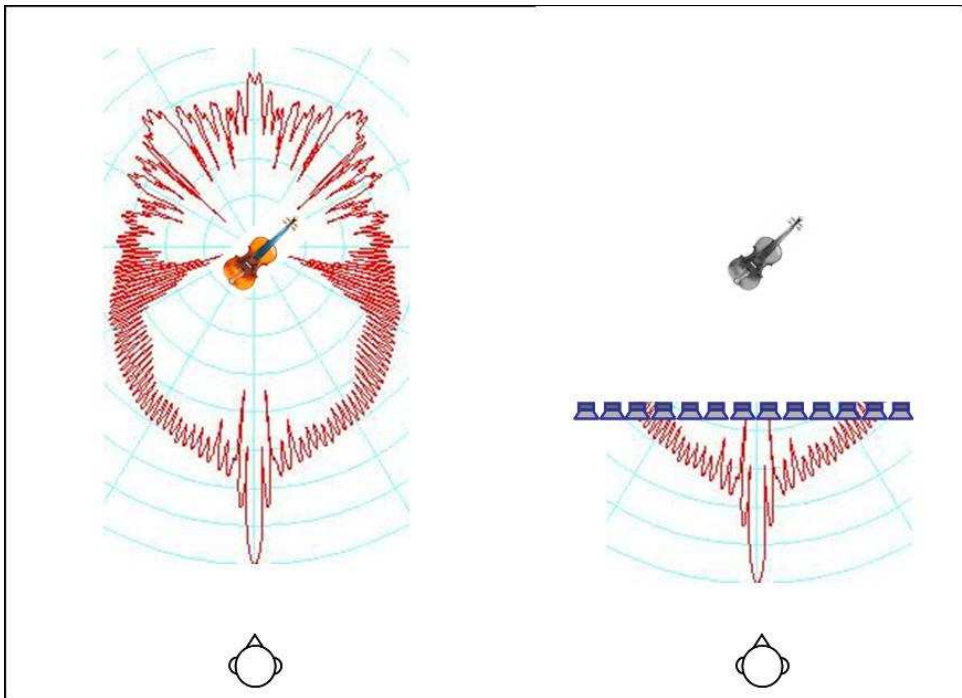


FIG. 3.6 – Mise en evidence du problème de fenêtrage pour la synthèse d’un diagramme de directivité de violon par un banc WFS.

cette réduction de visibilité aura des conséquences importantes sur l’interaction de la source virtuelle avec la salle, qui sont étudiées aux chapitres 4 et 5.

3.5 Echantillonnage spatial du banc WFS et ordre maximal de reproduction d’une figure de directivité

Nous avons vu à la partie 1.3.8 que l’échantillonnage spatial de la ligne de sources secondaires implique l’existence d’une fréquence de repliement au delà de laquelle les fronts d’ondes produits par WFS ne sont plus cohérents. De la même manière, l’échantillonnage de la ligne de sources secondaires implique une limite supérieure pour l’ordre maximal des harmoniques circulaires pouvant être reproduites par le banc de haut-parleurs pour un arrangement donné. Dans cette section nous développons un critère géométrique permettant de déterminer la validité du champ reproduit par WFS pour un ordre donné, en fonction de la position de la source virtuelle et du schéma de répartition des haut-parleurs.

3.5.1 Théorème d'échantillonnage dans l'espace des fonctions de directivité

Le théorème d'échantillonnage connu en théorie unidimensionnelle s'applique de manière équivalente aux fonctions de directivité 3D, et a été employé au chapitre 2 pour dériver un ordre maximal de décomposition en fonction du nombre de points d'échantillonnage en θ et ϕ (sec. 2.1.3). Dans cette section nous détaillons sa formulation de manière à pouvoir établir simplement un critère sur l'ordre maximal de recombinaison d'une source virtuelle directive en WFS.

Le théorème d'échantillonnage indique que tout signal continu et périodique peut être représenté de manière unique par une suite de N échantillons à condition que la fréquence maximale contenue dans le signal soit inférieure ou égale à la moitié de la fréquence à laquelle le signal a été échantillonné (condition de Shannon-Nyquist). On considère un signal $s(t)$ continu et périodique de période T ; selon la théorie de Fourier ce signal peut se décomposer suivant une base d'exponentielles complexes de la manière suivante :

$$s(t) = \sum_{n=-\infty}^{+\infty} a(n)e^{i\omega_n t} \quad (3.18)$$

avec $\omega_n = \frac{2\pi n}{T}$. Le critère de Shannon-Nyquist pour ce signal échantillonné en N points se traduit par la relation suivante :

$$\omega_{|n|} \leq \frac{\omega_{ech}}{2} \Rightarrow \frac{2\pi|n|}{T} \leq \frac{\pi N}{T} \Rightarrow |n| \leq \frac{N}{2}. \quad (3.19)$$

En d'autres termes, le signal $s(t)$ peut être développé convenablement en série de Fourier jusqu'à l'ordre $n_{max} = \frac{N}{2}$. Si on essaie de décomposer le signal suivant des fonctions de base $e^{i\omega_{n'}t}$ à oscillations trop rapides (i.e. telles que $n' > n_{max}$), on subit un repliement du spectre : les fonctions de base $e^{i\omega_{n'}t}$ se retrouvent avec une représentation discrète $e^{i\omega_{n'}(\frac{m}{N})}$ identique à celle des fonctions de base basses fréquences telles que $n = n' \text{ modulo } (n_{max} + 1)$. Les fonctions de base hautes fréquence sont donc dupliquées en fonctions basse fréquence, faussant ainsi l'estimation des composantes BF.

Ce phénomène peut se produire également dans l'espace des fonctions de directivité. On a rappelé à l'équation 3.3 la formule de décomposition d'une fonction de directivité suivant la base des harmoniques circulaires, qui sont d'après 3.2 des fonctions 2π -périodiques. La décomposition de ces fonctions obéit donc au théorème d'échantillonnage. Etant donné un échantillonnage régulier de la fonction de directivité Γ sur N points, la condition de Shannon-Nyquist

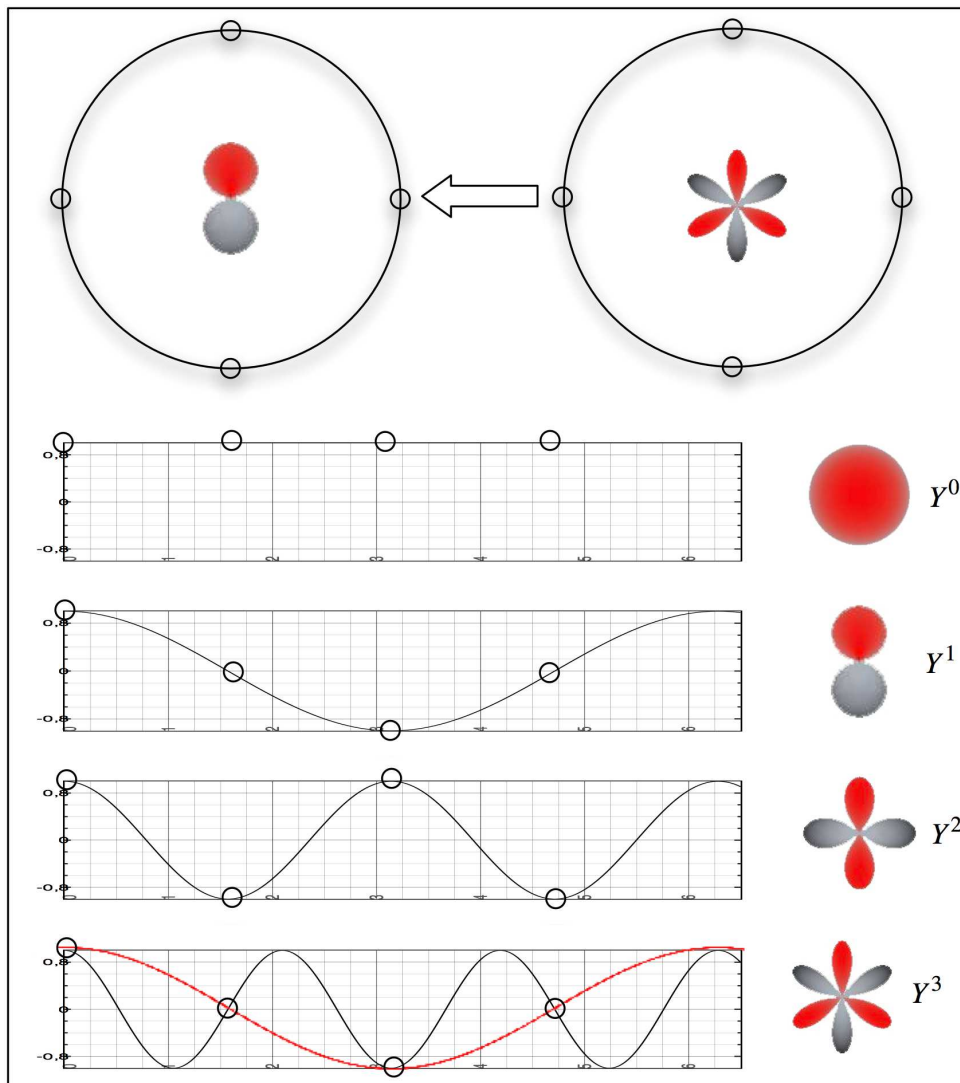


FIG. 3.7 – Mise en évidence du repliement de spectre spatial pour un échantillonnage régulier du cercle de mesure avec $N = 4$ points d'échantillonnage. Au delà de l'ordre $m = 2$, le sous-échantillonnage des harmoniques circulaires provoque un repliement de l'harmonique Y^3 sur l'harmonique Y^1 .

rappelée à l'équation 3.19 se traduit par :

$$M \leq \frac{N}{2} \quad (3.20)$$

avec M l'ordre maximal de décomposition admis pour la fonction de directivité Γ . Le fait d'introduire des harmoniques circulaires d'ordre $m' > M$ dans la décomposition fait apparaître un phénomène de repliement de spectre, illustré à la figure 3.7.

3.5.2 Ordre maximal de recomposition d'une fonction de directivité par WFS

La recomposition d'une fonction de directivité par le système WFS fait apparaître un problème d'échantillonnage et d'ordre maximal de reproduction que l'on peut rapprocher du cas de l'échantillonnage circulaire présenté à la section précédente 3.5.1 par le biais d'une manipulation géométrique simple. Pour cela, il suffit de voir que le front d'onde reproduit par la WFS forme, dans le plan horizontal, un cercle centré sur la source sonore. En prolongeant la ligne qui relie la source à chaque haut-parleur, on peut voir comment la directivité de la source est effectivement échantillonnée angulairement (Fig. 3.8).

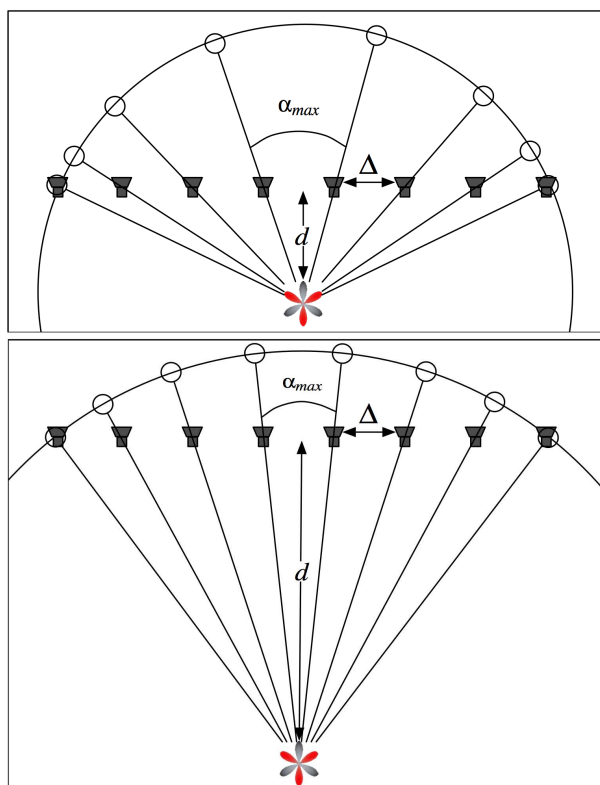


FIG. 3.8 – Mise en évidence du pouvoir de résolution offert par un banc WFS à haut-parleurs équirépartis (d'écartement Δ) pour la recomposition d'une fonction de directivité à une distance donnée. **Figure du haut** : La source virtuelle est synthétisée sur la médiatrice au banc à une distance d faible. **Figure du bas** : La source virtuelle est synthétisée sur la médiatrice au banc à une distance d élevée.

Effet de la distance sur l'ordre maximal de recomposition

L'observation de la figure 3.8 révèle que pour un espacement régulier des haut-parleurs et une source virtuelle centrée, la fonction de directivité de la source est en fait échantillonnée

irrégulièrement. Le pouvoir de résolution du banc est minimal au centre, ou l'angle α_{max} entre deux points d'échantillonnages successifs atteint son maximum, et constitue ainsi une limite inférieure pour la résolution angulaire. Par ailleurs, on peut voir que le pouvoir de résolution du banc de haut-parleurs vis-à-vis de la fonction de directivité augmente avec la distance au banc (Fig.3.8 haut et bas). En d'autres termes, l'ordre maximal pouvant être reproduit par le banc WFS augmente à mesure que la position de la source s'éloigne de celui-ci. Dans le cas où la source virtuelle est centrée (i.e. située sur la médiatrice au banc), on peut lier par une relation simple l'ordre maximal M pouvant être reproduit par le système à la distance d de la source au banc :

$$M = \frac{\pi}{\alpha_{max}} = \frac{\pi}{2 \cdot \arctan(\frac{\Delta}{2d})} \quad (3.21)$$

où Δ représente l'espacement entre les haut-parleurs. La figure 3.9 présente l'évolution de M en fonction de la distance au banc pour plusieurs espacements différents. On dispose donc d'un moyen simple de prédiction de l'ordre maximal de recombinaison offert par un banc WFS équi-réparti à différentes distances en fonction de l'espacement de haut-parleurs.

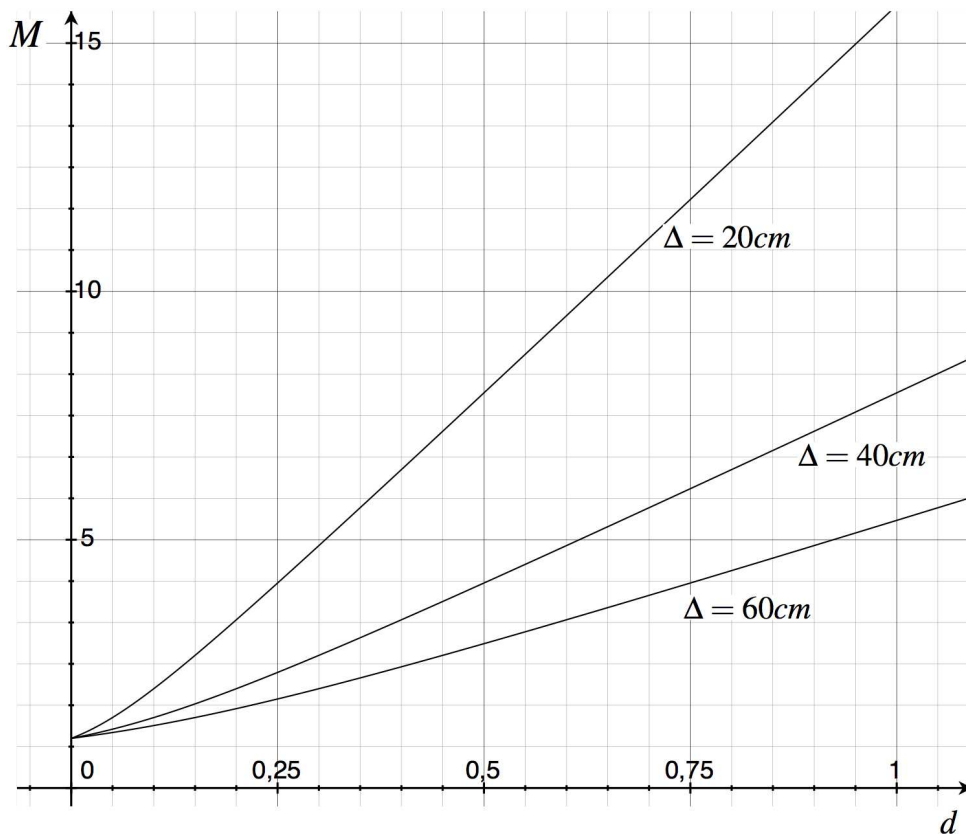


FIG. 3.9 – Ordre maximal de directivité M pouvant être reproduit par un banc WFS linéaire avec haut-parleurs espacés régulièrement étant donné une source virtuelle centrée (i.e. située sur la médiatrice au banc) en fonction de sa distance d au banc et l'espacement Δ entre les haut-parleurs.

Effet d'une répartition irrégulière de haut-parleurs sur l'ordre maximal de recombinaison

L'effet du choix d'une répartition irrégulière des haut-parleurs sur le pouvoir de résolution du banc WFS est mis en évidence à la figure 3.10. Cette figure montre que le choix d'une répartition irrégulière de haut-parleurs permet de gagner en pouvoir de résolution à une position donnée : en effet, on voit que l'angle maximal α_{max} entre deux points d'échantillonnage successifs est plus faible dans le cas d'une répartition irrégulière (Fig. 3.10, bas) que dans le cas d'une répartition régulière (Fig. 3.10, haut). Ceci traduit une augmentation de l'ordre maximal

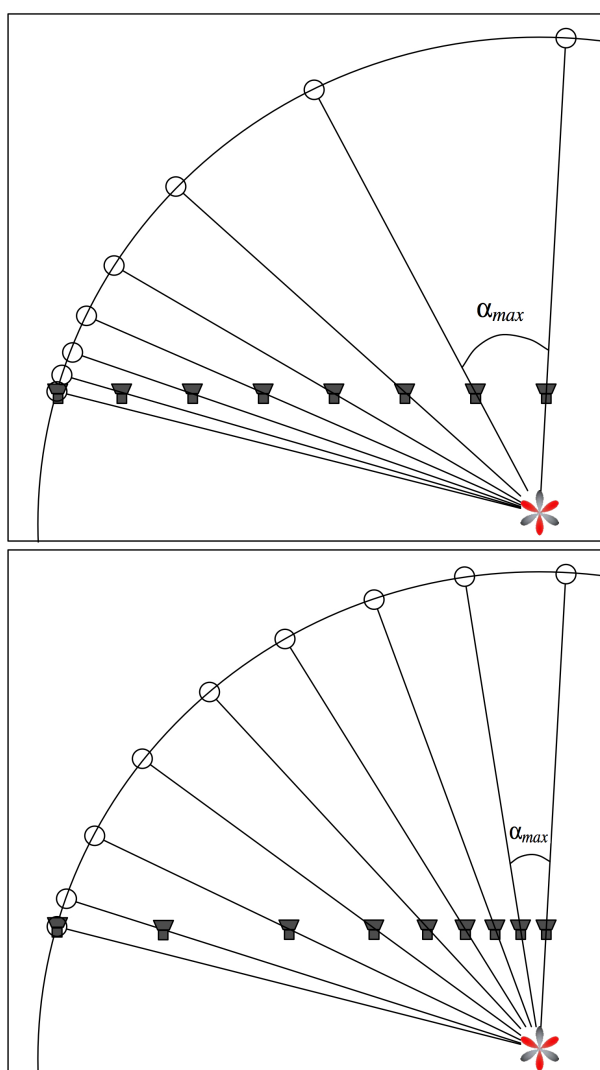


FIG. 3.10 – Mise en évidence du pouvoir de résolution offert par un banc WFS à haut-parleurs pour la recombinaison d'une fonction de directivité à une distance donnée. **Figure du haut** : Haut-parleurs équi-répartis. **Figure du bas** : Répartition irrégulière des haut-parleurs.

de directivité M que l'on peut synthétiser pour une source située à cette position. Cette augmentation s'accompagne cependant d'une diminution de M aux positions situées à l'extrémité

opposée du banc. Ce genre de disposition peut donc être intéressante dans des situations où l'on cherche à concentrer la précision de reproduction sur une zone donnée. Dans (Corteel, 2006b), l'auteur montre que les répartitions irrégulières de haut-parleurs permettent également de faire varier la fréquence de repliement spatial dans la zone d'écoute, et ainsi de définir des zones de reproduction privilégiées de ce point de vue.

Remarque sur la définition du critère d'ordre maximal de recomposition :

Le critère défini à l'équation 3.21 demeure un critère approximatif étant donné qu'il est proposé pour des échantillonnages irréguliers de fonctions de directivité. En raisonnant par analogie, on sait que si on se place dans le domaine temporel, l'échantillonnage irrégulier de la forme d'onde implique que la fréquence de Nyquist varie suivant la *nature à court terme* du signal échantillonné (Corteel, 2004). Dans le domaine spatial, l'échantillonnage irrégulier de la fonction de directivité implique que la fréquence de repliement spatial peut varier dans différents secteurs angulaires, suivant la *densité locale de points d'échantillonnage*. Il en découle que le critère donné en équation 3.21 fournit en réalité la valeur la plus basse de l'ordre maximum M reproductible par le banc linéaire. Il est probable que l'ordre maximal reproductible soit en réalité plus élevé dans certains secteurs angulaires entourant la source.

Précautions nécessaires pour la synthèse d'une directivité d'instrument réel.

Dans cette section on a montré que l'échantillonnage de la directivité effectuée par le banc de haut-parleurs implique l'existence d'un ordre maximal au delà duquel la recomposition de la source virtuelle ne plus être réalisée correctement. La recomposition d'une directivité instrumentale (eq. 3.3) requiert donc impérativement de limiter l'ordre de description de l'instrument au moment de sa caractérisation en champ libre.

C'est un cas analogue au problème de l'échantillonnage d'une forme d'onde temporelle. Dans le cas temporel, on résout le problème du repliement en appliquant un filtre analogique passe-bas anti-repliement en tête du convertisseur analogique/numérique dans la phase de décomposition. De ce fait, on est assuré qu'à la phase de recomposition le signal numérisé ne soit pas entaché d'artefacts de repliement.

Dans le cas de l'échantillonnage de la fonction de directivité, il faudrait ainsi théoriquement appliquer un filtrage "passe-bas spatial" avant la décomposition en harmoniques sphériques/cylindriques, de manière à s'assurer que le signal recomposé par la suite à l'ordre M ne subisse pas le repliement d'harmoniques d'ordre supérieur. Dans la méthode de décomposition de rayonnement en harmoniques sphériques proposée au chapitre 2, on a toutefois montré que ce filtrage passe-bas spatial se faisait "naturellement" à cause de l'atténuation des ordres élevés de la décomposition (section 2.1.2). En effet, pour une source d'étendue r_{min} ⁹, on a montré que

⁹ r_{min} étant le rayon de la plus petite sphère entourant entièrement la source

3.6. Evaluation de l'erreur de reconstruction du son direct associé à une source directive dans le plan horizontal

l'ordre maximal ayant une influence au niveau de la sphère de mesure est $l_{max} = kr_{min}$. En d'autres termes, à condition que l'échantillonnage de la sphère de mesure de la source soit choisi de manière suffisamment dense pour pouvoir correctement décrire l'ordre l_{max} à la fréquence $\omega = kc$, on est assuré de ne pas subir de repliement des ordres supérieurs dans la décomposition à l'ordre l_{max} .

Si on souhaite recomposer le rayonnement exact de la source mesurée en utilisant le système WFS il suffit alors de s'assurer que l'ordre maximal M que l'on est capable d'assurer d'un point de vue recombinaison à la position cible respecte la condition $M \leq l_{max}$ aux fréquences d'intérêt.

3.6 Evaluation de l'erreur de reconstruction du son direct associé à une source directive dans le plan horizontal

Dans cette section on propose d'évaluer la qualité du son direct associé à une source directive synthétisée dans le plan horizontal. Cette évaluation est conduite en simulant un système WFS composé de 48 haut-parleurs monopolaires idéaux (16.5cm d'espacement) mesuré sur un banc de 96 microphones omnidirectionnels (10cm d'espacement) placé 2m devant les haut-parleurs (Fig. 3.11).

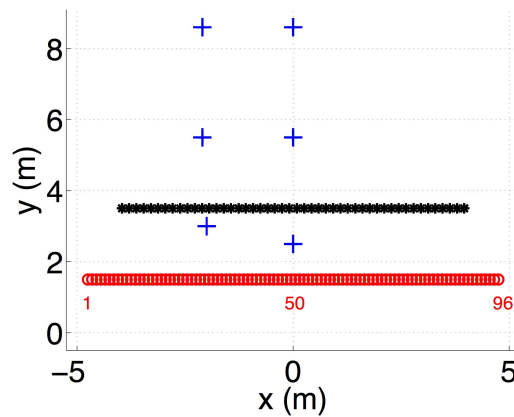


FIG. 3.11 – Configuration de simulation composée de 48 haut-parleurs monopolaires (astérisques noires '*') et 96 microphones omnidirectionnels (ronds rouges 'o'). Les microphones sont numérotés de 1 à 96 en partant de la gauche. Les positions des six sources virtuelles synthétisées sont représentées par des croix bleues '+'.

Le banc WFS est piloté par des filtres $h_{\Psi_m}^{Meq}$ préalablement calculés utilisant la méthode d'inversion multicanale (sec. 1.3.9) qui permettent de synthétiser des sources virtuelles Ψ_m situées aux 6 positions représentées par des croix bleues '+' sur la figure 3.11. A chacune de ces positions, on synthétise successivement 5 sources Ψ_m correspondant au 5 premières composantes

$\{Y^m\}_{m \in \{-2, -1, 0, 1, 2\}}$ de la base des harmoniques circulaires¹⁰, dont on mesure la réponse impulsionnelle x_{Ψ_m} sur l'ensemble des positions de microphone $\{\vec{r}^j\}_{j \in \{1, \dots, 96\}}$. La TFTD de cette réponse impulsionnelle est notée \mathbf{x}_{Ψ_m} .

Le champ sonore émis par le banc de haut-parleurs est comparé à une réponse impulsionnelle cible, dont la TFTD est définie de la manière suivante :

$$\mathbf{a}_{\Psi_m}(\vec{r}^j, \omega) = Att_{\Psi_m}(\vec{r}^j) \cdot Y^m(\phi(r_{\Psi_m}^{\vec{r}^j}, \vec{r}^j)) \cdot e^{-jk\|\vec{r}^j - r_{\Psi_m}^{\vec{r}^j}\|} \quad (3.22)$$

où $r_{\Psi_m}^{\vec{r}^j}$ représente la position de la source, $\phi(r_{\Psi_m}^{\vec{r}^j}, \vec{r}^j)$ représente l'angle entre la source et le microphone j , et Att_{Ψ_m} désigne l'atténuation qu'aurait la source virtuelle si elle était synthétisée par un banc linéaire infini (eq. 1.18). En d'autres termes, \mathbf{a}_{Ψ_m} correspond à la réponse d'un banc WFS idéal continu et de longueur infinie pour la synthèse de la source Ψ_m .

La comparaison entre la réponse cible \mathbf{a}_{Ψ_m} et la réponse effectivement synthétisé par le banc \mathbf{x}_{Ψ_m} se fait aux mêmes positions qui ont été utilisés pour conduire l'optimisation des filtres.

3.6.1 Evaluation de l'erreur complexe de reconstruction

Le premier critère utilisé pour évaluer la qualité de la synthèse du champ effectuée par le banc WFS est un critère d'erreur C_{Ψ_m} évalué dans le domaine complexe, qui place autant d'importance sur la reconstruction correcte de la phase que sur la reconstruction correcte de l'amplitude du champ synthétisé. On porte une attention particulière à la phase dans cette évaluation car le contrôle précis de la forme et de l'orientation des figures de directivité fait appel à des combinaisons linéaires des composantes élémentaires Y^m (cf. sec. 3.2.1). Le critère d'erreur C_{Ψ_m} utilisé est similaire au critère C_d de l'équation 2.29, et défini de la manière suivante :

$$C_{\Psi_m}(\vec{r}^j, \omega) = 10 \cdot \log_{10} \left(\frac{\langle \|\mathbf{d}_{\Psi_m}(\vec{r}^j, \omega)\| \rangle_{mics}}{\rho c \langle E_{\mathbf{a}}(\vec{r}^j, \omega) \rangle_{mics}} \right) \quad (3.23)$$

$$\text{avec } \mathbf{d}_{\Psi_m}(\vec{r}^j, \omega) = \mathbf{a}_{\Psi_m}(\vec{r}^j, \omega) - \mathbf{x}_{\Psi_m}(\vec{r}^j, \omega) \quad (3.24)$$

$$\text{et } E_{\mathbf{a}}(\vec{r}^j, \omega) = \frac{\mathbf{a}_{\Psi_m}(\vec{r}^j, \omega) \cdot \mathbf{a}_{\Psi_m}^*(\vec{r}^j, \omega)}{\rho c} \quad (3.25)$$

où $\langle \cdot \rangle_{mics}$ représente la moyenne sur toutes les positions de microphone, $\|\cdot\|$ représente la norme L2 d'un vecteur, et $E_{\mathbf{a}}(\vec{r}^j, \omega)$ représente l'énergie¹¹ du champ cible au niveau du microphone \vec{r}^j et à la pulsation ω . C_{Ψ} mesure donc la distance moyenne dans le plan complexe entre le cible et le champ effectivement synthétisé sur l'ensemble des positions de microphone,

¹⁰i.e. le quadrupole latéral Y^{-2} , le dipole latéral Y^{-1} , le monopole Y^0 , le dipole axial Y^1 , et le quadrupole axial Y^2 .

¹¹L'énergie est prise comme étant proportionnelle à la pression quadratique aux positions des microphones, ce qui correspond à une approximation de champ lointain (cf. Annexe A.3.3).

3.6. Evaluation de l'erreur de reconstruction du son direct associé à une source directive dans le plan horizontal

normalisée par l'énergie du champ cible qui traverse la ligne de microphones.

On conduit l'évaluation sous la fréquence de repliement spatial et à l'intérieur de la fenêtre de visibilité, car une reconstruction viable de la phase n'est pas possible en dehors de ce domaine spatio-fréquentiel en vertu des limitations de la WFS qui ont été rappelées au chapitre 1. Notons que dans cette configuration les 4 sources situées derrière le banc de haut-parleur dans la figure 3.11 ont une fenêtre de visibilité qui englobe entièrement le banc de microphones. Les 2 sources focalisées n'illuminent quand à elles qu'une partie réduite du banc de microphones, comme le montre la figure 3.12. Les microphones "non visibles" ne seront donc pas pris en compte dans le calcul de C_Ψ . Par ailleurs, la fréquence de repliement est calculée comme étant égale à 1200Hz¹².

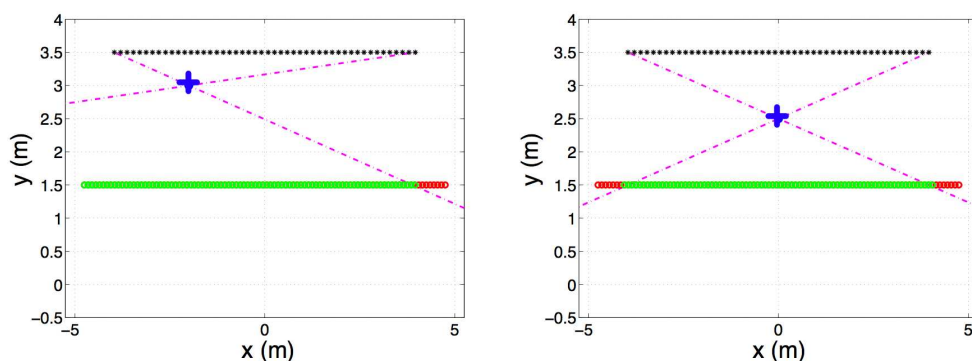


FIG. 3.12 – Problème du fenêtrage pour les sources focalisées. Les haut-parleurs sont représentés par des astérisques noirs. Les cercles verts représentent les microphones "visibles" et les cercles rouges représentent les microphones "non visibles" du point de vue de la source virtuelle représentée par une croix bleue. On peut observer que la source décentrée vers les x négatifs (figure de gauche) intercepte davantage de microphones que la source centrée (figure de droite), principalement du fait qu'elle est plus proche du banc de haut-parleurs.

Les résultats du calcul de C_Ψ sont donnés à la figure 3.13 pour les 3 sources virtuelles centrées situées le long de la médiatrice au banc ($x = 0$). L'observation de ces figures révèle que pour les 2 sources situées derrière le banc de haut-parleurs en $(0, 5.5)$ et $(0, 8.5)$, l'erreur C_Ψ demeure très faible sur l'ensemble des BF/MF (inférieure à -15dB en-dessous de 800Hz), puis augmente à mesure que l'on se rapproche de la fréquence de repliement. L'erreur de reconstruction devient plus élevée pour la source focalisée à 1m du banc en $(0, 2.5)$, en particulier pour les deux quadrupoles Y^{-2} et Y^2 qui enregistrent les erreurs les plus grandes en basses fréquences (jusque -8dB en-dessous de 800Hz). On remarque par contre que pour la source monopolaire, la courbe d'erreur demeure approximativement identique aux trois positions. En vertu du critère

¹²1200 Hz correspond à la fréquence de repliement minimale calculée sur l'ensemble des positions de sources et de microphones suivant l'équation 1.23. En réalité, cette fréquence excède 1200Hz pour certaines positions de sources et de microphones (cf. sec. 1.3.8)

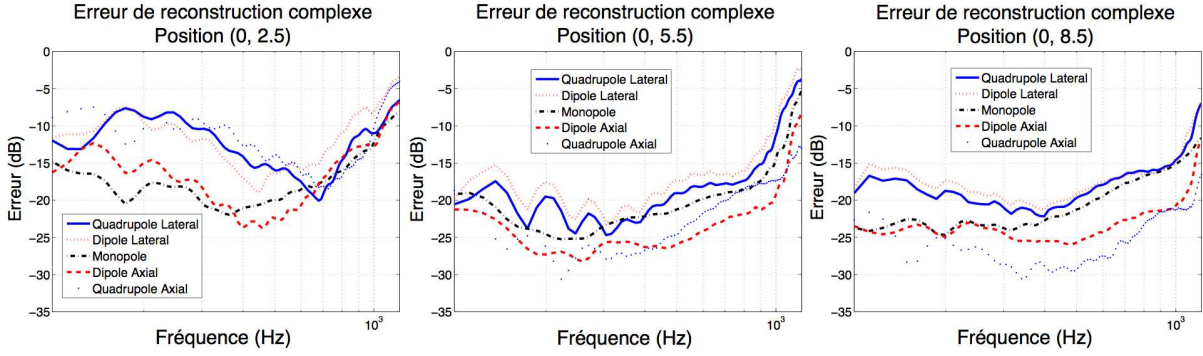


FIG. 3.13 – Evaluation de l’erreur de reconstruction complexe C_{Ψ} pour des sources directives synthétisées à trois profondeurs différentes le long de la médiatrice au banc de haut-parleurs. **Gauche** : Source à la position (0, 2.5). **Centre** : Source à la position (0, 5.5). **Droite** : Source à la position (0, 8.5)

développé en équation 3.21, le repliement de spectre spatial des harmoniques circulaires n’est a priori pas en cause, car étant donné l’écartement entre les haut-parleurs ($\Delta = 16.5\text{cm}$), à 1m devant le banc de haut-parleurs on peut reproduire une harmonique circulaire d’ordre inférieur ou égal à 19. Il est possible que la longueur des filtres utilisés (800 échantillons) soit insuffisante pour permettre la formation des lobes directifs à cette distance réduite. Cette conjecture resterait cependant à vérifier.

3.6.2 Evaluation de l’erreur de reconstruction énergétique

Au-delà de la fréquence de repliement, un contrôle précis de la phase de l’onde synthétisée devient impossible, et le critère complexe C_{Ψ} développé à la section précédente n’est plus adapté à l’analyse du champ sonore synthétisé par le banc de haut-parleurs. Un critère basé cette fois sur l’énergie qui traverse le banc de microphones est donc introduit de manière à compléter l’évaluation du son direct associé aux sources directives.

Soit $E_{\mathbf{x}}(\vec{r}^j, \omega)$ l’énergie effectivement émise par le banc de haut-parleur à la position \vec{r}^j et à la pulsation ω lors de la reconstruction d’une source virtuelle. Pour pouvoir illustrer les différences entre le champ cible et le champ effectivement reconstruit, on calcule la quantité $\frac{E_{\mathbf{x}}(\vec{r}^j, \omega)}{E_{\mathbf{a}}(\vec{r}^j, \omega)}$ à toutes les positions de microphone pour une source virtuelle dipolaire latérale Y^{-1} située 5m derrière le banc de haut-parleurs en (0, 8.5) (Fig. 3.14). Cette source présente un zéro au centre de la ligne de microphones, et en observant les trois graphiques spatio-fréquentiels de la figure 3.14, on constate que l’erreur principale commise par le banc dans la synthèse de la source concerne principalement les microphones centraux en hautes fréquences. En effet, d’un point de vue physique il est impossible de créer une zone d’annulation des ondes par interférences au-delà de la fréquence de repliement. On observe également des artefacts de

3.6. Evaluation de l'erreur de reconstruction du son direct associé à une source directive dans le plan horizontal

forme ondulée en BF dus aux ondes de diffraction émises par les bords du banc de haut-parleurs.

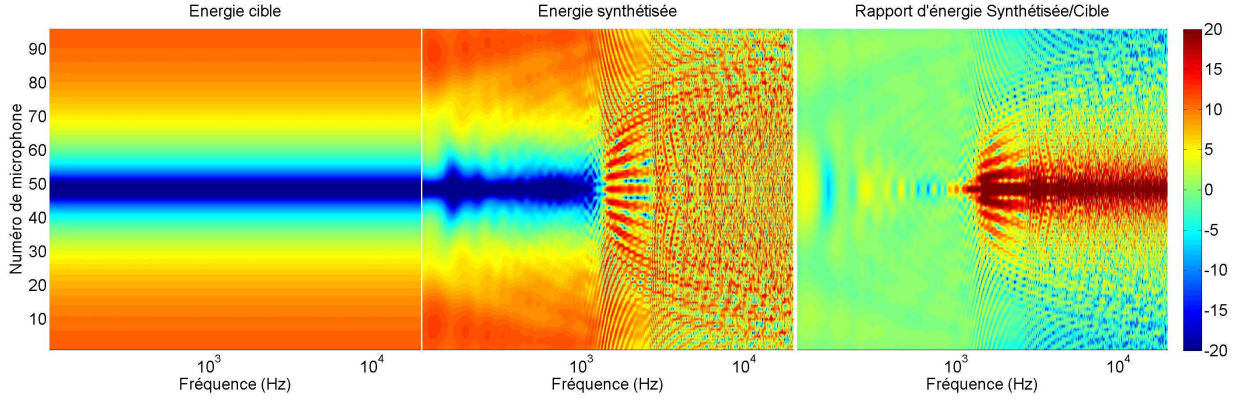


FIG. 3.14 – Graphiques spatio-fréquentiels de la synthèse d'un dipôle latéral (Y^{-1}) situé en $(0,8.5)$. Les microphones sont numérotés de 1 à 96, comme l'indique la figure 3.11. **Gauche :** Energie cible E_a (dB). **Centre :** Energie du champ synthétisé par le système WFS E_a (dB). **Droite :** Rapport $\frac{E_x(\vec{r}^j, \omega)}{E_a(\vec{r}^j, \omega)}$ (dB).

Au-dessus de la fréquence de repliement, les filtres sont calculés suivant la méthode d'égalisation individuelle avec compensation de l'erreur moyenne (eq. 1.28). Un critère global basé sur l'énergie moyennée sur l'ensemble des positions de microphones est donc adapté à l'évaluation de la qualité du champ synthétisé dans ce cas. On propose de définir l'erreur énergétique moyenne de reconstruction E_Ψ de la manière suivante :

$$E_\Psi = 10 \cdot \log_{10} \left(\frac{\langle E_x(\vec{r}^j, \omega) \rangle_{(mics, \omega)}}{\langle E_a(\vec{r}^j, \omega) \rangle_{(mics, \omega)}} \right). \quad (3.26)$$

Le calcul de E_Ψ est conduit sur la bande $[100-1200]$ Hz ainsi que sur la bande $[1200-20000]$ Hz de manière à pouvoir distinguer l'erreur commises en-dessous et au-dessus de la fréquence de repliement. La figure 3.15 donne les valeurs de E_Ψ en-dessous de la fréquence de repliement pour les trois positions de source centrées (graphique de droite) et décentrées (graphique de gauche) de la figure 3.11. La figure 3.16 donne les valeurs de E_Ψ au-dessus de la fréquence de repliement aux mêmes positions.

L'observation de la figure 3.15 montre que l'erreur énergétique globale E_Ψ demeure très proche de 0dB en-dessous de la fréquence de repliement, ce qui indique des bonnes performances pour la reconstruction dans cette zone fréquentielle¹³. Par contre, au-dessus de la fréquence de repliement (Fig. 3.16), on observe une dégradation des performances, et une fluctuation de l'erreur de reconstruction en fonction de la position et de l'orientation de la source.

¹³On notera que le critère E_Ψ est conçu de telle manière à ce que des valeurs proches de zéro indiquent un champ correctement reconstruit, alors que des grandes valeurs de $|E_\Psi|$ indiquent une mauvaise reconstruction du champ.

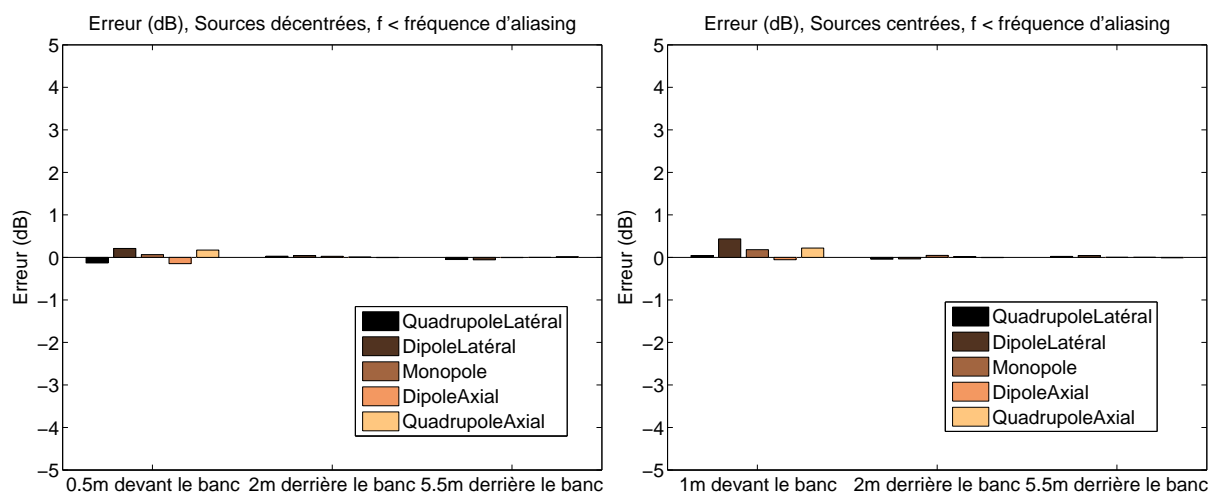


FIG. 3.15 – L’Erreur énergétique moyenne de reconstruction E_{Ψ} en-dessous de la fréquence de repliement ([100-1200]Hz) pour des sources virtuelles de directivité élémentaire. **Graphique de gauche** : Sources *décentrées* 2m à gauche de la médiatrice, i.e. en (-2, 3),(-2, 5.5),(-2, 8.5), c’est à dire resp. 0.5m devant le banc, 2m derrière le banc et 5.5m derrière le banc. **Graphique de droite** : Sources *centrées*, i.e. en (0, 2.5),(0, 5.5),(0, 8.5), c’est à dire resp. 1m devant le banc, 2m derrière le banc et 5.5m derrière le banc.

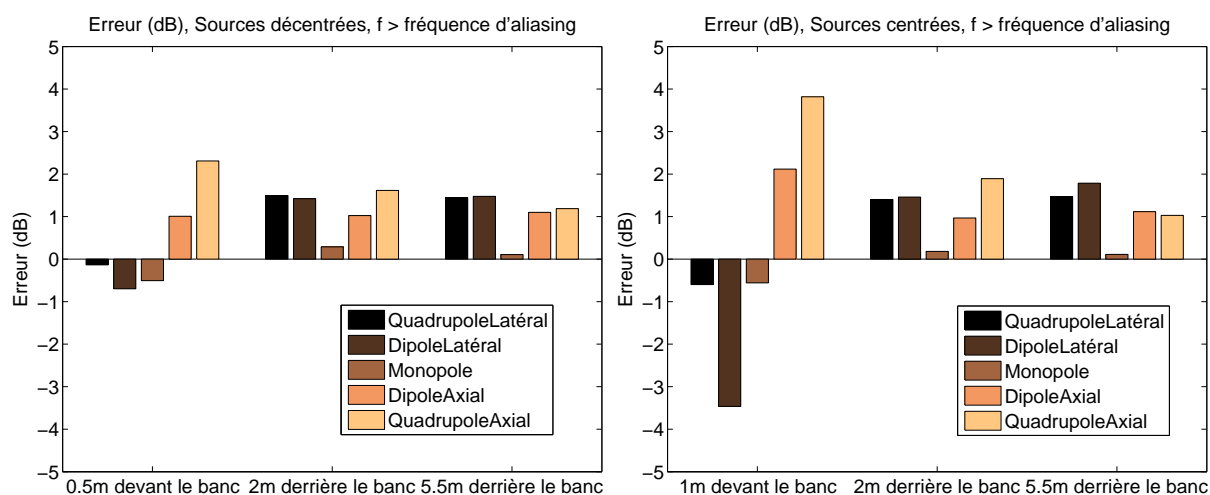


FIG. 3.16 – L’Erreur énergétique moyenne de reconstruction E_{Ψ} au-dessus de la fréquence de repliement ([1200-20000]Hz) pour des sources virtuelles de directivité élémentaire. **Graphique de gauche** : Sources *décentrées* 2m à gauche de la médiatrice, i.e. en (-2, 3),(-2, 5.5),(-2, 8.5), c’est à dire resp. 0.5m devant le banc, 2m derrière le banc et 5.5m derrière le banc. **Graphique de droite** : Sources *centrées*, i.e. en (0, 2.5),(0, 5.5),(0, 8.5), c’est à dire resp. 1m devant le banc, 2m derrière le banc et 5.5m derrière le banc.

Une source demeurant à la même position, mais dont l’orientation varie, peut exhiber des fluctuations d’erreur importantes. C’est le cas, par exemple, pour le quadrupole, dont l’erreur à 1m devant le banc devient beaucoup plus élevé lorsque celui-ci adopte une orientation axiale Y^2

3.6. Evaluation de l'erreur de reconstruction du son direct associé à une source directive dans le plan horizontal

($2dB < |E_{\Psi}| < 4dB$) que lorsqu'il adopte une orientation latérale Y^{-2} ($|E_{\Psi}| < 0.8dB$). En regardant la figure 3.17, on se rend compte que l'accroissement d'erreur provient du fait que le quadrupole axial présenté deux annulations le long de la ligne de microphones, au lieu d'une dans le cas du quadrupole latéral. Ces annulations étant impossibles à synthétiser au-dessus de la fréquence de repliement, elles génèrent des erreurs importantes au niveau des positions évaluées. De la même façon, l'évolution de l'erreur associée au quadrupole axial Y^2 varie en

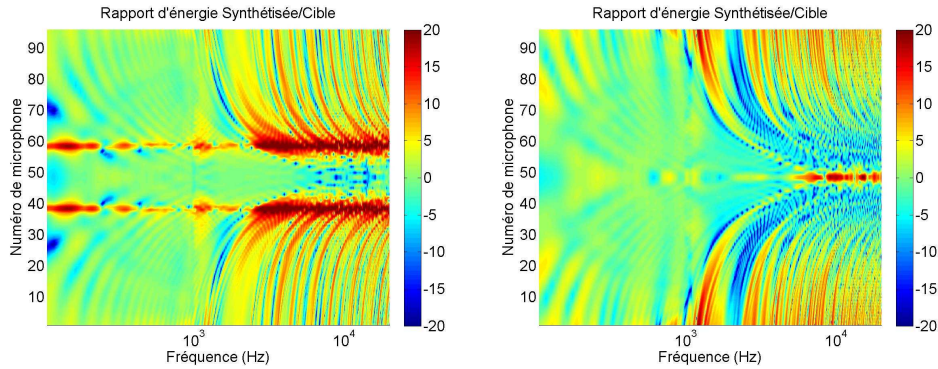


FIG. 3.17 – Rapport $\frac{E_x(r^j, \omega)}{E_a(r^j, \omega)}$ en dB pour un quadrupole synthétisé à une position centrée, 1m devant les haut-parleurs [en (0, 2.5)] pour deux orientation différentes : orientation axiale Y^2 (figure de gauche) et orientation latérale Y^{-2} (figure de droite).

fonction de la position de la source. On voit sur le graphique de droite de la figure 3.16 que l'erreur E_{Ψ} vaut 4dB à 1m devant le banc, puis 2dB à 2m derrière le banc, puis 1dB à 5m derrière le banc. L'observation de la figure 3.18 révèle effectivement que les zones d'annulations présentes le long de la ligne de microphones lorsque le quadrupole axial est situé 2m derrière le banc de haut-parleurs deviennent "invisibles" du point de vue de la ligne de microphones lorsque la source recule à 5m derrière le banc de haut-parleurs. Ceci est dû au fait qu'à partir d'une certaine distance, seule une portion très faible de la figure de directivité est visible au travers de la fenêtre que constitue le banc de haut-parleurs. La disparition progressive des zones d'annulation à mesure que la source virtuelle recule explique la chute de l'erreur mesurée.

3.6.3 Conclusions

Les résultats des simulations conduites dans cette section montrent que la qualité de reconstruction du champ est généralement bonne en dessous de la fréquence de repliement, à la fois en phase et en amplitude, ce qui laisse présager des bons résultats pour la synthèse de directivités instrumentales par combinaison des harmoniques de base¹⁴.

¹⁴Dans (Corteel, 2007) l'auteur montre en effet que l'on obtient des résultats sensiblement identiques lorsqu'on synthétise une directivité cardioïde directement et lorsqu'on la recompose par combinaison d'un filtre monopolaire et d'un filtre dipolaire.

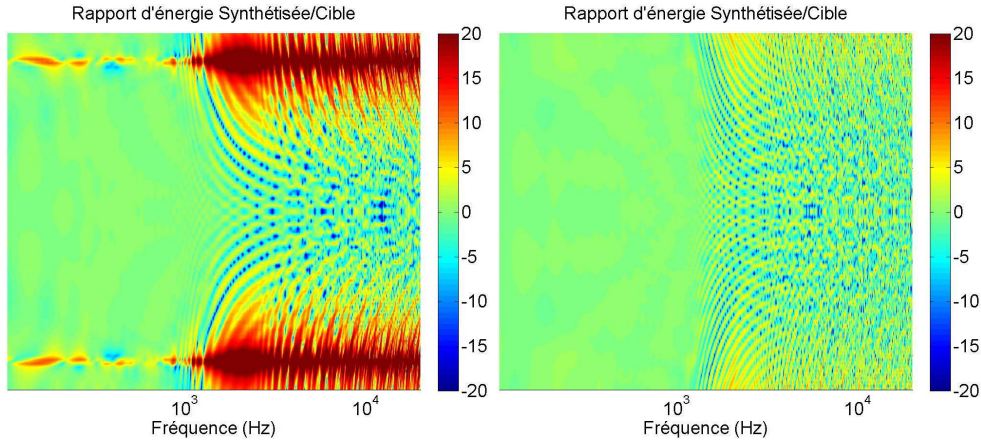


FIG. 3.18 – Rapport $\frac{E_{\mathbf{x}}(\vec{r}^j, \omega)}{E_{\mathbf{a}}(\vec{r}^j, \omega)}$ en dB pour un quadrupole axial synthétisé à une position centrée, pour deux profondeurs différentes : 2m derrière le banc de haut-parleurs en (0, 5.5) (figure de gauche) et 5m derrière le banc de haut-parleurs en (0, 8.5) (figure de droite).

Au-dessus de la fréquence de repliement, la qualité de reconstruction du champ devient fortement dépendante à la fois de la position et de la directivité de la source virtuelle, à cause du problème de fenêtrage de la directivité par le banc de haut-parleurs. Pour des positions/orientations de la source virtuelle faisant apparaître des annulations du champ dans la zone frontale contenant les microphones, l’erreur de reconstruction moyenne est plus forte que lorsque cette zone contient un lobe. La correction en énergie appliquée en hautes fréquences permet en effet d’assurer un niveau globalement cohérent avec le champ cible lorsque celui ci ne contient pas d’annulations. Par contre, dès que le champ cible contient des annulations, on est confronté à l’impossibilité de former des motifs interférentiels au-dessus de la fréquence de repliement.

3.7 Analyse du champ tridimensionnel effectivement rayonné par le banc WFS.

Les parties précédentes de ce chapitre ont démontré que le paradigme de synthèse modulaire de directivité par sommation d’harmoniques circulaires était bien adapté à la reproduction de diagrammes de directivité bidimensionnels dans le plan horizontal. On est donc à priori capable de reproduire le son direct d’une source instrumentale dans le plan horizontal (à l’atténuation près, cf. équation 1.18). Cependant, en pratique, la source instrumentale que l’on cherche à reproduire possède une directivité sphérique - elle rayonne à priori de manière hétérogène dans toutes les directions de l’espace. La convolution de cette directivité par la directivité du banc de haut-parleurs linéaire implique donc une perte importante d’informations sur les caractéristiques de la source, du fait de la “fenêtre de visibilité” restreinte qu’offre le banc WFS sur la

zone de reproduction. Au-delà de ce filtrage spatial, on peut voir que le banc rayonne lui aussi dans les trois dimensions, avec des propriétés de rayonnement qui lui sont propres. En effet, son champ rayonné possède une symétrie de révolution autour de l'axe du banc ¹⁵, ce qui lui confère des caractéristiques bien distinctes de celles de la source ponctuelle.

Dans cette section, on s'intéresse au champ tridimensionnel rayonné par le banc de haut-parleurs, en essayant de mettre en évidence comment la symétrie cylindrique du banc linéaire se répercute sur le champ associé à la source virtuelle WFS.

3.7.1 Analyse du champ en-dessous de la fréquence de repliement par décomposition sur la base des harmoniques sphériques

On propose dans un premier temps de décomposer le champ rayonné par un banc WFS constitué de haut-parleurs monopolaires sur la base des harmoniques sphériques. L'idée est de caractériser le spectre sphérique effectif d'une harmonique circulaire reproduite par WFS, de manière à mieux comprendre les modes tridimensionnels **réellement sollicités** par le banc lors de la synthèse d'une telle figure. On se place en-dessous de la fréquence de repliement du banc WFS pour cette étude.

Le banc de haut-parleurs simulé est composé de 24 monopoles espacés de 15cm. La décomposition de son champ suivant les harmoniques sphériques impose de choisir un échantillonnage suffisamment dense pour éviter le repliement des ordres supérieurs. En vertu de la section 2.1.2, l'ordre de décomposition minimal permettant de décrire le champ sortant rayonné par le banc de haut-parleurs est $l_{max} = kr_{min}$, avec r_{min} l'envergure de la source. Dans ce cas, le rayon du cercle le plus petit contenant le banc de haut-parleurs est $r_{min} = 1.725m$. L'échantillonnage sphérique régulier minimum (sec. 2.1.3) permettant d'éviter un repliement de spectre dans la description doit être choisi tel que :

$$N_{\theta} = l_{max} + 1 = kr_{min} + 1 \quad (3.27)$$

$$N_{\phi} = 2l_{max} + 1 = 2kr_{min} + 1 \quad (3.28)$$

avec N_{θ} le nombre minimum de points d'échantillonnage en site et N_{ϕ} le nombre minimum de points d'échantillonnage en azimuth.

On propose de mesurer le rayonnement du banc sur une sphère de microphones virtuels de rayon 2m centré sur le banc de haut-parleurs. Le nombre de points d'échantillonnage minimum est proportionnel à la fréquence maximale que l'on souhaite décrire, ce qui implique des tailles de structures de données à manipuler rapidement prohibitives d'un point de vue calcul. On prend ici une fréquence maximale de description égale à 1kHz ($l_{max} = 24$), ce qui implique de

¹⁵cf. chapitre 1 sec. 1.3.4

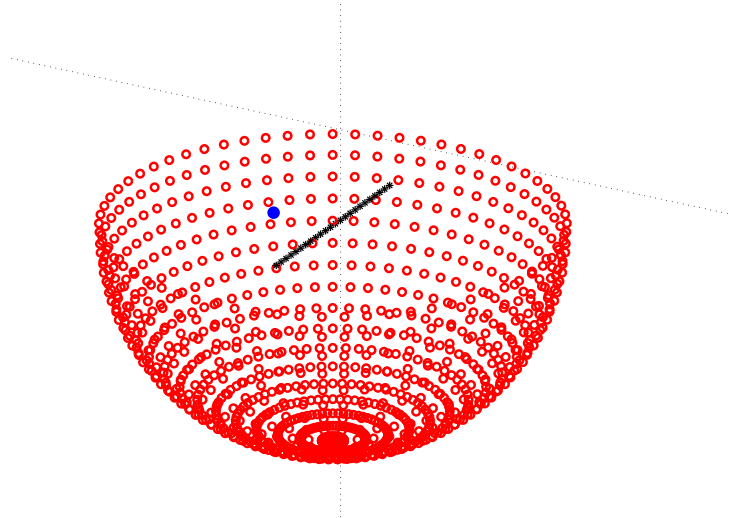


FIG. 3.19 – Configuration de mesure pour la décomposition du rayonnement du banc WFS suivant la base des harmoniques sphériques. Les 24 haut-parleurs sont représentés par des astérisques noires ; la source virtuelle est représentée par un point bleu. Les microphones de mesure sont représentés par des cercles rouges. Pour plus de visibilité, on représente seulement la demi-sphère inférieure de microphones.

prendre $N_\theta = 33$ et $N_\phi = 25$, soit 2145 microphones de mesures. On a représenté à la figure 3.19 une moitié de la sphère de mesure utilisée pour l’analyse.

La source virtuelle analysée est centrée et située à 1m derrière le banc de haut-parleurs (position représentée par un point bleu sur la figure 3.19). On fait varier la directivité de cette source suivant les harmoniques circulaires de base d’ordre $l \leq 2$. La décomposition du champ est faite jusque l’ordre 24 pour éviter tout repliement des ordres supérieurs. On notera cependant que l’observation à posteriori du spectre sphérique associé à chacune de ces figures de directivité de base révèle que l’énergie se concentre entièrement sur les ordres inférieurs ou égal à 12.

On a représenté le champ tridimensionnel rayonné et la décomposition du champ associé aux sources WFS de directivité élémentaires Y^{-2} , Y^{-1} , Y^0 , Y^1 , et Y^2 en Annexe C, figures C.1, C.2, C.3, C.4 et C.5, pour les fréquences 100Hz, 250Hz, 500Hz et 750Hz. Ces fréquences se situent toutes en-dessous de la fréquence de repliement spatial du banc de haut-parleurs étant donné la position de la source et la position des microphones.

Monopole Y^0

En observant la colonne de droite de la figure C.3, on voit que le champ associé à la source virtuelle Y^0 est principalement monopolaire en basses fréquences (100Hz, 250Hz). Cependant, à mesure que la fréquence augmente, l’énergie tend à se répartir sur d’autres modes. On observe d’ailleurs qu’à 500Hz, le mode associé à l’harmonique Y_2^2 prédomine par rapport au mode monopolaire. A 750Hz, l’énergie est répartie de manière plus homogène sur l’ensemble des modes associés aux harmoniques tesserales $\{Y_l^m \mid l, m \text{ pairs et } m > 0\}$ et

zonales $\{Y_l^0 \mid l \text{ pair}\}$, avec une domination des modes associés aux harmoniques sectorielles $\{Y_m^m \mid m \text{ pair et } m > 0\}$. L'observation des figures de directivité tridimensionnelles associées (Fig. C.3, colonne de gauche), montre que le champ rayonné tend à prendre un aspect cylindrique avec l'augmentation de la fréquence, ce qui explique l'apparition de modes non-monopolaires dans sa décomposition.

Dipole Y^{-1} et Quadrupole Y^{-2}

Nous proposons de dupliquer ici aux figures 3.21 et 3.22 les figures C.2 et C.1 qui se trouve déjà Annexe C, de manière à faciliter la lecture du commentaire sur les sources Dipole Y^{-1} et Quadrupole Y^{-2} dans cette section.

Les champs WFS associés aux harmoniques circulaires Y^{-1} (dipole latéral) et Y^{-2} (quadrupole latéral) ont des comportements tridimensionnels quasi similaires, ce qui se voit autant du point de vue des diagrammes de rayonnement (Fig. 3.22 et 3.21, colonne de gauche) que des coefficients modaux associés (Fig. 3.22 et 3.21, colonne de droite). Ces sources virtuelles tendent à exciter principalement le dipole Y_{-1}^1 en basses fréquences (100Hz, 250Hz). En fréquences moyennes (500Hz, 750Hz), l'énergie se répartit de manière à peu près homogène sur l'ensemble des modes associés aux harmoniques tesserales $\{Y_l^m \mid l, m \text{ impairs et } m < 0\}$.

La similarité observée entre les deux champs s'explique par une combinaison de fenêtrage de l'harmonique et de symétrie de rotation du banc (cf. Fig. 3.20). La figure 3.20 montre sché-

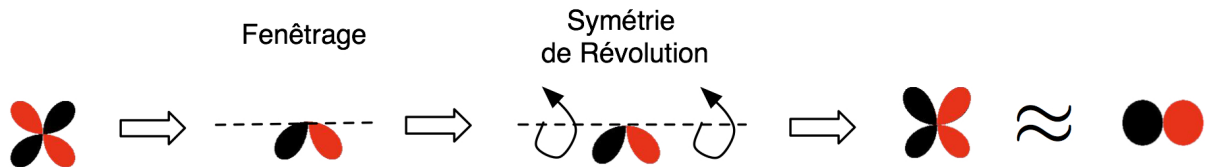


FIG. 3.20 – Mise en évidence de la redondance de l'harmonique Y^{-2} avec l'harmonique $-Y^{-1}$ provoquée par le fenêtrage et de la symétrie de révolution autour de l'axe du banc de haut-parleurs.

matiquement comment le la source virtuelle de directivité élémentaire Y^{-2} devient redondante avec la source Y^{-1} au cours de sa convolution par la directivité de rayonnement du banc de haut-parleurs.

Dipole Y^1 et quadrupole Y^2

Le champ WFS tridimensionnel associé à l'harmonique Y^1 (Fig. C.4) est quasi-identique au champ tridimensionnel associé à l'harmonique Y^0 (Fig. C.3), du fait de la convolution de son rayonnement par le banc linéaire. On peut remarquer, d'ailleurs, que l'harmonique sphérique sectorielle Y_1^1 , qui engendre en théorie l'harmonique Y^1 , est totalement absente du spectre

Source virtuelle Y^{-1} ●● centrée

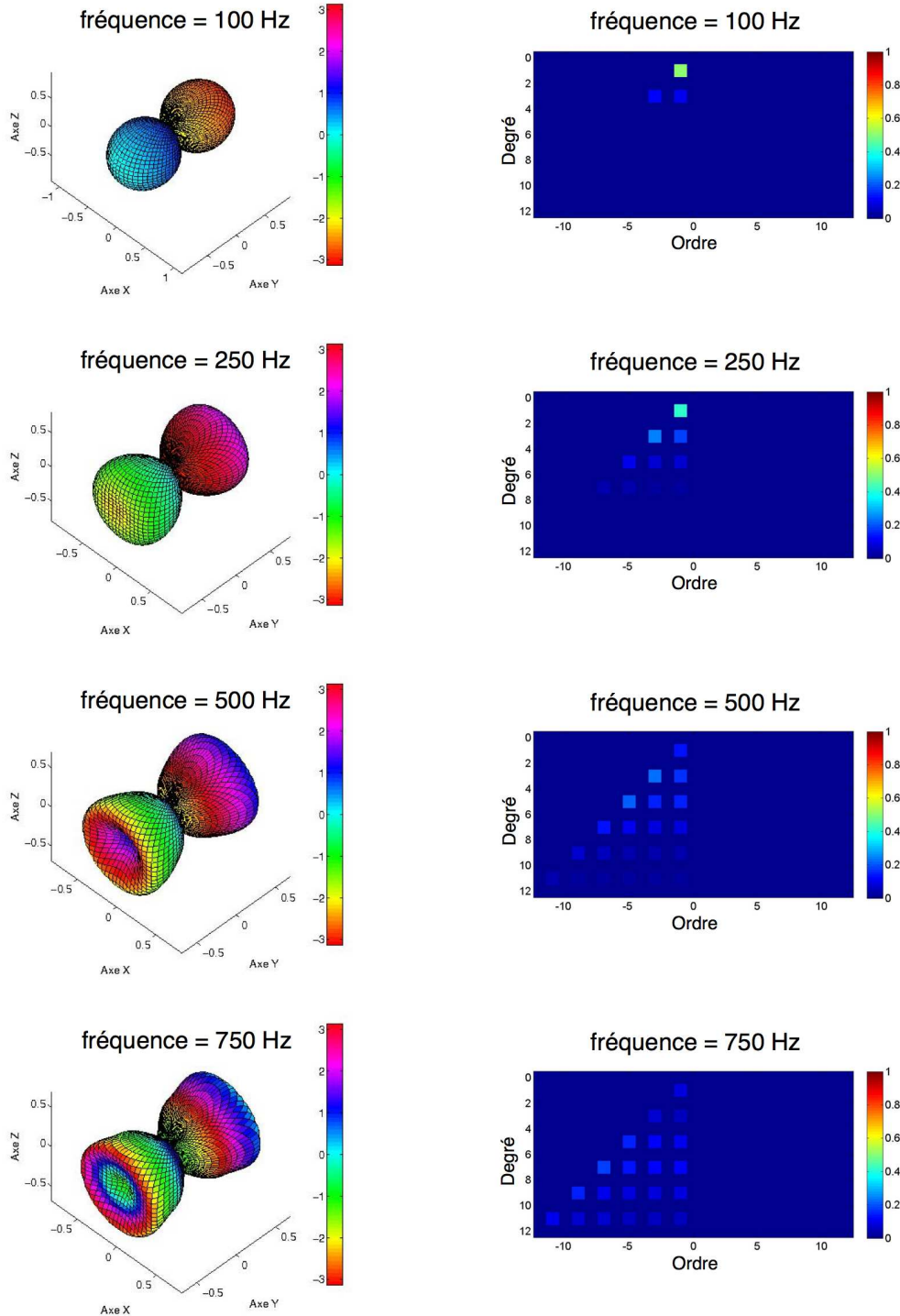


FIG. 3.21 – Décomposition en harmoniques sphériques du champ associé à un dipole latéral Y^{-1} synthétisé à une position centrée, 1m derrière le banc de haut parleurs. **Colonne de Gauche** : Graphiques représentant le module (échelle linéaire) et la phase du champ mesuré au niveau de la sphère de microphones. **Colonne de Droite** : Module des coefficients b_{lm} de la décomposition, niveau linéaire.

3.7. Analyse du champ tridimensionnel effectivement rayonné par le banc WFS.

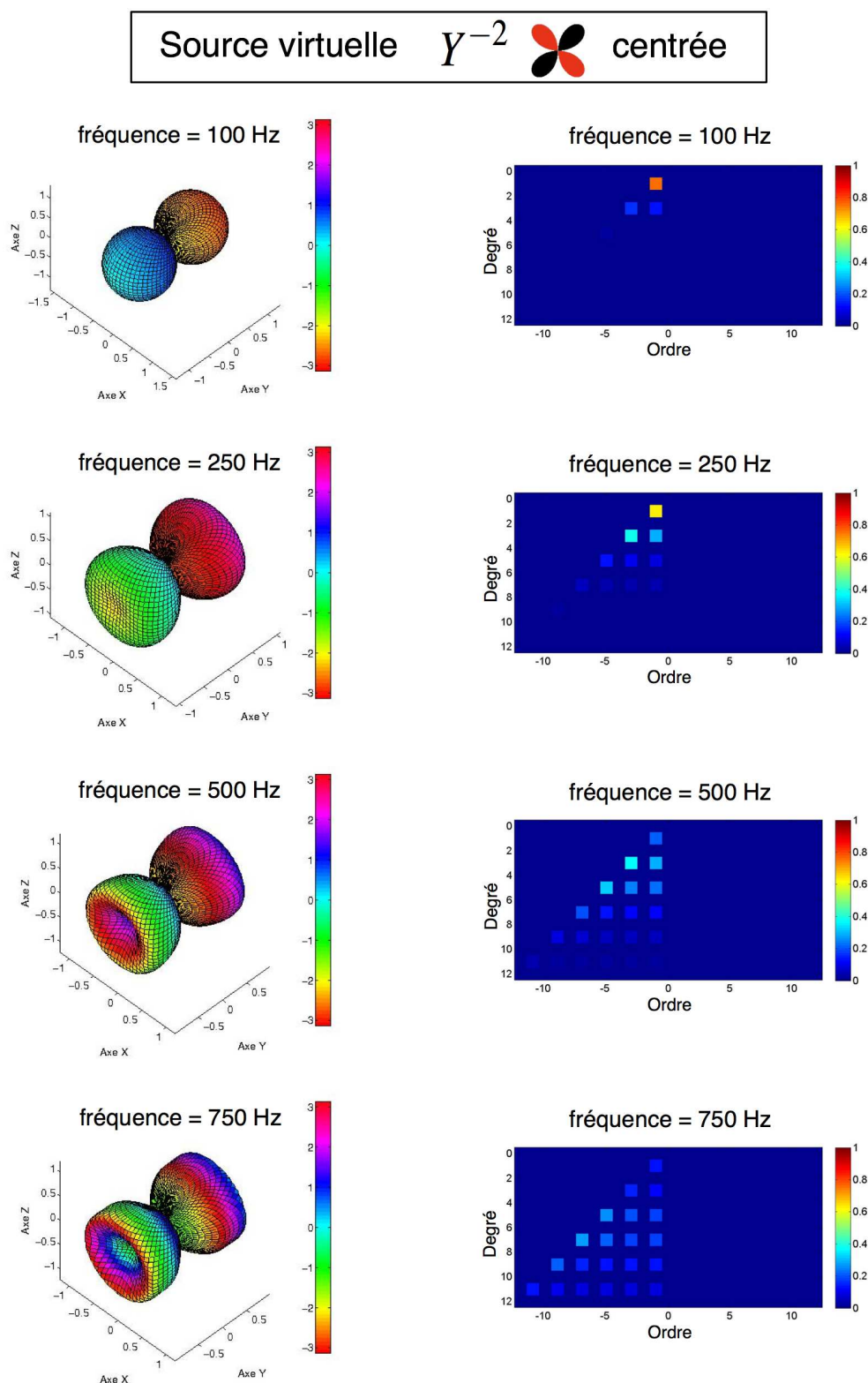


FIG. 3.22 – Décomposition en harmoniques sphériques du champ associé à un quadrupole latéral Y^{-2} synthétisé à une position centrée, 1m derrière le banc de haut parleurs. **Colonne de Gauche** : Graphiques représentant le module (échelle linéaire) et la phase du champ mesuré au niveau de la sphère de microphones. **Colonne de Droite** : Module des coefficients b_{lm} de la décomposition, niveau linéaire.

sphérique de l'harmonique Y^1 reproduite par WFS (Fig. C.4, colonne de droite). Ceci est dû à l'impossibilité de créer des oppositions de phase de part et d'autre du banc linéaire. L'harmonique WFS Y^2 présente un spectre sphérique similaire au spectre de Y^1 , avec une prédominance des modes $\{Y_l^m \mid l, m \text{ pairs et } m > 0\}$, et une répartition progressive de l'énergie sur les modes d'ordre supérieur à mesure que la fréquence augmente.

Malgré ces ressemblances du point de vue du spectre sphérique, les figures de directivité tridimensionnelles associées à Y^2 (Fig. C.4, colonne de gauche) montrent que l'indice de directivité DI_{Y^2} (eq. E.6) est supérieur à l'indice de directivité DI_{Y^1} . En effet, le lobe frontal de l'harmonique WFS Y^2 est visiblement plus fin que le lobe frontal de l'harmonique WFS Y^1 , ce qui dénote une meilleure efficacité de cette figure de directivité pour ce qui est de la focalisation de l'énergie acoustique dans la direction perpendiculaire au banc de haut-parleurs.

3.7.2 Rayonnement tridimensionnel du banc en hautes fréquences

En hautes fréquences, le banc WFS n'est plus en mesure de former des fronts d'ondes cohérents. Etant donné l'échantillonnage extrêmement dense qui serait nécessaire à la décomposition du champ au-dessus de 1500Hz pour un banc WFS ¹⁶, on ne calculera pas le spectre sphérique du banc à ces fréquences ; on propose plutôt simplement d'observer *comment évolue la directivité de rayonnement du banc entre les BF et les HF*, à partir de la mesure sur une sphère échantillonnée de façon beaucoup moins dense (15 points selon l'azimut ϕ et 11 points selon le site θ).

Le banc de haut-parleurs possède une géométrie identique à celui présenté dans la section précédente. On suppose dans un premier temps que les haut-parleurs sont des monopoles idéaux. La réponse du banc de monopoles sur la sphère de mesure est représentée à la figure 3.23, pour la synthèse d'une source virtuelle directive Y^2 centrée. L'observation de cette figure montre que le champ émis par le banc présente un indice de directivité non nul en-dessous de la fréquence de repliement. Cependant, à partir de 2kHz, le champ émis par le banc de haut-parleurs devient incohérent. Cette incohérence du champ se traduit par une diffusion spatiale de l'énergie - à une distance donnée du centre du banc, les microphones captent la même quantité d'énergie, quelle que soit la direction d'émission observée. L'indice de directivité passe donc d'une valeur positive en basses fréquences (pouvoir de focalisation) à une valeur nulle en hautes fréquences (aucun pouvoir de focalisation, production de champ omnidirectionnel "en moyenne").

L'omnidirectionnalité du champ émis par le banc de monopoles au-dessus de la fréquence de repliement subit la directivité propre des transducteurs constitutifs du banc si on remplace les

¹⁶Pour la configuration décrite dans la section précédente (sec. 3.7.1), 8385 points de mesure seraient nécessaires pour décrire le champ émis par le banc à 1.5kHz en évitant le repliement d'ordres supérieurs dans la description.

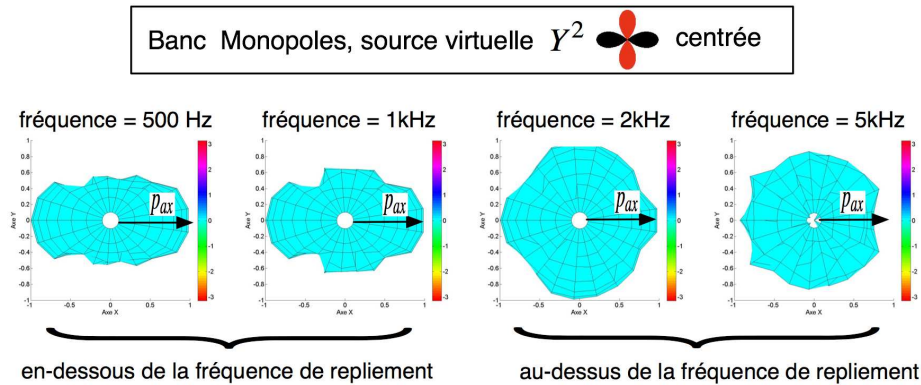


FIG. 3.23 – Magnitude (dB) de la réponse fréquentielle d’un banc WFS de 24 haut-parleurs monopolaires idéaux sur une sphère de microphones centrée sur le banc. On observe que l’indice de directivité du champ émis suivant l’axe acoustique p_{ax} tend vers 0 au-dessus de la fréquence d’aliasing.

haut-parleurs monopolaires par des haut-parleurs réels.

Pour s’en convaincre, on propage la réponse d’un banc constitué de haut-parleurs électrodynamiques de marque KEF¹⁷ au niveau de la même sphère de mesure que dans le cas du banc de monopoles (Fig. 3.24). On observe que le banc de haut-parleurs synthétise un champ globalement cohérent avec celui émis par le banc idéal en dessous de la fréquence de repliement. Cependant, à partir de 2kHz, les haut-parleurs constitutifs du banc deviennent hautement directs. A ce moment là, au lieu que l’énergie se répartisse également dans toutes les directions, elle devient concentrée principalement dans la direction frontale du banc de haut-parleurs (Fig. 3.24, droite). Celà étant dit, la directivité des haut-parleurs affecte également le champ émis par le banc en BF. Cependant, cet effet est moins visible du fait de leur omnidirectionnalité naturelle à ces fréquences.

Nous en déduisons qu’en HF, le modèle théorique monopolaire doit être impérativement remplacé par le modèle de prédiction décrit au chapitre 2 si l’on souhaite disposer d’une simulation réaliste de l’excitation de la salle par le banc WFS.

3.7.3 Extension possible de l’analyse du rayonnement du banc

Dans la section précédente, le choix a été fait de se placer "du côté de la source virtuelle" pour l’étude du rayonnement tridimensionnel émis par le banc. Etant donné ce cadre descriptif, nous avons donc choisi d’employer une disposition sphérique de microphones, et une décomposition du champ rayonné en modes sphériques. Pourtant, étant donné la géométrie linéaire du banc, l’emploi une configuration cylindrique de microphones comme celle montrée dans la

¹⁷La propagation est conduite en utilisant la méthode de prédiction développée au chapitre 2.

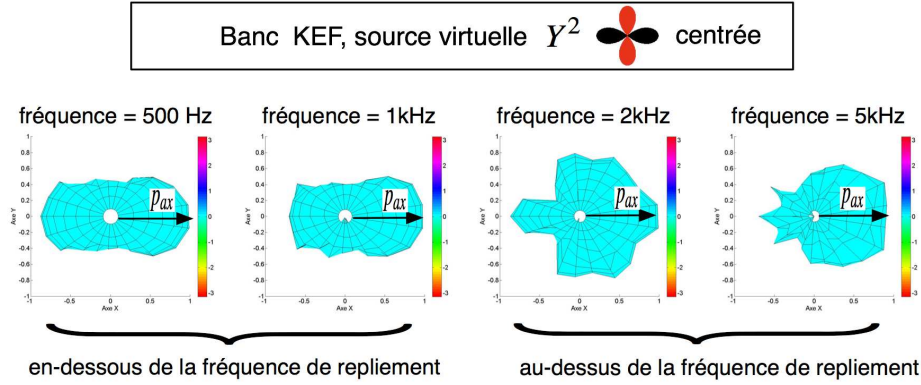


FIG. 3.24 – Magnitude (dB) de la réponse fréquentielle d'un banc WFS de 24 haut-parleurs électrodynamiques de type KEF sur une sphère de microphones centrée sur le banc. On observe que l'indice de directivité du champ émis suivant l'axe acoustique p_{ax} ne tend pas vers 0 au-dessus de la fréquence de repliement spatial du banc, contrairement au banc monopolaire, du fait de la directivité des transducteurs constitutifs du banc de haut-parleurs.

figure 3.25 pourrait être intéressante, de manière à étendre l'étude et se placer cette fois "du côté du banc de haut-parleurs" lui-même.

D'autres auteurs (Spors *et al.*, 2007; Gauthier & Berry, 2006) promeuvent l'idée de décomposer le champ émis par le banc WFS en harmoniques cylindriques suivant un repère (r, ϕ, z) , puis de travailler dans le domaine transformé, de manière à optimiser les calculs de compensation du contexte d'écoute lorsque pour un banc WFS placé en salle. Dans les références citées, les auteurs analysent des bancs WFS concentriques, et utilisent les harmoniques circulaires Y^m pour décrire le plan (r, ϕ) contenant le banc de haut-parleurs.

Pour l'étude d'un banc de haut-parleurs à géométrie linéaire, le repère cylindrique (r, ϕ, z) qui s'impose naturellement possède cette fois un axe z confondu avec la ligne de haut-parleurs, de telle manière à ce la propagation des ondes émises par le banc se fasse selon la direction radiale (Fig. 3.25). La décomposition modale du rayonnement du banc se fait par résolution de l'équation des ondes en coordonnées cylindriques, dont la solution se présente sous la forme (Williams, 1999) :

$$p(r, \phi, z, \omega) = \sum_{m=-\infty}^{\infty} Y^m(\phi) \frac{1}{2\pi} \int_{-\infty}^{\infty} A_m(k_z, \omega) e^{ik_z z} H_m^{(1)}(k_r r) dk_z \quad (3.29)$$

où $H_n^{(1)}(k_r r)$ représente la fonction de Hankel cylindrique sortante de première espèce et $A_n(k_z, \omega)$ est un coefficient de pondération réel. Les variations de directivité dans le plan médian des haut-parleurs (suivant ϕ dans la figure 3.25) sont donc décrites par les harmoniques circulaires Y^m ; les variations de directivité suivant l'axe z contenant le banc de haut-parleurs sont décomposables suivant les fonctions exponentielles $e^{ik_z z}$, comme les modes de vibration d'une corde.

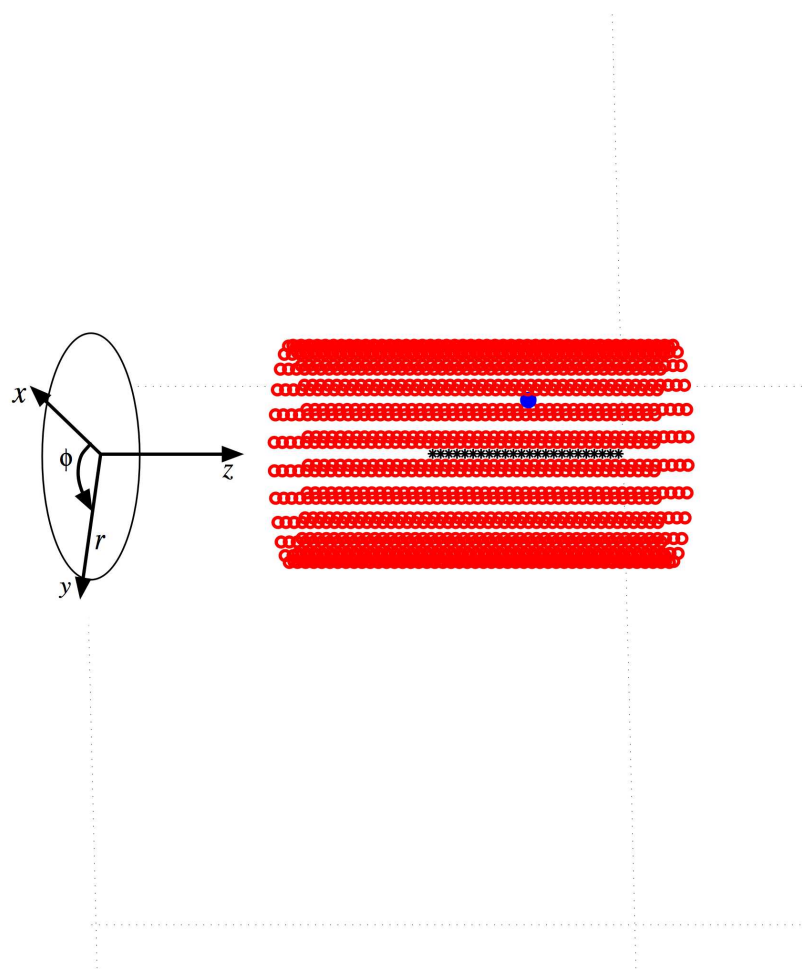


FIG. 3.25 – Disposition de microphones pour une décomposition du champ émis par un banc WFS suivant la base des harmoniques cylindriques. Les haut-parleurs sont représentés par des astérisques noires, les microphones par des cercles rouges et la source virtuelle par un point bleu.

Ce changement de coordonnées pour la décomposition peut s'apparenter à un changement de paradigme pour l'utilisation du système : on passe d'une description où l'on tente de reproduire une source ponctuelle rayonnant à travers la fenêtre que constitue le banc de haut-parleurs à une description où *le banc lui-même devient une source* linéaire dont on contrôle le rayonnement suivant différents modes. Cette description paraît prometteuse pour la tâche de trouver des modes de rayonnement du banc qui soient découplés les uns des autres, et pourrait constituer un mode intéressant d'excitation de la salle. Cette piste reste cependant à explorer dans le cadre de travaux futurs sur le contrôle de la directivité.

3.8 Discussion : De l'intérêt de pouvoir recréer une "Couleur Tonale Directionnelle"

Dans (Boulez, 1983), Pierre Boulez commente "l'impossibilité d'établir un orchestre idéal où tous les timbres pourraient être voisins l'un de l'autre aux moments requis par la partition." Ce problème, qui est lié au placement fixe des instruments d'un orchestre qui impose des "voisinnages" fixes de timbres au cours d'une pièce donnée, pourrait à priori être réglé par l'utilisation de haut-parleurs. Cependant, Boulez soulève le fait que "le haut-parleur anonymise la source réelle". En effet, bien que le haut-parleur permette de restituer de manière fidèle les spécifications temps-fréquence du timbre en un point de mesure donné, celui-ci impose par ailleurs sa propre directivité aux instruments qu'ils tentent de reproduire. L'instrument restitué perd alors son "individualité", à tel point que Boulez affirme qu'il "y aurait plus de ressemblance, d'une certaine manière, entre un piano amplifié et une harpe amplifiée, qu'entre un piano amplifié et un piano non-amplifié".

Weinreich (Weinreich, 1997) explique "l'anonymisation" de la source par le haut-parleur par le fait que le haut-parleur détruit les propriétés de "couleur tonale directionnelle" (Directional Tone Color) de l'instrument. Weinreich montre que cette caractéristique est responsable, entre autre, de changements de direction de provenance perçus lors de passages rapides joués par un violon solo produisant des effets de "brillance soudaine" (flashing brilliance). Elle permet également de distinguer un violon soliste dans un ensemble lorsque celui-ci effectue un vibrato. Weinreich postule enfin que la couleur tonale directionnelle est responsable de la qualité mystérieuse de "projection", qui permet aux instrumentistes de "commander l'attention" du public dans différentes parties de la salle. Le lien entre directivité et "projection" pour une voix de chanteur d'opéra fait d'ailleurs en ce moment l'objet d'un travail effectué par l'équipe Perception Située du Limsi (Katz *et al.*, 2006; Katz & d'Alessandro C., 2007).

Des test informels conduits dans le Forum Neues Musiktheater de Stuttgart en Mai 2005 (Fig. 3.26) ont montré que la simulation de couleur tonale directionnelle par WFS pouvait effectivement augmenter le semblant de réalisme dans la diffusion d'un enregistrement. Pour simuler la couleur tonale directionnelle, un enregistrement d'alto fut filtré en trois bandes, chaque bande étant associée à une figure de directivité donnée.

L'orientation des figures de directivité était asservi aux caractéristiques du signal entrant, de manière à simuler le rayonnement modal d'un instrument réel. Les propriétés de directivité attribuées à l'alto virtuel n'étaient pas issues de mesures en chambre sourde, ce qui limitait la pertinence du champ reproduit¹⁸.

¹⁸Le chapitre 2 expose une méthode de mesure de rayonnement de haut-parleur pouvant à priori permettre de décomposer le rayonnement d'une source sonore dans le but de diffuser celle-ci via un système WFS. Son utilisation pour l'analyse de rayonnement de sources non-électroacoustiques pose cependant un certain nombre de

3.8. Discussion : De l'intérêt de pouvoir recréer une "Couleur Tonale Directionnelle"



FIG. 3.26 – Séance d'écoute d'un système WFS lors d'une série d'ateliers organisés conjointement par l'Ircam et sonic emotion ag. au Forum Neues Musiktheatre de Stuttgart en Mai 2005.

Malgré tout, un sentiment de présence et de réalisme augmentés de l'instrument ont été ressentis lors des écoutes par rapport à la situation où le signal instrumental était diffusé via une source à directivité statique. (Jacques *et al.*, 2005) remarque à ce propos que la synthèse des caractéristiques de directivité instrumentales ne nécessitent pas d'être recréées avec une précision objective élevée, du moment que l'auditeur n'est pas parfaitement familier avec la directivité de l'instrument en question ; il suffit de générer une directivité "subjectivement convaincante" de manière à assurer une image sonore vivante.

Une thèse commençant actuellement dans l'équipe Acoustique des Salles de l'Ircam aura (entre autre) pour but d'étudier dans un cadre plus formel les possibilités offertes par les systèmes WFS pour le contrôle des variations temporelles des modes de directivité d'un instrument réel.

problèmes, qui sont abordés dans le chapitre suivant, section 4.2.1.

3.9 Conclusion

Dans cette partie, nous avons détaillé les effets que produisent des variations de directivité d'un instrument réel en salle : l'effet de "brillance soudaine", la "couleur directionnelle tonale", la "projection". Ces variations participent, avec le timbre, à forger l'identité propre à chaque instrument de l'orchestre. Malheureusement, à en croire Boulez, cette identité se perd au moment de la diffusion via le haut-parleur, qui "anonymise la source réelle". Cette propriété de "dissolution de l'identité la source virtuelle" par le haut-parleur classique le rend ainsi peu attractif pour ce qui est de compléter les timbres dans différentes sections de l'orchestre lorsque la partition le demanderait. Dans ce contexte, la nécessité de développer une méthode permettant de *rendre son identité* à l'instrument, par le biais d'un travail sur la directivité de rayonnement, est mise en exergue.

Le travail de reproduction d'une directivité instrumentale avait déjà été initié avec la Timée, qui est le modèle d'une source à directivité variable permettant de "redonner" en quelque sorte "une identité à l'instrument" virtuel. Un des objectifs principaux de ce chapitre a été de développer des outils permettant de réaliser une tâche similaire, cette fois dans le cadre de la WFS. La WFS présente l'inconvénient, vis à vis de la Timée, d'un contrôle de directivité restreint au plan horizontal, et d'une symétrie cylindrique de rayonnement imposée, mais elle offre cependant la possibilité de retrouver la liberté de localisation de la source virtuelle sur scène, à la fois en largeur et en profondeur. On possède donc, dans une certaine mesure, la possibilité d'*augmenter les sections de l'orchestre sur scène, par l'ajout d'un ou plusieurs instruments virtuels, de manière à y compléter le timbre au moment requis par la partition.*

Pour permettre d'avancer vers cet objectif, nous avons repris le formalisme déjà mis en place dans le cadre de travaux précédents (Caulkins *et al.*, 2003; Corteel, 2007; Warusfel *et al.*, 2004b), et complété celui-ci, en montrant comment les opérations de manipulation de la directivité utilisées couramment dans le format Ambisonique (rotation, focalisation) pouvaient être transcrites au cadre de la WFS. Les outils que nous avons formalisé permettent d'abord, a priori, de *redonner une identité aux instruments virtuels synthétisés en WFS*, via un contrôle précis de la directivité (tout du moins partiellement, du fait de la limitation du contrôle au plan horizontal).

Ces outils de manipulation de directivité peuvent toutefois également être envisagés

comme un *moyen de contrôler l'excitation de la salle*. En effet, comme nous le montrons dans le chapitre suivant, l'absence d'un système de diffusion de réverbération à l'arrière de la salle nécessite de réfléchir à des moyens de recréer un champ diffus en utilisant uniquement le système frontal placé en nez de scène. Les sources hyperdirectives (Diracs Directifs), dont nous avons décrit la synthèse dans ce chapitre, seront ainsi envisagés comme moyen d'alimenter le champ diffus de la salle, par focalisation de l'énergie tardive vers les murs latéraux.

Dans la suite du chapitre, nous avons introduit un *critère permettant d'approximer simplement l'ordre maximal de recombinaison d'une source directive en WFS*, basé sur des considérations géométriques montrant le lien entre la répartition des haut-parleurs et la fréquence de repliement de spectre des fonctions de directivité reproduites en WFS. Ce critère peut être utilisé pour délimiter des zones sur la scène dans lesquelles on peut restituer un instrument suivant son ordre de description.

Nous avons ensuite *évalué la qualité de reconstruction du son direct des sources directives élémentaires WFS dans le plan horizontal*. Nous avons montré, par une analyse de l'erreur utilisant un critère dans le domaine complexe, que le champ est globalement bien restitué en dessous de la fréquence de repliement de spectre, à la fois en phase et en amplitude, ce qui montre que le paradigme de synthèse de directivité par combinaison linéaire des figures de base est viable en-dessous de la fréquence de repliement. Au-dessus de la fréquence de repliement, nous avons montré que la qualité de reconstruction du champ devenait dépendante vis à vis de la position et de la directivité de la source virtuelle, à cause d'un problème de fenêtrage de la directivité par le banc de haut-parleurs. Dès que le champ cible contient des annulations, le banc WFS n'est plus capable de le synthétiser correctement au-dessus de la fréquence de repliement. Nous montrerons dans le chapitre suivant comment ce problème se répercute sur la puissance émise par le banc de haut-parleurs dans la salle pendant des déplacements et rotations de la source virtuelle sur la scène.

Dans la section d'après, nous avons examiné les *propriétés du champ effectivement rayonné par le banc dans les trois dimensions*. Nous avons observé que malgré le fait que la WFS octroie un contrôle sur le plan horizontal en-dessous de la fréquence de repliement, le rayonnement tridimensionnel des directivités élémentaires WFS n'est pas conforme aux figures de directivité tridimensionnelles que nous voudrions idéalement reproduire. En effet, nous avons montré par simulation d'un banc WFS idéal que la symétrie de révolution du banc crée des redondances entre certaines figures de directivité synthétisées par WFS (Y^{-2} redondante avec Y^{-1} , par exemple); par

ailleurs, elle empêche de recréer des variations de directivité suivant les axes perpendiculaires au banc de haut-parleurs, ce qui limite le spectre spatial des sources qui peuvent être effectivement recrées à travers la “fenêtre” du banc WFS. Ces particularités seront à prendre en compte au moment du transcodage de directivités tridimensionnelles d’instruments réels pour la synthèse sur banc WFS.

Nous avons montré par ailleurs que le rayonnement d’un banc WFS constitué de monopoles tendait à devenir monopolaire en hautes fréquences, du fait de l’impossibilité de synthétiser des ondes cohérentes au-dessus de la fréquence de repliement. Toutefois, en remplaçant les haut-parleurs idéaux monopolaires de la simulation précédente par des haut-parleurs réels, nous avons observé que la directivité de rayonnement du banc au-dessus de la fréquence de repliement tendait à se concentrer plutôt dans la direction frontale, du fait de la directivité naturelle associée aux haut-parleurs réels. Cet effet était moins notable en BF, car les haut-parleurs dont nous disposons sont omnidirectionnels à ces fréquences. Néanmoins, *nous en déduisons que le modèle théorique monopolaire doit être impérativement remplacé par le modèle de prédiction du chapitre 2 si on souhaite obtenir une estimation réaliste de l’excitation de la salle par le banc WFS.*

Enfin, nous avons proposé une *extension du formalisme de contrôle de la directivité du banc, basé sur une décomposition de son rayonnement sur la base des harmoniques cylindriques.* Cette extension est envisagée comme un moyen prospectif de fournir des nouveaux modes d’excitation de la salle via un système WFS. Nous noterons au passage l’existence des techniques de “monopole synthesis” (Giron, 1996), ainsi que de synthèse de sources étendues à partir de nuages de sources ponctuelles (Menzies, 1999; Malham, 2001), qui n’ont pas encore, à la connaissance de l’auteur, été explorées en WFS.

Dans le chapitre suivant, nous investiguons les conséquences de ces propriétés de rayonnement tridimensionnel du banc WFS que nous avons mises en évidence dans ce chapitre. En effet, la non-conformité de ces propriétés avec celles d’une source réelle implique que l’effet de salle engendré par le banc va devoir être compensé, si on souhaite assurer le réalisme des instruments virtuels reproduits par WFS dans une situation de concert.

CHAPITRE 4

Caractérisation et contrôle de l'effet de salle tardif pour la Wave Field Synthesis

4.1 Introduction

Nous avons établi, en introduction du chapitre précédent, que la reproduction réaliste d'une source instrumentale par voie électroacoustique exige d'assurer une composante de son direct et un effet de salle associé qui soient le plus précis possible à toutes les positions d'écoute dans la salle. L'enjeu est de pouvoir "augmenter" la réalité du concert instrumental en recréant la présence d'instruments virtuels sur scène, à côté de sources instrumentales réelles (sec. 3.1, Fig. 3.1).

Contrôle du son direct

Plusieurs approches permettant d'optimiser la synthèse de la composante de son direct d'une source virtuelle WFS en conditions réelles ont été exposé au chapitre 1, basées sur un filtrage multicanal du système de haut-parleurs. Une extension de la technique, permettant de contrôler la directivité des sources virtuelles d'une manière simple et flexible a été exposée au chapitre 3.

La WFS en tant que système de renforcement sonore

Le problème du contrôle du champ réverbéré dans des salles de concert équipées par un système WFS a déjà été abordé de nombreuses fois dans le passé. Initialement, la recherche a été

conduite avec l'idée d'utiliser la WFS en tant que système de renforcement sonore des sources sur scène, couplé à un système électroacoustique pour la synthèse de réflexions précoces et de réverbération tardive (Vogel, 1993). L'approche s'est avérée efficace lors une implémentation conduite au Malmö Stadsteater, et a permis notamment d'améliorer le niveau du son direct des sources sur scène de 6dBA sans causer d'effet de Larsen perceptible. Dans cette étude, la contribution du système de renforcement sonore WFS à l'effet de salle n'était cependant pas prise en compte.

WFS en tant que système d'auralisation de réverbération

Dans (Sonke, 2000), l'auteur montre que la propriété d'isotropie d'un champ réverbéré peut être recréée dans le plan horizontal d'une manière perceptuellement convaincante par l'emploi d'un ensemble d'ondes planes synthétisées en WFS. Le concept d'un système d'acoustique variable permettant de reproduire l'acoustique d'une salle "référence" dans la salle de reproduction est développé, suivant une analyse directionnelle du champ sonore basée sur les travaux de (Hulsebos, 2004).

Sonke affirme qu'en supposant que les réponses dans les différentes directions analysées sont décorréelées¹, le processus d'*auralisation* de la salle de référence consiste simplement à injecter de l'énergie réverbérée supplémentaire, via des ondes planes WFS, dans chaque direction de la salle de reproduction où il en manque vis-à-vis de la référence². Notons que les réponse énergétiques de l'ensemble des ondes planes utilisées pour compenser la réverbération manquante sont mesurées et prises en compte dans le processus. En d'autres termes, l'interaction du système de diffusion de réverbération avec la salle d'écoute est caractérisée. Cependant, la synthèse du son direct est considérée de manière séparée, et sa contribution naturelle à l'effet de salle n'est pas prise en compte dans le processus.

Compensation d'effet de salle en WFS

Dans les travaux conduits sur la compensation d'effet de salle en WFS, il semble que l'accent principal ait été placé sur le contrôle des réflexions précoces (Corteel & Nicol, 2003; Spors *et al.*, 2004a; Corteel, 2004; Spors *et al.*, 2007), et non la réverbération tardive. En effet, on peut noter que le contrôle des réflexions précoces dans une salle permet, par extension, un contrôle de l'effet de salle tardif. Cependant, les études citées soulignent le fait que l'annulation parfaite des réflexions précoces dans une salle est impossible via un système WFS, principalement à

¹Ce qui est à priori le cas pour le champ réverbéré, mais pas nécessairement pour le son direct et les premières réflexions

²En absence d'un système d'annulation d'écho associé, le concept proposé n'est valable que si la salle de référence est au moins aussi réverbérante que la salle de reproduction.

cause de leur restriction au plan horizontal. Cette restriction empêche une réduction efficace de l'énergie acoustique rayonnée en dehors du plan horizontal.

Problématique et objectif du chapitre

Etant donné qu'un système WFS classique ne permet pas d'annuler l'énergie qu'il rayonne en dehors du plan horizontal, il apparaît que *la composante de son direct synthétisée par le banc WFS contribue de manière intrinsèque à la réverbération tardive*. Cette contribution doit alors être caractérisée et compensée si on souhaite assurer que l'effet de salle global produit par le système WFS soit cohérent avec l'effet de salle cible. L'importance de cette caractérisation devient particulièrement critique si l'on considère le cas où l'on reproduit des sources sonores directives, pour lesquelles l'énergie émise dans la salle dépend à la fois de leur position et de leur orientation. Ce problème est illustré de manière simple à la figure 4.1.

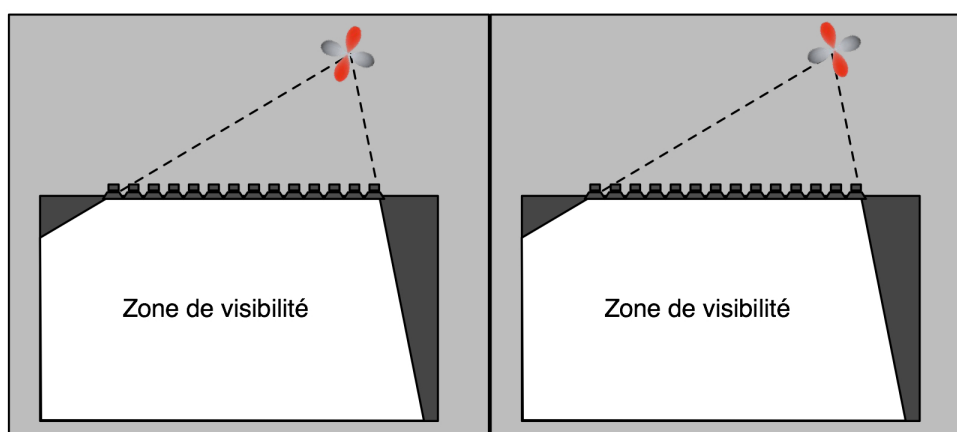


FIG. 4.1 – Mise en évidence du problème de fenêtrage dans le cas d'une source directive.

Le schéma de gauche de la figure 4.1 montre une source de directivité quadrupolaire Y^2 orientée de telle manière à ce que la fenêtre de visibilité de la source englobe entièrement son lobe principal. Dans la situation représentée à droite, la source est orientée différemment, et cette fois la fenêtre de visibilité est centrée sur un zéro de la source. Des variations importantes dans la quantité de puissance émise par le banc de haut-parleurs sont donc à prévoir lors de la synthèse du son direct associé à ces deux orientations de la même source. Ceci est vrai également pour des changements de position d'une source, ce qui a été mis en évidence dans (Caulkins *et al.*, 2004).

Comme la quantité de puissance émise dans la salle détermine les caractéristiques de la réverbération tardive³, les variations de puissance émise observées doivent être compensées

³L'équation A.8 donnée en Annexe A.2 montre que la densité d'énergie en régime établi E_∞ (liée au niveau de la réverbération) et le temps de réverbération Tr_{60} dépendent tous les deux de la puissance W émise par la source.

si on souhaite assurer une continuité perceptive de la réverbération lors de déplacements et rotations de la source virtuelle sur scène.

L'enjeu de ce chapitre est de fournir plusieurs méthodes permettant de contrôler la qualité acoustique associée aux sources virtuelles reproduites sur scène.

Dans la première méthode proposée, on vise à ajuster cette qualité acoustique dans les 4 sections temporelles OD , $R1$, $R2$ et $R3$ suivant une cible idéale de reproduction définie à une position de référence dans la salle.

Dans la seconde méthode dite "simplifiée", on cherche simplement à caractériser et compenser les variations de puissance totale émise par le banc WFS lors de changements de position/orientation de la source virtuelle synthétisée, de manière à rétablir une cohérence dans l'évolution de la réverbération tardive au cours de la diffusion en salle.

4.2 Proposition d'une méthode de compensation de l'effet de salle tardif pour la WFS

Cette section est consacrée à la description d'une méthode pour compenser l'effet de salle tardif associé à une source virtuelle WFS, basée sur l'architecture d'une méthode de compensation de contexte décrite dans (Jot *et al.*, 1998). Le principe de la méthode consiste à modifier les enveloppes énergétiques des signaux envoyés par les haut-parleurs du système de diffusion en fonction d'une *qualité acoustique* (QA) "cible" à réaliser, d'une *qualité acoustique* "contexte", et d'une *qualité acoustique* "live" au niveau d'une position de référence \vec{r}_{ref} située dans la salle d'écoute.

Dans la formulation originale de la méthode de compensation, la QA "cible" décrit la qualité acoustique que l'on cherche à obtenir au point de référence. La QA "contexte" se déduit de la caractérisation de chaque haut-parleur employé pour la diffusion d'effet de salle au sein du lieu d'écoute. Enfin, la QA "live" est réservée à la situation de concert *mixte* où des instrumentistes jouent simultanément sur scène avec le système de diffusion électroacoustique. Dans ce cas, la QA "live" permet de caractériser l'effet de salle initial engendré par le/les instrumentistes à la position de référence.

L'adaptation de cette méthode à la compensation de l'effet de salle associé à une source virtuelle WFS exige de redéfinir les QA "cible" et "contexte", ce que nous proposons de faire dans les deux sections suivantes.

4.2.1 Définition d'une qualité acoustique "cible"

On propose de distinguer deux QA "cibles" possibles : une QA "cible" définie de manière perceptive, et une QA "cible" définie de manière physique.

Définition d'une qualité acoustique "cible" perceptive

La définition d'une QA "cible" perceptive se fait en utilisant l'opérateur perceptif du Spatialisateur, décrit en section 1.4.3. En utilisant l'interface graphique représentée en figure 1.20, on règle la qualité acoustique via les 9 paramètres perceptifs disponibles décrits en section 1.4.3⁴.

Le réglage de la qualité acoustique induit une modification des paramètres bas niveau du Spatialisateur, c'est à dire l'énergie OD , $R1$, $R2$ et $R3$ dans les différentes sections temporelles et bandes fréquentielles (OD , $R1$, $R2$ et $R3$ sont définies en eqs. 1.32, 1.33, 1.34 et 1.35).

Ainsi, par exemple, le paramètre perceptif *source presence* est associé à un critère objectif $DirE$ défini de la manière suivante :

$$DirE = 10 \cdot \log (\max (OD + R1 + R2), (R2)) + \min ((OD + R1), (0.18 \cdot R2)) \quad (4.1)$$

Lorsque l'on manipule le paramètre perceptif de *source presence*, le processeur d'effet de salle se charge alors de calculer un nouvel ensemble de contributions OD , $R1$, $R2$ et $R3$ dans les différentes sections temporelles.

A l'issue du réglage, on dispose d'un ensemble de valeurs OD_{cible} , $R1_{cible}$, $R2_{cible}$ et $R3_{cible}$ dans chaque bande qui représentent la simulation de la QA "cible" perceptive.

Définition d'une qualité acoustique "cible" physique

Principe

Contrairement à la définition d'une qualité acoustique "perceptive" utilisant le Spatialisateur, qui s'adresse plutôt à une source virtuelle, on désire prendre en compte ici le caractère "physique" de la reproduction d'une source instrumentale, notamment l'effet de sa directivité. L'objectif dans la définition d'une QA "cible" physique, est de pouvoir mesurer l'effet de salle produit par un l'instrument réel que nous visons à reproduire, et ce quelle que soit sa position sur la scène.

L'idée dans cette méthode est donc de caractériser l'ensemble des positions ou orientations que peuvent potentiellement occuper un instrument réel sur la scène. Cette caractérisation sous-entend la possibilité de mesurer le champ réverbéré associé à l'instrument à une position de

⁴Ces 9 paramètres sont : *source presence*, *source warmth*, *source brilliance*, *room presence*, *running reverberance*, *envelopment*, *late reverberance*, *heaviness*, *liveness*

référence dans la salle d'écoute, en tenant compte de sa directivité.

On propose ici une méthode modulaire de caractérisation de la QA cible, qui tire partie de la possibilité de recréer le champ 3D émis par un instrument en utilisant une source à directivité contrôlée type "Timée" (voir section 3.1 pour une présentation générale de ce type de source). La Timée est utilisée pour effectuer une mesure de réponse impulsionnelle de figures de directivité élémentaires à toutes les positions souhaitées sur la scène. Alors, à condition de savoir exprimer le rayonnement de l'instrument réel que l'on souhaite reproduire par une somme de figures de directivité élémentaires pondérées, on peut simuler *a posteriori* son rayonnement dans la salle.

On a donc par extension un moyen de recomposer une QA cible physique de l'effet de salle associé à n'importe quel instrument réel à la position de référence dans la salle *a posteriori*, à partir d'une seule campagne de mesures conduite utilisant la Timée.

Nous allons décrire à présent en détail les étapes nécessaires pour reconstituer la QA cible d'un instrument réel, c'est à dire :

- Analyse et décomposition du rayonnement de l'instrument en figures de directivité élémentaire
- Recomposition du rayonnement de l'instrument en salle utilisant la Timée, de manière à dériver la QA cible physique.

Analyse du rayonnement de la source sonore simulée

La première étape de la définition d'une QA "cible" physique consiste à analyser le champ émis par la source sonore réelle que l'on souhaite simuler par WFS. Si la source à simuler est de type électroacoustique, la mesure de caractérisation peut se faire directement en utilisant la méthode de décomposition du rayonnement décrite au chapitre 2.

Par contre, si la source à simuler est une source acoustique (comme, par exemple, un instrument de musique), la méthode de mesure de réponse impulsionnelle par séquences pseudo-aléatoires (MLS) ou sinus balayés exposée en Annexe A.1.2 ne peut pas être utilisée, et la méthode du chapitre 2 doit être adaptée. D'une part, contrairement aux sources électroacoustiques, les sources acoustiques ne peuvent pas, en général, être excitées de manière répétable, sauf par des procédés d'excitation de type robotique (bouche artificielle, pots vibrants, disklavier). D'autre part, les signaux émis par la plupart des sources acoustiques sont variés dans le temps. Enfin, il se peut que leurs diagrammes de directivité soient dépendants du signal émis (sec. 3.8).

Dans (Giron, 1996), l'auteur propose de faire l'hypothèse que la fonction de directivité en champ lointain de l'instrument acoustique est *indépendante de la phase et de l'amplitude du signal original*. En reprenant la méthode de caractérisation qu'il propose, on peut donc décom-

poser le champ de pression $p(r, \theta, \phi, \omega)$ rayonné par la source sonore en un signal $p_{ax}(r, \omega)$ émis suivant "l'axe acoustique" de la source, et un *filtre de facteur directionnel* $D(\theta, \phi, \omega)$ (appelé également *facteur directionnel* dans (Kinsler *et al.*, 2000))⁵.

A l'issue de cette phase de caractérisation, en supposant que la directivité de l'instrument soit bien indépendante du signal d'entrée, on dispose d'une description complète du rayonnement tridimensionnel de l'instrument. Ce format peut ensuite être converti en une suite de coefficients de Fourier sphériques $a_{lm}(\omega)$ (cf. Fig. 4.2), à condition que la source soit suffisamment compacte pour pouvoir approcher tous les filtres de facteur directionnels D par des filtres à minimum de phase (Giron, 1996). Remarquons que les coefficients a_{lm} de la décomposition n'ont pas de dépendance en r dans ce cas, car le filtre de facteur directionnel D ne dépend pas du rayon.

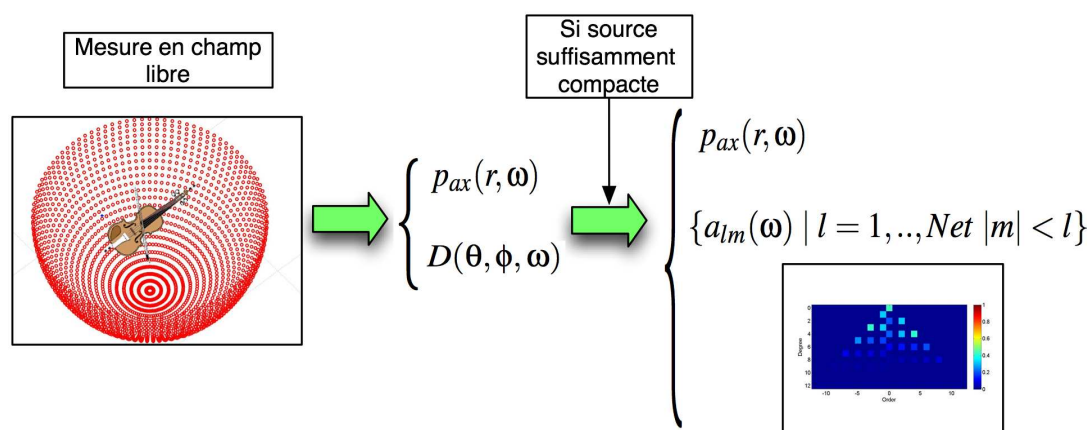


FIG. 4.2 – Décomposition du rayonnement de l'instrument réel en un filtre de facteur directionnel $D(\theta, \phi, \omega)$ et un champ de pression sur l'axe acoustique $p_{ax}(r, \omega)$. Si la source est suffisamment compacte, on peut dériver une décomposition $a_{lm}(\omega)$ du filtre de facteur directionnel suivant la base des harmoniques sphériques.

Giron montre que cette méthode de caractérisation est efficace pour la prédiction du rayonnement d'une voix parlée. L'extension de cette méthode à la caractérisation d'une voix chantée paraît également raisonnable étant donné l'invariance qui a été trouvée dans les figures de directivité de la voix d'un chanteur professionnel (Katz *et al.*, 2006; Katz & d'Alessandro C., 2007), même lorsque le chanteur "projette" sa voix.

Giron observe cependant que l'application de cette méthode à la prédiction du rayonnement d'un piano ne donne pas de bons résultats. Il suggère que la resynthèse haute qualité d'une source étendue exige le développement d'un modèle de rayonnement dépendant du signal émis. Ceci implique d'abord de pouvoir estimer correctement la phase de l'instrument au niveau de la sphère de mesure. A moins de pouvoir reproduire l'excitation de l'instrument de manière

⁵La signification détaillée de ces termes est donnée en annexe à l'équation A.56.

répétable (par un pot vibrant, ou l'utilisation d'un Disklavier), une mesure précise de la phase nécessite donc d'enregistrer le signal émis par la source *simultanément* sur toutes les positions d'une sphère d'échantillonnage entourant la source. Dans l'installation de mesure décrite en 2.2, cette opération est impossible.

D'autre part, l'analyse précise du rayonnement nécessite soit une préconnaissance des hauteurs jouées pendant la mesure, soit le développement de méthodes permettant d'analyser le pitch de l'instrument au cours de l'enregistrement. Giron préconise d'ailleurs d'investiguer des méthodes de synthèse physique du rayonnement avant d'envisager la mise en oeuvre de telles méthodes de traitement de signal pour la caractérisation des sources sonores étendues.

Principe de la simulation du rayonnement de l'instrument réel simulé

En supposant qu'on a pu caractériser convenablement le rayonnement de l'instrument réel suivant la méthode exposée au paragraphe précédent, la deuxième phase de la définition d'une QA "cible" physique consiste à utiliser une source à directivité contrôlée type "Timée" pour synthétiser la réponse de l'instrument réel à la position de référence.

On suppose que la source à directivité contrôlée (que l'on nommera "Timée" par économie de langage) permet de synthétiser des sources de directivité élémentaire jusqu'à un ordre $L_{max} > N$, avec N l'ordre maximal de décomposition de l'instrument mesuré à l'étape précédente. Cette synthèse est réalisée en appliquant un ensemble de filtres $\{H_{lm}^i(\omega) \mid i = 1, \dots, N_C\}$ aux N_C haut-parleurs constitutifs de la Timée, et permet a priori de reconstituer le champ de pression qui serait émis par une source de directivité élémentaire $y_l^m(\theta, \phi)$ au niveau d'une sphère de rayon r_0 .

Pour recréer le champ émis qui serait émis par l'instrument réel orienté dans la direction α_S , on doit pondérer chacun des filtres élémentaires $\{H_{lm}^i(\omega) \mid i = 1, \dots, N_C, l = 1, \dots, N, |m| < l\}$ par des facteurs $a_{lm}(\alpha_S, \omega)$ déduits de la rotation de $a_{lm}(\omega)$ d'un angle α_S en utilisant relation de rotation d'une figure de directivité donnée en 3.6. Les filtres ainsi pondérés permettent de reproduire le rayonnement original de l'instrument, à condition de les alimenter par le signal $p_{ax}(r_0, t)$ mesuré selon l'axe de l'instrument. Cette opération est illustrée à la figure 4.3.

Recomposition de la qualité acoustique "cible"

On place la Timée à la position \vec{r}_S que l'on souhaite caractériser. Soient $\{s_i(\vec{r}_S, \vec{r}_{ref}, t) \mid i = 1, \dots, N_C\}$ l'ensemble des réponses impulsionnelles des haut-parleurs individuels constitutifs de la Timée sont mesurées la position de référence \vec{r}_{ref} , et $\{s_i(\vec{r}_S, \vec{r}_{ref}, \omega) \mid i = 1, \dots, N_C\}$ leur transformées de Fourier. Les réponses impulsionnelles de ces sources électroacoustiques sont mesurées suivant la méthode décrite en Annexe A.1. La réponse impulsionnelle "cible"

4.2. Proposition d'une méthode de compensation de l'effet de salle tardif pour la WFS

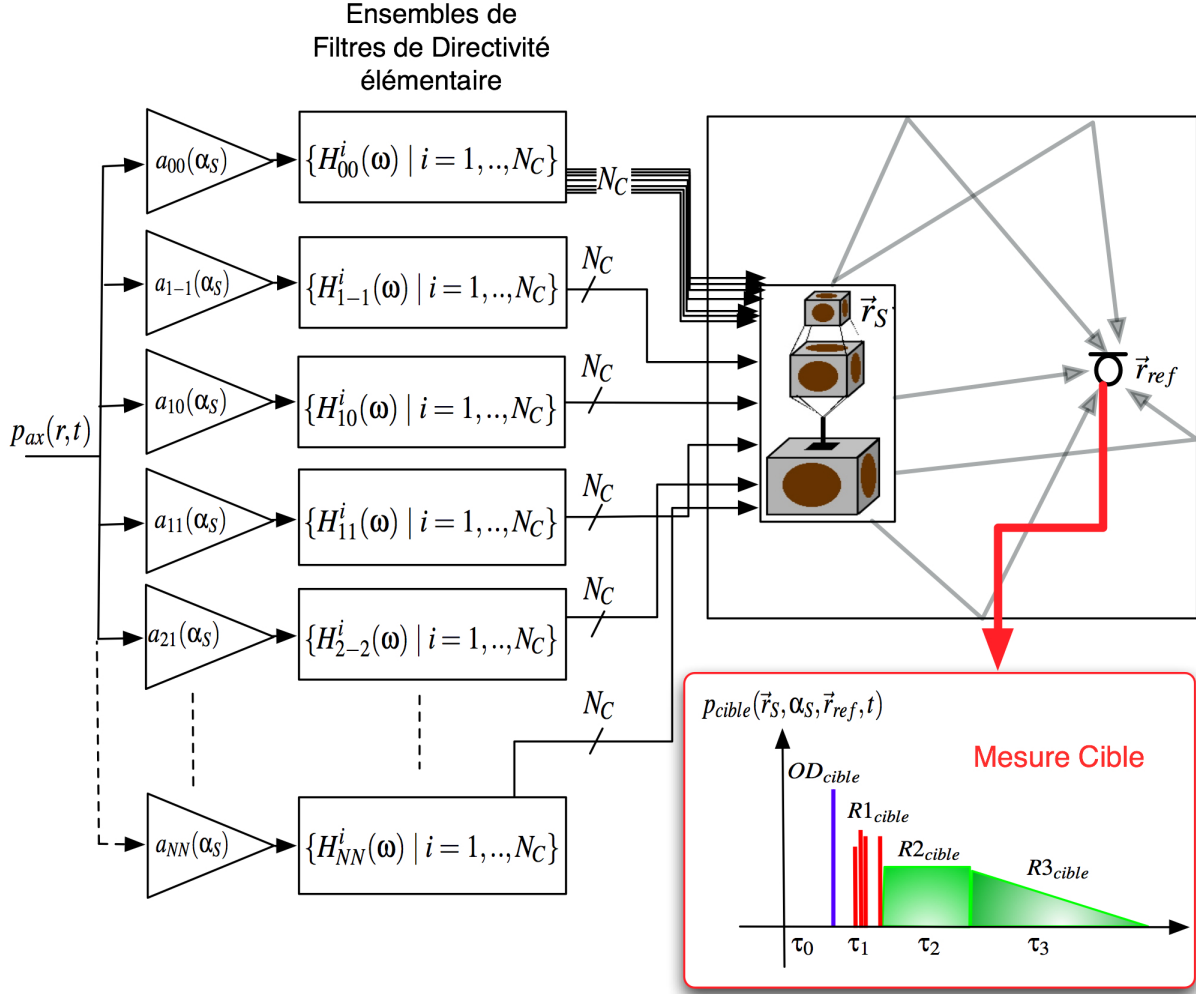


FIG. 4.3 – Recomposition de la qualité acoustique "cible" pour un instrument situé en \vec{r}_S à la position de référence \vec{r}_{ref} , utilisant la Timée. Pour ce faire, on filtre la pression $p_{ax}(r_0, t)$ mesurée sur l'axe de l'instrument réel par une combinaison linéaire des directivités de base H_{lm}^i de la Timée placée en \vec{r}_S , pondérées par les coefficients $a_{lm}(\alpha_S)$ du spectre Fourier sphérique de l'instrument réel orienté suivant α_S .

$p_{cible}(\vec{r}_S, \alpha_S, \vec{r}_{ref}, t)$ à la position de référence est ensuite recomposée de la manière suivante :

$$p_{cible}(\vec{r}_S, \alpha_S, \vec{r}_{ref}, t) = \text{IFFT} \left(\frac{\sum_{i=1}^{N_C} \sum_{l=1}^N \sum_{m=-l}^{m=l} \mathbf{s}_i(\vec{r}_{ref}, \omega) H_{lm}^i(\omega) a_{lm}(\alpha_S, \omega)}{R_d(\omega)} \right), \quad (4.2)$$

où $R_d(\omega)$ représente la réponse en champ diffus du microphone de mesure employé.

La réponse p_{cible} est ensuite décomposée en sections temporelles pour effectuer le calcul de l'énergie OD_{cible} , $R1_{cible}$, $R2_{cible}$ et $R3_{cible}$ suivant les équations 1.32, 1.33, 1.34 et 1.35. L'ensemble des valeurs OD_{cible} , $R1_{cible}$, $R2_{cible}$ et $R3_{cible}$ dans chaque bande constituent la QA "cible" physique.

Conclusion

A l'issue de cette phase de mesure, on est a priori capable de recomposer la QA "cible" associée à n'importe quelle source acoustique à la position caractérisée, pour peu que celle-ci ait été analysée suivant la méthode donnée à la section précédente.

Les mesure des réponses impulsionnelles de la Timée doivent ensuite être répétées en la plaçant à toutes les positions auxquelles on souhaite reproduire le rayonnement de la source instrumentale (idéalement l'ensemble des positions de la base de données de filtres WFS, représentées à la figure 1.28). Cette mesure se fait "une fois pour toutes" dans une salle donnée, après quoi on peut reconstituer l'ensemble des QA cibles associées à n'importe quel instrument dans la salle.

Remarque sur le choix du type microphone de mesure utilisé

Il n'y a pas de raison a priori de contraindre le choix du type de microphone utilisé pour conduire la mesure de la QA cible "physique". On pourrait par exemple utiliser un microphone directif de type HOA⁶, permettant d'isoler l'énergie en provenance de différentes directions, de manière à définir une cible énergétique restreinte au plan horizontal. On peut également imaginer définir une QA "cible" à partir de la moyenne de plusieurs mesures monophoniques à des positions de référence disséminées dans la salle. Le principe général de la méthode de compensation que nous proposons demeure le même, quel que soit le dispositif microphonique envisagé.

Dans la suite de cette section, nous supposons par simplicité que le microphone placé à la position de référence est monophonique.

4.2.2 Définition d'une qualité acoustique "contexte"

La QA "contexte" caractérise la réponse du système de diffusion à la position de référence \vec{r}_{ref} . Sa définition donnée dans (Jot *et al.*, 1998) s'appuie sur l'organisation spatio-temporelle spécifique des signaux de sortie du processeur d'effet de salle utilisée dans le Spat~, qui est détaillée en section 1.4.3.

Comme le montre la figure 1.21, les 4 sections temporelles OD , $R1$, $R2$ et $R3$ sont réparties sur trois groupes de haut-parleurs : un groupe "central" affecté à la reproduction le son direct OD , un groupe "latéral" affecté aux réflexions primaires $R1$, et un groupe "diffus" affecté aux réflexions secondaires $R2$ et à la réverbération tardive $R3$. A chacun de ces groupes correspond un élément de la QA "contexte", qui se subdivise donc en un contexte "central", un contexte "latéral" et un contexte "diffus".

⁶Un microphone de ce type est décrit dans le chapitre suivant à la section 5.2

Il existe plusieurs possibilités pour la définition du contexte central, latéral, et diffus. On peut, par exemple, définir un contexte central et latéral qui varient au cours du temps, pour tenir compte des changements de position de la source virtuelle. Parmi les définitions de contexte données dans (Jot *et al.*, 1998), on compte une technique qui consiste à calculer le contexte central et latéral une fois pour toutes, en prenant comme référence une source virtuelle centrée du point de vue du dispositif électroacoustique. Il paraît judicieux, dans le cadre existant du système d'effet de salle associé à la WFS, de partir de cette définition pour définir les contextes central, latéral et diffus.

Suivant la configuration de la salle, et les moyens technologiques disponibles, il se peut que la disposition périphérique des haut-parleurs montrée en figure 1.23 ne soit pas envisageable. Nous proposons donc de développer ci-dessous deux définitions possibles pour la QA "contexte". La première définition s'applique à la situation où l'on dispose effectivement d'un système de diffusion périphérique, comme celui représenté à la figure 1.23, et la deuxième permet de tenir compte de l'éventuelle absence de haut-parleurs en périphérie de la salle.

Définition d'une qualité acoustique "contexte" pour un système périphérique

La première définition proposée pour la QA "contexte" s'applique à la situation où le dispositif de rendu d'effet de salle est installé sur toute la circonférence de la salle d'écoute, avec un système WFS principal installé en front de scène. Cette configuration est représentée à la figure 4.4.

Transcodage de la directivité de l'instrument réel pour la WFS

On doit déterminer dans un premier temps l'ensemble de coefficients $a_m(\alpha_S)$ permettant de recomposer le rayonnement de l'instrument cible par WFS dans le plan horizontal. En vertu des équations 3.2 et B.9, on peut écrire que :

$$Y^m(\phi) = \begin{cases} \frac{\sqrt{\pi}}{\bar{P}_m^{|m|}(0)} Y_m^m(\theta = \frac{\pi}{2}, \phi) & \text{pour } m > 0 \\ \frac{\sqrt{2\pi}}{\bar{P}_m^0} Y_m^m(\theta = \frac{\pi}{2}, \phi) & \text{pour } m = 0 \end{cases} \quad (4.3)$$

On en déduit que le spectre de Fourier circulaire $a_m(\alpha_S)$ de l'instrument simulé, restreint au plan horizontal, s'obtient de la manière suivante à partir des coefficients de son spectre sphérique $a_{lm}(\alpha_S)$:

$$a_m(\alpha_S, \omega) = \begin{cases} \frac{\sqrt{\pi}}{\bar{P}_m^{|m|}(0)} a_{lm}(\alpha_S, \omega) & \text{pour } m > 0 \\ \frac{\sqrt{2\pi}}{\bar{P}_m^0} a_{lm}(\alpha_S, \omega) & \text{pour } m = 0 \end{cases} \quad (4.4)$$

Ayant établi cette règle de conversion, on est donc en mesure de recomposer, par une somme de filtres WFS élémentaires, un instrument virtuel WFS $\Psi(\vec{r}_S, \alpha_S)$ qui approxime, dans le plan

horizontal, la directivité de l'instrument réel placé en \vec{r}_S et orienté suivant α_S .

Description de l'implémentation

Dans le but de simplifier les notations, on pose que le système entier (frontal + haut-parleurs périphériques) est constitué de N_W haut-parleurs.

Les haut-parleurs du système frontal sont utilisés pour synthétiser l'instrument virtuel $\Psi(\vec{r}_S, \alpha_S)$. Pour effectuer cette synthèse, on leur applique une somme pondérée de filtres WFS élémentaires, notés $\{\mathbf{h}_{i,\Psi^m}(\vec{r}_S, \omega) \mid i = 1, \dots, N_W\}$, qui réalisent à la fois la formation du front d'onde WFS associé à la source virtuelle Ψ^m de directivité élémentaire Y^m et l'égalisation du système de haut-parleurs. Etant donné leur position, les haut-parleurs de la périphérie ne participent pas à la synthèse de la source Ψ , et les filtres $\mathbf{h}_{i,\Psi^m}(\omega)$ pour i correspondant à ces haut-parleurs sont tous nuls.

Le système frontal réalise également la synthèse des haut-parleurs virtuels WFS $\{\Psi_{HP}^k \mid k = 1, 2, 3\}$ par le biais de filtres notés $\{\mathbf{h}_{i,\Psi_{HP}^k}(\omega) \mid i = 1, \dots, N_W, k = 1, \dots, 3\}$. Là encore, les filtres associés aux haut-parleurs périphériques sont nuls car inutilisés pour la synthèse de ces sources.

Le système périphérique permet de synthétiser les haut-parleurs virtuels $\{\Psi_{HP}^k \mid k = 4, \dots, 7\}$ par le biais de filtres notés $\{\mathbf{h}_{i,\Psi_{HP}^k}(\omega) \mid i = 1, \dots, N_W, k = 4, \dots, 7\}$, qui sont nuls pour les indices i correspondant aux haut-parleurs frontaux pour les mêmes raisons que citées précédemment.

Calcul du contexte central, latéral et diffus

La figure 4.4 représente la définition des trois composantes de la QA "contexte", dans le cas où on dispose d'un système de synthèse d'effet de salle périphérique. Etant donné cette configuration, la QA "contexte" peut se subdiviser en :

- Un contexte central, défini par la contribution de l'instrument virtuel $\Psi(\vec{r}_S, \alpha_S)$, mesuré à la position de référence. La réponse impulsionnelle associée à ce contexte se calcule de la manière suivante :

$$p_{central}(\vec{r}_S, \alpha_S, \vec{r}_{ref}, t) = \text{IFFT} \left(\frac{\sum_{i=1}^{N_W} \mathbf{c}_i(\vec{r}_{ref}, \omega) \left[\sum_{m=1}^N a_m(\alpha_S, \omega) \cdot \mathbf{h}_{i,\Psi^m}(\vec{r}_S, \omega) \right]}{R_d(\omega)} \right). \quad (4.5)$$

où $\{\mathbf{c}_i(\vec{r}_{ref}, \omega) \mid i = 1, \dots, N_W\}$ représente l'ensemble des réponses impulsionnelles de chaque haut-parleur i à la position de référence \vec{r}_{ref} .

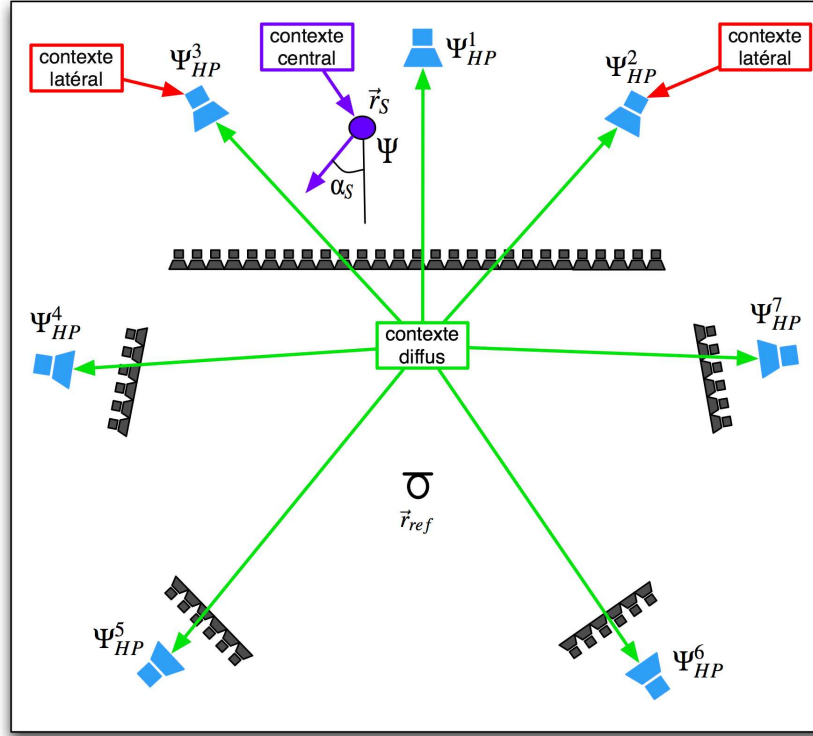


FIG. 4.4 – Définition des contextes "central", "latéral" et "diffus" pour une configuration de rendu d'effet de salle périphérique. On représente en bleu clair les haut-parleurs virtuels $\{\Psi_{HP}^i \mid i = 1, \dots, 7\}$ employés pour diffuser l'effet de salle. L'instrument virtuel $\Psi(\vec{r}_S, \alpha_S)$ est représentée par un point violet. La position de référence pour la qualité acoustique contexte \vec{r}_{ref} est représentée au centre de la figure.

La réponse $p_{central}$ est ensuite décomposée en sections temporelles pour effectuer le calcul de l'énergie $OD_{central}$, $R1_{central}$, $R2_{central}$ et $R3_{central}$ suivant les équations 1.32, 1.33, 1.34 et 1.35. On remarquera que d'après cette définition, le contexte central dépend à la fois de la position et de l'orientation de l'instrument virtuel.

- Un contexte latéral, défini comme la moyenne des mesures produites par les haut parleurs virtuels Ψ_{HP}^2 et Ψ_{HP}^3 situés à gauche et à droite du haut-parleur virtuel central. La réponse impulsionnelle associée à ce contexte se calcule de la manière suivante :

$$p_{lateral}(\vec{r}_{ref}, t) = \text{IFFT} \left(\frac{\sum_{i=1}^{N_W} \mathbf{c}_i(\vec{r}_{ref}, \omega) \left[\sum_{k=2}^3 \mathbf{h}_{i, \Psi_{HP}^k}^k(\omega) \right]}{2 \cdot R_d(\omega)} \right). \quad (4.6)$$

On dérive le contexte latéral $OD_{lateral}$, $R1_{lateral}$, $R2_{lateral}$ et $R3_{lateral}$ à partir de $p_{lateral}$ suivant les équations 1.32, 1.33, 1.34 et 1.35.

- Un contexte diffus, défini comme la moyenne des mesures de tous les haut-parleurs virtuels $\{\Psi_{HP}^k \mid k = 1, \dots, 7\}$, dont la réponse impulsionnelle associée se dérive par la formule suivante :

$$p_{diffus}(\vec{r}_{ref}, t) = \text{IFFT} \left(\frac{\sum_{i=1}^{N_W} \mathbf{c}_i(\vec{r}_{ref}, \omega) [\sum_{k=1}^7 \mathbf{h}_{i, \Psi_{HP}^k}(\omega)]}{7 \cdot R_d(\omega)} \right). \quad (4.7)$$

Le contexte diffus OD_{diffus} , $R1_{diffus}$, $R2_{diffus}$ et $R3_{diffus}$ est calculé à partir de p_{diffus} de manière analogue aux deux cas précédents.

Définition d'une qualité acoustique "contexte" pour un système frontal

Dans le cas où la configuration de haut-parleurs ne couvre pas la partie arrière de la conférence de salle, la méthode de synthèse d'effet de salle décrite au chapitre en section 1.4.3 ne peut pas s'appliquer convenablement. Il devient alors nécessaire d'introduire une méthode permettant de synthétiser un champ diffus en utilisant uniquement le banc frontal.

Pour réaliser cette opération, on propose d'utiliser des haut-parleurs virtuels *directifs* supplémentaires pour synthétiser les réflexions tardives $R2$ et la réverbération tardive $R3$. Dans le cadre de la méthode proposée ici, nous proposons d'employer des Diracs Directifs d'ordre élevé (cf. sec. 3.2.1), orientés de telle manière à diriger l'énergie tardive $R2$ et $R3$ synthétisée par le processeur d'effet de salle vers les murs de la salle d'écoute (cf. Fig. 4.5). Ce choix particulier d'orientation des sources permet d'utiliser les murs de la salle comme sources de réverbération supplémentaires, sans rajouter au niveau du son direct.

Description de l'implémentation

Les Diracs Directifs utilisés pour diffuser l'effet de salle sont placés en \vec{r}_D , à l'extrémité d'un triangle isocèle dont la base est confondue avec le banc de haut-parleurs. L'angle de déviation de chaque Dirac avec la médiatrice au banc vaut α_D . Le choix d'une position centrée et éloignée pour les Diracs Directifs est fait de manière à assurer une fréquence de repliement spatial élevée pour la synthèse des fronts d'ondes associés à ces sources (cf. sec. 1.3.8), ainsi qu'un ordre maximal de reproduction de directivité élevé (cf. sec. 3.5.2).

La synthèse des Diracs Directifs se fait en combinant les filtres WFS de directivité élémentaire à la position \vec{r}_D , sachant que le spectre de Fourier circulaire a_m^δ du Dirac est donné en section 3.2.1, relation 3.11.

Les haut-parleurs virtuels Ψ_{HP}^1 , Ψ_{HP}^2 , et Ψ_{HP}^3 , ainsi que la source virtuelle Ψ_i ne sont pas

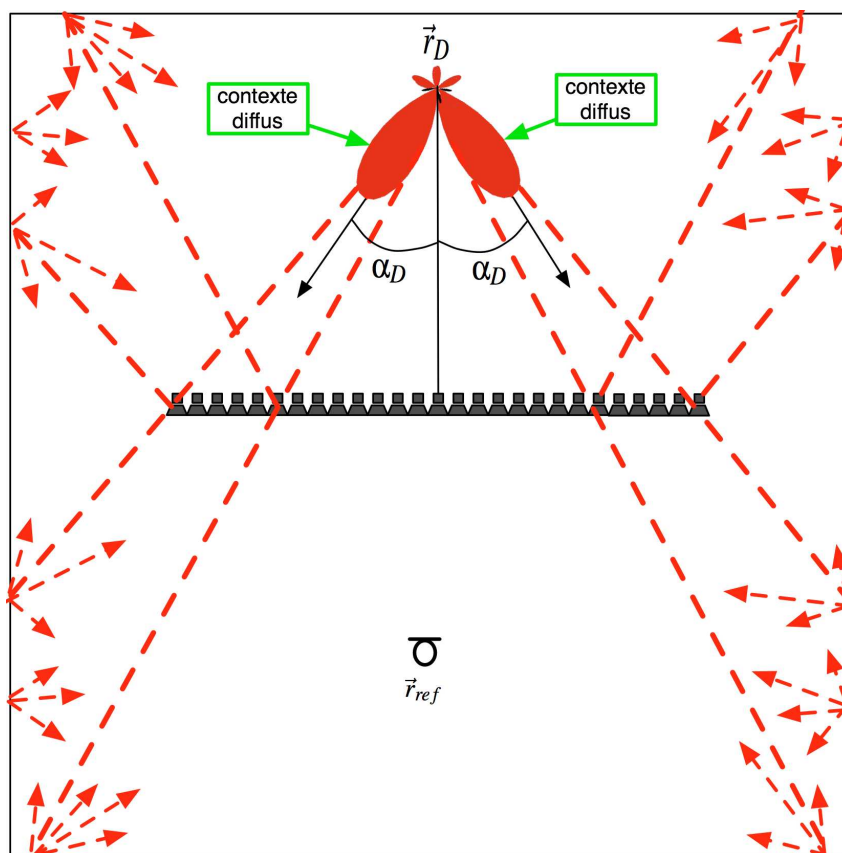


FIG. 4.5 – Définition du contexte "diffus" pour un système WFS strictement frontal. Dans cette configuration, deux Diracs Directifs d'ordre 5 sont utilisés pour diffuser l'effet de salle tardif vers les murs latéraux de la salle.

représentés sur la figure 4.5, pour des raisons de lisibilité de la figure. Ces sources demeurent cependant actives dans cette implémentation, et les définitions du contexte "central" et "latéral" sont identiques à celles données dans le cas précédent.

L'énergie tardive $R2$ et $R3$ est diffusée suivant les 3 haut parleurs virtuels disponibles $\{\Psi_{HP}^k \mid k = 1, 2, 3\}$, ainsi que les deux Diracs Directifs d'ordre 5 montrés en figure 4.5. Le choix d'un Dirac Directif d'ordre 5 a été fait car il fournit un faisceau suffisamment concentré pour minimiser le son direct envoyé vers la zone d'écoute, tout en couvrant une zone suffisamment large des murs latéraux de la salle pour maximiser la zone qui rayonne l'effet de salle tardif par réverbération (cf. Fig. 4.5).

Calcul du contexte diffus

Pour cette configuration exclusivement frontale, on définit le contexte diffus comme la moyenne des mesures des trois haut-parleurs virtuels disponibles $\{\Psi_{HP}^k \mid k = 1, \dots, 3\}$ et des deux Diracs Directifs supplémentaires.

La réponse impulsionnelle p_{diffus} associée à ce contexte est obtenue de la manière suivante :

$$p_{diffus}(\vec{r}_{ref}, t) = \text{IFFT} \left(\frac{\sum_{i=1}^{N_W} \mathbf{c}_i(\vec{r}_{ref}, \omega) \left[\sum_{m=1}^5 [a_m^\delta(\alpha_S, \omega) + a_m^\delta(-\alpha_S, \omega)] \cdot \mathbf{h}_{i, \Psi^m}(\vec{r}_D, \omega) + \sum_{k=1}^3 \mathbf{h}_{i, \Psi_{HP}^k}(\omega) \right]}{5 \cdot R_d(\omega)} \right) \quad (4.8)$$

La réponse p_{diffus} est ensuite décomposée en sections temporelles pour effectuer le calcul de l'énergie OD_{diffus} , $R1_{diffus}$, $R2_{diffus}$ et $R3_{diffus}$ suivant les équations 1.32, 1.33, 1.34 et 1.35. On remarquera au passage que l'utilisation d'un système périphérique, comme celui présenté à la section précédente, ne réduit pas a priori l'intérêt d'utiliser des sources directives pour rajouter de l'énergie diffuse supplémentaire dans la salle. En effet, la multiplication des sources de réverbération permet de réduire l'amplification individuelle des canaux de réverbération. Ceci a pour effet de réduire les chances que l'on perçoive des colorations dans la zone d'écoute dues au système de réverbération artificielle (Kuttruff, 1991).

Evaluation subjective de la méthode de diffusion directive de l'effet de salle

Aucune évaluation formelle de cette méthode de diffusion d'effet de salle n'a pu être faite à ce jour. Cependant, trois sessions de travail conduites courant 2006-2007 avec les réalisateurs en informatique musicale de l'Ircam ont permis de constater l'intérêt de cette méthode de diffusion d'effet de salle d'un point de vue subjectif. Ces sessions ont notamment révélé la possibilité de contrôler l'éloignement subjectif de la source virtuelle en jouant sur la directivité de sources virtuelles diffusées via un système WFS uniquement frontal. Ces résultats subjectifs restent cependant à être confirmés par des tests perceptifs dans un cadre formel.

4.2.3 Bilan énergétique et calcul de compensation

Ayant défini les qualité acoustique "cible", "live" et "contexte", le bilan énergétique permettant de déduire les composantes OD , $R1$, $R2$ et $R3$ à fournir par le processeur d'effet de salle de manière à assurer une compensation de l'effet de salle peut s'exprimer sous la forme matricielle manière suivante :

$$Q_k^c = Q_k^l + \sum_{j=0}^3 T_{jk}^c \cdot Q_j^s. \quad (4.9)$$

avec :

$$Q^c = \begin{bmatrix} OD_{cible} \\ R1_{cible} \\ R2_{cible} \\ R3_{cible} \end{bmatrix}, \quad Q^l = \begin{bmatrix} OD_{live} \\ R1_{live} \\ R2_{live} \\ R3_{live} \end{bmatrix}, \quad \text{et} \quad Q^s = \begin{bmatrix} OD \\ R1 \\ R2 \\ R3 \end{bmatrix}.$$

La matrice T_{jk}^c est une matrice qui décrit les *transferts d'énergie* réalisés entre la section j du processeur d'effet de salle Spat~et la section k effectivement reproduite dans la salle. Cette matrice se calcule en remarquant que les transferts d'énergie dans la salle sont réalisés par la convolution des enveloppes énergétiques de la qualité acoustique contexte et la sortie Q^s du Spat~. Cette opération de convolution énergétique est illustrée pour un choix particulier de bornes temporelles à la figure D.1 de l'annexe D.

La matrice T_{jk}^c est une matrice triangulaire inférieure (les coefficients $\{T_{jk}^c, j > k\}$ sont nuls par principe de causalité). L'inversion de T_{jk}^c permet d'obtenir les valeurs Q^s que doit fournir le Spat~de manière à réaliser la qualité acoustique cible Q^c à la position de référence. Ces valeurs, qui représentent des composantes énergétiques, doivent être strictement positives (il est en effet impossible "d'enlever" de l'énergie dans la salle par l'opération de convolution énergétique entre les signaux produits par le Spat~et le système de diffusion).

L'inversion de la matrice T_{jk}^c est décrite en Annexe D. L'expression des composantes énergétiques que doit fournir le Spat~pour réaliser la compensation dans le cas décrit en Annexe (eq. D.1) est identique à celle donnée dans la formulation originale du brevet du Spatialisateur (Jot *et al.*, 1998), à l'exception près que dans une situation de rendu WFS, le contexte central dépend de la position \vec{r}_S et orientation α_S de la source virtuelle Ψ . En effet, dans le cas général, la première colonne de T_{jk}^c correspond simplement à la réponse impulsionnelle de la source virtuelle, pondérée par le gain OD fourni par le Spat.

4.2.4 Conclusion

Dans cette section, nous avons introduit une méthode de compensation pour la WFS dont l'objectif est de reproduire une qualité acoustique cible dans la salle de reproduction.

A partir d'une qualité acoustique "contexte", qui décrit la configuration employée pour diffuser l'effet de salle, et d'une qualité acoustique "live" qui décrit la contribution éventuelle d'instrumentistes réels sur scène, la méthode permet de calculer l'enveloppe énergétique des signaux à fournir au processeur d'effet de salle de manière à reproduire une qualité acoustique "cible" de qualité acoustique à une position de référence dans la salle.

Nous avons proposé deux manières pour définir la QA "cible" :

- La première, dite perceptive, fait intervenir l'opérateur perceptif du Spatialisateur, et per-

met au créateur de contenu de fixer une QA "cible" en réglant directement les 9 paramètres perceptifs du Spatialisateur.

- La seconde, dite physique, tente de reproduire la qualité acoustique naturellement associée à un instrument réel qui serait placé sur scène. Le calcul de cette QA "cible" physique se fait en deux étapes. Dans la première phase, on analyse de rayonnement de l'instrument réel que l'on souhaite simuler via le système WFS. Dans la deuxième phase, on emploie une source à directivité contrôlée pour recomposer le rayonnement de l'instrument simulé dans la salle, et ainsi déterminer par mesure la qualité acoustique associée naturellement à celui-ci lorsqu'il occupe une position/orientation donnée sur scène. Cette qualité acoustique représente alors la QA "cible" physique. On peut par cette méthode reproduire l'ensemble des qualités acoustiques associées aux positions de la base de données de sources virtuelles montrée en 1.28.

La méthode proposée permet idéalement d'ajuster l'acoustique de la salle de reproduction dans les 4 sections temporelles de réponse impulsionnelle, à condition que l'acoustique cible soit plus réverbérante que les QA "live" et "contexte".

Nous avons décrit par ailleurs une méthode consistant à diffuser l'énergie tardive produite par le processeur d'effet de salle via des sources virtuelles directives. Ces sources permettent d'envoyer un faisceau d'énergie tardive vers les murs de la pièce, de manière à alimenter le champ diffus tout en minimisant l'énergie du son direct envoyé vers la zone d'écoute.

La section suivante introduit une méthode simplifiée de compensation de l'effet de salle tardif, basée cette fois sur une caractérisation de la puissance rayonnée par le système.

4.3 Elaboration d'une méthode de compensation de l'effet de salle tardif simplifiée

4.3.1 Introduction

Dans cette section, nous proposons une méthode de compensation simplifiée, qui vise non plus à ajuster l'acoustique de la salle suivant une cible précise définie à une position d'écoute référence, mais uniquement à assurer une continuité perceptive de la réverbération tardive associée aux sources virtuelles sur scène. Pour pouvoir assurer cette continuité perceptive, l'idée est de compenser les variations de puissance émise par le banc WFS lors de déplacements et rotations de la source virtuelle. En effet, en vertu de l'équation A.8 donnée en Annexe A.2, la densité d'énergie en régime établie E_∞ (liée au niveau perçu de réverbération) dépend de la puissance W émise dans la salle par le système.

Le principe de la méthode "simplifiée" consiste donc à caractériser les variations de la puissance émise par le banc WFS lors de changements de position/orientation de la source virtuelle synthétisée, de manière à pouvoir ensuite compenser ceux-ci en réinjectant de l'énergie diffuse dans la salle par le biais du système de rendu d'effet de salle. En Annexe A.3, nous décrivons un ensemble de techniques permettant de déterminer la puissance émise par une source sonore. Dans les sections suivantes, nous discutons donc des avantages/inconvénients des méthodes existantes pour l'estimation de la puissance émise par un banc WFS lors de la synthèse d'une source virtuelle, de manière à déterminer la méthode optimale pour la caractérisation de la puissance émise par le banc WFS.

Des résultats de l'estimation de la puissance émise par un banc WFS sont donnés et analysés pour 3 de ces méthodes (estimation par simulation sur un banc de haut-parleurs monopolaires, estimation in situ par relief de décroissance et estimation par prédiction de rayonnement). On décrit enfin le principe général de la méthode de compensation, en montrant comment inscrire celle-ci dans l'architecture existante du système WFS.

4.3.2 Mesure de puissance en chambre réverbérante

Description générale

On trouve en annexe A.3.1 une description de la méthode d'estimation de puissance d'une source sonore en chambre réverbérante telle qu'elle est décrite dans la norme ISO 3741. Cette méthode est bien adaptée à l'estimation de puissance des sources électroacoustiques comme des haut-parleurs.

La puissance individuelle W_i d'un haut-parleur est obtenue en moyennant un ensemble de mesures de la pression quadratique moyenne dans la salle. La formule exacte permettant de dériver la puissance du haut-parleur est donnée en équation A.17.

Discussion : application de la méthode à l'estimation de puissance du banc WFS

Les dimensions typiques d'une chambre réverbérante sont généralement trop faibles pour y faire rentrer des bancs WFS, dont la longueur peut atteindre plusieurs dizaines de mètres. On est donc a priori contraint de mesurer les haut-parleurs séparément, ou par petits groupes.

Si on suppose que les ondes émises par les haut-parleurs du banc WFS sont *incohérentes* (cf. glossaire), on peut calculer la puissance totale W_Ψ émise par le banc WFS pour la synthèse de la source $\Psi(\vec{r}_S, \alpha_S)$ par une simple sommation des puissances individuelles W_i de ses haut-parleurs constitutifs :

$$W_\Psi = \sum_{i=1}^{i=N} W_i. \quad (4.10)$$

Cette formule de sommation de puissance est généralement acceptée lorsqu'on considère des ensembles instrumentaux en salle (Cremer *et al.*, 1978). La formule 4.10 est également valable, a priori, pour estimer la puissance émise par un banc WFS au-dessus de la fréquence de repliement, étant donné que le champ devient incohérent au delà de cette fréquence (cf. section 1.3.8). En effet, en HF la distance entre les haut-parleurs est supérieure à plusieurs longueurs d'ondes et les contributions individuelles de chaque haut-parleurs sont incohérentes. Par contre, en-dessous de la fréquence de repliement spatial, les ondes émises par les haut-parleurs individuels du banc sont cohérentes et la méthode n'est plus applicable.

4.3.3 Mesure de puissance par Intensimétrie

Description générale

Les principes de la mesure de puissance par intensimétrie sont décrits en Annexe A.3.3. Les méthodes dites intensimétriques sont souvent utilisées pour des mesures en milieu industriel, car elles permettent de dériver la puissance d'une source en champ proche, ce qui permet d'obtenir des rapports signal sur bruit acceptables dans des cas où la source qui est mesurée (machine lourde par exemple) est placée dans un environnement bruyant, où les sources environnantes ne peuvent pas être éteintes pendant la mesure.

Un inconvénient de la méthode réside dans le fait qu'il faille répéter les mesures intensimétriques avec des sondes d'écartement inter-capsule variables pour pouvoir estimer le spectre de puissance en entier. Cette contrainte est due à un problème de biais lié aux erreurs d'estimation de la phase en BF, et des erreurs de diffusion/diffraction/approximation de différences finies en HF.

Discussion : application de la méthode à l'estimation de puissance du banc WFS

Dans le cas qui nous intéresse, l'amélioration du rapport signal sur bruit rendu possible par l'utilisation de méthodes intensimétriques n'est pas un argument fort pour l'utilisation de la méthode, étant donné que l'on caractérise des sources électroacoustiques pour lesquelles on possède déjà des techniques (MLS, sweep) permettant d'obtenir des rapports signal sur bruit élevés.

Par ailleurs, en vertu de la section précédente, une estimation de la puissance W_{Ψ} rayonnée par le banc WFS par sommation des mesures de puissance individuelle W_i des haut-parleurs en chambre sourde obtenues par méthode intensimétrique n'est pas valide en-dessous de la fréquence de repliement spatiale.

On peut par contre envisager de conduire une caractérisation intensimétrique in situ du banc de haut-parleurs installé dans la salle de concert ; il faut pour cela mesurer l'intensité

sortante sur toutes les positions d'une surface échantillonnée entourant le banc de haut-parleurs. A chaque position, on fixe la sonde intensimétrique et on enregistre la réponse impulsionnelle individuelle de tous les haut-parleurs constitutifs du banc sur les deux capsules de la sonde. Il est alors possible, par linéarité, de reconstituer a posteriori la réponse de n'importe quelle source virtuelle $\Psi(\vec{r}_S, \alpha_S)$, et estimer sa puissance associée W_Ψ par intégration de l'intensité sortante, suivant la relation A.42.

Un inconvénient de cette méthode réside cependant dans le fait qu'il faille répéter la mesure si on souhaite changer les transducteurs, ou encore la répartition de ceux-ci (en changeant l'écartement inter haut-parleur, par exemple). Etant donné la difficulté de manipulation implicite à une mesure sur une surface fermée entourant entièrement le banc, le fait de devoir potentiellement répéter ces mesures peut être rédhibitoire à l'utilisation de la méthode intensimétrique pour la caractérisation de la puissance du banc. Pour cette raison, nous n'avons pas opté pour l'utilisation de cette méthode dans le cadre de ce travail.

4.3.4 Méthode d'estimation par simulation d'un banc de haut-parleurs monopolaires

Dans la méthode décrite dans cette section, on simule la mesure de puissance d'un banc WFS composé de haut-parleurs monopolaires sur un réseau de microphones omnidirectionnels placés en champ lointain.

Cette simulation réalisée dans des conditions idéales n'a pas pour objectif de fournir une estimation exploitable de la puissance émise par le banc de haut-parleurs réel. Elle est conçue plutôt pour cerner les mécanismes qui sous-tendent les variations de puissance émise par le banc WFS.

La configuration de simulation proposée comprend un banc WFS linéaire constitué de 48 haut-parleurs (espacés de 16.5cm) que l'on mesure sur un réseau de 128 microphones omnidirectionnels. Les microphones sont positionnés suivant un cercle de rayon 10m dans le plan horizontal contenant le banc de haut-parleurs (fig. 4.6). Pour des raisons de symétrie de révolution du champ émis par le banc de monopoles, chaque microphone est représentatif d'une portion d'une sphère 3D entourant le banc de haut-parleurs. En pondérant chaque microphone par l'angle solide qu'il représente, on peut alors estimer l'intégrale de la pression quadratique sur la sphère, ce qui permet de déterminer la puissance émise par le banc en vertu des équations A.42 et A.47 données en Annexe A.

La configuration représentée en figure 4.6 peut être divisée en deux zones. La première zone, délimitée par des microphones verts (clairs), comprend la ligne de microphones qui a été utilisée pour informer la procédure d'égalisation multicanale. Cette portion sera désignée comme étant "dans l'angle solide visible", par référence au fait qu'elle comprend la fenêtre de visibilité du

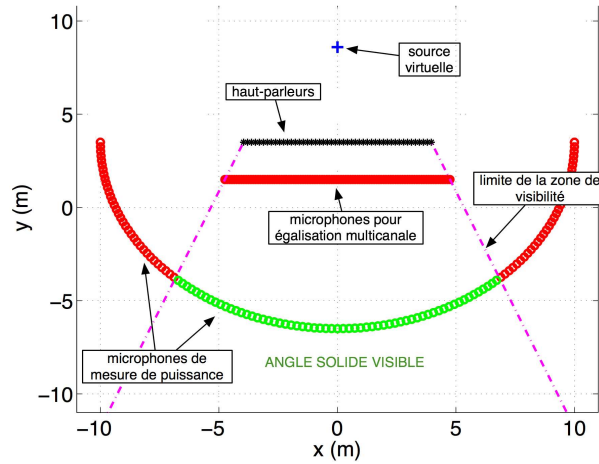


FIG. 4.6 – Configuration employée pour mesurer la puissance émise par le banc de haut-parleurs.

banc de haut-parleurs (la notion de *visibilité* est introduite en section 1.3.7). La seconde zone, délimitée par des microphones rouges (foncés), est située en dehors de la zone optimisée par la procédure de filtrage multicanale. Cette zone sera décrite comme étant "en dehors de l'angle solide visible".

Une comparaison de la puissance émise par le banc de haut-parleurs pour la synthèse de 5 sources virtuelles de directivité élémentaire est représentée sur la figure 4.7. La source virtuelle est synthétisée à la position (0,8.5), ce qui correspond à une source centrée située à 5m derrière le banc de haut-parleurs. Cette position correspond également à une des positions analysées lors de l'évaluation de l'erreur de reconstruction du son direct (section 3.6, figure 3.11).

Le graphique de gauche de la figure 4.7 représente la puissance totale émise par le banc WFS. Le graphique central représente la puissance émise dans l'angle solide visible. Le graphique de droite représente la puissance émise en dehors de l'angle solide visible.

Il est intéressant de noter tout d'abord qu'en-dessous de la fréquence de repliement de spectre spatial il est possible, dans une certaine mesure, de contrôler la directivité du banc de manière à concentrer l'énergie à l'intérieur de l'angle solide visible. En effet, en regardant la figure 4.7, (centre), on voit que la puissance émise par le banc WFS dans l'angle solide visible pour la synthèse d'un monopole, d'un dipole axial, ou d'un quadrupole axial est approximativement la même. Par contre, en dehors de l'angle solide visible, la puissance émise pour la synthèse du dipole axial (resp. du quadrupole axial) chute de 3dB (resp. de 10dB). Dans le contexte de diffusion en salle, on peut donc s'attendre à ce que le banc de haut-parleurs produise un rapport variable de son direct/son réverbéré suivant la directivité de la source qui est synthétisée.

On constate cependant, en regardant la puissance totale émise par le banc (Fig. 4.7, gauche),

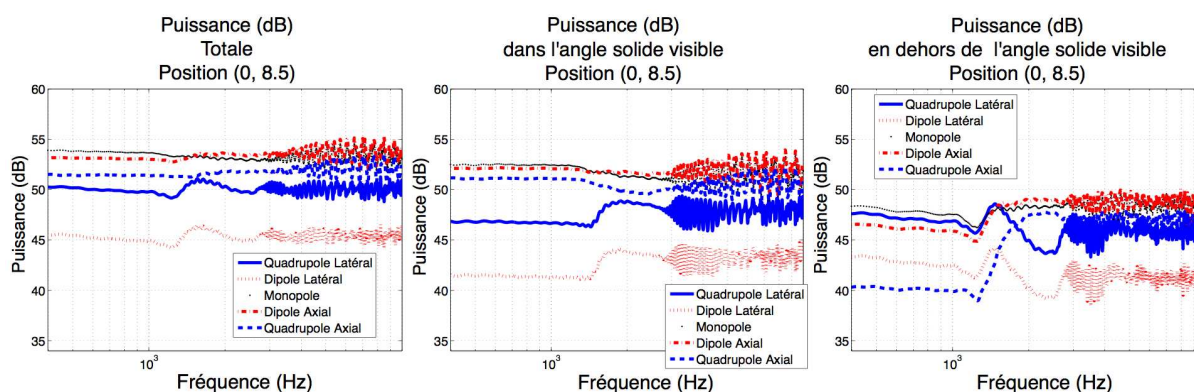


FIG. 4.7 – Puissance émise par le banc WFS de haut-parleurs monopolaires pour la synthèse de sources de directivité élémentaire situées 5m derrière le banc de haut-parleurs à une position centrée. **Droite** : Puissance émise en dehors de l'angle solide visible. **Centre** : Puissance émise dans l'angle solide visible. **Gauche** : Puissance Totale.

que le banc émet moins de puissance pour la synthèse d'un dipole latéral ou d'un quadrupole que pour la synthèse d'un dipole axial ou d'un quadrupole axial. Or, les filtres utilisés pour la synthèse des sources virtuelles sont normalisés de manière à assurer que ces sources produisent un son direct de même niveau $|p_{ax}|^2$ suivant leur axe principal de rayonnement. Etant donnée cette normalisation, si le dipole sphérique était parfaitement reproduit par le banc de haut-parleurs, il exhiberait une différence de puissance constante de 4.8dB par rapport au monopole (en effet l'indice de directivité idéal d'un dipole vaut $DI_{dipole} \approx 4.8dB$, cf. glossaire équation E.5). La source virtuelle dipolaire (axiale *et* latérale) devrait donc idéalement exhiber une différence de niveau de 4.8dB lorsqu'on compare sa puissance totale à celle de la source virtuelle monopolaire. Au lieu de ça, on observe une différence de niveau de 1dB entre le dipole axial et le monopole, et 8dB entre le dipole latéral et le monopole. En d'autres termes, *suivant les directivités particulières que l'on cherche à reproduire, le fenêtrage et la symétrie de rotation imposés par le banc affectent plus ou moins la quantité de puissance émise par la source virtuelle.*

Cette observation contredit les propriétés de sources sonores réelles, dont la puissance est une quantité intrinsèque, par définition indépendante de l'orientation. Ici, on voit que la puissance mesurée pour la synthèse d'une source virtuelle directive varie suivant que la source présente un lobe un zéro dans l'angle solide visible. Plus précisément, la quantité de puissance émise par le banc de haut-parleurs dépend de l'étendue des zones d'annulation le long de la ligne de microphones de contrôle⁷ (représentée à la figure 4.7). La variation de puissance observée est particulièrement remarquable pour la source dipolaire, qui exhibe une chute de puissance de 8dB entre l'orientation axiale et l'orientation latérale.

⁷Cette ligne de contrôle est utilisée pour caractériser la composante de son direct du banc dans le calcul des filtres WFS par processus d'égalisation multicanale (sec. 1.3.9)

De manière générale, ces résultats montrent que l'orientation de la source directive influe sur la quantité de puissance effectivement émise par le banc WFS dans la salle. Ceci confirme la nécessité de compenser la puissance émise dans la salle au moment de la reproduction.

Dans les deux sections suivantes, nous présentons des méthodes qui ont été utilisées en pratique pour estimer la puissance émise par le banc WFS.

4.3.5 Méthode d'estimation de puissance par prédiction de rayonnement

Description générale

Dans cette section nous décrivons une méthode de caractérisation de la puissance émise par le banc de haut-parleurs utilisant la méthode de prédiction de rayonnement décrite au chapitre 2. L'idée de la méthode est de prévoir le champ rayonné par le banc WFS sur une sphère entourant celui-ci (figure 4.8), par propagation du rayonnement de ses haut-parleurs constitutifs.

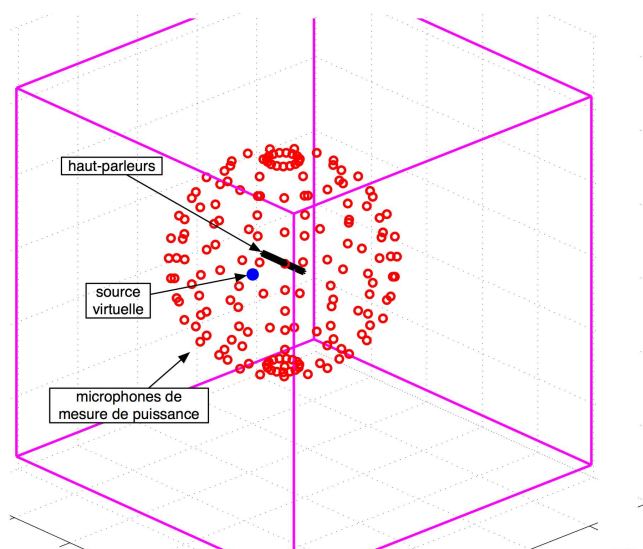


FIG. 4.8 – Configuration pour l'estimation de la puissance d'un banc WFS de haut-parleurs électrodynamiques KEF par méthode de prédiction de rayonnement. La réponse du banc de haut-parleurs de longueur 7.05m est propagée sur une sphère de microphones de rayon 14.1m. La source virtuelle est représentée par un point bleu, les haut-parleurs sont représentés par des astérisques noirs, et les microphones sont représentés par des cercles rouges.

Le banc WFS dont nous souhaitons caractériser le rayonnement est composé de haut-parleurs électrodynamiques de marque KEF. Ces haut-parleurs sont fabriqués par procédé industriel, et leurs caractéristiques de rayonnement sont très similaires d'un haut-parleur à l'autre. Cette similitude a été observée lors de mesures de rayonnement faites sur l'ensemble des 48 haut-parleurs en champ libre (fig. 2.10). Deux réponses fréquentielles de haut-parleurs KEF pris parmi les 48 disponibles à sont représentées à la figure 4.8.

Etant donné la similarité entre les haut-parleurs disponibles, nous supposons que la caractérisation d'un seul de ces haut-parleurs permet de donner une estimation acceptable du champ rayonné par les 47 autres haut-parleurs du banc.

Dans le chapitre 2, nous avons décrit la mesure d'un de ces haut-parleurs placé en configuration de banc, pour tenir compte des interférences dues à la présence de haut-parleurs adjacents (figure 2.14). Nous proposons d'utiliser la mesure de caractérisation faite avec haut-parleur placé en configuration de banc comme base pour la prédiction du rayonnement du banc complet.

La décomposition en harmoniques sphériques de ce haut-parleur est utilisée pour propager la réponse individuelle de 48 haut-parleurs placés en ligne, avec un espacement de 15cm entre chaque, sur une sphère de rayon 14.1m. La sphère est échantillonnée régulièrement par les microphones (15 positions d'azimut et 11 positions d'élévation).

En observant que le banc de haut-parleurs a une envergure de 7.2 mètres, il peut être inclus dans une sphère de rayon $R = 3.525\text{m}$. Le rayon de 14.1m est choisi pour la sphère de mesure car il représente un rayon de $4R$, ce qui d'après (BrielKjaer, 1984) est une distance acceptable pour considérer que l'on est en champ lointain de la source étendue, et donc pouvoir estimer l'intensité à partir de la pression quadratique suivant l'équation A.57.

Etant donné l'étude conduite en section 2.4.2, la prédiction du rayonnement d'un haut-parleur KEF produit des résultats très similaires à des mesures faites en champ libre jusqu'à 1.5kHz. On pourra donc a priori reconstituer correctement le champ interférentiel produit par le banc de haut-parleurs au niveau de la sphère d'estimation lors de la synthèse du champ d'ondes WFS en-dessous de la fréquence de repliement spatial, qui vaut au maximum 1.5kHz pour cette configuration. Au-delà de la fréquence de repliement, la puissance du champ émis par le banc sera estimée par sommation des puissances individuelles de ses haut-parleurs constitutifs.

Résultats

On estime la puissance émise par le banc de KEF pour la synthèse de 5 sources de directivité élémentaire positionnées en (0,8.5) sur la figure 3.11, c'est à dire à 5m derrière le banc de haut-parleurs à une position centrée. Les résultats de cette estimation sont donnés à la figure 4.9

En comparant la figure 4.9 avec la figure 4.7 (graphique de gauche), on voit que les puissances totales associées aux sources de directivité élémentaire présentent des différences de niveau similaires à celles observées dans la simulation du banc idéal. Ainsi, par exemple, on retrouve une différence d'environ 8dB entre la puissance associée au dipole latéral et celle associée au dipole axial ; de même, on retrouve une différence d'environ 4dB entre la puissance associée au monopole et celle associée au quadrupole latéral. Par ailleurs, de la même manière que dans les simulations, on observe que l'orientation de la figure de directivité affecte forte-

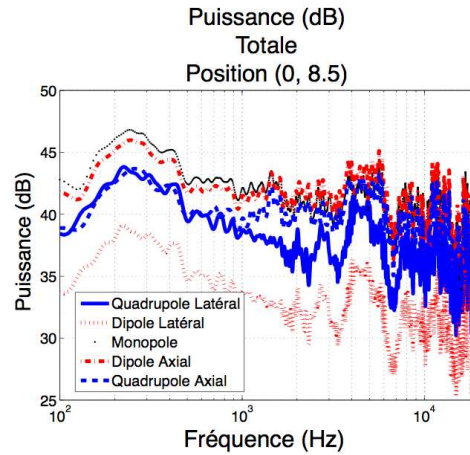


FIG. 4.9 – Estimation (utilisant la méthode de prédiction) de la puissance émise par un banc WFS de haut-parleurs MAP pour la synthèse d'un ensemble de sources virtuelles de directivité élémentaire $\{Y^m\}_{m \in \{-2, -1, 0, 1, 2\}}$ situées à 5m derrière le banc de haut-parleurs à une position centrée.

ment la puissance émise par le banc de haut-parleurs.

Par contre, le spectre de puissance estimé pour le banc de KEF présente des variations en fonction de la fréquence qui n'apparaissent pas dans le cas du système idéal. La raison des variations du spectre de puissance observées tient aux caractéristiques de rayonnement 3D des haut-parleurs utilisés. En effet, même si le champ direct émis par le banc de KEF a été égalisé de manière à présenter un spectre plat à l'intérieur de la fenêtre de visibilité dans le plan horizontal par le processus d'égalisation multicanale (sec. 1.3.9), celui-ci émet un champ sonore 3D dont on ne peut pas contrôler simultanément le spectre de puissance. Ces variations observées dans le spectre de puissance du banc de KEF sont attribuables aux propriétés non omnidirectionnelles des haut-parleurs utilisés dans les mesures. Ainsi, la décroissance en HF observée dans le spectre de puissance des sources virtuelles reproduites sur le banc de KEF tient au fait que ces haut-parleurs tendent à devenir très directifs à ces fréquences.

4.3.6 Méthode d'estimation de puissance in situ utilisant le relief de décroissance

Description générale

Il se peut dans certaines situations que l'on ne puisse pas conduire les mesures sphériques en champ libre des transducteurs d'un banc dont on souhaiterait caractériser la puissance (pas de chambre anéchoïque à disposition par exemple). Il peut également arriver que les différences inter individuelles dans les caractéristiques de rayonnement d'un ensemble de haut-parleurs

4.3. Elaboration d'une méthode de compensation de l'effet de salle tardif simplifiée

soient telles qu'une mesure sphérique en chambre sourde de *tous* les haut-parleurs soit indispensable à l'estimation de la puissance rayonnée par le banc, et que cette opération représente un temps de manipulation trop important vis à vis des délais à respecter. Dans ce cas, la méthode d'estimation de puissance par prédiction de rayonnement ne peut pas être utilisée.

Nous proposons donc méthode alternative d'estimation *in situ* de la puissance du banc WFS, qui s'appuie sur l'analyse du relief de décroissance énergétique de la réverbération (la méthode est décrite en détail en Annexe A.3.2). Cette méthode est attractive du point de vue de la caractérisation de la puissance du banc WFS car elle permet, à partir d'une seule mesure *in situ* de chacun des haut-parleurs, d'estimer la puissance émise par le banc WFS entier pour la synthèse de n'importe quelle source virtuelle.



FIG. 4.10 – Campagne de mesures de haut-parleurs MAP (panneaux blancs) conduite dans l'Espace de Projection en avril 2005. A droite de la photographie, on peut voir des microphones Schoeps employés pour mesurer la réponse du banc de haut-parleurs.

On s'appuie pour cela sur le fait que les systèmes étudiés en acoustique des salles peuvent généralement être considérés comme étant linéaires et invariants dans le temps (LIT). Il faut malgré tout s'assurer que les conditions de température, pression et humidité ne varient pas pendant la mesure de tous les haut-parleurs constitutifs du banc.

En respectant ces conditions, on peut alors reconstituer la réponse impulsionnelle de n'importe quelle source virtuelle à une position donnée dans la salle *a posteriori*, à partir d'une seule campagne de mesure de tous les haut-parleurs individuels du banc. Ceci implique en outre la possibilité de caractériser *a posteriori* la puissance de toutes les sources virtuelles de la base de données WFS (fig. 1.28) à partir de la seule campagne de mesure en salle, en utilisant la

méthode d'intégration rétrograde du relief de décroissance.

Résultats

Nous avons réalisé une campagne de mesures dans l'Espace de Projection⁸ mettant en jeu un système de 48 haut-parleurs de type MAP (Multi Actuator Panel) dans le but d'estimer la puissance émise par le banc WFS par la méthode d'intégration rétrograde du relief de décroissance. Nous mesurons la réponse impulsionnelle d'un ensemble de sources de directivité élémentaire $\{Y^m\}_{m \in \{-2, -1, 0, 1, 2\}}$ sur un microphone Schoeps situé dans la salle (cf. Fig. 4.10). Cette mesure est ensuite utilisée pour estimer la puissance émise par le banc de haut-parleurs pour la synthèse de chacune des sources de directivité élémentaire.

Nous estimons la puissance émise pour une position de source virtuelle située en (0, 8.5) sur la figure 3.11, soit à 5m derrière le banc de haut-parleurs à une position centrée. Les résultats de l'estimation de puissance sont donnés à la figure 4.11.

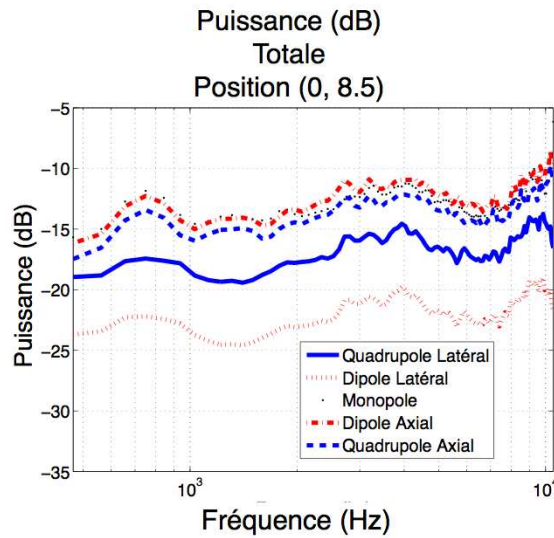


FIG. 4.11 – Estimation (utilisant le relief de décroissance) de la puissance émise par un banc WFS de haut-parleurs MAP pour la synthèse d'un ensemble de sources virtuelles de directivité élémentaire $\{Y^m\}_{m \in \{-2, -1, 0, 1, 2\}}$ situées à 5m derrière le banc de haut-parleurs à une position centrée.

En comparant la figure 4.11 avec la figure 4.7 (graphique de gauche) et la figure 4.9, on voit que les puissances associées aux sources de directivité élémentaire synthétisées par le banc de MAP présentent des différences de niveau similaires aux deux cas précédents. On trouve par exemple un écart de niveau de d'environ 10dB entre le dipole latéral et le dipole axial

⁸L'Espace de Projection, qui est la salle de concert utilisée pour les mesures, possède une acoustique réglable, permettant notamment de contrôler son temps de réverbération.

(cette différence valait 8dB dans les cas précédents) ; on trouve une différence de 5dB entre le quadropole latéral et le monopole (cette différence valait 4dB dans le cas précédents).

Par contre, le spectre de puissance des sources virtuelles restituées par le banc de MAP exhibe des variations importantes suivant la fréquence, contrairement au spectre de puissance plat des sources virtuelles reproduites par banc de monopoles idéaux. En effet, de la même manière que dans le cas des haut-parleurs KEF (Fig. 4.9), ces haut-parleurs possèdent des caractéristiques de directivité tridimensionnelles non idéales. Par conséquent, le fait d'avoir égaliser leur rayonnement dans la zone de visibilité frontale n'assure pas un spectre de puissance plat.

Malgré tout, les spectres de puissance des sources virtuelles reproduites par le banc MAP n'exhibent pas la chute observée entre 500Hz et 10000Hz pour celles reproduites via le banc de KEF. Nous pensons que ceci est dû au fait que les haut-parleurs MAP ont des comportements plus monopolaires que les haut-parleurs électrodynamiques classiques en hautes fréquences (Corteele *et al.*, 2007). Il faudrait cependant estimer la puissance émise par les MAP au-dessus de 10kHz en utilisant la méthode d'estimation de puissance par prédiction pour pouvoir confirmer ces tendances.

Discussion

La méthode d'intégration rétrograde du relief de décroissance est, comme nous l'avons dit, attractive, car elle permet de caractériser la puissance de toutes les sources virtuelles élémentaires de la base de données à partir d'une seule mesure in situ du banc de haut-parleurs.

On notera cependant que la méthode requiert un temps de réverbération suffisamment long et un rapport signal à bruit suffisamment élevé pour que l'estimation de la puissance par intégration rétrograde de la queue de réverbération puisse produire des résultats réalistes. Nous avons constaté que pour un réglage absorbant de l'acoustique de l'Espace de Projection, la méthode ne produisait pas des résultats utilisables. Même dans le cas où la salle était réglée sur un mode réverbérant (Tr_{60} moyen d'environ 3s), l'estimation de la puissance par la méthode du relief de décroissance n'a pas produit des résultats utilisables au-dessus de 10kHz. Ceci est dû au fait qu'en HF, les temps de réverbération deviennent trop courts (absorption forte de l'air et des parois) pour pouvoir conduire l'intégration rétrograde du signal de réverbération.

Par ailleurs, la méthode fonctionne uniquement au-dessus de la *fréquence de Schroeder* (cf. glossaire) du fait de l'hypothèse de régime de champ diffus établi qui est sous-jacente à la méthode. Ceci constitue un problème en basses fréquences pour des salles petites (Baskind, 1999). Dans le cadre de cette étude, ce problème ne s'est cependant pas posé, étant donné que l'Espace de Projection possède une fréquence de Schroeder d'environ 50Hz, qui est inférieure à la fréquence de coupure du banc WFS utilisé (environ 100Hz). Malgré tout, nous avons constaté l'impossibilité d'estimer le champ en-dessous de 450Hz. Nous supposons que ceci est dû à un

rapport signal à bruit trop faible dans cette zone fréquentielle, lié à la faible puissance des transducteurs employés en BF, mais ce problème serait à explorer davantage pour confirmer cette hypothèse.

Nous concluons de cette étude qu'une caractérisation de la puissance par la méthode in situ exposée ici devra être complétée par une caractérisation individuelle des haut-parleurs utilisant d'autres méthodes (chambre réverbérante, mesure intensimétrique, mesure sur une sphère en chambre anéchoïque) si on souhaite estimer correctement la puissance émise par le banc WFS sur l'ensemble du spectre audible.

4.3.7 Conclusion

L'étude menée dans cette section a montré que la méthode d'estimation de puissance par prédiction de rayonnement présente plusieurs avantages par rapport aux deux autres méthodes qui ont été testées en pratique. Cette méthode est d'abord a priori plus intéressante que la méthode par simulation de banc idéal, car elle permet de prendre en compte la directivité des transducteurs utilisés. Elle est également plus performante que la méthode d'estimation in situ utilisant le relief de décroissance, car elle permet de conduire une estimation sur une plus large bande. Nous préconisons donc d'utiliser la méthode basée sur la prédiction dans le cadre de l'estimation de la puissance d'un banc de haut-parleurs.

4.3.8 Principe de mise en oeuvre de la compensation

Les sections précédentes ont confirmé que la puissance émise par le banc WFS variait effectivement suivant la position/orientation de la source virtuelle synthétisée. On doit donc compenser ces variations si on souhaite assurer une continuité perceptive de la réverbération tardive.

Nous proposons d'exposer, dans cette section, les principes d'une méthode de compensation de l'effet de salle associé à une source virtuelle dans des conditions réelles, qui n'aura toutefois pas été implémentée en pratique faute de temps nécessaire.

Le cadre fixé pour cette méthode est le suivant : on cherche à simuler le rayonnement d'un instrument réel, qui serait situé sur scène en \vec{r}_S et orienté suivant α_S , par un instrument virtuel WFS Ψ . On voudrait assurer par ailleurs que le système WFS émette un spectre de puissance W_Ψ constant et égal au spectre de puissance W_S de la source instrumentale réelle cible, quelles que soient la position/orientation de l'instrument virtuel Ψ , de manière à assurer une continuité perceptive de la réverbération tardive.

Analyse de la source instrumentale cible

Le rayonnement de la source instrumentale réelle est analysé et décomposé en harmoniques sphériques suivant la méthode décrite en section 4.2.1, puis transcodé en harmoniques circulaires $a_m(\alpha_S)$ en utilisant l'équation 4.4, de manière à pouvoir être simulée par le système WFS.

On estime en plus le spectre de puissance W_S de la source instrumentale réelle à partir de la pression sur l'axe p_{ax} et du filtre de facteur directionnel D en utilisant la relation A.57 donnée en annexe, ou bien directement à partir de son spectre sphérique a_{lm} en utilisant la relation A.69.

Bilan de puissance et compensation

On supposera pour simplifier que le système de diffusion d'effet de salle possède N canaux délivrant une puissance W_d . Ces canaux sont alimentés par N canaux décorrélés d'énergie $R3$ produits par le processeur d'acoustique virtuelle (la définition de $R3$ est donnée en équation 1.35).

On suppose que la source virtuelle Ψ est alimentée par un signal $S(\omega)$. Le bilan de puissance émise dans la salle à un instant donné peut s'écrire :

$$|S(\omega)|^2 \cdot W_\psi(\omega) + R3(\omega) \cdot N \cdot W_d(\omega) = W_S(\omega) \quad (4.11)$$

Ainsi, si

$$\forall \omega, |S(\omega)|^2 \cdot W_\psi(\omega) < W_S(\omega) \quad (4.12)$$

alors il suffit de régler $R3$ de telle manière à ce que

$$R3(\omega) = \frac{W_S(\omega) - |S(\omega)|^2 \cdot W_\psi(\omega)}{N \cdot W_d(\omega)} \quad (4.13)$$

pour pouvoir assurer un niveau de puissance émise dans la salle qui soit conforme à la puissance $W_S(\omega)$ cible. *A l'issue de ce processus, on possède donc une valeur de $R3_{comp}$ permettant de compenser l'effet de salle associé au contexte d'une source virtuelle placée en \vec{r}_S et orientée suivant α_S .*

Par contre, toutes les quantités du bilan sont strictement positives ; on ne peut pas avoir $R3(\omega) < 0$.

Donc, si

$$\exists \omega \text{ tel que } |S(\omega)|^2 \cdot W_\psi(\omega) > W_S(\omega) \quad (4.14)$$

on ne peut pas a priori assurer un spectre de puissance émise conforme au niveau absolu de la source instrumentale cible (ce problème est illustré sur le graphique du haut de la figure 4.12).

Malgré tout, en tolérant de faire une erreur sur le niveau absolu de puissance reproduite, on peut artificiellement augmenter le niveau de l'ensemble du spectre de puissance W_S de l'instrument réel cible. On obtient alors un spectre W'_S qui possède la même couleur spectrale que le le spectre W_S et qui respecte la condition 4.14. Cette opération est illustrée sur les graphiques du milieu et du bas de la figure 4.12)

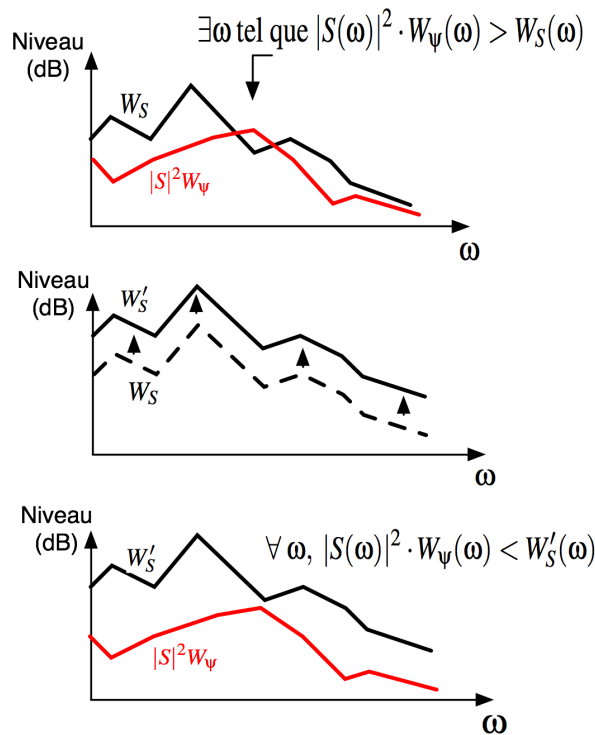


FIG. 4.12 – Illustration de la possibilité d'augmenter artificiellement le niveau global du spectre de puissance W_S de l'instrument cible, dans l'idée de pouvoir faire une compensation qui respecte la couleur spectrale de la réverbération.

On est donc a priori capable, par compensation, de s'assurer que le banc émette la puissance W'_S dont la couleur spectrale respecte celle de l'instrument cible. Par contre, le niveau de puissance émise dans la salle sera dans ce cas supérieur à celui qui serait émis par l'instrument réel. Etant donné que le niveau de son direct n'est pas affecté par le processus de compensation, l'émission trop élevée de puissance va donc se répercuter sur le rapport de son direct / son réverbéré ; un auditeur situé dans la salle risque donc de percevoir que l'instrument virtuel se situe plus loin que la position cible. Les effets et l'intérêt de cette manipulation resteraient à vérifier en pratique.

4.3.9 Base de données pour la compensation

A supposer que les conditions du problème permettent de compenser l'effet de salle associé à toutes les positions/directivités de sources virtuelles possibles (illustrées la figure 1.28), on peut consigner les valeurs de $R3_{comp}$ dans une base de données parallèle à la base de données utilisée pour la synthèse du son direct, suivant un schéma donné en figure 4.13.

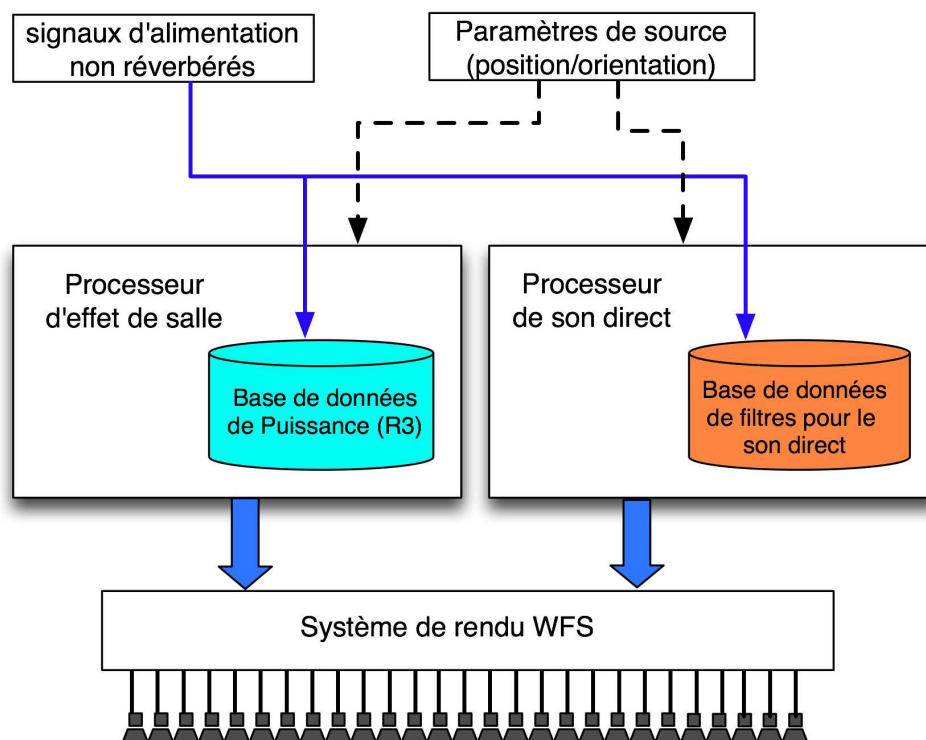


FIG. 4.13 – Implémentation pratique de la compensation. Les valeurs des paramètres énergétiques $R3_{comp}$ utilisés pour réaliser la compensation de l'effet de salle sont consignés dans une base de données en parallèle à la base de données qui conservent les filtres de rendu du son direct. A chaque changement de position/orientation de la source virtuelle, le processeur d'effet de salle charge une nouvelle valeur de $R3_{comp}$ permettant de compenser les variations de puissance émise par le banc lors de la synthèse du son direct.

Au moment du rendu de la scène sonore, les changements de position/orientation de la source virtuelle sont envoyés au processeur d'effet de salle, qui charge une nouvelle valeur de $R3_{comp}$. Le processeur d'effet de salle peut alors modifier les signaux envoyés vers le système de rendu d'effet de salle de manière à assurer une cohérence entre la puissance émise dans la salle et le son direct synthétisé par le système de rendu de son direct WFS.

4.4 Conclusion

Dans ce chapitre, nous avons abordé le problème de la cohérence de l'effet de salle tardif associé à une source virtuelle. En effet, bien qu'il soit possible de reproduire une source virtuelle directive qui approxime de manière acceptable le son direct d'un instrument réel dans le plan horizontal, le fenêtrage de la directivité de l'instrument induit une *erreur sur le spectre de puissance effectivement rayonné par le banc de haut-parleurs*. Il s'agit donc de compenser le spectre de puissance émis, si l'on souhaite *assurer une continuité perceptive dans l'effet de salle tardif associé aux sources virtuelles*.

Dans la première section, une méthode de compensation basée sur l'architecture décrite dans (Jot *et al.*, 1998) a été exposée. A partir d'une qualité acoustique "contexte", qui décrit la configuration employée pour diffuser l'effet de salle, et d'une qualité acoustique "live" qui décrit la contribution éventuelle d'instrumentistes réels sur scène, cette méthode permet de calculer l'enveloppe énergétique des signaux à fournir au processeur d'effet de salle, de manière à reproduire une qualité acoustique "cible" à une position de référence dans la salle.

La qualité acoustique cible peut être décrite de manière "perceptive" via l'opérateur perceptif du Spatialisateur. Elle peut également être décrite de manière "physique", à partir de mesures en salle. Nous avons décrit un procédé modulaire permettant de simuler la QA "cible" d'un instrument de musique sur un ensemble de positions dans une salle. Ce procédé requiert une analyse préalable du rayonnement tridimensionnel de l'instrument en champ libre. Il utilise également les données d'une campagne de mesures réalisées dans la salle utilisant une source à directivité contrôlée.

Nous avons décrit par ailleurs une méthode consistant à diffuser l'effet de salle par le biais de sources virtuelles hyper-directives. Ces sources permettent d'envoyer un faisceau d'énergie tardive vers les murs de la pièce, de manière à alimenter le champ diffus, tout en minimisant l'énergie du son direct envoyé vers la zone d'écoute.

Dans la seconde section, une méthode de compensation dite "simplifiée" a été exposée, ont l'objectif est de compenser uniquement les variations de puissance émise par la source virtuelle lors de déplacements ou rotations sur scène, de manière à assurer une cohérence de l'effet de salle tardif diffusé dans la salle. Cette méthode requiert une caractérisation de la puissance émise par le banc WFS lors de la synthèse des sources virtuelles.

Plusieurs méthodes de mesure de la puissance ont été exposées, et nous avons discuté de leur pertinence pour le problème de la caractérisation de puissance émise par un banc de haut-parleurs. Nous avons montré, en simulant une mesure de puissance d'un banc WFS de monopoles, qu'il était possible, en-dessous de la fréquence de repliement spatial, de modifier la quantité de puissance émise par le banc à l'intérieur et à l'extérieur de l'angle solide visible en jouant sur la directivité des sources virtuelles. On peut donc s'attendre qu'en situation de concert, le banc produise un rapport variable de son direct/son réverbéré suivant la directivité de la source qui est synthétisée. Par contre, les simulations ont montré le fait que le système était incapable de fournir une quantité de puissance qui soit indépendante de l'orientation de la source virtuelle reproduite.

Les tendances relevées dans les simulations ont ensuite été confirmées par la mesure de puissance d'un banc de haut-parleurs réel. Nous avons d'abord exposé une méthode d'estimation de la puissance émise par un banc de haut-parleurs électrodynamiques qui utilise la méthode de prédiction de rayonnement développée dans le cadre de cette thèse et décrite au chapitre 2. Cette méthode a été retenue comme prometteuse vis à vis des autres méthodes étudiées, car elle permet de prendre en compte la directivité des HP et fournit une estimation large bande de la puissance émise. Il resterait toutefois à vérifier son efficacité dans une situation pratique de compensation de contexte, ce qui n'a pas pu être testé faute de temps. Nous avons également décrit une méthode d'estimation de la puissance par des mesures en salle, basée sur une intégration rétrograde du relief de décroissance intégrée. Cette méthode est moins performante que la méthode par prédiction du rayonnement au niveau de la largeur de bande de l'estimation qu'elle fournit. Elle demeure toutefois intéressante dans le cas où l'on ne disposerait pas des mesures en champ libre des transducteurs nécessaires à l'estimation de puissance par prédiction.

A la fin de la deuxième section, un principe de compensation de l'effet de salle utilisant les résultats de la caractérisation de puissance émise par le banc WFS a été exposé, suivi de la description d'une architecture permettant d'intégrer la compensation au système de rendu WFS.

La méthode de diffusion d'effet de salle par sources directives, décrite dans la première section de chapitre, permet potentiellement de remplacer, ou tout du moins compléter, un système périphérique de diffusion de l'effet de salle. Cette méthode, qui s'appuie sur la possibilité de varier l'excitation de la salle en jouant sur la directivité des sources sonores reproduites, a pu être testé en situation réelle, et a montré son intérêt lors de

séances d'écoutes informelles. Dans le chapitre suivant, nous proposons d'exposer des éléments de validation du contrôle de l'excitation de la salle par la WFS, par le biais de mesures en salle utilisant un microphone à haute résolution spatiale.

Use of a high spatial resolution microphone to characterize the early reflections generated by a WFS loudspeaker array

Terence Caulkins¹, Arnaud Laborie², Etienne Corteel¹, Remy Bruno², Sebastien Montoya², and Olivier Warusfel¹

Introduction en français

Ce chapitre, qui reprend l'article (Caulkins *et al.*, 2006) écrit par l'auteur en collaboration avec Arnaud Laborie, Etienne Corteel, Remy Bruno, Sebastien Montoya, et Olivier Warusfel, vise à décrire un travail de caractérisation du banc de haut-parleurs placé en conditions réelles de diffusion.

L'idée qui motive ce travail est la validation du contrôle de la directivité offerte par le banc de haut-parleurs. En effet, bien que nous ayons eu l'occasion de vérifier l'intérêt de l'utilisation de la directivité pour diffuser l'effet de salle tardif en situation d'écoute informelle, il manque des éléments de validation objective par la mesure dans cette étude. *On propose d'apporter des éléments de validation du contrôle de la directivité faite par le banc WFS dans ce chapitre, en s'appuyant sur des mesures des réflexions précoces engendrées par le banc de haut-parleurs placé en salle.*

¹IRCAM, 1 Pl. Igor Stravinsky, Paris, FRANCE

²TRINNOV AUDIO, 30-32 Ave. de la République, Villejuif, FRANCE

Des méthodes de caractérisation du champ précoce utilisant des réseaux microphoniques ont déjà été développées dans le passé, en tant qu'outils pour *l'auralisation* (cf. glossaire) (Hulsebos *et al.*, 2001). Il a été montré que les géométries circulaires de réseaux microphoniques pouvaient être utilisées pour extraire une décomposition en harmoniques cylindriques d'ordre élevé du champ sonore au point de mesure (Hulsebos, 2004). Cependant, lorsqu'on utilise des réseaux bidimensionnels, le problème de la discrimination entre les réflexions provenant du sol/plafond et les réflexions provenant des mur latéraux se pose, comme le soulève (Spors *et al.*, 2004a). Cette capacité de discrimination est cruciale à la bonne interprétation du champ mesuré en salle, car sinon des réflexions qui proviennent de la dimension verticale viennent entacher la caractérisation du plan horizontal.

Pour pallier ce problème nous proposons une nouvelle méthode pour caractériser les composantes précoces de l'effet de salle produit par le banc WFS. La méthode combine l'utilisation d'un réseau microphonique 3D "à haute résolution spatiale" avec un modèle géométrique simple de la salle d'écoute de manière à identifier les différentes composantes réfléchies. La caractérisation que nous proposons est focalisée sur le plan horizontal, du fait de son importance d'un point de vue perceptif. Ceci est rendu possible par l'utilisation du réseau microphonique, qui permet de réduire l'influence des composantes d'effet de salle verticales en formant des lobes de directivité dans le plan horizontal.

La première section de l'article décrit le principe du réseau microphonique, et comment il peut être utilisé pour extraire une information suivant une direction donnée dans un champ sonore tridimensionnel. Dans une seconde section, un modèle simple de sources images est utilisé pour simuler la mesure in situ d'une source idéale et d'un banc WFS constitué de monopoles sur un réseau microphonique placé dans l'Espace de Projection. Ce modèle idéal est ensuite utilisé pour analyser les différences entre le champ précoce produit par la source idéale et le banc WFS qui synthétise une source virtuelle à la même position. Dans la section finale, les résultats de la simulation sont confrontés à des mesures réelles conduites dans l'Espace de Projection avec un banc WFS de haut-parleurs électrodynamiques et le microphone haute résolution spatiale.

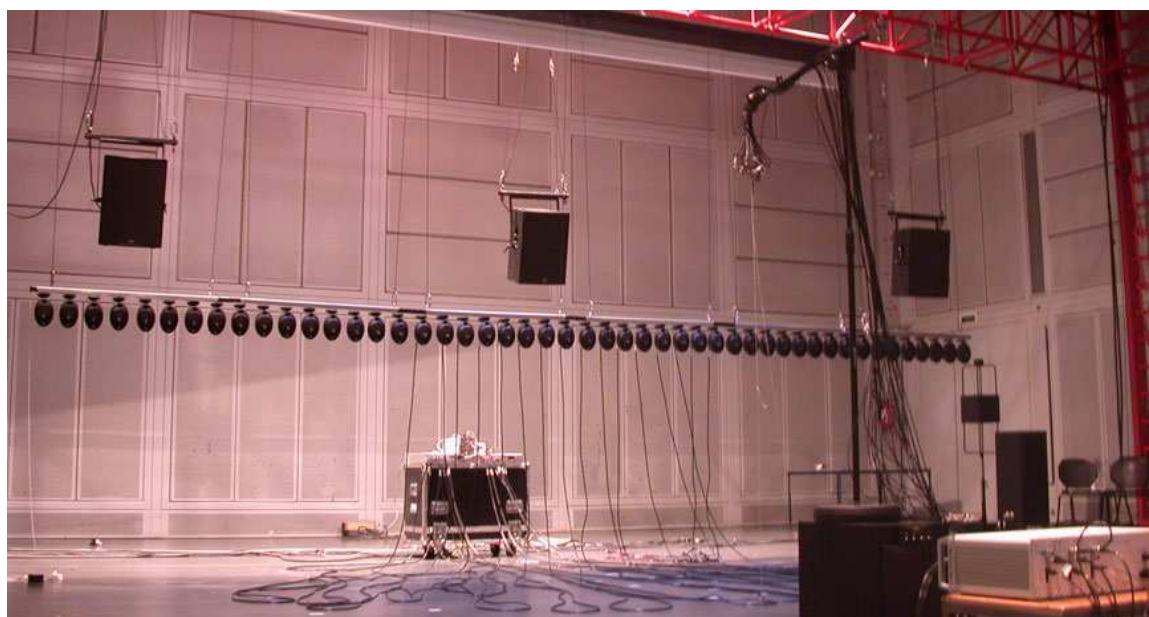


FIG. 5.1 – Campagne de mesure de caractérisation d’un banc WFS dans la salle de concert de l’Ircam. Le réseau microphonique qui figure au premier plan est un microphone à haute résolution spatiale développé par Trinnov Audio.

Abstract

This article deals with a characterization of the early reflections generated by a linear Wave Field Synthesis (WFS) array. Techniques based on the use of planar microphone arrays suffer from an inability to discriminate side-wall reflections from ceiling-floor reflections. An approach combining the use of a high spatial resolution microphone with a geometric model is investigated in order to overcome this problem.

5.1 Introduction

A current topic of research in surround sound concerns the application of Wave Field Synthesis to the enhancement of live concert performances, whereby virtual sound sources are rendered on stage alongside real instruments. This juxtaposition of real and virtual sources on stage implies the possibility for the audience to assess the quality of the WFS rendered sound field by direct comparison with a real source sharing the same physical space.

On one hand, the **direct sound field** produced by real sources will be compared to the direct sound field of virtual sources rendered by WFS. Previous research has yielded a content-dependent multi-channel equalization scheme which allows an optimization of the direct sound field within the horizontal plane situated in front of the loudspeaker array (Corteel *et al.*, 2002).

This algorithm, which supposes free field conditions, generates a set of filters to be applied to the loudspeakers in order to produce the best possible least-squares approximation of a real source's direct sound field. The proposed framework also offers the means to render the direct sound fields of virtual sources displaying variable directivity patterns, based on a linear decomposition of the sound field onto a basis of spatial eigenfunctions (Caulkins *et al.*, 2003; Corteel, 2004).

On the other hand, the audience will be able to judge the realism of virtual source reproduction by comparing the **room effect associated to real and virtual sources**. Setting aside the case of extended sources, it can be approximated that real instruments on stage will exhibit point source radiation patterns. This is not the case for virtual sources reproduced using a linear WFS array ; though their sound field in the horizontal plane may resemble the target source's sound field, the array itself radiates as a finite line source in 3D. This entails a different mode of interaction with the listening room, and differences are to be expected when comparing the room effect associated to a real source and the room effect that is naturally produced by a WFS array.

In (Caulkins & Warusfel, 2006) a study of the late reverberance component associated to WFS sources is carried out. For a real source radiating sound into a concert hall, the total power emitted by the source above the Schroeder frequency does not depend on its position or orientation. It follows that the late reverberance will remain steady during source movements and rotations on stage. However, the authors show that the power emitted by a WFS array is subservient to both positioning and orientation of the virtual source, mainly because of the finite extent of the window through which the source illuminates the listening room. A method is proposed allowing to predict the power associated to a virtual source synthesized by a WFS array for any position and orientation of the source. When directivity and/or spectral characteristics of the transducers cannot easily be modeled, in-situ measurements of the array synthesizing a given virtual source can provide the power estimation through time-frequency analysis proposed in (Jot *et al.*, 1997).

The goal of the present article is to expose a complimentary characterization of the room effect produced by a WFS array, in which the focus is shifted towards the **early components of the impulse response**. Methods for characterizing the early components of an impulse response using microphone arrays have been developed in the past as tools for auralization (Hulsebos *et al.*, 2001). It was shown that a circular array geometry could be used to perform a high order cylindrical harmonic decomposition of the sound field in a concert hall (Hulsebos, 2004). However when considering 2D microphone arrays the problem of discriminating floor/ceiling reflections from wall reflections arises, as stated in (Spors *et al.*, 2004a). This discrimination is crucial for proper sound field characterization, since the reflections stemming from the vertical dimension are "re-mapped" upon the 2D plane and can therefore constitute a source of error in the interpretation of the array measurements.

In light of this problem, the authors investigate a new method for characterizing the early components of the room effect associated to a WFS array. This method combines the use of a 3D high spatial resolution microphone array (Laborie *et al.*, 2003) with a simple geometric model of the listening room in order to identify different reflection components. The proposed characterization will focus on the horizontal plane, because of its importance in regard to spatial perception. This can be properly done since the use of a 3D array mitigates the influence of vertical components.

The first section of this paper will be dedicated to describing the functionalities of the microphone, and how it can be used to extract directional information in a 3D sound field. In the second section, a simple image source model is used to simulate the in-situ measurement of an ideal source and an ideal WFS array by the 3D microphone array placed in the Espace de Projection concert hall in IRCAM. The model is employed to analyze the discrepancies that can be found between the early room effect produced by an ideal point source and the one produced by a WFS array synthesizing a virtual source at the same position in the room. In the final section, results of the simulation are confronted with real measurements of a 48 channel WFS array in the Espace de Projection.

5.2 Description of the High Spatial Resolution Microphone Array

5.2.1 Presentation

The microphone used in this study is a third-order, full-sphere acoustic field microphone, as described in (Laborie *et al.*, 2003). Its principle is based on the decomposition (Bruneau, 1998; Skudrzyk, 1971) of an acoustic field $p(r, \theta, \phi, t)$ into a series of Fourier-Bessel coefficients $p_{l,m}(t)$. This decomposition is given in the frequency domain by

$$P(r, \theta, \phi, f) = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^l P_{l,m}(f) j^l h_l(kr) y_l^m(\theta, \phi) \quad (5.1)$$

where :

- $k = 2\pi f/c$, c is the speed of sound, approximately 340 m/s,
- $P_{l,m}(f)$ is the Fourier transform of $p_{l,m}(t)$,
- $P(r, \theta, \phi, f)$ is the Fourier transform of $p(r, \theta, \phi, t)$,
- $h_l^{(1)}(x)$ is the spherical bessel fonction of order l (Laborie *et al.*, 2003),
- $y_l^m(\theta, \phi)$ is the real valued spherical harmonic (l, m) (Laborie *et al.*, 2003).

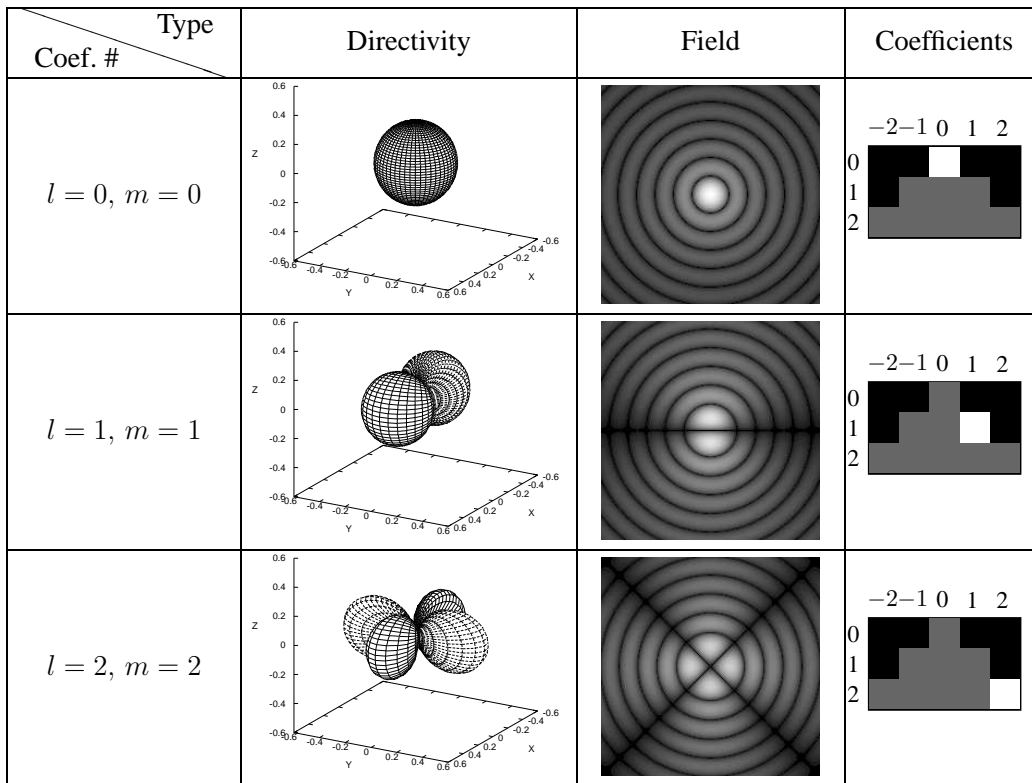


FIG. 5.2 – Various representations of acoustic fields.

Using this formalism, it is possible to scan the full 3D acoustic field using the discrete set of signals $p_{l,m}(t)$. In real-world applications of this decomposition, the series is truncated above a given maximum value for l , which is referred to as the order L of the decomposition.

The Fourier-Bessel coefficients can be interpreted in three different ways, as represented on figure 5.2 :

- a directivity function giving the source distribution in space,
- a field representation that gives the pressure level for any point in space,
- a spectral representation of the amplitude of the Fourier-Bessel coefficients arranged in a triangle.

The microphone used in this article is a third-order Fourier-Bessel microphone, which means that it delivers an estimation of the Fourier-Bessel coefficients of the acoustic field in which it is immersed up to order 3.

5.2.2 Principle

The design of the microphone is based on spatial sampling (Laborie *et al.*, 2003). This general approach considers a microphone consisting in an array of capsules of known characteristics (by measurements or modeling). Each capsule extracts a portion of the information contained in

the acoustic field being measured. This is represented mathematically by the following *sampling relation*, written in the frequency domain :

$$\mathbf{c} = B\mathbf{p} \quad (5.2)$$

where

- B is the *sampling matrix* dependent only on microphone characteristics (position and directivity of its constituent capsules),
- \mathbf{p} is a vector containing the Fourier-Bessel coefficients of the acoustic field being measured,
- \mathbf{c} is a vector containing the signals delivered by the capsules.

This matrix relation expresses the fact that it is possible to relate the signals delivered by the capsules to the acoustic field in which they are immersed.

Estimation of the Fourier-Bessel coefficients from the capsule signals is the reverse problem of the spatial sampling process. By *inverting* matrix B , it is possible to determine an estimation $\hat{\mathbf{p}}$ of the unknown Fourier-Bessel coefficients \mathbf{p} from the capsule signals \mathbf{c} . This operation is called *encoding* and can be expressed as :

$$\hat{\mathbf{p}} = E\mathbf{c} \quad (5.3)$$

where E is the *encoding matrix*. This matrix can be determined from B using least squares methods, for example. (Laborie *et al.*, 2003) gives a more precise description of the principle of this microphone.

5.2.3 Spatial resolution of the microphone array

The principle described above allows, in theory, for the design of an unlimited order microphone, provided that the array being used is large enough and sufficient number of capsules. Trinnov has manufactured a 24-capsule array allowing to obtain an order 3 decomposition of the sound field. The capsules are omnidirectional and irregularly distributed inside a 20 cm diameter ball.

Artifacts appear in the acoustic field estimation for the following reasons :

- The acoustic field representation is truncated at a given order, here 3. A consequence of this is a limited angular resolution in the resulting directivities. In other words, a plane wave, which should, in the directivity approach, appear as an angular Dirac is not seen as an infinitely narrow source, but as a finite width lobe. Smaller secondary lobes also appear in certain directions even though no source is present.
- The array measurement, and the encoding process, introduce artifacts in the estimation.

Use of a high spatial resolution microphone to characterize the early reflections generated by a WFS loudspeaker array

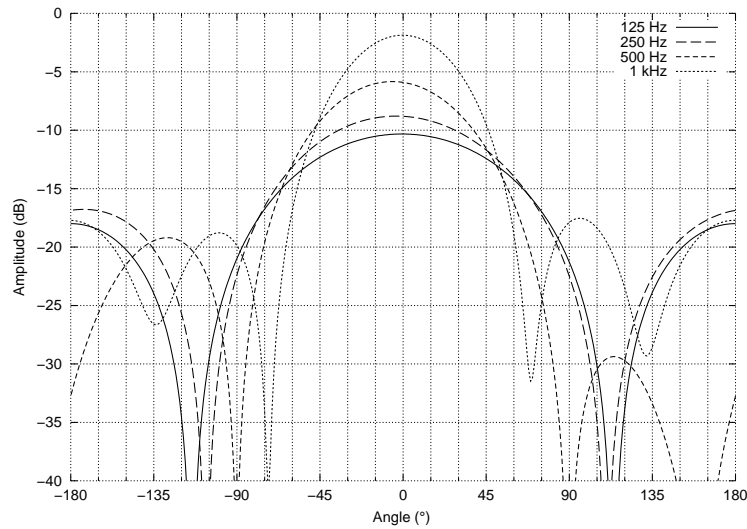


FIG. 5.3 – Angular response of the microphone to a frontal plane wave.

The 16 coefficients corresponding to the third-order estimated acoustic field do not correspond to the exact 16 first coefficients of the actual acoustic field.

- The microphone array design is not perfect (position, frequency response and directivity of the capsules, acoustic masking effect of the array structure...).

Models allow to quantify these imperfections, which depend on frequency. Figure 5.3 represents the directivity provided by the microphone in the horizontal plane for several frequencies when it captures a plane wave incoming from the front direction. It outlines the better performances of the microphone for medium frequencies than for low frequencies, as the main lobe (around 0°) is narrower and the secondary lobes are simultaneously smaller. These performances can be represented for incidence directions other than 0° and for several frequencies using a matrix representation, as in figures 5.4, 5.5, 5.6 and 5.7. Each column of this matrix gives the directivity determined by the microphone for a plane wave in a certain direction. The first column is for a source at $+180^\circ$, the last column is for a source at -180° . Colored portions correspond to amplitudes above a 25 dB threshold. Here again, better performances for high frequencies are outlined, since the diagonal, which corresponds to the main lobe, is narrower at 1 kHz, and secondary lobes are lower. An ideal infinite-order microphone would of course have a directivity transfer matrix consisting of ones on its diagonal and zeros everywhere else.

The directivity transfer matrices can be interpreted in two different ways :

- As stated above, each column gives the directivity response that the microphone delivers when it is immersed in a plane wave coming from the direction corresponding to the considered column. Therefore, the curves on figure 5.3 correspond to the columns labeled 0° of the matrices represented on figures 5.4, 5.5, 5.6 and 5.7.

5.2. Description of the High Spatial Resolution Microphone Array

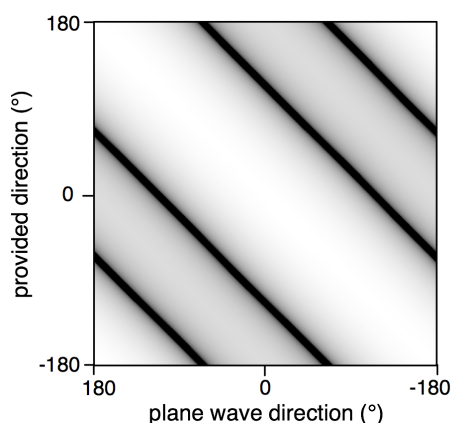


FIG. 5.4 – Directivity transfer matrix of the microphone at 125 Hz

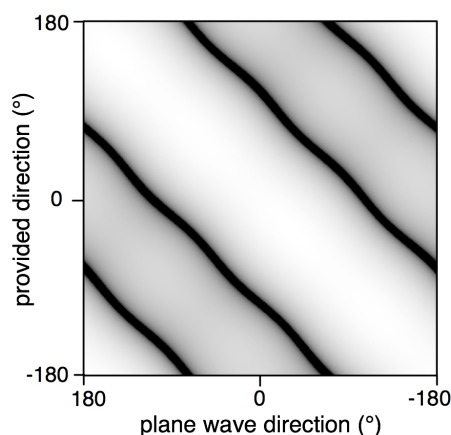


FIG. 5.5 – Directivity transfer matrix of the microphone at 250 Hz.

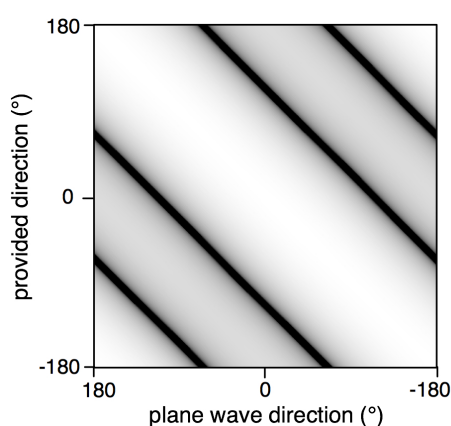


FIG. 5.6 – Directivity transfer matrix of the microphone at 500 Hz.

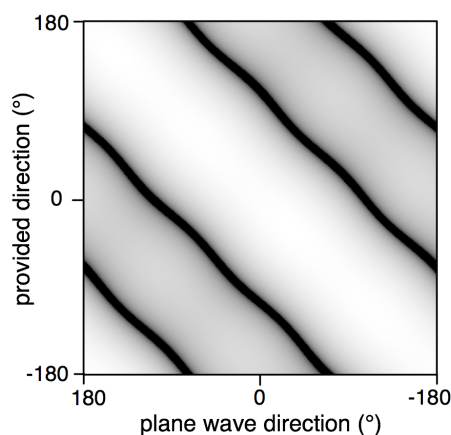


FIG. 5.7 – Directivity transfer matrix of the microphone at 1 kHz.

- Each row corresponds to the directivity of the virtual microphone obtained when the actual directivity provided by the microphone is spatially sampled in the direction corresponding to the considered row. For example, the signal obtained by considering the value of the directivity, delivered by the microphone, in the 0° direction (in the horizontal plane) is the same as the signal delivered by a virtual microphone of which the directivity would be the row of the matrix at 0° . In the matrixes presented here, only the horizontal plane is displayed; however, the actual directivity of each virtual microphone obtained by sampling the provided directivity is in fact three-dimensional.

Figure 5.8 gives the virtual directivity at 1kHz obtained by considering the 0° direction in the horizontal plane. The curve labeled *Horizontal* gives the directivity in the horizontal plane while the curve labeled *Vertical* gives the directivity in the vertical (median) plane going through the axis of the directivity lobe for a 0° direction. For the vertical curve, 0°

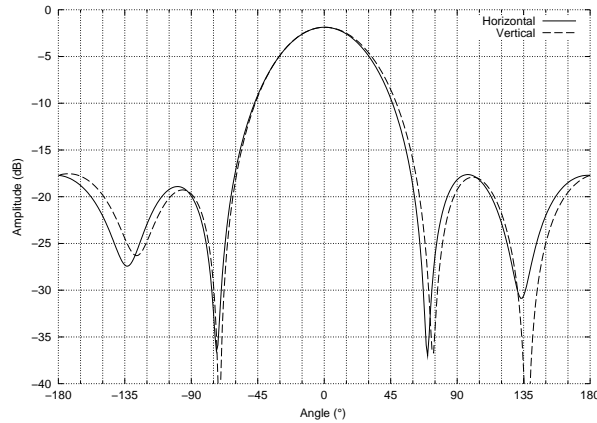


FIG. 5.8 – Frontal sample of directivity delivered by the microphone at 1kHz in the horizontal and vertical plane.

is the frontal direction, $+90^\circ$ is the ceiling, $+180^\circ$ and -180° are the back, and -90° is the floor.

5.3 Characterizing the early room effect associated to a WFS array from simulated measurements made on the high resolution microphone

This section proposes a simulated measurement of the early room effect generated by an ideal source radiating in a rectangular concert hall, and the early room effect produced by a WFS array synthesizing the same source. The first objective is to verify that the microphone is capable of discriminating specular reflections produced by the ideal source, given a simple image source model of a concert hall. Having verified this, the second objective will be to compare the behavior of the WFS array with that of the ideal source, so as to characterize the particularities of the early room effect that can be associated to virtual WFS sources.

5.3.1 Description of the simulated measurement setup

The simulated measurement setup involves an ideal source and an ideal WFS array synthesizing the same source at the same position in a rectangular hall that exhibits the same dimensions as the Espace de Projection concert hall in IRCAM (23.5m (Length) \times 15.5m (Width) \times 11m (Height)).

The WFS array is modeled as a linear and evenly spaced (16.5cm) distribution of 48 mono-

5.3. Characterizing the early room effect associated to a WFS array from simulated measurements made on the high resolution microphone

pole sources. It occupies a central position with respect to the left and right walls, and is situated 5.5m in front of the front wall and 1m78 above the ground (cf figure 5.9). The virtual source is situated 1m behind the array and 2m to the right of its right bisector.

A characterization of the two sound fields is simulated upon a virtual 24 capsule microphone array exhibiting the same geometrical arrangement as the one described in the previous section. The microphone array is centered at a position situated 5.3m in front of the WFS array and 2.1m to the left of its right bisector, at the same height (1m78 above the ground). An image source model is employed to synthesize the impulse response of the ideal source and the ideal WFS array at the 24 capsule positions. This model will be used in order to help the identification of the various reflections off of the walls of the listening room. Specular reflections (up to the third order image source) are recreated using this model. The walls are modeled with full band, real valued absorption coefficients that approximate the boundary properties of the Espace de Projection during the in-situ measurements described in section 5.4 (front wall 0.8, left wall 0.001, back wall 0.3, right wall 0.7, floor 0.001, ceiling 0.2).

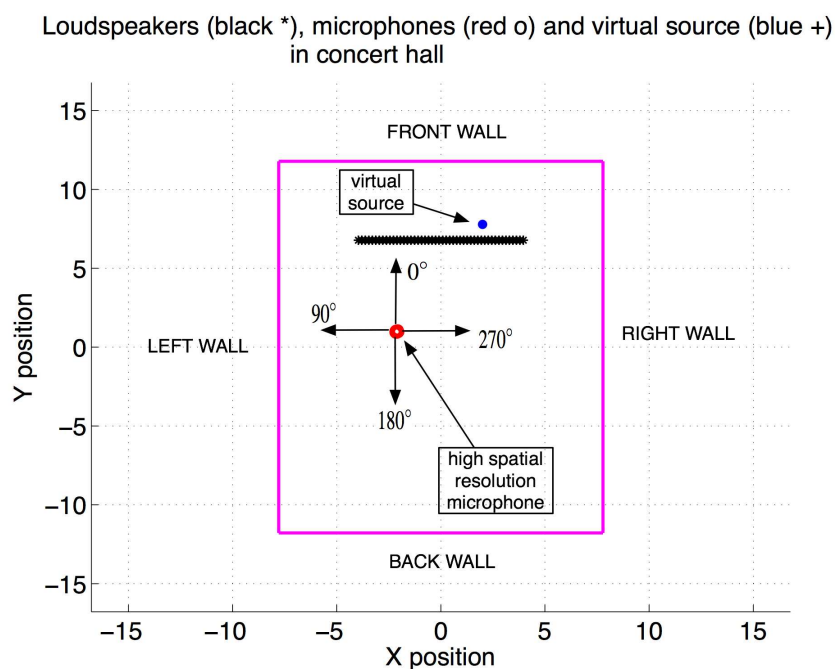


FIG. 5.9 – Configuration for the simulated measurements of an ideal source and a WFS array synthesizing a virtual source at the same position in the concert hall. Source position is represented by a blue cross. Microphone array is represented by a set of red circles. Directional encoding by the microphone is done following a positive angular progression represented on the black axes.

The 24 simulated impulse responses are encoded onto a set of third order spherical harmonics and subsequently decoded in order to form the response of 36 hyper-directional micro-

phones arranged in the horizontal plane, at 10° intervals (figure 5.9 shows 4 of the 36 available directions).

5.3.2 Ideal source measurement

The simulated measurement of an ideal source situated at the position described in figure 5.9 is carried out on the high spatial resolution microphone and translated into 36 directional responses in the horizontal plane, as described in the previous section. Three ideal sources are considered : a monopole (omnidirectional) source, a dipole source (degree -1 spherical harmonic) and a dipole source with its zero plane perpendicular to the previous one (degree 1 spherical harmonic). The pressure field $p_{ideal}(r, \theta, \phi)$ emitted by the ideal source is based on the far field approximation of a directive point source and described by the following equation :

$$p_{ideal}(r, \theta, \phi) = \frac{e^{-jkr} \cdot \cos(n\theta) \sin(n\phi)}{r} \quad (5.4)$$

with n the degree of the source, r the distance from the source, θ the azimuthal angle (contained in the plane parallel to the floor of the hall) and ϕ the polar angle.

The result of the measurement and subsequent directional decoding of the **ideal monopole source** is represented as a function of time on figure 5.10. Letters are placed at the expected (azimuth, time) coordinates of first and second order reflections according to the image source model applied to the ideal source.¹ :

- Capital letters 'F', 'L', 'B', 'R', 'G', 'C' represent angular/temporal coordinates of first order reflections off of (resp.) the Front wall, Left wall, Back wall, Right wall, Ground and Ceiling. *Direct sound is represented by a capital 'D'*.
- Lowercase letters represent second order reflections with each letter representing a consecutive reflection off of a given boundary of the listening room. For example, *lg* represents a reflection off of the left wall and then the ground; *bc* represents a reflection off of the back wall then the ceiling.
- Image sources with identical positions are represented only once. In this configuration we have $gf = fg, gl = lg, gb = bg, gr = rg, cf = fc, cl = lc, cb = bc, cr = rc, fl = lf, fr = rf, bl = lb, \text{ and } br = rb$
- Image sources *gc* and *cg* are at different elevation angles. However they present the same (azimuth, time) coordinates so only *gc* is represented.

Impulse responses of the ideal source are filtered below 1000Hz, which is the aliasing frequency of the WFS array for the chosen source position at the measurement position (see (Corteel, 2004) for an aliasing frequency calculation method taking into account loudspeaker

¹Image source positions have been slightly shifted in some cases to improve legibility of the figures.

5.3. Characterizing the early room effect associated to a WFS array from simulated measurements made on the high resolution microphone

spacing as well as virtual source and listener positioning). This is done in order to be able to compare results with those given for an ideal WFS array in section 5.3.3.

Observation of figure 5.10 shows good agreement between the image source model and the directional encoding performed by the microphone array in terms of the expected (azimuth, time) coordinates of first and second order image sources. The monopole source being measured is situated at 340° from the microphone's point of view, as can be seen on figure 5.9. The direct sound can therefore be expected to arrive from this direction, as is indicated by the 'D' on the figure. The decoded impulse response is properly centered on the expected (azimuth, time) coordinates of the direct sound. The effect of the directivity lobes of the microphone described on figure 5.3 can be seen to cause a smearing of the impulse response over the azimuthal plane, whereby energy is seen to be incoming from the rear directions as compared to the expected directions of arrival (strictly comprised between angles $[0^\circ - 70^\circ]$ and $[250^\circ - 360^\circ]$). The level of these rear components remains 20dB below the level of direct sound for the $[70^\circ - 250^\circ]$ angular sector. The measured level for the rear components could nonetheless be made lower by application of a high pass filter on the decoded data (secondary lobes of the microphone are larger for low frequencies, as is visible on figure 5.3).

Ceiling reflections 'C', lc , rc and bc are greatly diminished (25dB below the direct sound) in comparison to other contributions due to the focusing of the microphone onto the horizontal plane. Ground reflections conserve high levels because of a smaller polar angle of incidence for the first order ground reflection 'G' ($\phi \approx -22^\circ$) compared to the polar angle of incidence for the first order ceiling reflection 'C' ($\phi \approx 42^\circ$), as well as a smaller absorption coefficient of the ground (0.001 vs 0.2 for the ceiling).

In figures 5.12 (top) and 5.13 (top), the impulse responses of two **ideal dipole sources** (degree 1 and degree -1) are represented along the 36 incoming directions at the microphone position. The dipole of degree -1 is oriented in such a way that its zero plane encompasses the microphone position. This explains the absence of any direct component on figure 5.12 (top). However, the same figure shows a strong primary reflection 'L' incoming from the left wall at 70° (at 49.2ms) and a slightly weaker primary reflection 'R' incoming simultaneously from the right wall at 290° as predicted by the image source model. In other words, though the microphone is not illuminated directly by the source, it receives the early reflections produced by the concert hall.

Conversely, the case of the degree 1 dipole shown on figure 5.13 (top) shows an opposite tendency. This source has its lobe oriented straight at the microphone position. It is therefore observed that the direct sound component (as well as its reflection off the ground) is very strong. However the primary reflection 'L' on the left wall is greatly diminished in comparison with the monopole depicted on figure 5.10 because this wall receives very little direct sound from the source. The same reasoning can be applied to the second order reflections lc (and cl , not repre-

Use of a high spatial resolution microphone to characterize the early reflections generated by a WFS loudspeaker array

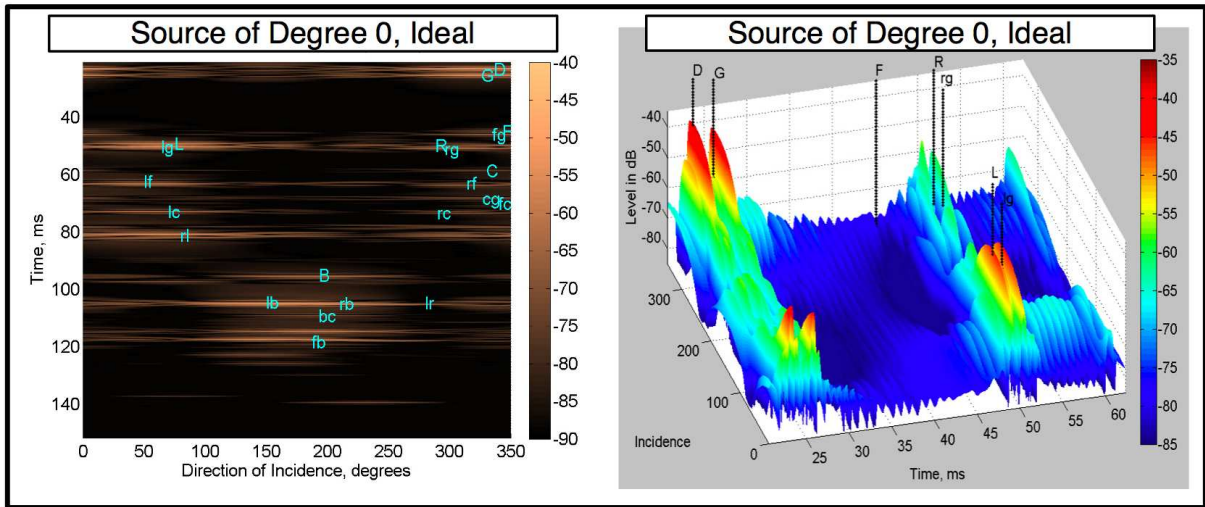


FIG. 5.10 – Amplitude of the impulse response (dB) of an **ideal monopole point source** decoded onto 36 directions (with 10° increments) as a function of time (ms). Letters are placed at the expected (azimuth, time) coordinates of the first and second order reflections according to the image source model for an ideal source (see section 5.3.2 for details). **Left** : Top view, [0 ; 150]ms interval. **Right** : Side view, [0 ; 65]ms interval.

sented) that are extremely attenuated for the degree 1 dipole source (20 dB drop as compared to the monopole source).

5.3.3 Measurement of an ideal WFS array

In this section the ideal measurements of the ideal point source will be compared with ideal measurements of a WFS array using the setup described in section 5.3.1.

A set of individual filters needed to be applied to the 48 monopole loudspeaker array in order to synthesize the direct sound field of an ideal point source occupying the same position as the source studied in the previous section is calculated according to the multichannel equalization procedure. Image sources of each individual loudspeaker of the array reflected off of the 6 walls of the listening room are formed in order to construct first, second, and third order 'image arrays'. The individual filters calculated for each loudspeaker are then applied to their corresponding image sources, and the result of the convolution is measured upon the microphone array. Impulse responses are low-pass filtered below the aliasing frequency, calculated to be equal to 1000Hz for this source position.

Figure 5.11 (top) shows the results of the simulation measurement of the ideal WFS array synthesizing a **virtual monopole point source** at the same position as the ideal source studied in the previous section. The expected image source positions of an ideal point source in the

5.3. Characterizing the early room effect associated to a WFS array from simulated measurements made on the high resolution microphone

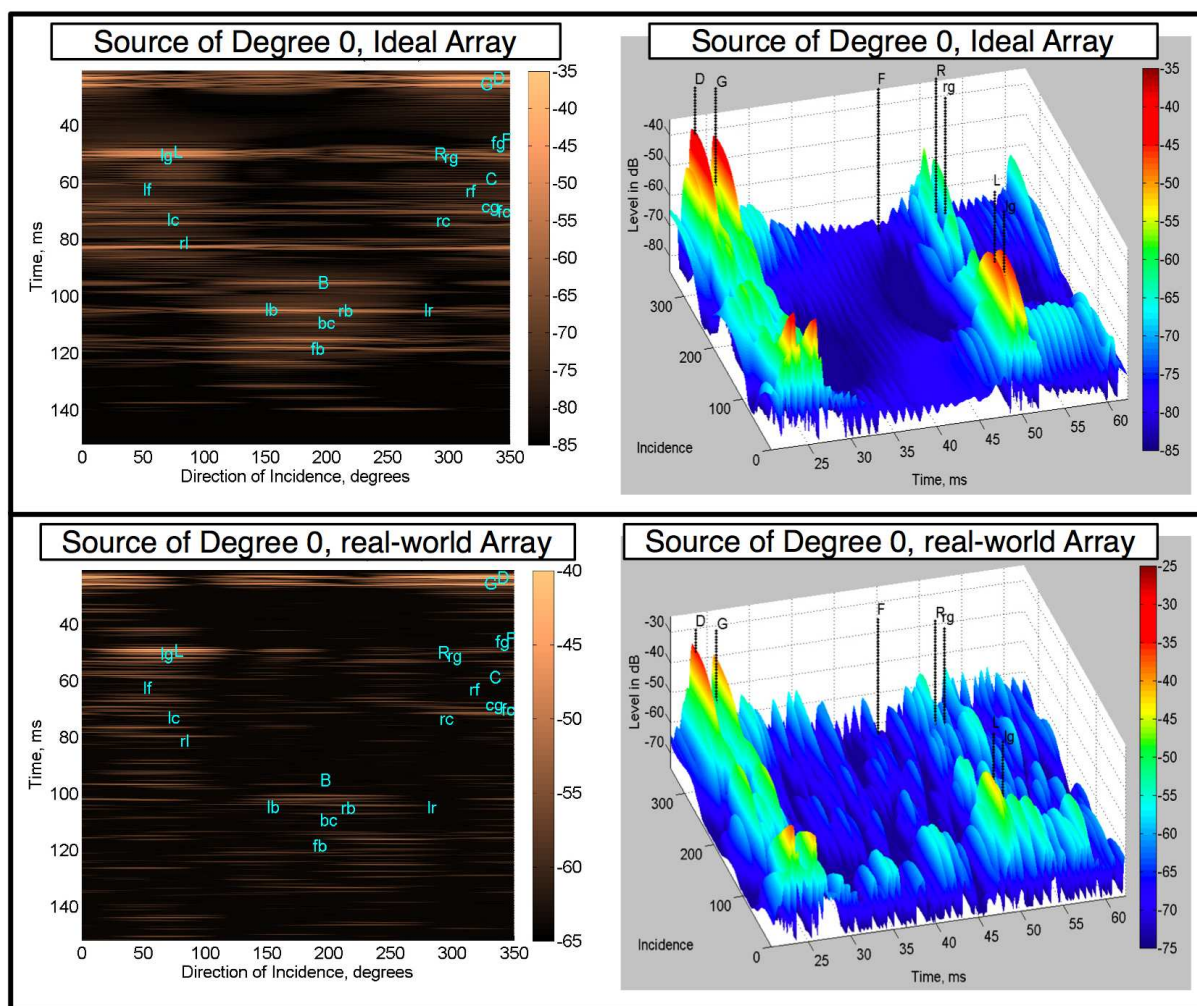


FIG. 5.11 – Amplitude of the impulse response (dB) of a **virtual monopole point source synthesized on the ideal WFS array** (top) and **virtual monopole point source synthesized on the real-world WFS array** (bottom), decoded onto 36 directions at the microphone position (with 10° increments) as a function of time (ms). Letters are placed at the expected (azimuth, time) coordinates of the first and second order reflections according to the image source model for an ideal point source. **Left** : Top view, [0 ; 150]ms interval. **Right** : Side view, [0 ; 65]ms interval.

(azimuth, time) plane of the figure are represented by their corresponding letters ('D', 'G', etc..). Their positions are strictly identical to those represented on figure 5.10.

Effect of cylindrical symmetry of the WFS array

A quick scan of figure 5.11 (top) reveals that certain reflections are slightly delayed compared to their expected time of arrival for an ideal source. This is true for the first order reflections

Use of a high spatial resolution microphone to characterize the early reflections generated by a WFS loudspeaker array

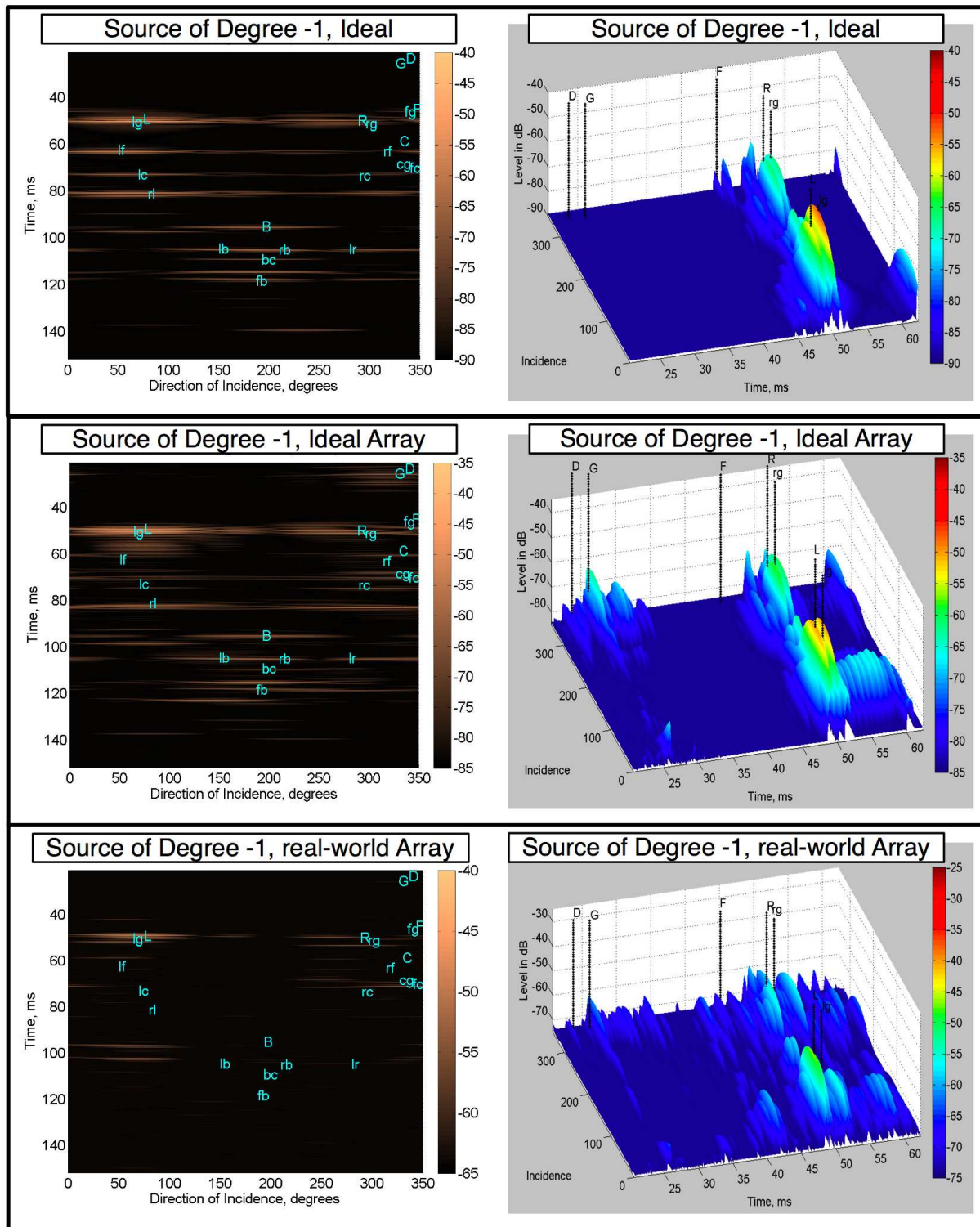


FIG. 5.12 – Amplitude of the impulse response (dB) of an **ideal dipole point source of degree -1** (top), **virtual dipole point source of degree -1 reproduced on an ideal WFS array**(middle) and **virtual dipole point source of degree -1 reproduced on the real-world WFS array** (bottom), decoded onto 36 directions by 10° increments as a function of time (ms). **Left** : Top view, [0 ; 150]ms interval. **Right** : Side view, [0 ; 65]ms interval.

5.3. Characterizing the early room effect associated to a WFS array from simulated measurements made on the high resolution microphone

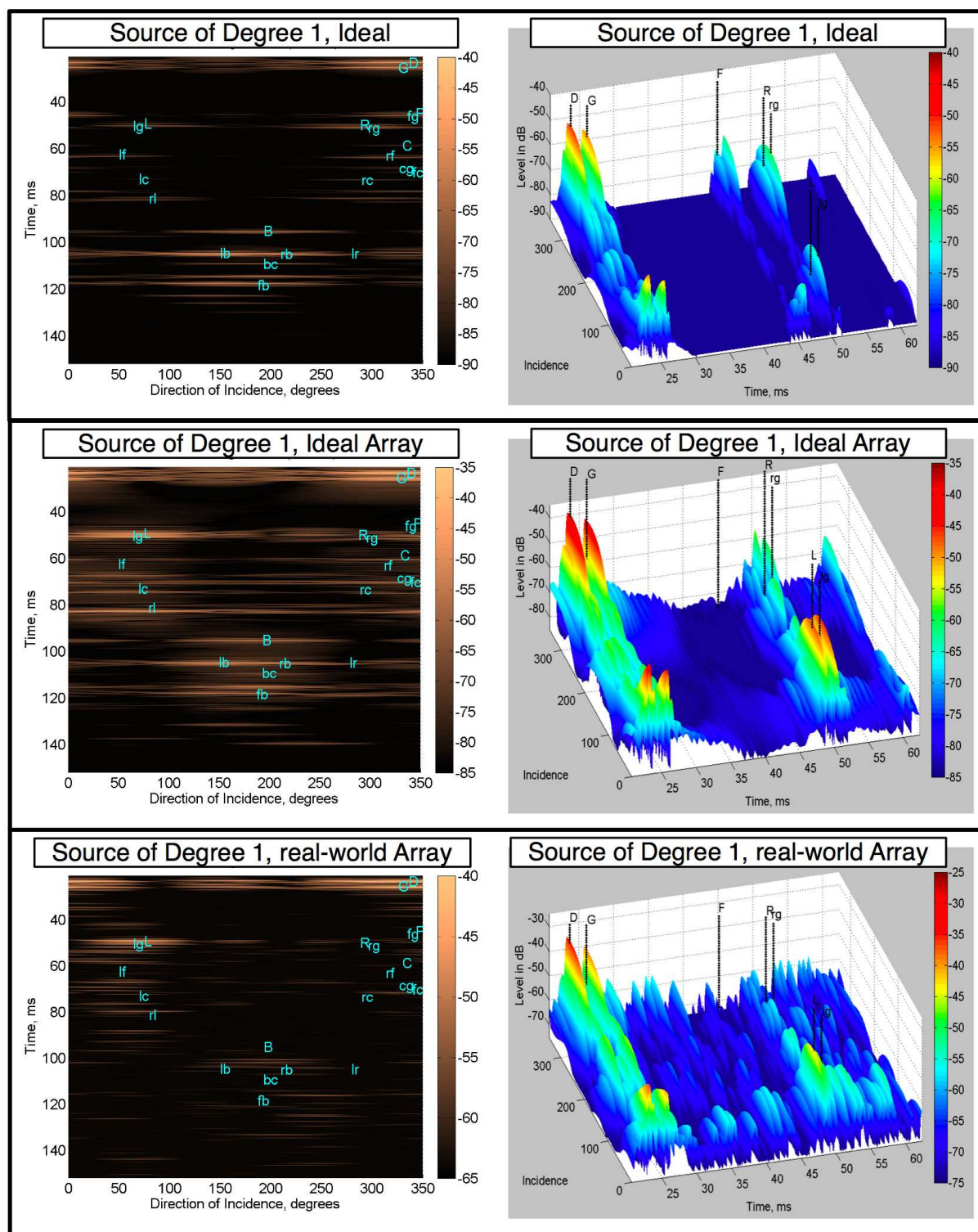


FIG. 5.13 – Amplitude of the impulse response (dB) of an **ideal dipole point source of degree 1** (top), **virtual dipole point source of degree 1 reproduced on an ideal WFS array**(middle) and **virtual dipole point source of degree 1 reproduced on the real-world WFS array** (bottom), decoded onto 36 directions by 10° increments as a function of time (ms). **Left** : Top view, [0 ; 150]ms interval. **Right** : Side view, [0 ; 65]ms interval.

Use of a high spatial resolution microphone to characterize the early reflections generated by a WFS loudspeaker array

off of the front wall 'F' and ceiling 'C', as well as for second order reflections off of these walls (notably cg and fc). However the direct sound 'D' and first order reflections off of the left wall 'L', right wall 'R' and back wall 'B' display proper arrival times as compared to the ideal source.

To explain this behavior one must take into account the cylindrical geometry of the loudspeaker array. Since each individual source of the loudspeaker array is an ideal monopole, the array displays a cylindrical symmetry around its axis. This in turn implies that the wavefront synthesized for a source situated "behind the array" will be perceived "behind the array" on both sides of the loudspeaker array, as can be seen on figure 5.14. From this observation, it is clear

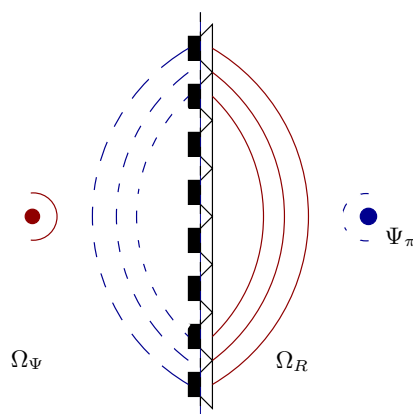


FIG. 5.14 – Illustration of how a given wavefront associated to a source ψ_0 virtually situated in the sub-plane Ω_ψ causes the formation of a symmetrical wavefront corresponding to a virtual source ψ_π situated at a symmetrical position within the sub-plane Ω_ψ

that the front and back walls receive contributions that arrive with different delays. Since the back wall, left wall and right walls are situated in the targeted reproduction plane they receive contributions that remain coherent with the target source position. However, the front wall receives the contribution from a virtual source situated 1m to the other side of the array, i.e. *inside the audience area*, instead of from the expected position situated 1m behind the loudspeaker array (and therefore *outside of the audience area*). This false location of the virtual source also occurs in the vertical direction : the ceiling of the concert hall is in fact illuminated by a wavefront corresponding to virtual source situated 1m *below* the array instead of 1m behind the array in the horizontal plane ; the ground of the concert hall receives the contribution from a virtual source situated 1m *above* the array. The consequence of these symmetry errors is that the images of virtual point sources reflected off of the front wall, ceiling and ground are in fact further away than those of the ideal point source. This discrepancy is illustrated in figure 5.15 for the front reflection. The arrival of the reflections off of these walls is therefore delayed, as can be observed on figure 5.11 (top).

5.3. Characterizing the early room effect associated to a WFS array from simulated measurements made on the high resolution microphone

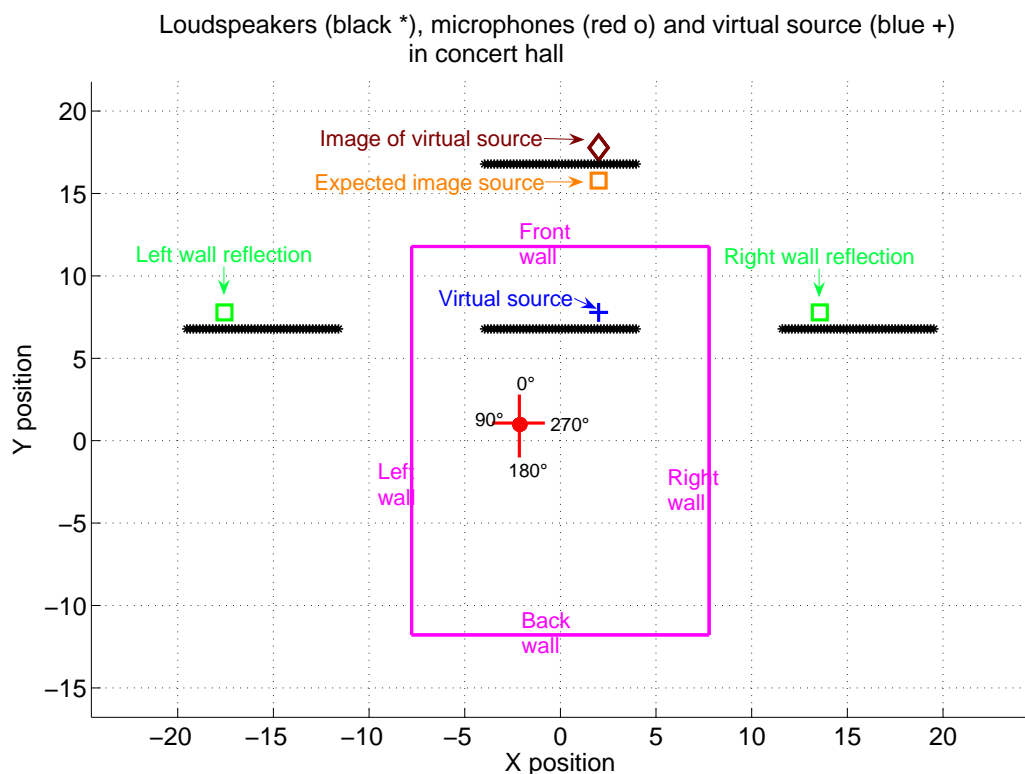


FIG. 5.15 – Illustration of the difference between the expected position of the first order reflection of an ideal source and the actual position of the image of the WFS virtual source. The positions of the image sources associated to left and right walls are correct ; however the position of the image source associated to the front wall deviates from the expected position.

Variations on the early room effect through directive source synthesis

As stated in the introduction, new developments in WFS allow for the synthesis of virtual sources displaying variable directivity patterns, based on the use of cylindrical harmonics (Cor-tee, 2004). One of the prescribed goals of this type of directivity rendering can be to approach the directivity patterns of real instruments, as described in (Warusfel *et al.*, 2004b). Another interesting prospect offered by directive source rendering lies in the possibility to create *natural variations in the manner in which the loudspeaker array illuminates the concert hall*. In other words, by varying the directivity of a virtual source, one may naturally modify the room effect perceived by the audience.

Simulations of the WFS array synthesizing a virtual dipole point source of degree -1 and degree 1 in the model concert hall are shown on figures 5.12 (middle) and 5.13 (middle). By comparing the WFS reproduction of the degree -1 dipole (figure 5.12 middle) with the ideal degree -1 dipole (figure 5.12 top), it appears that the array is able to recreate the desired at-

uation in the direct sound at the microphone position (level of direct sound 'D' 25dB below the level of the primary reflection 'L'). Some artifacts, due notably to the fact that the ground image source 'G' is out of place, remain in the ground reflection component. However, measurements show the possibility for the array to create side reflections while reducing the direct sound. This type of manipulation of the sound field is possible strictly below the aliasing frequency.

The rendering of the degree 1 dipole (figure 5.13 middle) shows that the lateral reflection 'L' that had been greatly reduced for the ideal source is present for in the array synthesis. This is due to the fact that the zero plane of the degree 1 point source array is not properly reproduced since situated outside of the bounds of the visibility window of the array. Therefore, the primary reflection of the degree 1 source off of the left wall does not exhibit the necessary level attenuation present in the 'L' reflection of figure 5.13 (top).

The reproduction of a *plane wave* synthesized in a direction perpendicular to the axis of the array and decoded on the microphone array is represented figure 5.16 (top). It appears first of all that the plane wave source favors the propagation of acoustic energy along the front/back direction (0° and 180°) while reducing the illumination of side walls. Left-wall reflections, which display high levels in the $[20^\circ - 150^\circ]$ angular quadrant for the monopole source, have been strongly attenuated (25dB attenuation of the peak level of the left hand side reflections compared to the 'L' reflection of the monopole source).

Figure 5.16 (top) also indicates the presence of acoustic energy at the microphone position incoming between 20 and 40 ms within the $[350^\circ - 140^\circ]$ angular sector. This acoustic wave corresponds to the left-wall reflection of the diffracted wave emitted by the sides of the WFS array as the plane wave passes through the finite window of the array. The fact that the diffracted wave is mainly a low frequency phenomenon (below 300Hz) causes an apparent spreading of its angular location since the lobes of the microphone are wider at low frequencies (see figure 5.3 for frequency dependence of microphone lobes).

5.4 Measurements of a WFS loudspeaker array in the Espace de Projection concert hall

This section describes measurements carried out in the Espace de Projection concert hall in IRCAM using a prototype of the third order microphone array developed by Trinnov. The WFS array is constituted of 48 electrodynamic loudspeakers. The layout of the loudspeakers and microphone array is identical to the one described on figure 5.9. Results for the measurement of a virtual monopole source, a dipole degree -1 source, a dipole degree 1 source and a plane wave are shown on the bottom of figures 5.12, 5.13, 5.16, and 5.11. The dynamics of the representation have been greatly reduced in comparison with the specular model (25dB vs

5.4. Measurements of a WFS loudspeaker array in the Espace de Projection concert hall

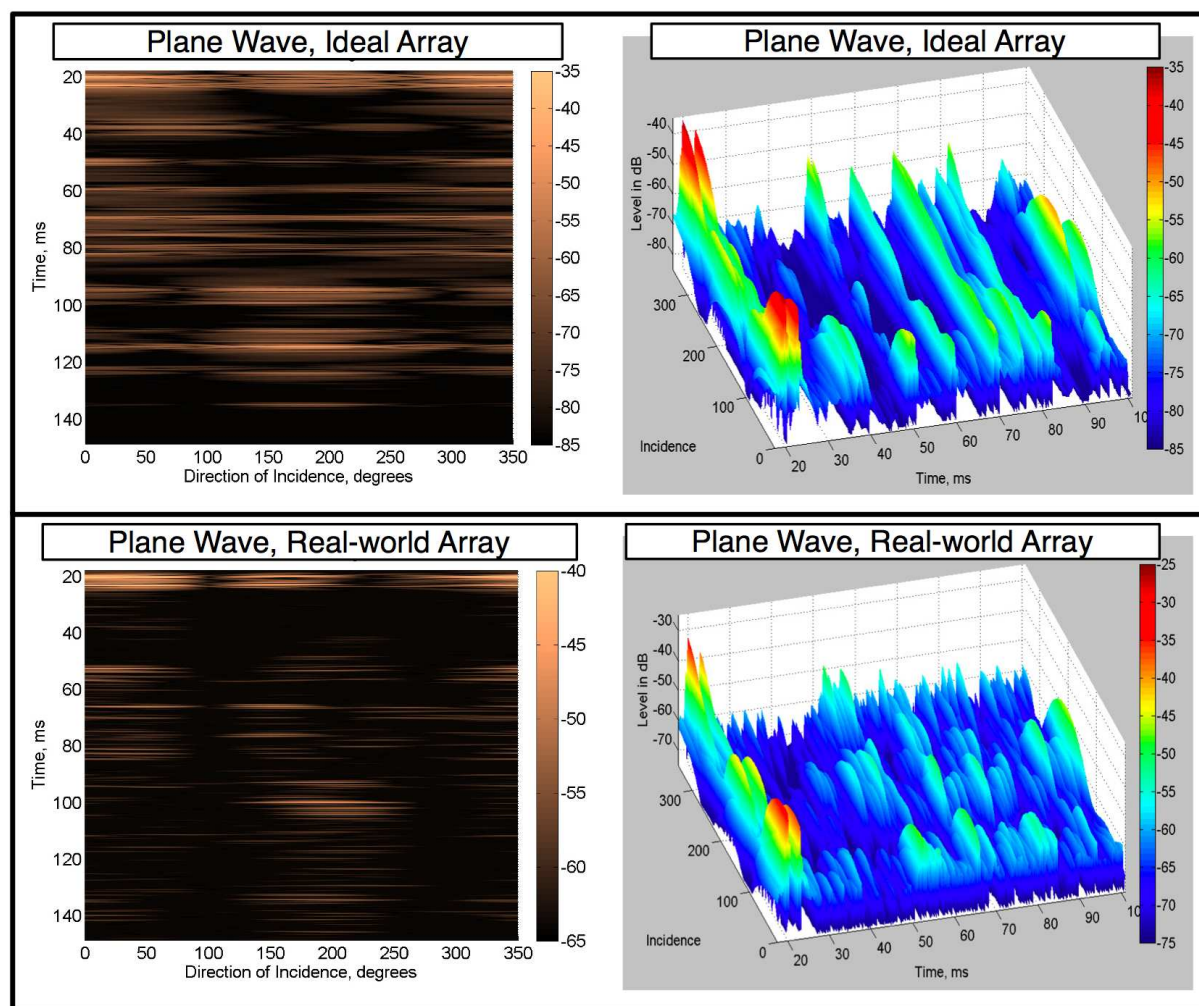


FIG. 5.16 – Amplitude of the impulse response (dB) of a **virtual plane wave synthesized in a direction perpendicular to the axis of the ideal WFS array** (top) and **virtual plane wave synthesized in a direction perpendicular to the axis of the real-world WFS array** (bottom), decoded onto 36 directions at the microphone position (with 10° increments) as a function of time (ms). **Left** : Top view, [17 ; 150]ms interval. **Right** : Side view, [17 ; 100]ms interval.

50dB for the ideal simulations) notably because of the presence of measurement noise makes the interpretation of the figures more difficult.

Figure 5.11 (bottom) shows that the main specular components 'D' and 'L' remain coherent with the ideal source for the real-world monopole source reproduction. Synthesis of the degree -1 dipole (fig. 5.12 bottom) illustrates the potential for canceling out the direct sound while conserving lateral reflection components using the WFS array. Finally, figure 5.16 (bottom) shows that the real world array synthesizing a plane wave successfully manages to concentrate acoustic energy in the front-back direction of the concert hall while attenuating reflections incoming from lateral directions at the microphone position.

The observation of these figures shows a large amount of diffuse energy (absent in the simulations) that appears as a result of the diffusiveness of the walls ; however specular components remain spatially and temporally coherent with the simulation measurements carried out in the previous section. These results confirm that the WFS array is indeed capable of varying the room effect produced in situ through the use of directive sources.

5.5 Conclusions

This article demonstrated the interest of using a high spatial resolution microphone to characterize the early room effect associated to a linear WFS array placed in a concert hall. *The array displayed the capacity to reduce contributions originating from outside of the horizontal plane, therefore furnishing the means by which to interpret the measured sound field.* The following characteristic aspects of the early room effect associated to a WFS array were underlined :

- Reflections off of surfaces parallel to the axis of the array can be expected to reach the audience area later than expected because of the cylindrical symmetry of the wavefront emitted by the WFS array.
- The synthesis of directive sources allows to create lateral reflections while eliminating the direct sound. The synthesis of plane wave sources allows to accentuate the direct sound level while attenuating lateral reflections. **The WFS array can therefore provide a natural means¹ by which to vary the room effect produced within the audience area.**

The objective results shown in this study form the potential basis for a perceptual analysis of the early room effect produced by WFS arrays. Objective criteria associated to the perception of room acoustics such as Lateral Energy Fraction and IACC are intrinsically linked to previous microphone technologies (figure 8, dummy head). *A topic for further research can be found in the conception of more advanced room acoustics criteria, based on the improved spatial resolution offered by the microphone used in this study.*

Dans le contexte de la problématique proposée, cet article valide le fait que la Wave Field Synthesis permet effectivement de faire varier l'effet de salle par un jeu sur la directivité des sources virtuelles reproduites. En effet, nous avons montré la possibilité d'atténuer le son direct tout en conservant des réflexions latérales via la synthèse d'une source à directivité dipolaire, ou au contraire de conserver le son direct en atténuant les réflexions latérales via la synthèse d'une onde plane. La méthode d'analyse

¹Without the use of a dedicated reverberation enhancement system

développée dans l'article a également permis de mettre en évidence les problèmes liés à la symétrie de révolution du banc dans un contexte en salle, qui provoque un retard supplémentaire de toutes les réflexions provenant des surfaces de la salle qui sont parallèles à l'axe du banc de haut-parleurs.

Nous pensons que les résultats de cette analyse objective de l'effet de salle dans le plan horizontal pourraient fournir une base intéressante pour une analyse perceptive de l'effet de salle engendré par le réseau WFS.

De manière plus générale, il apparaît que les critères existants qui permettent d'objectiver la perception de l'effet de salle sont intrinsèquement liés aux technologies de prises de son existantes (figure 8, tête artificielle). Nous pensons que de nouvelles voies de recherche en perception de l'acoustique des salles sont à présent ouvertes grâce à l'apparition des microphones à haute résolution spatiale.

Conclusion

Le travail mené tout au long de cette thèse a été consacré à la caractérisation et l'extension des possibilités offertes par un système de Wave Field Synthesis pour la situation de concert. Le cadre choisi a été celui du concert "mixte", dans lequel le banc de haut-parleurs WFS est placé en nez de scène et synthétise des instruments virtuels partageant l'espace acoustique de la salle avec des instrumentistes. Ce cadre particulier est né des besoins du compositeur, qui déplore "*l'impossibilité d'établir un orchestre idéal, où tous les timbres pourraient être voisins l'un de l'autre aux moments requis par la partition.*" (Boulez, 1983). La volonté affichée de ce travail est donc de fournir, par voie électroacoustique, un moyen de compléter les sections de l'orchestre aux moments requis par la partition, en utilisant le système WFS de manière à "augmenter" l'environnement de la scène par l'ajout d'instruments virtuels.

On peut observer, étant donné la situation, que si le banc WFS permettait de synthétiser à l'identique le *son direct* émis par un instrument réel à n'importe quelle position sur la scène, alors il n'y aurait, pour ainsi dire, "rien d'autre à faire". En effet, dans ce cas l'effet de salle engendré par l'instrument virtuel serait indiscernable de celui produit par un instrument réel, et l'objectif de créer l'illusion acoustique d'un instrument réel sur scène serait atteint automatiquement. Cependant, nous sommes forcés d'observer que le champ effectivement rayonné par le système WFS n'est pas conforme à celui que nous cherchons à reproduire. Par conséquent, les auditeurs situés dans la salle sont susceptibles de percevoir des différences entre les instruments réels et virtuels en termes du son direct qu'ils émettent, et de l'effet de salle qui lui est associé.

L'approche globale que nous avons adoptée, étant donné cette problématique de base, a consisté d'une part à *identifier la partie* du champ rayonné par la source virtuelle qui est *cohérente* avec la source sonore réelle cible, d'autre part à *rajouter l'information manquante*, en

s'appuyant sur la salle.

Avant d'aborder cette problématique, nous avons toutefois fait le choix de poser le contexte existant au début de la thèse. Dans le chapitre 1, nous avons donc fait le point sur l'état de l'art de la synthèse du son direct et de l'effet de salle en WFS. L'objectif du chapitre a été de présenter cette technique de reproduction sonore, qui se situe aux confluent de l'acoustique, du traitement de signal, et de la psychoacoustique. Nous sommes volontairement resté concis sur certains points, en essayant de renvoyer à des références plus complètes lorsque cela convenait. Sur d'autres points, nous avons au contraire approfondi la description, de manière à introduire les outils théoriques nécessaires à la construction des méthodes et concepts développés dans le reste du document.

Au cours des chapitres suivants, nous avons distingué deux axes principaux pour le travail : d'une part *l'analyse et le contrôle du rayonnement du banc*, et d'autre part *la caractérisation et compensation de l'effet de salle*. Ces deux axes ont été envisagés dès le début de la thèse en parallèle.

Dans le deuxième chapitre, nous nous sommes intéressés au développement d'une méthode de prédiction de rayonnement d'un banc WFS, à partir d'une mesure d'un seul ou plusieurs de ses haut-parleurs constitutifs. Cette méthode répond aux besoins de la caractérisation du système, voire de sa calibration, que ce soit du point de vue du son direct ou de l'effet de salle. En effet, nous avons constaté très vite, au début de la thèse, que la caractérisation du système en puissance par des méthodes classiques (en vue de calibrer son effet de salle tardif associé) posait des problèmes pratiques importants, du fait de l'envergure du banc de haut-parleurs, de la cohérence de son champ émis, et de la variabilité de sa géométrie suivant le contexte d'utilisation choisi. Nous nous sommes donc penchés sur le développement d'une méthode de prédiction de rayonnement permettant de réaliser cette tâche. Un bénéfice supplémentaire de la méthode réside dans la simplification de la phase de calibration du son direct, qui nécessite actuellement de mesurer le banc entier en champ libre, ce qui pose des problèmes pratiques importants.

Nous avons exploré trois approches différentes permettant d'extrapoler le champ émis par un haut-parleur à partir d'une mesure en chambre sourde, suivant la présence ou non d'obstacles dans l'environnement de mesure. La méthode la plus adaptée pour notre situation a ensuite été utilisée pour extrapoler le rayonnement de deux transducteurs typiques utilisés dans le contexte de la Wave field Synthesis (un de type électrodynamique et un de type MAP). En comparant cette extrapolation avec la mesure de ces transducteurs en champ libre, nous avons vérifié que la méthode était efficace jusqu'à la fréquence de repliement moyenne associée aux bancs WFS que nous utilisons (environ 1.5kHz). En d'autres termes, la méthode est bien adaptée à la cali-

bration du système. Cependant, nous avons constaté par simulation que la méthode de mesure permettrait de monter plus haut en fréquence (jusqu'à 3kHz) en utilisant les 24 canaux de mesure disponibles sur une sphère de rayon plus faible. Le réaménagement du système de mesure suivant ces critères appartient aux travaux prospectifs que nous voyons pour ce travail. Nous voyons également comme travaux prospectifs suite à cette étude son application au calcul des filtres d'égalisation.

Au troisième chapitre, nous avons abordé le problème de l'analyse et du contrôle de directivité de rayonnement du banc de haut-parleurs. L'idée, d'une part, était d'apporter une solution permettant de compléter les timbres de l'orchestre par le biais du système WFS, en évitant le problème de "dissolution de l'identité de la source réelle par le haut-parleur". Pour tenter de résoudre ce problème, nous avons repris et approfondi un ensemble de techniques permettant d'associer une directivité de rayonnement aux sources virtuelles reproduites par WFS, en adaptant des outils de manipulation de directivité dérivés du formalisme Ambisonique (rotation, synthèse de sources hyperdirectives) à la WFS. Nous avons ensuite introduit un critère permettant d'approximer simplement l'ordre maximal de recomposition d'une source directive en WFS, puis évalué la qualité de reconstruction du son direct des sources directives. Enfin, nous avons utilisé le formalisme de décomposition du champ sonore développé au chapitre 2 pour analyser le rayonnement tridimensionnel effectif d'un banc linéaire WFS.

Parmi les suites possibles du travail présenté au chapitre 3, nous voyons deux aspects principaux qui pourraient être intéressants dans le cadre proposé de la situation de concert. D'une part, la synthèse du rayonnement d'instruments réels, dont les caractéristiques de directivité sont asservies au signal émis. Cette voie de recherche fait partie des axes proposés pour le travail de Joseph Sanson, qui commence actuellement une thèse dans l'équipe Acoustique des Salles de l'Ircam dont le but entre autres sera d'étudier dans un cadre plus formel les possibilités offertes par un système WFS pour le contrôle des variations temporelles des modes de directivité d'un instrument réel. D'autre part, nous voyons un travail prospectif possible dans la décomposition du champ rayonné telle qu'elle est décrite en section 3.7.1, qui pourrait permettre de former des fonctions de directivité tridimensionnelles du banc WFS qui soient orthogonales entre elles. Cette orthogonalité dans la directivité de rayonnement du banc fournirait alors des modes d'excitation de la salle qui soient découplés les uns des autres, permettant par exemple de gérer plus finement la répartition temps-espace-fréquence de l'effet de salle précoce. Cette piste paraît intéressante à explorer dans le cadre de travaux futurs.

Dans le chapitre 4, nous avons abordé le problème de la cohérence de l'effet de salle tardif associé à la source virtuelle. En effet, le fenêtrage de la directivité par le banc de haut-parleurs implique que les sources virtuelles émettent un spectre de puissance variable dans la salle suivant leur position ou orientation sur la scène. Nous avons donc exposé deux méthodes permettant de compenser les niveaux d'énergie dans la salle, de manière à rétablir une continuité

perceptive dans l'effet de salle tardif associé aux sources virtuelles.

Nous avons introduit une première méthode de compensation qui adapte l'architecture décrite dans (Jot *et al.*, 1998) au cadre de la WFS. Elle vise à compenser l'effet de salle en un point de référence de la salle, basé sur une qualité acoustique cible qui peut être définie suivant des critères perceptifs ou physiques. Dans l'approche utilisant un critère physique, on tente de resynthétiser l'effet de salle associé à un instrument réel placé sur la scène. Nous avons alors décrit une méthode modulaire permettant de resynthétiser la qualité acoustique cible associée à n'importe quel instrument de musique à la position de référence. Cette méthode s'appuie sur une caractérisation d'un ensemble de positions dans la salle utilisant une source à directivité contrôlée, ainsi que l'analyse en chambre sourde du rayonnement des instruments que l'on souhaite reproduire, ce que nous avons décrit brièvement. Au niveau de la compensation de l'effet de salle, la méthode s'appuie soit sur un système périphérique d'enceintes, soit sur l'utilisation de sources hyperdirectives permettant de diffuser la partie tardive de l'effet de salle vers les murs de la pièce, de manière à recréer un champ diffus provenant de l'arrière et des côtés de la zone d'écoute. La méthode de compensation demeurerait à être testée en pratique, ce qui n'a pas été possible dans le temps imparti pour cette thèse. Malgré tout, nous avons pu constater l'intérêt de la technique de diffusion d'effet de salle utilisant des sources hyperdirectives lors d'écoutes informelles et sessions de travail avec des compositeurs et réalisateur en informatique musicale. Une validation de cette technique par des critères objectifs reste également être effectuée.

Une deuxième méthode de compensation "simplifiée" a également été décrite, basée sur une caractérisation de la puissance effectivement émise par le système. Nous avons analysé un ensemble de méthodes existantes permettant (ou non) de réaliser cette caractérisation de puissance, au terme de quoi nous avons proposé une nouvelle méthode basée sur la prédiction du rayonnement du banc de haut-parleurs sur une sphère entourant celui-ci. Cette méthode a été retenue comme la plus efficace des méthodes étudiées, car elle permet de prendre en compte la directivité 3D des transducteurs du banc WFS tout en fournissant une estimation large bande de sa puissance émise. A la fin du quatrième chapitre, nous avons décrit un concept de compensation de l'effet de salle utilisant les résultats de la caractérisation de puissance émise par le banc WFS, et décrit une architecture de système permettant d'intégrer la compensation au système de rendu WFS. Il resterait à valider cette méthode de compensation par des mesures en divers points de la salle de concert, de manière à vérifier que l'on est bien capable d'assurer une compensation efficace en zone étendue.

Le chapitre 5 a été l'occasion, enfin, de valider le contrôle de la directivité du banc de haut-parleurs en conditions de diffusion en salle. En utilisant un microphone à haute résolution spatiale en combinaison avec un modèle simple de la salle de concert, nous avons pu vérifier que le banc de haut-parleurs permettait de faire varier l'effet de salle précoce effectivement produit dans le plan horizontal de la zone d'écoute. Nous avons en effet constaté que le contrôle

Conclusion

de directivité en WFS permettait de réduire le niveau du son direct tout en conservant des réflexions latérales, ou au contraire de conserver le son direct tout en atténuant le niveau des réflexions latérales dans la salle.

Les résultats des recherches conduites dans le cadre de cette thèse, en collaboration avec l'ensemble des membres de l'équipe d'Acoustique des Salles de l'Ircam, seront mis en pratique dans les années à venir dans le cadre de pièces musicales qui seront jouées en concert avec le système. La première commande officielle de pièce utilisant le système en concert a d'ailleurs été lancée à l'Ircam et devrait être diffusée courant 2008. Nous espérons par ailleurs que dans le futur ces techniques seront utilisées en dehors de l'Ircam, ainsi que dans d'autres domaines du spectacle vivant.

Techniques de Mesure en Acoustique des Salles

A.1 Mesure d'une réponse impulsionnelle de salle

A.1.1 Principe général

Cette première section décrit l'opération de mesure d'une réponse impulsionnelle, qui est un outil indispensable pour l'*auralisation* et l'analyse de l'acoustique d'une salle. En effet, en supposant que l'on cherche à décrire la réponse en un point M d'une salle à une excitation quelconque en un point P, le procédé le plus simple consiste à convoluer l'excitation avec la réponse impulsionnelle de la salle entre P et M.

Soit $h_{salle}^{P \rightarrow M}(t)$ la réponse impulsionnelle du canal acoustique entre P et M. $h_{salle}^{P \rightarrow M}(t)$ décrit théoriquement la réponse de la salle à une impulsion de Dirac $\delta(t)$ (signal de durée théoriquement nulle et d'énergie non-nulle finie). En pratique, pour simuler une impulsion de Dirac $\delta(t)$ il est commun d'utiliser un haut-parleur. Etant donné que tout haut-parleur produit une distorsion linéaire du signal d'entrée, on doit donc pré-mesurer sa réponse impulsionnelle $h_{HP}(t)$ avant la mesure en salle de manière à pouvoir éliminer son influence par la suite.¹ Cette mesure préliminaire se fait en général en chambre anéchoïque, avant la campagne de mesure en salle.

En supposant que les conditions de température et d'humidité ne varient pas au cours de la mesure, et que le microphone ne bouge pas, on peut considérer que le système mesuré est LIT (Linéaire et Invariant dans le Temps). Dans ce cas on peut affirmer que la réponse mesurée en M

¹On supposera pour les besoins de cette explication que le microphone de mesure placé en M est idéal, i.e. qu'il ne produit aucune distorsion en amplitude ou en phase du signal mesuré.

h_{mes} résulte d'une opération de convolution entre la réponse du haut-parleur et la réponse de la salle :

$$h_{mes}(t) = \int_{-\infty}^{+\infty} h_{salle}^{P \rightarrow M}(t') h_{HP}(t - t') dt' \quad (\text{A.1})$$

Soit H_{mes} la transformée de Fourier de h_{mes} , $H_{salle}^{P \rightarrow M}$ la transformée de Fourier de $h_{salle}^{P \rightarrow M}$ et H_{HP} la transformée de Fourier de h_{HP} . En vertu de l'interchangeabilité de l'opération de convolution dans le domaine temporel et l'opération de multiplication dans le domaine de Fourier, l'équation A.1 devient alors :

$$H_{mes}(\omega) = H_{salle}^{P \rightarrow M}(\omega) \cdot H_{HP}(\omega) \quad (\text{A.2})$$

En supposant que la réponse fréquentielle H_{HP} du haut-parleur ne comporte pas de zéros dans la gamme de fréquences étudiées, on obtient alors la réponse de la salle $H_{salle}^{P \rightarrow M}$ en divisant H_{mes} par H_{HP} (opération dite de "déconvolution").

A.1.2 Choix du signal de mesure

En pratique, les mesures en salle comportent systématiquement un bruit additionnel, et l'utilisation d'une impulsion de type bruit blanc concentrée en temps ne permet pas d'obtenir un niveau de signal sur bruit suffisant en utilisant un haut-parleur traditionnel. On emploie par conséquent des signaux large bande, qui tendent à répartir l'énergie sur un plus grand support temporel de manière à diminuer l'influence du bruit à une fréquence donnée. Les signaux les plus utilisés à cet égard sont les séquences pseudo-aléatoires MLS [Maximum Length Sequences (Schroeder, 1979)] et les sinus balayés linéairement ou logarithmiquement en fonction de la fréquence (sweep linéaires et logarithmiques).

Dans (Hulsebos, 2004), l'auteur fait une comparaison de différents signaux d'excitation pour essayer de déterminer le plus adapté à la mesure en salle. Les signaux d'excitation à spectre plat (MLS blancs, bruit blanc à phase aléatoire, sweep linéaire) sont écartés du fait qu'ils privilégient le rapport signal sur bruit (S/B) en hautes fréquences plutôt qu'en basses fréquences. Or la plupart des bruits de mesure (électriques ou ambiants) ont typiquement un spectre rose ; ainsi l'utilisation de signaux d'excitation à spectre rose permet d'augmenter le rapport S/B dans les zones fréquentielles d'intérêt. Par ailleurs, les signaux à spectre plat conduisent à un déséquilibre en termes de sollicitation des excitateurs du haut-parleur (10-15% provenant du woofer contre 80-85% provenant du tweeter), ce qui peut mener à une surchauffe du tweeter. De ce fait, l'auteur préconise l'utilisation de signaux d'excitation à spectre rose comme les MLS rose, bruits roses à phase aléatoire et sweep logarithmique.

Parmi tous les signaux étudiés, le sweep logarithmique est montré comme étant le plus efficace en termes de rapport S/B (5-6dB plus efficace qu'un MLS rose et 7-8dB plus efficace que le bruit rose à phase aléatoire). Notons que ce signal est également plus robuste aux variations

temporelles (déplacements du microphone par exemple) du fait que son énergie dans les hautes fréquences est concentrée sur un support temporel faible.

En lumière de ces résultats, toutes les mesures réalisées dans la salle de concert de l'Ircam (l'Espace de Projection) ont été conduites en utilisant des sweeps logarithmiques (chapitre 4 et 5).

A.1.3 Déconvolution circulaire

Soit $h_{sweep}(t)$ la réponse impulsionnelle d'un signal sweep logarithmique et $h_{sweepMes}(t)$ le signal mesuré à la position M lorsqu'un haut-parleur émet le signal sweep à la position P. On suppose que l'on a fait l'acquisition de ces signaux et leur conversion A/D dans des conditions respectant le critère de Shannon. On emploie alors l'algorithme de transformée Fourier rapide (Fast Fourier Transform - FFT) et son inverse (IFFT) pour calculer en pratique la réponse h_{mes} qui figure dans l'équation théorique A.1 :

$$h_{mes}(t) = \text{IFFT} \left(\frac{\text{FFT}(h_{sweepMes}(t))}{\text{FFT}(h_{sweep}(t))} \right) \quad (\text{A.3})$$

Cette opération se nomme "déconvolution circulaire" du fait des propriétés de circularité de la transformée de Fourier. La circularité implique qu'il y a un risque de repliement temporel si le signal d'excitation est trop court par rapport au signal mesuré. On doit donc choisir le stimulus de telle sorte que sa durée excède le temps de réverbération de la salle. De plus, on choisira cette durée de telle manière à ce que la réponse s'étale sur un nombre d'échantillons qui soit un multiple de 2, pour des raisons d'efficacité de l'algorithme FFT (typiquement entre 2^{14} et 2^{17} échantillons pour une fréquence d'échantillonnage de 48kHz).

A.1.4 Analyse de distorsions

Dans le cas où le signal mesuré contient des distorsions non-linéaires (i.e. si le haut-parleur distord pendant la mesure) l'emploi d'une déconvolution linéaire² du sweep logarithmique permet d'isoler les composantes de distorsion, qui se retrouvent alors décalées en amont du signal utile lors de l'opération (Farina, 2000).

²Convolution dans le domaine temporel par le filtre inverse h_{sweep}^{inv} calculé de manière à ce que $h_{sweep}^{inv}(t) * h_{sweep}(t) = \delta(t)$

A.1.5 Généralisation de la méthode classique de mesure de réponse impulsionnelle

Dans la relation A.2 on suppose implicitement une directivité monopolaire, à la fois pour le haut-parleur et le microphone. En réalité, le champ émis par une source typique (haut-parleur ou instrument) dans une salle de concert possède une directivité de rayonnement complexe, et la tête d'un auditeur placé dans une salle produit un filtrage directionnel du son incident (représenté par un ensemble discret de fonctions de transfert nommées Head Related Transfer Functions - HRTF). Par conséquent la mesure conventionnelle d'une source monopolaire captée par un microphone monopolaire ne suffit pas pour décrire intégralement le filtrage temps-fréquence-espace qui est propre à l'écoute dans une salle donnée.

Gerzon en 1975 fut le premier à suggérer l'utilisation de techniques de captation multicanales (Ambisonique) pour constituer des archives de l'acoustique de salle de concerts connues (Gerzon, 1975). Cette idée a été reprise et étendue récemment par Farina et al. (Farina *et al.*, 2006) qui propose d'utiliser à la fois des sources directives et des microphones directifs pour enregistrer les réponses de salles. En effet, l'état de l'art en matière de réseaux de haut-parleurs permet de recréer des sources directives qui approchent les premiers ordres la base des harmoniques sphériques (Warusfel & Misdariis, 2001; Kassakian & Wessel, 2004; Warusfel *et al.*, 2004b). On dispose également de réseaux microphoniques permettant de capter et décomposer la réponse impulsionnelle en un point sur la base des harmoniques sphériques jusque des ordres élevés (Meyer & Elko, 2002; Abhayapala & Ward, 2002; Daniel & Moreau, 2004; Laborie *et al.*, 2004a; Rafaely, 2005). On peut ainsi imaginer stocker une image beaucoup plus précise que celle fournie par un couple haut-parleur/microphone omnidirectionnels utilisés traditionnellement.

En possédant de telles traces de l'image spatio-temporelle d'une salle aux positions d'émission et de réception, il est possible d'envisager une auralisation de haut niveau de la salle de concert. En effet, la directivité d'un instrument de musique ou d'un haut-parleur peut être décomposé selon la base des harmoniques sphériques, (Giron, 1996; Caulkins & Girier, 2007), impliquant la possibilité de décomposer un enregistrement en un flux omni, un flux dipole, quadripole etc..et de recréer par la suite sa réponse impulsionnelle spatiale à la position enregistrée dans la salle par simple convolution. Notons que la directivité mesurée pour l'instrument est dépendante de la fréquence, ce qui implique une décomposition différente pour différents modes de jeu (Meyer, 1972; Weinreich, 1997; Derogis, 1999). Ainsi nous sommes dès aujourd'hui en mesure de pouvoir recréer ex-situ une impression spatiale fidèle du champ perçu à une position d'écoute donnée dans une salle de concert lors d'un concert ou plusieurs sources instrumentales jouent sur scène³.

³à condition d'avoir enregistré les réponses impulsionnelles d'ordre élevé à toutes les positions occupées par

Au moment de la restitution, le codage en ambisonique ordre élevé de la réponse impulsionnelle à la position d'écoute permet d'envisager une diffusion du contenu par le biais d'une méthode type synthèse de champ sonore (WFS ou Ambisonique), ou par le biais d'une diffusion de type binaurale recomposée à partir d'une somme d'harmoniques sphériques pondérées (Evans *et al.*, 1998).

A.2 Mesure de réverbération

La réverbération est un élément clé de l'acoustique des salles et joue un rôle important à divers niveaux. De manière très simple, la réverbération décrit la décroissance de l'énergie dans une salle suite à l'extinction d'une source sonore. Le critère objectif le plus souvent utilisé pour décrire la réverbération est le *temps de réverbération*, noté Tr_{60} . Ce critère, mis au point par Wallace Clement Sabine au début du siècle (Sabine, 1900), désigne le temps mis pour que le niveau d'énergie acoustique présente dans une salle chute de 60dB. Le Tr_{60} est considéré encore aujourd'hui comme étant le critère le plus important de description d'une salle du fait qu'il est théoriquement⁴ indépendant de la position de l'auditeur dans la salle, qu'il peut être mesuré ou prédit précisément avec peu de moyens, et que l'on dispose de données précises et abondantes sur des temps de réverbérations dans la plupart des salles de concert existantes (Kuttruff, 1991).

Pour pouvoir décrire la réverbération, il est de commun usage d'employer des modèles stochastiques. En effet, étant donné que le nombre de réflexions présentes dans une salle croît avec le carré du temps (c.f. glossaire : *densité de réflexions*), la description de la réverbération est quasi-irréalisable par un modèle déterministe. En prenant l'exemple d'un modèle déterministe par sources images, Baskind montre que pour décrire la réverbération en un point d'une salle de $250m^3$ au bout de 1s, il faudrait pouvoir calculer les contributions d'environ 600000 sources distinctes (Baskind, 2003). Cette tâche est difficile, voire impossible, dans le contexte actuel du traitement informatique, d'où la nécessité de faire appel à des modèles statistiques pour décrire l'effet de salle.

En pratique, il est nécessaire de faire appel à l'analyse statistique de l'effet de salle à partir du moment où il y règne un champ diffus. Dans le domaine temporel, on montre que l'état de champ diffus est atteint après un certain temps dénommé *temps de mélange* (Polack, 1988). Dans le domaine fréquentiel, il a été montré qu'un champ diffus règne à partir du moment où le nombre de modes excités à une fréquence donnée devient supérieur à 3, c'est à dire au delà de la *fréquence de Schroeder* (Schroeder, 1987). Notons toutefois que la notion de temps de mélange et de fréquence de Schroeder ne sont pas des limites *strictes*, mais plutôt des ordres de

les instruments sur scène

⁴En pratique pour que le Tr_{60} soit indépendant de la position, il faut que la salle soit suffisamment diffusante, et qu'il n'y ait pas de volumes couplés comme des balcons ou des cages de scène dans la salle.

grandeur qui renseignent sur la salle.

Le processus de réverbération décrit le vidage d'une réserve d'énergie (une salle remplie de son) dans un puits d'énergie (les frontières absorbantes et le contenu absorbant de la salle). Ce type de processus est décrit par une fonction exponentielle décroissante, et dans le cas d'une salle qui se vide de son énergie on a :

$$E(t) = E_0 \cdot e^{-\frac{t}{\tau}} \quad (\text{A.4})$$

où E représente la densité d'énergie (énergie par unité de volume, exprimée en J/m^3) à l'instant t , E_0 représente la densité d'énergie à l'instant initial et τ représente le taux de décroissance de l'énergie dans la salle. Alors que la durée de réverbération dépend du niveau initial d'énergie dans la salle, la constante τ caractérise intrinsèquement la salle. Elle peut être exprimée en termes du temps ($t_2 - t_1$) mis pour que la densité d'énergie présente dans la salle diminue de E_1 à E_2 (Cremer *et al.*, 1978) :

$$\tau = \frac{t_2 - t_1}{\ln(E_1/E_2)} \quad (\text{A.5})$$

Dans des salles de concert typiques, les rapports d'énergie E_1/E_2 peuvent atteindre des ratios aussi élevés que 1,000,000 :1 si on prend E_2 proche du niveau de bruit de fond. C'est pour cette raison que Sabine a choisi de définir la donnée caractéristique Tr_{60} comme étant le temps que met l'énergie sonore pour diminuer à un millionième de sa valeur initiale. Il est important de noter que le Tr_{60} définit non pas un 'temps' mais plutôt un 'taux de décroissance' du son.

En introduisant le rapport $E_2/E_1 = 10^6$ dans l'équation A.5, on obtient la relation entre τ et Tr_{60} :

$$\tau = \frac{Tr_{60}}{6 \cdot \ln(10)} = \frac{Tr_{60}}{13.8}. \quad (\text{A.6})$$

Considérons une salle excitée en continu par une source sonore. L'équation qui décrit le bilan de puissance dans la salle à un instant donné peut s'écrire :

$$V \frac{dE}{dt} = W - \frac{VE}{\tau}, \quad (\text{A.7})$$

où E représente la densité d'énergie présente dans la salle, W représente la puissance émise par la source sonore en régime stationnaire et V le représente le volume de la salle. On notera que le bilan A.7 est valable uniquement dans le cas où l'on suppose l'additivité des énergies (ou puissances) envoyées dans la salle.

Lorsque la salle atteint un régime stationnaire, la valeur de la densité d'énergie E_∞^5 est telle

⁵On emploie la notation E_∞ pour indiquer que théoriquement cette densité d'énergie n'est atteinte qu'au bout d'un temps théoriquement infini. Dans cet état asymptotique, l'énergie est absorbé par la salle aussi vite qu'elle est créé par la source.

que (Cremer *et al.*, 1978) :

$$E_{\infty} = \frac{W\tau}{V} = \frac{W}{V} \frac{Tr_{60}}{13.8}. \quad (\text{A.8})$$

Cette équation montre la dépendance du niveau d'énergie en régime stationnaire vis à vis du temps de réverbération et du volume de la salle. En vertu de la fameuse "loi de Sabine" :

$$Tr_{60} = 0.163 \cdot \frac{V}{\sum_k \alpha_k S_k}, \quad (\text{A.9})$$

où $\sum \alpha_k S_k$ représente l'absorption totale effectuée par les surfaces S_k de la salle de coefficients d'absorptions α_k , si l'on rajoute du matériau absorbant dans la salle alors on fait baisser son temps de réverbération. Or, en vertu de l'équation A.8, cette baisse de temps de réverbération s'accompagnera inévitablement d'une baisse de niveau pour en régime stationnaire.

De même, le niveau d'énergie en régime stationnaire est inversement proportionnel au volume de la salle ; ainsi si l'on souhaite maintenir un niveau convenable il s'agit d'augmenter le temps de réverbération pour des salles plus grandes.

Soit E_- la décroissance de densité d'énergie suite à une interruption soudaine de la source alors que la salle est en régime stationnaire. En vertu de l'équation A.4 et A.8 on peut écrire :

$$E_-(t) = \frac{W\tau}{V} \cdot e^{-\frac{t}{\tau}}. \quad (\text{A.10})$$

Il a longtemps été coutume d'employer des bruits blancs pour faire des mesures en salle, car ces signaux permettent de respecter l'hypothèse d'additivité des puissances sous-jacente aux équations A.7, A.9, A.8, et A.10. L'inconvénient réside dans le fait qu'en pratique, il est impossible de mesurer directement l'énergie présente dans la salle : la mesure de l'énergie se déduit forcément d'une mesure de pression. Or, étant donné que l'on emploie des signaux aléatoires, les amplitudes et phases initiales des modes excités dans la salle au moment de l'extinction de la source varient d'une mesure à une autre (Schroeder, 1965). On observe donc des variations de la courbe de décroissance, bien que la position de la source et du microphone de mesure ne varient pas.

La technique employée pendant longtemps pour pallier le problème de variabilité dans les courbes de décroissance énergétiques consistait à répéter les mesures un grand nombre de fois et moyenner les courbes obtenues, de manière à lisser les variations et s'approcher de la moyenne d'ensemble de la décroissance énergétique. Cette méthode de mesure de réverbération contraignante en temps et peu précise a pu par la suite être remplacée par une mesure unique lorsque Schroeder montra le lien entre la décroissance d'énergie suite à une excitation en régime stationnaire et la décroissance d'énergie suite à une excitation impulsionnelle ?? . Au lieu d'effec-

tuer des moyennes d'ensemble, Schroeder remarqua que la véritable décroissance énergétique (Energy Decay Curve notée EDC) s'obtient avec une précision bien supérieure en effectuant une intégration rétrograde du carré de la réponse impulsionnelle du système (Integrated Square Impulse Response Method - ISIR). Ce résultat, dont le détail de calcul se trouve dans (Schroeder, 1965), s'écrit de la manière suivante :

$$EDC_h(t) = \langle E_-(t) \rangle = W_b \int_t^\infty h^2(\tau) d\tau \quad (\text{A.11})$$

où $\langle E_-(t) \rangle$ désigne la moyenne d'ensemble de la décroissance d'énergie mesurée dans la salle, h représente la réponse impulsionnelle du système complet (filtre passe-bande du bruit, amplificateurs, transducteurs, salle), W_b désigne la puissance par unité de bande fréquentielle du signal de bruit blanc théoriquement interrompu au temps $t = 0$ et EDC_h désigne la courbe de décroissance intégrée obtenue à partir de h . Ainsi la moyenne d'ensemble de la décroissance d'énergie à un temps t suivant un régime stationnaire est égale au carré de la réponse impulsionnelle intégrée entre t et “ $+\infty$ ” (en pratique, quelque secondes), à condition que le spectre de la réponse impulsionnelle soit identique à la densité spectrale de puissance du bruit employé pour faire l'excitation en régime stationnaire.

Schroeder propose initialement d'utiliser des signaux de type “noise burst” pour réaliser les mesures de réponses impulsionnelles en salle. Par la suite, l'utilisation de signaux de type Maximum Length Sequence (MLS) ou, plus tard, de type sweep linéaires/logarithmiques pour effectuer les mesures de réponses impulsionnelles (c.f. sec. A.1.2) permet d'améliorer les rapports signal/bruit insuffisants en BF observés lors de l'utilisation de signaux de type “noise-burst” dans le cadre de caractérisation de courbes de décroissance énergétique (Schroeder, 1979).

La courbe de décroissance énergétique EDC_h , une fois mesurée, permet d'estimer le temps de réverbération de la salle par le biais d'une régression linéaire sur la pente dans la zone temporelle de la réponse correspondant à une décroissance logarithmique. Ainsi la norme ISO (ISO3382, 1993) définit le temps de réverbération comme une régression linéaire sur la pente de la courbe de décroissance intégrée de la réponse impulsionnelle dans la plage énergétique de $-5dB$ à $-35dB$ par rapport au niveau maximal. Cette définition s'écarte de la définition donnée par Sabine (l'intégration part de $-5dB$ au lieu de $0dB$), car la formule définie par Sabine suppose implicitement que la salle est idéalement diffuse.

A.3 Mesure de la puissance émise par une source acoustique

Cette annexe regroupe des techniques connues de mesure de la puissance émise par une source acoustique.

A.3.1 Mesure de puissance en chambre réverbérante

Soit $\{M_k(\vec{r}_k)\}_{k \in \{1,2,\dots,N\}}$ un ensemble de positions de mesure dans une chambre réverbérante. On suppose qu'une source sonore émet un signal stationnaire de puissance W (que l'on cherche à estimer) dans la chambre, et que la salle est en régime stationnaire au moment de la mesure. Par construction, la chambre réverbérante favorise la formation d'un champ diffus dans lequel le champ acoustique en un point M_k de la salle peut être considéré comme un assemblage d'ondes planes de direction de provenance équiprobable, chaque onde étant caractérisée par sa direction définie par les angles θ et ϕ , son amplitude de pression $A(\vec{r}_k, \theta, \phi)$ et son intensité $\frac{|A(\vec{r}_k, \theta, \phi)|^2}{2\rho c}$.

En d'autres termes, la pression totale en M_k s'écrit sous la forme d'une superposition de chacune des ondes planes :

$$p(\vec{r}_k, t) = \int_0^{2\pi} d\phi \int_0^\pi A(\vec{r}_k, \theta, \phi) e^{-i\vec{k} \cdot \vec{r}_k + i\omega t} \sin(\theta) d\theta \quad (\text{A.12})$$

et la densité d'énergie en régime stationnaire $E_\infty(M_k)$ s'écrit comme la somme des densités d'énergie des ondes planes (Bruneau, 1998) :

$$E_\infty(\vec{r}_k) = \frac{1}{2\rho c^2} \int_0^{2\pi} d\phi \int_0^\pi |A(\vec{r}_k, \theta, \phi)|^2 \sin(\theta) d\theta = \frac{p_{rms}^2(\vec{r}_k)}{\rho c^2}, \quad (\text{A.13})$$

où p_{rms} représente la pression quadratique moyenne (Root Mean Square) mesurée en M_k . La pression quadratique moyenne est utilisée pour mesurer l'amplitude de pression dans des champs stationnaires, qui nécessitent d'être caractérisés par des moyennes temporelles de durée longue (en pratique, quelques secondes) :

$$p_{rms}(\vec{r}_k) = \left(\overline{p^2(\vec{r}_k, t)} \right)^{\frac{1}{2}} = \sqrt{\frac{1}{T} \int_0^T p^2(\vec{r}_k, t) dt}, \quad (\text{A.14})$$

où $p(\vec{r}_k, t)$ représente la pression instantanée mesurée en M_k à l'instant t .

En admettant l'hypothèse de champ parfaitement diffus selon laquelle les ondes planes constituant le champ de pression sont distribuées de manière parfaitement aléatoires (en direction et en phase), l'amplitude $A(\vec{r}_k, \theta, \phi)$ est indépendante de \vec{r}_k , θ et ϕ . Dans ces conditions, et d'après l'équation A.13, on a alors :

$$E_\infty = \frac{p_{rms}^2}{\rho c^2}. \quad (\text{A.15})$$

De ce fait, et en vertu de l'équation A.8, on obtient :

$$W = \frac{13,8V}{\rho c^2 T r_{60}} p_{rms}^2, \quad (\text{A.16})$$

Cependant, étant donné que le champ n'est jamais parfaitement diffus dans la salle, on observe des non-linéarités dans les courbes de décroissances mesurées (Kawakami & Yamaguchi, 1981), ce qui invite à employer des moyennes de mesures à diverses positions dans la salle pour améliorer l'estimation des quantités observables. Ainsi, selon la norme ISO 3741, le niveau de puissance acoustique d'une source sonore se déduit d'une mesure de pression par tiers d'octave à plusieurs positions dans une chambre réverbérante.

En faisant la moyenne sur les N points de mesure disponibles, on obtient :

$$W = \frac{13,8V}{\rho c^2} \frac{1}{N} \sum_{k=1}^N \frac{p_{rms}^2(\vec{r}_k)}{T r_{60}(\vec{r}_k)}, \quad (\text{A.17})$$

L'évaluation de la moyenne spatiale de p_{rms} doit être restreinte aux zones de la pièce où un champ diffus existe (c.f. glossaire : *diffusivité*). Le standard international recommande donc de choisir les points de mesure le plus loin possible de la source de manière à réduire le rapport son direct/champ réverbéré, et loin de toute paroi et/ou diffuseur (au moins une demi-longueur d'onde). Malgré ces précautions on peut remarquer une variabilité importante dans les mesures de pression quadratique moyenne (Penna, 1993) à différentes positions, ce qui invite à investiguer d'autres méthodes pour estimer la puissance de la source acoustique.

A.3.2 Mesure de puissance basée sur le relief de décroissance énergétique de la réverbération.

Etant donné les variabilités observables dans les mesures de pression quadratique moyenne faites en chambre réverbérante, on propose d'utiliser les propriétés statistiques des courbes de décroissances énergétiques mises en évidence dans la partie A.2 de manière à améliorer l'estimation de la puissance de la source sonore. En vertu de l'équation A.11, on peut remplacer la mesure de p_{rms}^2 en régime stationnaire par une valeur estimée en extrapolant la courbe de décroissance intégrée $EDC_h(t)$ jusqu'au temps zéro. Cette extrapolation se fait par le biais d'une régression linéaire sur la pente dans la zone de la réponse qui exhibe une décroissance logarithmique. Ainsi, la restriction spatiale sur les position de mesure (loin de la source et de tout obstacle) nécessaire dans le cadre de la méthode précédente peut-être levée. De plus, les mesures éventuellement multiples à un endroit donné pour déterminer la pression quadratique moyenne en régime établi peuvent être remplacées par une mesure unique de la réponse impulsionnelle, typiquement filtrée par octave ou tiers d'octave. Malgré tout, la qualité de cette estimation faite

selon la méthode ISIR (Integrated Square Impulse Response Method) demeure limitée par un certain nombre de facteurs :

- Bien que la précision fréquentielle soit suffisante pour une caractérisation perceptive, elle demeure insuffisante pour des procédés d’analyse/synthèse de la réverbération. Comme le Tr_{60} tend à s’allonger dans les basses fréquences, l’effet du filtrage par bandes aura tendance à causer une surestimation du temps de réverbération (du fait que la mesure produira essentiellement la valeur du Tr_{60} du côté de la borne inférieure de chaque bande) et par conséquent une sous-estimation de la densité spectrale de puissance de la source. La qualité de l’estimation est donc assujettie aux propriétés des filtres passe-bande employés.
- La réponse impulsionnelle contient du bruit de mesure qui introduit des distorsions dans la courbe de décroissance énergétique qui tendent à fausser l’estimation du Tr_{60} et de la puissance de la source.

Une extension de la méthode ISIR (Integrated Square Impulse Response Method) de Schroeder a donc été proposée par Jot et al. dans (Jot *et al.*, 1997). Elle permet de passer de la description par bandes à une représentation fine de l’enveloppe temps-fréquence de la réverbération, nommée “Relief de Décroissance Énergétique” (Energy Decay Curve ou EDR). On montre dans cette section les étapes principales permettant d’estimer, à partir de cette représentation du signal, une valeur précise de la puissance de la source $W(f)$ et du temps de réverbération de la salle $Tr_{60}(f)$ en fonction de la fréquence.

Choix d’une distribution temps-fréquence pour le calcul du relief de décroissance

Soit $\rho_h(t, f)$ la distribution temps-fréquence énergétique de la réponse impulsionnelle $h(t)$ de la salle. Le relief de décroissance $EDR_h(t, f)$ est défini par :

$$EDR_h(t, f) = \int_t^{+\infty} \rho_h(\tau, f) d\tau. \quad (\text{A.18})$$

Idéalement, la représentation ρ_h doit conférer à l’EDR définie ainsi un certain nombre de propriétés (Jot *et al.*, 1997) dont les suivantes ont été jugées utiles d’être rappelées ici. D’une part, l’intégration de $EDR_h(t, f)$ selon les fréquences devrait donner la courbe de décroissance d’énergie $EDC_h(t)$ donnée dans l’équation A.11. Ceci est vrai si en intégrant $\rho_h(t, f)$ le long de l’axe des fréquences on retrouve bien $h^2(t)$:

$$\int_{-\infty}^{+\infty} \rho_h(t, f) df = h^2(t) \Rightarrow \int_{-\infty}^{+\infty} EDR_h(t, f) df = EDC_h(t) \quad (\text{A.19})$$

De manière converse, une propriété souhaitable de $EDR_h(t, f)$ serait qu'elle fournisse le carré de la transformée de Fourier H de h lorsqu'on l'intègre selon l'axe temporel, ce qui s'obtient facilement à condition que ρ préserve la distribution marginale dans le domaine des fréquences :

$$\int_{-\infty}^{+\infty} \rho_h(t, f) dt = |H(f)|^2 \Leftrightarrow EDR_h(-\infty, f) = |H(f)|^2. \quad (\text{A.20})$$

Enfin, il est souhaitable que ρ préserve la causalité des signaux, i.e. que $\rho_h(t, f) = 0$ pour $t < 0$ si $h(t)$ est un signal causal. Dans ce cas l'équation A.20 devient :

$$\int_0^{+\infty} \rho_h(t, f) dt = |H(f)|^2 \Leftrightarrow EDR_h(0, f) = |H(f)|^2, \quad (\text{A.21})$$

ce qui signifie que la valeur de l'EDR au temps 0 donne directement la puissance du signal au temps initial, d'où l'intérêt d'employer ce modèle pour l'estimation de puissance.

On montre que les propriétés A.19, A.20, et A.21 sont respectées si et seulement si ρ_h est une distribution de Levin, ce qui équivaut à définir le relief de décroissance comme suit : (Grace, 1981) :

$$EDR_h(t, f) = \left| \int_{\tau=t}^{+\infty} h(\tau) e^{-2i\pi f t} d\tau \right|^2 \quad (\text{A.22})$$

En pratique, la distribution de Levin n'est employée pour calculer l'EDR uniquement sur des signaux relativement courts. Pour des réponses impulsionnelles longues, l'utilisation de la Transformée de Fourier à Court Terme (TFCT) est préférable du fait des temps de calculs très optimisés pour cette représentation très largement utilisée dans le domaine du traitement de signal. Ce choix simplificateur induit un lissage et une décimation en temps et en fréquence par rapport à la version idéale de l'EDR, mais on montrera que le lissage s'avère utile dans l'estimation de la puissance de la source et du temps de réverbération de la salle.

Par ailleurs, il est important de noter que quel que soit la distribution ρ_h choisie, l'EDR est par essence une représentation stochastique qui n'est donc valable que dans un plan temps/fréquence borné par la fréquence de Schroeder et le temps de mélange. Elle suppose également que l'enveloppe temps-fréquence de la distribution varie lentement en t et f (Jot *et al.*, 1997).

Choix d'un modèle de signal de réverbération

Le choix d'un modèle de signal de réverbération est important car il va déterminer à quel point on s'approche de la réalité lors de l'estimation de la puissance de la source $W(f)$ et du temps de réverbération $Tr_{60}(f)$ de la salle, qui sont les paramètres réels que nous cherchons à estimer.

On choisit en pratique une généralisation du modèle de réverbération de Moorer (Moorer,

1978) suggérée dans (Baskind, 1999) obtenu en décomposant un bruit $n(t)$ de densité spectrale $P(f)$ non constante selon un très grand nombre de canaux fréquentiels de très faible largeur auxquels on associe une exponentielle décroissante.

L'opération de décomposition du signal évoque la transformée de Fourier; en effet, si le signal est une réalisation d'un bruit $n(t)$, alors on a :

$$N(f) = \int_{-\infty}^{+\infty} n(t)e^{-2i\pi ft} df, \quad (\text{A.23})$$

où $N(f)$ représente l'amplitude de la composante de $n(t)$ à la fréquence f dans la transformée de Fourier. L'opération de transformée de Fourier inverse peut écrire :

$$n(t) = \int_{-\infty}^{+\infty} N(f)e^{+2i\pi ft} df \quad (\text{A.24})$$

L'idée dans le modèle proposé est de multiplier chaque composante $N(f)$ par une exponentielle décroissante avant la reconstitution du signal. La réponse impulsionnelle de la salle prend donc l'allure suivante :

$$h(t) = \begin{cases} \int_{-\infty}^{+\infty} N(f)e^{+2i\pi ft} e^{-\delta(f)t} df & \text{si } t > 0 \\ 0 & \text{sinon} \end{cases} \quad (\text{A.25})$$

Dans cette équation, le facteur $N(f)$ contient toute l'information sur le spectre initial de la réponse impulsionnelle et indique de quelle façon l'énergie a été transmise par la salle au point d'écoute considéré. Le coefficient d'atténuation $\delta(f)$ d'un mode se calcule de la manière suivante :

$$20 \log(e^{-\delta(f)Tr_{60}(f)}) = -60(\text{dB}) \Rightarrow \delta(f) = \frac{3 \ln(10)}{Tr_{60}(f)}. \quad (\text{A.26})$$

L'équation A.25 ne modélise que la réverbération tardive, et par conséquent n'a aucune validité en dessous du temps de mélange. Néanmoins, on peut montrer que l'estimation de la puissance de la source s'obtient en prolongeant la réverbération tardive obtenue à partir de ce modèle par régression sur le relief jusqu'à l'instant initial.

Pour montrer cela, on introduit l'enveloppe temps-fréquence $ENV_h(t, f)$ caractéristique du modèle stochastique de réverbération utilisé, définie de la manière suivante :

$$ENV_h(t, f) = \langle \rho_h(t, f) \rangle = P(f)e^{-2\delta(f)t}. \quad (\text{A.27})$$

Baskind (Baskind, 1999) montre que si on se place dans des conditions où la densité spectrale de puissance $P(f)$ de la source et le coefficient d'atténuation $\delta(f)$ varient lentement (ce qui n'est vraie qu'au dessus de la fréquence de Schroeder), alors en vertu de l'équation A.18, la

moyenne d'ensemble du relief de décroissance peut s'écrire :

$$\begin{aligned}
 \langle EDR_h(t, f) \rangle &= \int_t^{+\infty} ENV_h(\tau, f) d\tau \\
 &\approx \frac{P(f)}{2\delta(f)} e^{-2\delta(f)t} \\
 &\approx \frac{Tr_{60}(f)}{6 \ln(10)} ENV_h(t, f).
 \end{aligned} \tag{A.28}$$

En prenant $t = 0$ dans l'équation A.28, et en s'aidant de la définition de l'EDR (eq. A.22) on voit alors que la densité spectrale de puissance initiale peut s'obtenir à partir de la réponse fréquentielle $H(f)$ de $h(t)$, à condition d'avoir préalablement déterminé le temps de réverbération $Tr_{60}(f)$:

$$\begin{aligned}
 P(f) &= ENV_h(0, f) \\
 &= \frac{6 \ln(10)}{Tr_{60}(f)} \langle EDR_h(0, f) \rangle \\
 &= \frac{6 \ln(10)}{Tr_{60}(f)} \left\langle \left| \int_{\tau=0}^{+\infty} h(\tau) e^{-2i\pi f\tau} d\tau \right|^2 \right\rangle \\
 &= \frac{6 \ln(10)}{Tr_{60}(f)} \langle |H(f)|^2 \rangle
 \end{aligned} \tag{A.29}$$

Notons que cette expression n'a de sens que si chaque isofréquence⁶ du relief de décroissance considéré décroît exponentiellement dès l'instant 0, ce qui n'est jamais le cas dans une EDR calculée à partir d'une réponse impulsionnelle réel (Jot, 1992). Cependant, dans le cas général, la valeur de $EDR_h(0, f)$ à prendre en compte correspond à l'ordonnée à l'origine de la droite de régression utilisée pour calculer le temps de réverbération.

On peut noter également que bien que les variations temporelles du signal mesuré soient atténuées lors de l'intégration rétrograde, il subsiste des variations fréquentielles aléatoires⁷ dans l'EDR. Pour pouvoir d'estimer correctement Tr_{60} et $P(f)$ il faudrait donc théoriquement faire la moyenne des EDR obtenus pour plusieurs réalisations de $h(t)$ de manière à s'approcher de sa moyenne d'ensemble $\langle EDR_h(0, f) \rangle$. Cependant, étant donné l'hypothèse de départ de variation lentes de Tr_{60} et $P(f)$ varient lentement avec la fréquence, il suffit de moyenniser l'EDR sur des bandes de fréquences dans lesquelles ces quantités varient peu, ce qui permet d'estimer la moyenne d'ensemble de l'EDR à partir d'une seule réalisation du signal $h(t)$.

⁶coupe à fréquence constante

⁷Leur densité de probabilité suit une loi de Rayleigh(Jot *et al.*, 1997)

Elimination du bruit

L'utilisation de la TFCT en tant que méthode de calcul de l'EDR (sec. A.3.2) permet de supprimer le bruit de fond par fréquence car elle permet d'estimer celui-ci au début et à la fin de la réponse. En supposant que le bruit de fond est décorrélé de la réponse, sa densité spectrale de puissance $P_{bruit}(f)$ peut être soustraite de tous les spectres de puissances avant l'opération de rétro-intégration. Du fait que la partie tardive de la réverbération induit un biais dans l'estimation initiale du bruit, on peut ensuite réestimer les paramètres du bruit une fois que les évaluations de $P(f)$ et $Tr_{60}(f)$ ont été effectuées. Par conséquent, un processus itératif peut être mis en oeuvre de manière à estimer plus précisément les valeurs de $P_{bruit}(f)$, $P(f)$ et $Tr_{60}(f)$ (Jot *et al.*, 1997). Ce processus itératif a été implémenté dans la librairie de fonctions Matlab⁸ utilisée pour faire les analyses présentées au chapitre 4.

Interprétation physique des résultats

Dans le cas qui nous intéresse, on étudie une source qui émet une impulsion d'énergie $W(f)$ dans la salle à l'instant initial $t = 0$. On fait l'hypothèse que l'énergie présente dans la salle décroît exponentiellement à partir de l'instant initial. De plus, bien qu'on ne suppose pas une densité d'énergie constante dans la salle pendant la partie précoce de la réponse de la salle, on suppose qu'elle devient uniformément répartie dans la partie tardive de la réponse. Ainsi, à condition d'être au delà du *temps de mélange* on peut donc affirmer que :

$$\langle \omega(t, f) \rangle = \frac{W(f)}{V} e^{-2\delta(f)t} \quad (\text{A.30})$$

où $\omega(t, f)$ représente la densité d'énergie présente dans la salle.

Dans le modèle, le spectre de puissance initial $P(f)$ communiqué aux modes propres de la salle ne peut dépendre que des directivités de la source et du récepteur (mais pas de leurs positions). Plus précisément, $P(f)$ est lui-même le produit de la directivité de la source moyennée suivant les directions d'émission (ou *spectre de rayonnement moyen* (Warusfel, 1990)) et la directivité du récepteur moyennée suivant les directions d'incidence (Jot, 1992).

Il convient donc d'introduire les deux grandeurs suivantes :

- la “fonction de transfert en champ diffus” du haut parleur $S_d(f)$, définie de la manière suivante (Jot *et al.*, 1997) :

$$S_d(f)^2 = \rho c^2 W(f), \quad (\text{A.31})$$

Cette relation permet d'exprimer la puissance de la source en fonction de l'énergie du signal électrique qui lui a été fourni (supposé être une impulsion de Dirac).

⁸Room Acoustics Toolbox - RAT

- la “fonction de transfert en champ diffus” du microphone $R_d(f)$ reliant le spectre de puissance initiale $P(f)$ et la densité d’énergie, définie de la manière suivante :

$$R_d(f)^2 = \frac{P(f)}{\rho c^2 \langle \omega(t, f) \rangle} \quad (\text{A.32})$$

En combinant ces équations, on voit que :

$$P(f) = \frac{S_d(f)^2 R_d(f)^2}{V}, \quad (\text{A.33})$$

ce qui montre la symétrie et réciprocité des deux transducteurs dans le modèle. En remaniant ces expressions, on obtient finalement une expression de la puissance de la source en fonction des données accessibles du problème :

$$W(f) = \frac{P(f)V}{\rho c^2 R_d(f)^2}. \quad (\text{A.34})$$

Conclusion

La méthode EDR présentée ici permet ainsi d’estimer la puissance de la source dans une salle quelconque, pourvue qu’elle soit pourvue d’un temps de réverbération suffisamment long. On se libère donc de la contrainte de déplacer les sources mesurées dans une chambre réverbérante ; au contraire, les mesures peuvent se faire in situ dans la salle de reproduction. On notera que cette méthode peut être appliquée à d’autres tâches, telles que la représentation de la qualité acoustique de la salle, la caractérisation et l’égalisation des transducteurs en champ diffus, la restauration (débruitage) des réponses impulsionnelles mesurées, ainsi que la réverbération artificielle et les effets de distance dans les systèmes de mixage audio.

A.3.3 Mesure de puissance basée sur l’intensité acoustique

Théorie

On développe le plus souvent l’expression de l’intensité acoustique à partir du principe de conservation de l’énergie dans un volume V de fluide :

$$\begin{aligned} \frac{\partial}{\partial t} \iiint_V E dV &= - \iiint_V \vec{\nabla} \cdot (p(\vec{r}, t) \vec{v}(\vec{r}, t)) dV \\ &= - \iint_S p(\vec{r}, t) \vec{v}(\vec{r}, t) \cdot \vec{n} dS \\ &= - \iint_S \vec{I}_i(\vec{r}, t) \cdot \vec{n} dS, \end{aligned} \quad (\text{A.35})$$

A.3. Mesure de la puissance émise par une source acoustique

où p représente la pression acoustique, \vec{v} représente la vitesse particulière instantanée, E représente la densité d'énergie⁹ présente dans le volume V , \vec{I}_i représente l'intensité instantanée, et \vec{n} représente un vecteur normal à la surface S entourant le volume V . L'équation A.35 signifie que le taux de variation de l'énergie interne totale dans l'élément de volume V est égal au flux d'énergie qui traverse la surface S entourant ce volume.

Si on suppose que le champ est monochromatique, il est commode de définir la pression acoustique et la vitesse particulière sous forme de nombres complexes, de la manière suivante :

$$p(\vec{r}, t) = A(\vec{r})e^{i[\omega t - \Phi(\vec{r})]} \quad (\text{A.36})$$

$$\begin{aligned} \vec{v}(\vec{r}, t) &= -\frac{1}{i\omega\rho}\vec{\nabla}p(\vec{r}, t) \quad (\text{eq. d'Euler}) \\ &= \frac{1}{\omega\rho}\left[A(\vec{r})\vec{\nabla}\Phi(\vec{r}) + i\vec{\nabla}A(\vec{r})\right]e^{i[\omega t - \Phi(\vec{r})]}, \end{aligned} \quad (\text{A.37})$$

où $A(\vec{r})$ et $\Phi(\vec{r})$ sont des nombres réels. L'intensité instantanée $\vec{I}_i(\vec{r}, t)$, qui représente un flux d'énergie par unité de surface et par unité de temps, se calcule alors à partir du produit des parties réelles de p et \vec{v} :

$$\begin{aligned} \vec{I}_i(\vec{r}, t) &= \Re e(p(\vec{r}, t)) \cdot \Re e(\vec{v}(\vec{r}, t)) \\ &= \frac{1}{\omega\rho}A^2(\vec{r})\vec{\nabla}\Phi(\vec{r})\cos^2(\omega t - \Phi) - \frac{1}{2\omega\rho}A(\vec{r})\vec{\nabla}A(\vec{r})\sin[2(\omega t - \Phi)] \end{aligned} \quad (\text{A.38})$$

L'interprétation de l'équation A.38 est facilitée si on introduit les quantités $\vec{I}(\vec{r})$ et $\vec{J}(\vec{r})$ définies de la manière suivante :

$$\begin{cases} \vec{I}(\vec{r}) = \frac{1}{2\rho\omega}A^2(\vec{r})\vec{\nabla}\Phi(\vec{r}) = \frac{1}{2}\Re e(p(\vec{r}) \cdot \vec{v}^*(\vec{r})) \\ \vec{J}(\vec{r}) = -\frac{1}{2\rho\omega}A(\vec{r})\vec{\nabla}A(\vec{r}) = \frac{1}{2}\Im m(p(\vec{r}) \cdot \vec{v}^*(\vec{r})) \end{cases} \quad (\text{A.39})$$

où $p(\vec{r})$ et $\vec{v}(\vec{r})$ représentent les composantes spatiales de la pression et vitesse, i.e.

$$\begin{cases} p(\vec{r}) = A(\vec{r})e^{-i\Phi(\vec{r})} \\ \vec{v}(\vec{r}) = \frac{1}{\omega\rho}\left[A(\vec{r})\vec{\nabla}\Phi(\vec{r}) + i\vec{\nabla}A(\vec{r})\right]e^{-i\Phi(\vec{r})}. \end{cases}$$

On a alors :

$$\vec{I}_i(\vec{r}, t) = 2\vec{I}(\vec{r})\cos^2(\omega t - \Phi) + \vec{J}(\vec{r})\sin[2(\omega t - \Phi)]. \quad (\text{A.40})$$

Dans cette expression, $\vec{I}(\vec{r})$ correspond à ce qu'on appelle communément *intensité acoustique*.

⁹ $E = \frac{1}{2}\rho|\vec{v}|^2 + \frac{1}{2\rho c^2}p^2$

On donne également le nom d'*intensité active* à cette quantité de manière à la distinguer de l'*intensité réactive* $\vec{J}(\vec{r})$.

En calculant la moyenne temporelle de l'équation A.40 sur une période, on voit que :

$$\overline{\vec{I}_i(\vec{r}, t)} = \frac{1}{T} \int_0^T \vec{I}_i(\vec{r}, t) dt = \vec{I}(\vec{r}). \quad (\text{A.41})$$

Ainsi dans un champ monochromatique, l'intensité active (ou intensité acoustique) n'est autre que la moyenne temporelle de l'intensité instantanée. Puisque le terme contenant l'intensité réactive $\vec{J}(\vec{r})$ s'annule en moyenne dans le temps, on considère dans beaucoup d'ouvrages que l'intensité réactive ne participe pas aux transferts nets d'énergie dans le champ (Bruneau, 1998; BruelKjaer, 1993). Cependant, d'après l'équation A.40 il est clair que l'intensité réactive participe au flux d'énergie d'un point de vue instantané. Mann et.al. (Mann *et al.*, 1987) soulignent l'importance de distinguer intensité moyennée dans le temps et l'intensité instantanée, car ces deux quantités ont une interprétation physique distincte, dont il faut tenir compte dans l'analyse énergétique de champs dits "réactifs" où la composante $\vec{J}(\vec{r})$ n'est plus négligeable devant $\vec{I}(\vec{r})$. Les champs proches de sources sonores et les champs diffus sont des exemples typiques de champs réactifs.

La puissance $W(\omega)$ (en Watts) rayonnée par une source peut être caractérisée à partir du vecteur intensité acoustique \vec{I} (en $W \cdot m^{-2}$) selon la formule suivante :

$$W(\omega) = \iint_S \vec{I} \cdot \vec{n} dS \quad (\text{A.42})$$

où S est une surface fermée entourant la source et \vec{n} est un vecteur normal à S . Le niveau de

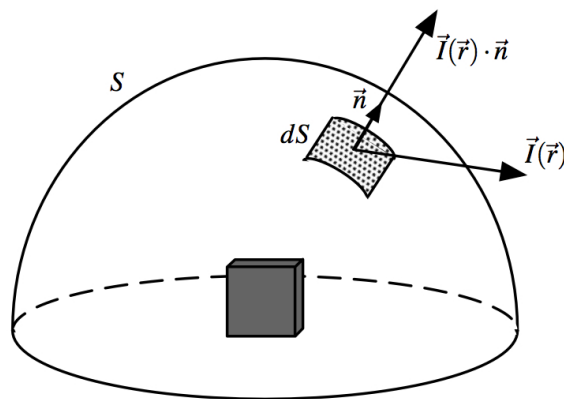


FIG. A.1 – Mesure de puissance rayonnée par intégration de l'intensité acoustique.

A.3. Mesure de la puissance émise par une source acoustique

puissance L_W est défini de la manière suivante :

$$L_W = 10 \log \left(\frac{W}{W_0} \right) \quad (\text{A.43})$$

où $W_0 = 10^{-12}W$ représente le niveau de référence de la puissance dans l'air.

Puissance de sources multipolaires idéales

On se place dans un repère sphérique (c.f. figure 2.1) et on considère une source monopolaire centrée qui émet une onde sphérique harmonique. Cette source produit un champ de pression indépendant de la direction (θ, ϕ) , qui peut s'écrire (Kinsler *et al.*, 2000) :

$$p(r, t) = \frac{i\omega\rho Q}{4\pi r} e^{i(\omega t - kr)} \quad (\text{A.44})$$

où Q représente le débit volumique en m^3/s de la source. La composante longitudinale de la vitesse particulière, qui s'obtient via l'équation d'Euler, s'exprime de la manière suivante :

$$v_r(r, t) = \frac{p(r, t)}{\rho c} \left(1 + \frac{1}{ikr} \right). \quad (\text{A.45})$$

En champ lointain de la source ($kr \gg 1$), pression et vitesse longitudinale sont donc liés par la relation suivante, caractéristique d'une onde plane :

$$v_r = \frac{p(r, t)}{\rho c} \quad (\text{A.46})$$

Dans le cas d'une source ponctuelle monopolaire en champ lointain, l'intensité acoustique I_r dans la direction longitudinale vaut donc :

$$I_r = \frac{\overline{p^2}}{\rho c} = \frac{p_{rms}^2}{\rho c} = \frac{\hat{p}^2}{2\rho c} \quad (\text{A.47})$$

où \hat{p} représente l'amplitude de l'onde harmonique, étant donné que pour une onde harmonique $p_{rms} = \hat{p}/\sqrt{2}$. En intégrant cette expression sur une surface sphérique S , pour laquelle l'intensité sortante $I_n = \vec{I} \cdot \vec{n}$ se confond avec la composante longitudinale de intensité I_r , on obtient alors la valeur de la puissance émise par une source ponctuelle :

$$W_{monopole}(\omega) = \frac{\rho Q^2 \omega^2}{8\pi c} \quad (\text{A.48})$$

On voit ici que pour un débit volumique fixe, la puissance émise par un monopole est proportionnelle à ω^2 .

Dans le cas d'une source dipolaire constituée de deux monopoles en opposition de phase, éloignés d'une distance d , on peut exprimer la pression en champ lointain de la manière suivante (Kinsler *et al.*, 2000) :

$$p(r, \theta, t) = -i \frac{Q \rho d \omega^2}{4\pi r} \cos \theta e^{i(\omega t - kr)} \quad (\text{A.49})$$

La puissance émise par le dipole vaut donc :

$$W_{dipole}(\omega) = \frac{Q^2 \rho d^2 \omega^4}{6\pi c^3} \quad (\text{A.50})$$

Cette fois, pour un débit volumique donné, la puissance émise par le dipole varie en ω^4 , ce qui signifie que le dipole est moins efficace que le monopole (à débit égal) en basses fréquences. De même, on peut montrer qu'à débit constant, la puissance émise par un quadripole (source à quatre lobes de directivité) varie selon ω^6 , ce qui le rend en fait un radiateur encore moins efficace de basses fréquences (Russell *et al.*, 1999).

Mesure pratique de puissance par méthode intensimétrique

Dans une situation pratique, on ne connaît pas le champ émis par la source sonore de manière analytique, et on cherche à estimer sa puissance émise par le biais du champ intensimétrique qui l'entoure. On distingue deux situations, selon que la mesure s'effectue en champ libre et lointain ou qu'elle s'effectue en champ proche de la source.

Le champ lointain d'une source dont la plus grande dimension vaut L est défini par : (Bera- nek, 1954; Kinsler *et al.*, 2000) :

$$\begin{cases} r \gg L \\ kr \gg 1 \end{cases} \quad (\text{A.51})$$

En supposant que le diagramme de rayonnement de la source sonore est indépendant du signal émis¹⁰, le champ de pression émis par la source en champ lointain peut s'écrire (Kinsler *et al.*, 2000) :

$$p(r, \theta, \phi, \omega) = \Gamma(\theta, \phi, \omega) \frac{e^{-ikr}}{r} \quad (\text{A.52})$$

Si on fixe une direction de référence (θ_{ax}, ϕ_{ax}) pour le diagramme d'émission de la source, que l'on nommera "axe acoustique" de la source, la pression p_{ax} le long de cet axe peut s'écrire :

$$p_{ax}(r, \omega) = p(r, \theta_{ax}, \phi_{ax}, \omega) = A(\omega) \frac{e^{-ikr}}{r}. \quad (\text{A.53})$$

$p_{ax}(r)$ représente donc la pression émise en champ lointain par la source suivant son axe acous-

¹⁰Cette hypothèse n'est pas forcément vérifiée pour certaines sources acoustiques, cf. (Giron, 1996).

tique. On peut alors exprimer la fonction de directivité Γ de la manière suivante :

$$\Gamma(\theta, \phi, \omega) = A(\omega) \cdot \frac{p(r, \theta, \phi, \omega)}{p_{ax}(r, \omega)}. \quad (\text{A.54})$$

En introduisant le concept de *Filtre de Facteur Directionnel D* (Giron, 1996), on a :

$$D(\theta, \phi, \omega) = \frac{\Gamma(\theta, \phi, \omega)}{A(\omega)}. \quad (\text{A.55})$$

L'équation A.52 devient alors :

$$p(r, \theta, \phi, \omega) = p_{ax}(r, \omega)D(\theta, \phi, \omega). \quad (\text{A.56})$$

On remarquera que le filtre de facteur directionnel D est normalisé de telle manière à ce que sa valeur maximale soit égale à 1.

- **Estimation de l'intensité en champ libre et lointain**

Lorsqu'on se situe dans des conditions de champ libre, et suffisamment loin de la source pour éviter le champ proche, on peut approximer la valeur de l'intensité par le carré de la pression, et déterminer la puissance rayonnée par la source suivant l'équation suivante (Beranek, 1954) :

$$W(\omega) = \frac{r_0^2}{\rho c} \int_0^{2\pi} \int_0^\pi p_{rms}^2(r_0, \theta, \phi, \omega) \sin \theta d\theta d\phi \quad (\text{A.57})$$

$$= \frac{r_0^2}{\rho c} \frac{1}{p_{ax}(r_0, \omega)^2} \int_0^{2\pi} \int_0^\pi D(\theta, \phi, \omega) \sin \theta d\theta d\phi, \quad (\text{A.58})$$

où r_0 représente le rayon de la sphère de mesure. L'opération d'intégration de l'intensité se fait en maillant la sphère de mesure par un ensemble discret de points. On pondère les points de mesure par un facteur Ω_i représentatif de la *zone d'influence* du ce point dans le maillage (le calcul de la zone d'influence d'un point de maillage est décrit dans la section 2.1.3).

- **Estimation de l'intensité en champ proche.**

Dans beaucoup de situations pratiques, il est impossible de se placer suffisamment loin de la source pour pouvoir estimer correctement sa puissance émise par simple mesure de pression suivant l'équation A.57. En effet, si on fait la mesure en salle, on doit se placer au-delà du champ proche de la source et en dehors du champ réverbéré. Ceci n'est possible qu'à condition de disposer d'une salle suffisamment grande à parois très absorbantes (et de tenir compte de la réflexion au sol). De même, en extérieur ou dans un

milieu bruyant il est possible que le rapport signal sur bruit soit trop faible pour conduire correctement l'estimation de puissance par le biais de la pression quadratique. Dans ces conditions, on est contraint d'effectuer une mesure en champ proche de la source. Or, en champ proche, l'intensité réactive \vec{J} n'est plus négligeable devant l'intensité active \vec{I} ; l'intensité instantanée \vec{I}_i peut se mettre à circuler, et on observe la formation de vortex dans la carte intensimétrique (Mann *et al.*, 1987), représentés de manière schématique dans la Figure A.2. Dans ce cas, la pression peut varier énormément entre deux positions

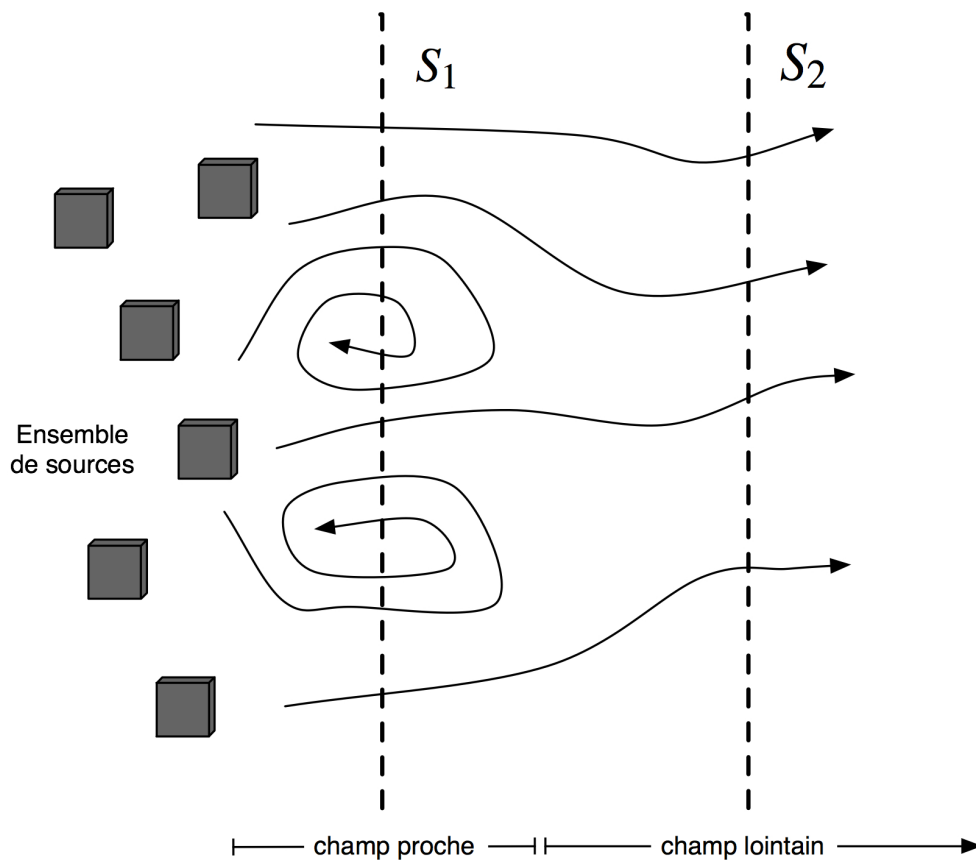


FIG. A.2 – Champ intensimétrique émis par un ensemble de sources acoustiques. En champ proche de la source, l'énergie circule et une mesure de puissance selon la surface S_1 nécessite forcément l'emploi d'une sonde intensimétrique. En champ lointain, l'énergie ne circule plus ; la pression et la vitesse particulière sont en phase et l'estimation de puissance sur la surface S_2 peut se faire par une simple mesure de la pression.

très proches, et l'estimation correcte du champ intensimétrique sortant nécessite l'utilisation d'une sonde intensimétrique. Il existe deux types de sondes intensimétriques : les sondes de type pression-pression ou 'p-p' et les sondes de type pression-vitesse ou 'p-v'¹¹. Une sonde intensimétrique de type p-p mesure la pression en deux points A et B espacés

¹¹Pour une description des sondes de type p-v voir par exemple (Jacobsen & DeBree, 2004))

A.3. Mesure de la puissance émise par une source acoustique

de Δr , et fournit une estimation de l'intensité acoustique selon l'axe de la sonde (de A vers B dans la figure A.3). Dans le cadre de l'estimation de la puissance d'une source,

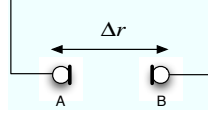


FIG. A.3 – Schéma d'une sonde intensimétrique de type p-p

il s'agit d'intégrer la valeur de l'intensité acoustique sortante sur une surface fermée S entourant la source (eq. A.42, fig. A.1). Le champ intensimétrique sortant $I_n = \vec{I} \cdot \vec{n}$ se mesure en orientant la sonde selon la normale \vec{n} à la surface S . La composante de vitesse particulaire sortante $v_n = \vec{v} \cdot \vec{n}$, qui est proportionnelle au gradient de pression, s'estime alors par différences finies, tandis que la pression est estimée par une moyenne simple des deux signaux de pression, d'où

$$\hat{I}_n = \overline{\hat{p}v_n} = \frac{p_A(t) + p_B(t)}{2} \int_{-\infty}^t \frac{p_B(\tau) - p_A(\tau)}{\rho \Delta r} d\tau \quad (\text{A.59})$$

où \hat{I}_n représente l'estimation de l'intensité acoustique sortante de la surface S au point de mesure.

L'intégration de l'intensité sortante permettant de déterminer la puissance de la source se fait soit par un balayage continu de la sonde le long de la surface de mesure S , soit par des mesures sur un ensemble discret de points définissant un maillage de cette surface (BruelKjaer, 1993). Dans le deuxième cas, on pondère les points de mesure par un facteur Ω_i représentatif de la *zone d'influence* de ce point dans le maillage (le calcul de la zone d'influence d'un point de maillage est décrit dans la section 2.1.3).

Les limitations les plus importantes de ce type de mesure sont liées à l'approximation de différences finies, la diffusion et la diffraction sur la sonde de mesure, et le biais lié aux erreurs de phase dans l'instrumentation.

En hautes fréquences, l'erreur liée à l'approximation de différences finies prédomine. Pour une onde plane incidente selon l'axe, on montre (Bortfeldt & Kramer, 1991) que cette erreur dépend de l'écartement entre les deux capsules, et varie selon la formule :

$$\frac{\hat{I}_n}{I_n} = \frac{\sin k\Delta r}{k\Delta r} \quad (\text{A.60})$$

La diffusion et diffraction sur la sonde dépendent fortement de sa géométrie. On montre qu'une configuration avec des capsules face-à-face séparées d'une entretoise solide est particulièrement favorable du point de vue de la diminution de la diffraction (Jacobsen & DeBree, 2004). De plus, en choisissant l'entretoise de telle manière à ce qu'elle ait

une longueur environ égale au diamètre de la capsule, les effets de diffraction tendent à compenser l'erreur due à l'approximation de différences finies. Ceci permet typiquement de couvrir jusqu'à une octave de plus en hautes fréquences lorsqu'on emploie une sonde $\frac{1}{2}$ pouce avec entretoises de $12mm$, ce qui implique la possibilité de couvrir la plus grande partie de la bande audible (i.e. [50-10kHz]) en utilisant cette seule configuration de sonde (Jacobsen *et al.*, 1998).

Le manque d'appariement en phase entre les deux microphones de mesure, si il n'est pas compensé, constitue un facteur limitant pour l'estimation de l'intensité en basses fréquences. En effet, ce manque d'appariement introduit des erreurs dans l'estimation de la variation de phase subie par l'onde lors de la propagation entre la capsule A et la capsule B. Ces erreurs se traduisent alors par un biais sur l'estimation de l'intensité, qui est d'autant plus significatif que la différence entre le niveau de pression et le niveau d'intensité au point de mesure est grande (BruelKjaer, 1993). On peut montrer que l'erreur totale sur l'estimation de la puissance sera inférieure à 1dB à condition que :

$$\Delta_{pI} < \delta_{pI_0} - 7dB, \quad (A.61)$$

où

$$\Delta_{pI} = 10 \log \left[\frac{\iint_S \frac{\overline{p^2}}{\rho c} dS}{\iint_S \vec{I} \cdot \vec{n} dS} \right] \quad (A.62)$$

représente l'Indice de Pression-Intensité Global¹² et

$$\delta_{pI_0} = 10 \log \left(\frac{\frac{\overline{p_0^2}}{\rho c}}{|I_0|} \right) \quad (A.63)$$

représente l'Indice de Pression-Intensité Résiduel¹³ (Jacobsen, 1997). p_0 et I_0 représentent les valeurs références de la pression et de l'intensité, définies au début du document dans les notations. Notons que l'indice δ_{pI_0} peut se déterminer à l'aide d'un pistophone au début ou même au cours de la session de mesure. Lors d'une mesure, si la condition donnée dans l'équation A.61 n'est pas remplie à cause de la présence de sources de bruit externes à la surface fermée de mesure, on recommande de rapprocher la surface de mesure de la source (ce qui implique d'augmenter le nombre de points de mesure pour compenser d'éventuelles erreurs sur la position des points de mesure). Si on souhaite plus de précision dans les basses fréquences, on recommande d'augmenter l'espacement Δr entre les capsules de manière à minimiser l'erreur sur l'estimation de la phase.

¹²Global Pressure-Intensity Index

¹³Residual Pressure-Intensity Index

A.3.4 Méthode basée sur la décomposition en harmoniques sphériques

On se place dans la situation où l'on cherche à estimer la puissance sonore émise par une source sonore dont on a conduit la décomposition du champ rayonné sur la base des harmoniques sphériques (opération décrite en section 2.1.1).

En appliquant l'équation A.42 à une surface de mesure sphérique S de rayon r_0 , on peut exprimer la puissance émise par la source sous la forme suivante :

$$W(\omega) = \frac{1}{2} \iint \Re[p(r_0, \theta, \phi, \omega) \cdot v_r^*(r_0, \theta, \phi, \omega)] r_0^2 \sin \theta d\theta d\phi. \quad (\text{A.64})$$

D'après l'équation 3.13, la décomposition du champ de pression émis par une source située à l'intérieur de la sphère de mesure peut s'écrire :

$$p(r, \theta, \phi, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^n b_{mn}(\omega) h_n^{(1)}(kr) Y_n^m(\theta, \phi). \quad (\text{A.65})$$

En vertu de l'équation d'Euler, la vitesse dans la direction radiale s'écrit donc :

$$v_r(r, \theta, \phi, \omega) = \frac{1}{i\rho ck} \frac{\partial p}{\partial r} = \frac{1}{i\rho c} \sum_{n=0}^{\infty} \sum_{m=-n}^n b_{mn}(\omega) h_n^{(1)}(kr) Y_n^m(\theta, \phi). \quad (\text{A.66})$$

En appliquant l'équation 2.5 dans la relation A.64 on obtient :

$$W(\omega) = \frac{r_0^2}{2} \sum_{n=0}^{\infty} \sum_{m=-n}^n |b_{mn}(\omega)|^2 \Re\left[\frac{h_n(kr_0)h_n'(kr_0)^*}{i\rho}\right] \quad (\text{A.67})$$

En vertu de la relation sur le Wronskien (Williams, 1999), qui, sachant que $h_n = j_n + in_n$ (eq. 2.2), dicte que

$$j_n(x)n_n'(x) - j_n'(x)n_n(x) = \frac{1}{x^2}, \quad (\text{A.68})$$

on obtient enfin l'expression permettant d'exprimer la puissance émise par une source en fonction de la somme quadratique des coefficients de sa décomposition en harmoniques sphériques :

$$W(\omega) = \frac{1}{2\rho ck^2} \sum_{n=0}^{\infty} \sum_{m=-n}^n |b_{mn}(\omega)|^2. \quad (\text{A.69})$$

Notons que pour cette expression de la puissance émise en champ lointain, on n'observe pas de couplage entre les harmoniques sphériques aux différents ordres ; le vecteur d'intensité active \vec{I} en champ proche s'exprime quant à lui à l'aide d'harmoniques sphériques couplées (Williams, 1999).

Les Harmoniques Sphériques

Au chapitre 2, on montre que la solution générale de l'équation des ondes exprimée en coordonnées sphériques ¹ peut s'écrire :

$$p(r, \theta, \phi, k) = \sum_{l=0}^{+\infty} \sum_{m=-l}^l \left(b_{lm}^{(1)}(k) h_l^{(1)}(kr) + b_{lm}^{(2)}(k) h_l^{(2)}(kr) \right) Y_l^m(\theta, \phi) \quad (\text{B.1})$$

où l et m sont des entiers, $0 \leq m \leq |l|$.

Cette équation peut se découper en deux termes :

- un terme de **propagation** $(b_{lm}^{(1)}(k) h_l^{(1)}(kr) + b_{lm}^{(2)}(k) h_l^{(2)}(kr))$, qui est explicité dans la section 2.1.1, équations 2.2, 2.3 2.4.
- un terme de **directivité** $Y_l^m(\theta, \phi)$.

Dans cette annexe on apporte quelques détails supplémentaires concernant le terme de directivité $Y_l^m(\theta, \phi)$. Ce terme, qui décrit la dépendance angulaire du champ sonore décrit par l'équation B.1, représente une famille de fonctions nommées harmoniques sphériques complexes, définies de la manière suivante :

$$Y_l^m(\theta, \phi) = \frac{1}{\sqrt{2\pi}} \tilde{P}_l^{|m|}(\cos \theta) e^{im\phi}, \quad (\text{B.2})$$

où \tilde{P}_l^m est un polynôme de Legendre associé normalisé.

¹Les conventions choisies pour le repère sphérique sont explicitées à la figure 2.1.

On adopte ici les conventions de nommage telles que Y_l^m est une harmonique sphérique dite de **degré** l et **d'ordre** m ; un degré croissant décrira une complexité croissante de l'harmonique selon l'angle θ . Cette convention de nommage est également adoptée dans (Daniel, 2000; Wikipedia, 2007; Mathworks, 2006). Notons que dans le reste du document, lorsque le degré et l'ordre ne seront pas évoqués conjointement, on parle par abus de langage d'harmoniques sphériques d'ordre m plutôt que de degré m .

En adoptant une convention de normalisation donnée dans (Ritchie & Kemp, 1999; Laborie, 2000), on définit le polynôme de Legendre associé normalisé de la manière suivante :

$$\tilde{P}_l^m(x) = \sqrt{\frac{2l+1}{2}} \sqrt{\frac{(l-m)!}{(l+m)!}} (1-x^2)^{\frac{m}{2}} \frac{d^m}{dx^m} P_l(x) \quad \text{pour } m, l \in \mathbb{N} \quad (\text{B.3})$$

$$\text{avec } P_l(x) = \frac{1}{l!2^l} \left(\frac{d^l}{dx^l} (x^2-1)^l \right). \quad (\text{B.4})$$

L'expression exacte des polynômes de Legendre P_l se calcule par récurrence ; voir par exemple (Williams, 1999; Daniel, 2000) pour une expression analytique des premiers polynomes de cette famille.

La normalisation particulière du polynôme de Legendre associé adoptée ici permet d'assurer l'orthonormalité de la famille des \tilde{P}_l^m pour le produit scalaire $\langle \cdot, \cdot \rangle_{[-1,1]}$, i.e.

$$\int_{-1}^1 \tilde{P}_l^m(x) \tilde{P}_{l'}^{m'}(x) dx = \delta_{ll'} \quad (\text{B.5})$$

où $l, l' \in \mathbb{N}$, $0 \leq m \leq l$ et $0 \leq m' \leq l'$.

Etant donné l'orthonormalité des exponentielles complexes pour le produit scalaire $\langle \cdot, \cdot \rangle_{2\pi}$, c'est à dire le fait que

$$\frac{1}{2\pi} \int_0^{2\pi} e^{im\phi} e^{-im'\phi} d\phi = \delta_{mm'}, \quad (\text{B.6})$$

on peut démontrer la propriété d'orthonormalité des harmoniques sphériques complexes pour le produit scalaire de l'espace sphérique $\langle \cdot, \cdot \rangle_{4\pi}$ donné à l'équation 2.5. En effet,

$$\begin{aligned} \langle Y_l^m(\theta, \phi) Y_{l'}^{m'}(\theta, \phi) \rangle_{4\pi} &= \int_{\theta=0}^{\pi} \int_{\phi=-\pi}^{\pi} Y_l^m(\theta, \phi) Y_{l'}^{m'}(\theta, \phi)^* \sin\theta d\theta d\phi \\ &= \int_{\theta=0}^{\pi} \tilde{P}_l^m(\cos\theta) \tilde{P}_{l'}^{m'}(\cos\theta) \sin\theta d\theta \times \frac{1}{2\pi} \int_{\phi=-\pi}^{\pi} e^{im\phi} e^{-im'\phi} d\phi \end{aligned}$$

et donc, en effectuant le changement de variable $x = \cos\theta$, on retrouve bien que

$$\langle Y_l^m(\theta, \phi) Y_{l'}^{m'}(\theta, \phi) \rangle_{4\pi} = \int_{-1}^1 \tilde{P}_l^m(x) \tilde{P}_{l'}^{m'}(x) dx \times \frac{1}{2\pi} \int_0^{2\pi} e^{im\phi} e^{-im'\phi} d\phi = \delta_{ll'} \delta_{mm'}.$$

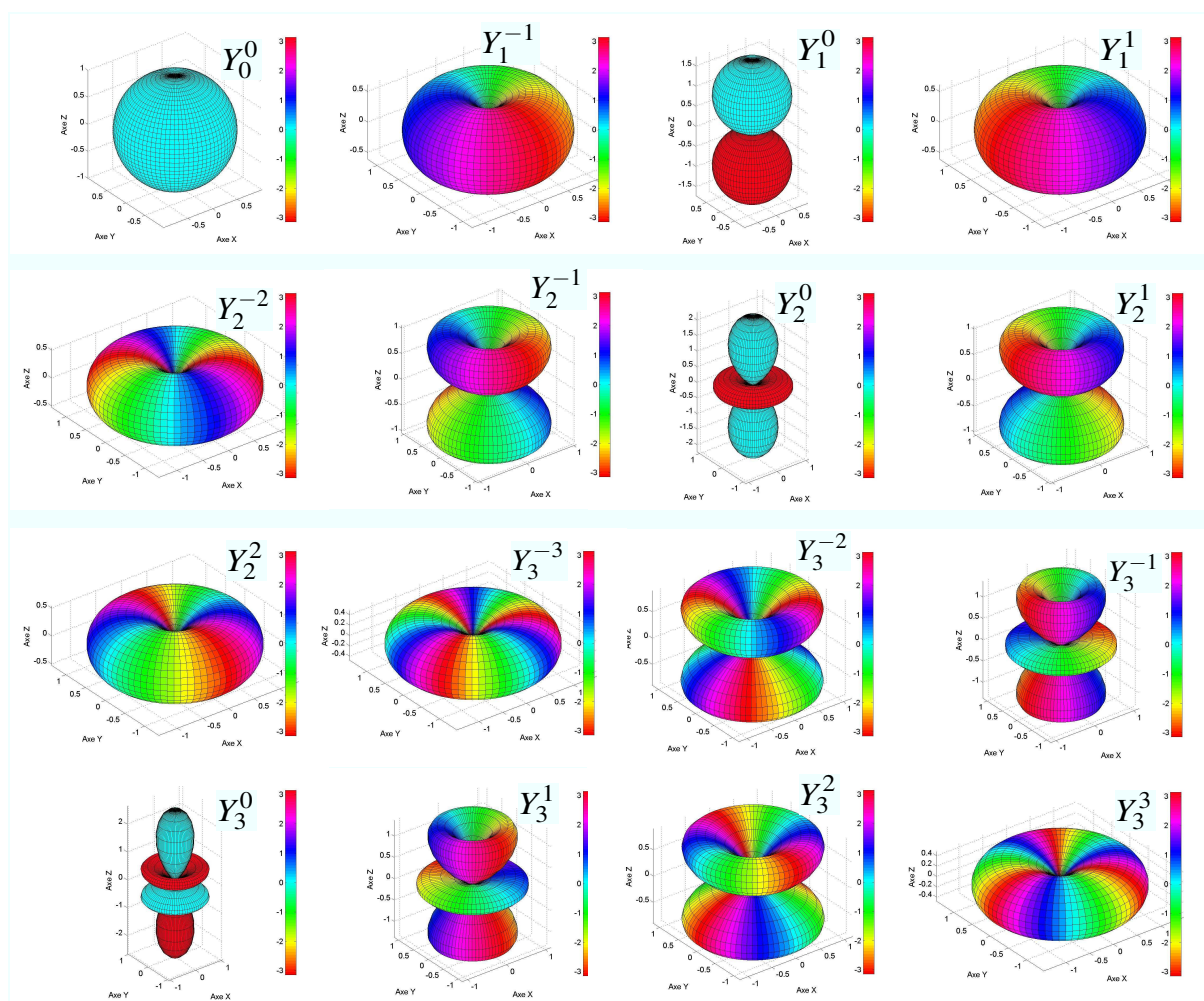


FIG. B.1 – Harmoniques sphériques complexes de degré 0, 1, 2 et 3.

Les harmoniques sphériques complexes sont représentées jusqu'au degré 3, à tous les ordres, sur la figure B.1.

Les harmoniques sphériques d'ordre 0 et de degré quelconque (Y_0^0 , Y_1^0 , Y_2^0 , etc.) n'ont pas de dépendance azimuthale et sont nommées *harmoniques axiales* par référence à leur symétrie autour de l'axe polaire. On trouve également le terme *harmonique zonale* dans la littérature, du fait que les cercles $P_l = 0$ découpent la sphère en zones (fig. B, gauche). Les harmoniques sphériques telles que $|m| = l$ divisent la sphère en secteurs angulaires : on les nomme donc *harmoniques sectorielles* (fig. B, droite). Les harmoniques restantes possèdent des lignes nodales à la fois en azimuth et en site, et semblent découper la sphère en tesserae (ou mosaïques). On les nomme donc *harmoniques tesserales* (fig. B, centre).

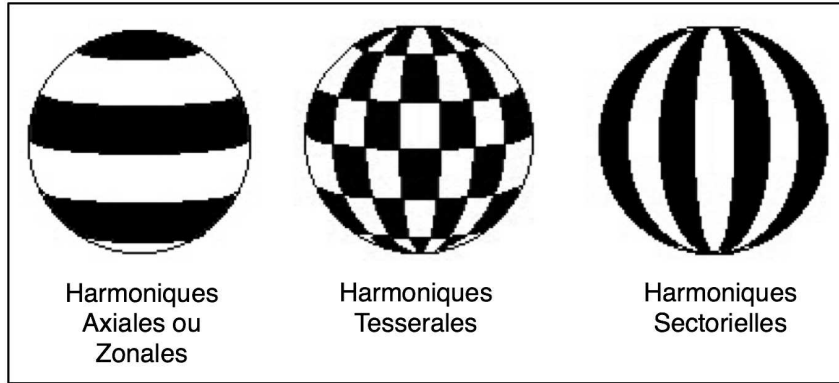


FIG. B.2 – Harmoniques Axiales/Zonales, Tesserales et Sectorielles.

Harmoniques sphériques réelles

En observant que seules les exponentielles complexes contribuent à la partie imaginaire des harmoniques sphériques, on peut montrer que les harmoniques sphériques présentent une symétrie hermitienne, i.e.

$$Y_l^{-m}(\theta, \phi)^* = \frac{1}{\sqrt{2\pi}} \tilde{P}_l^{|m|}(\cos(\theta)) (e^{im\phi})^* = Y_l^m(\theta, \phi). \quad (\text{B.7})$$

En vertu de cette symétrie hermitienne, toute fonction de directivité réelle $f_{\Re e}(\theta, \phi)$ décomposée sur la base des harmoniques sphériques présente une symétrie hermitienne dans le domaine de Fourier :

$$b_{l,-m} = \langle f_{\Re e}, Y_l^{-m} \rangle_{4\pi} = \langle f_{\Re e}, Y_l^{m*} \rangle_{4\pi} = \langle f_{\Re e}, Y_l^m \rangle_{4\pi}^* = b_{lm}^* \quad (\text{B.8})$$

De ce fait, il devient intéressant de définir une base d'harmoniques sphériques réelles, calculées à partir des harmoniques complexes par combinaisons linéaires, ce qui donne (Laborie, 2000) :

$$Y_l^m(\theta, \phi) = \begin{cases} \frac{1}{\sqrt{\pi}} \tilde{P}_l^{|m|}(\cos \theta) \cos(m\phi) & \text{pour } m > 0 \\ \frac{\tilde{P}_l^0}{\sqrt{2\pi}} & \text{pour } m = 0 \\ \frac{1}{\sqrt{\pi}} \tilde{P}_l^{|m|}(\cos \theta) \sin(m\phi) & \text{pour } m < 0 \end{cases} \quad (\text{B.9})$$

Cette base hérite des propriétés d'orthogonalité des harmoniques complexes et peut être utilisée pour décomposer n'importe quel champ sonore. Elle présente également l'avantage de fournir des coefficients b_{lm} réels lorsque la fonction de directivité décomposée est réelle, ce qui peut s'avérer utile lorsqu'on cherche à caractériser l'énergie rayonnée par une source sonore.

Cette base a été employée dans le cadre du travail présenté aux chapitres 1, 2, 4. Elle est représentée jusqu'au degré 3, à tous les ordres, sur la figure B.3.

Il est intéressant de noter qu'en fixant $\theta = \frac{\pi}{2}$, les harmoniques sectorielles réelles engendrent

Annexe B : Les Harmoniques Sphériques

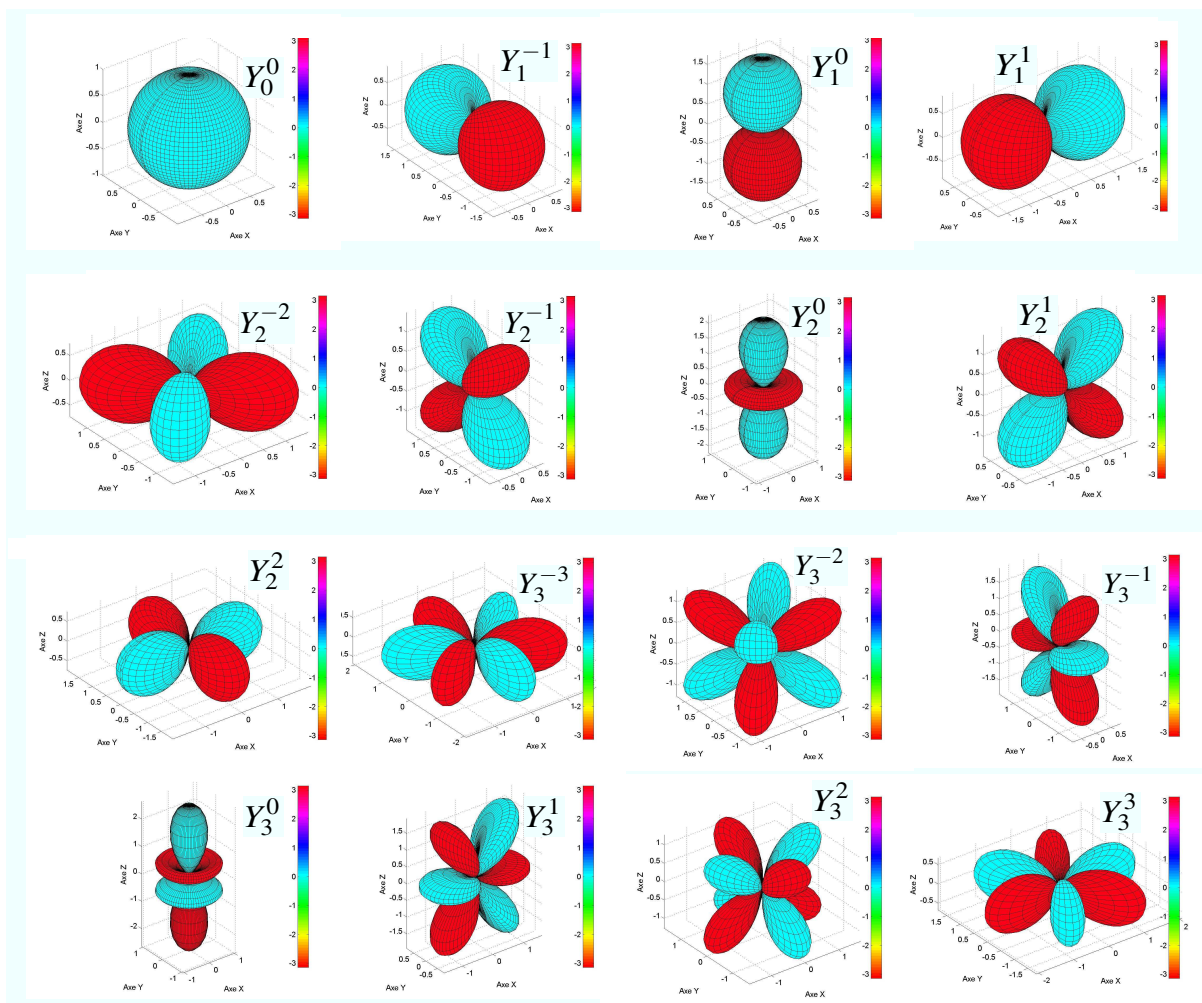



FIG. B.3 – Harmoniques sphériques réelles de degré 0, 1, 2 et 3.

l'ensemble des harmoniques circulaires Y^m défini à l'équation 3.2.

ANNEXE C

Décomposition en harmoniques sphériques du champ
associé à source virtuelle WFS

Source virtuelle Y^{-2}  centrée

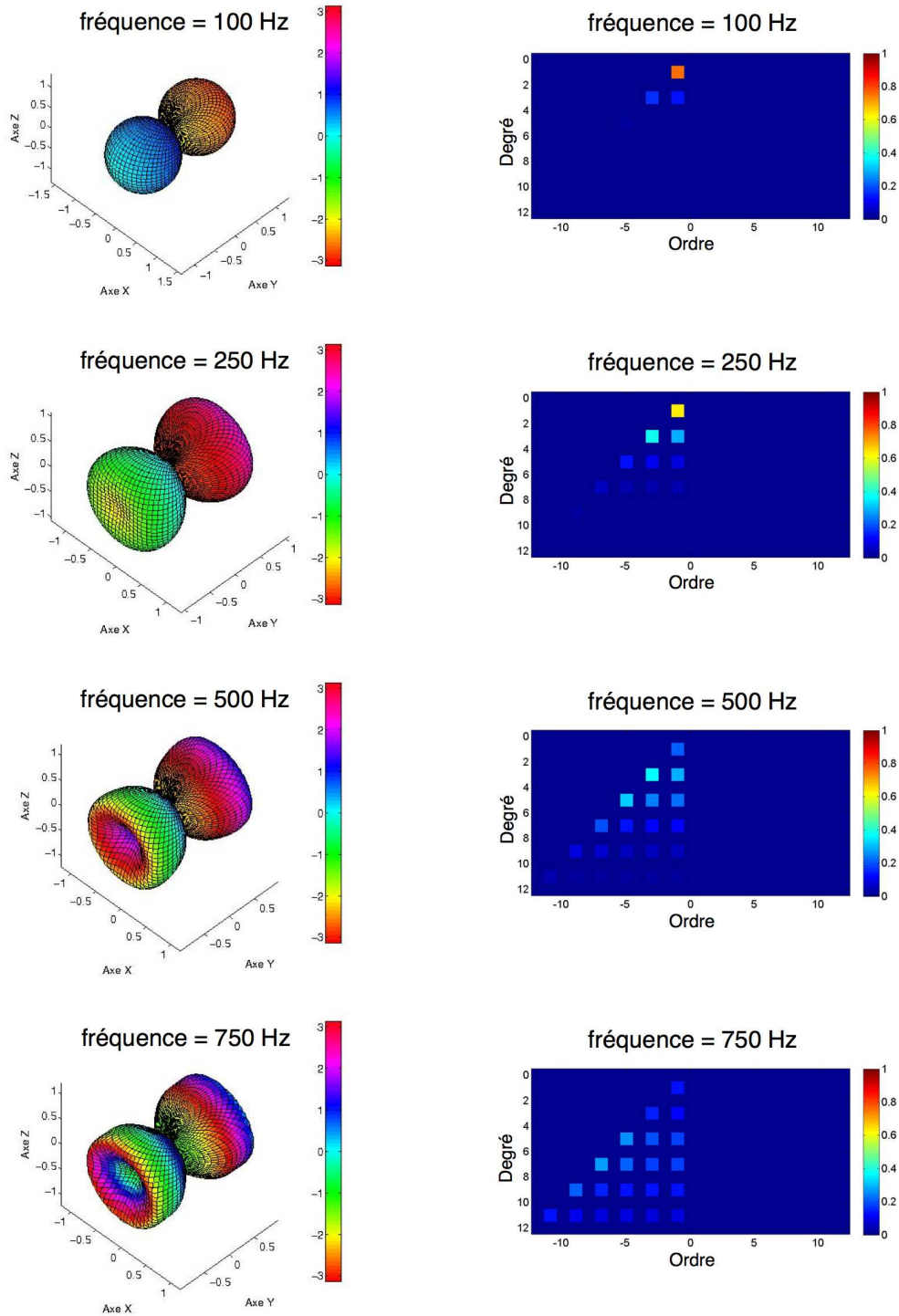


FIG. C.1 – Décomposition en harmoniques sphériques du champ associé à un quadrupole latéral Y^{-2} synthétisé à une position centrée, 1m derrière le banc de haut parleurs. **Colonne de Gauche** : Graphiques représentant le module (échelle linéaire) et la phase du champ mesuré au niveau de la sphère de microphones. **Colonne de Droite** : Module des coefficients b_{lm} de la décomposition, niveau linéaire.

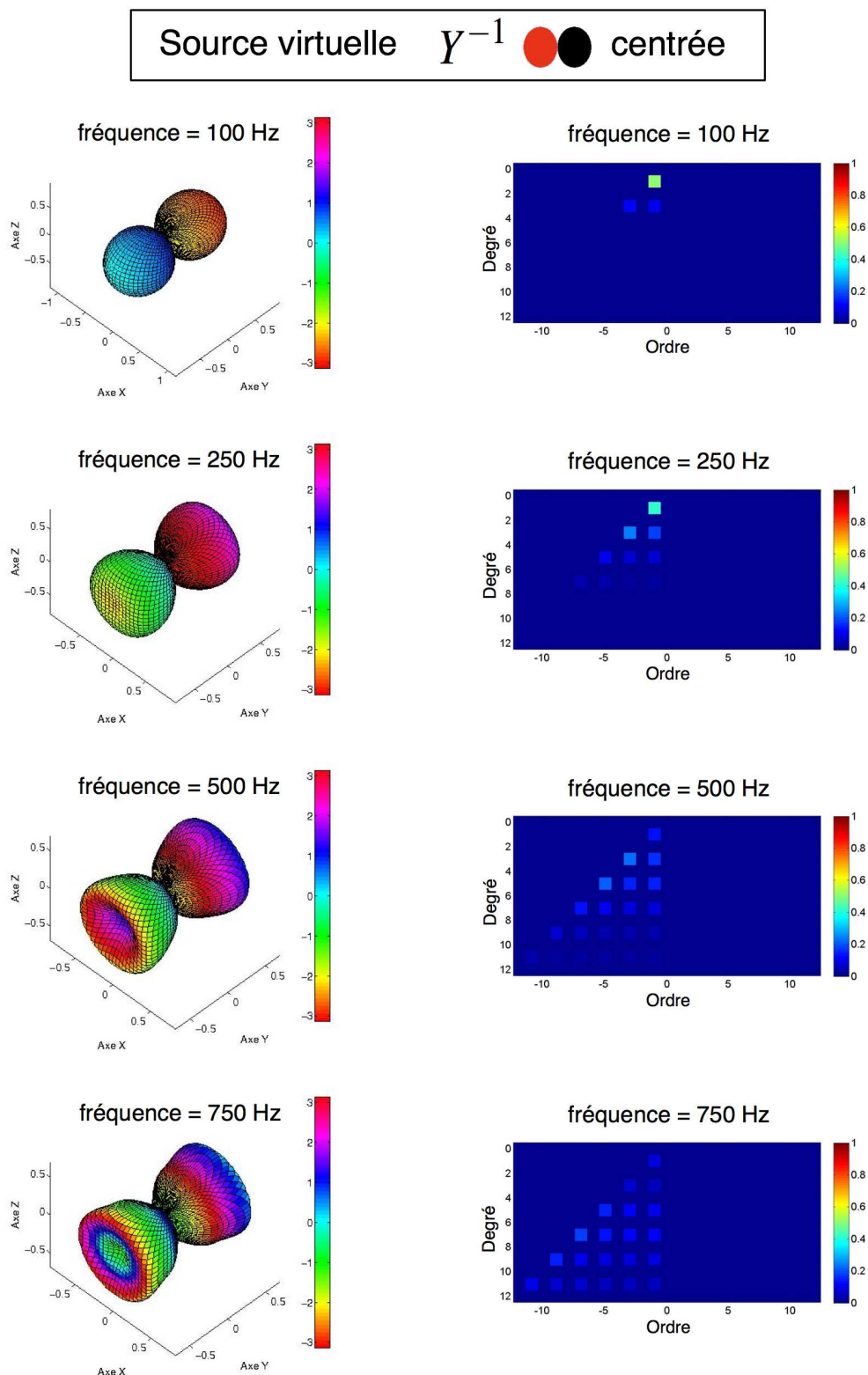



FIG. C.2 – Décomposition en harmoniques sphériques du champ associé à un dipôle latéral Y^{-1} synthétisé à une position centrée, 1m derrière le banc de haut parleurs. **Colonne de Gauche :** Graphiques représentant le module (échelle linéaire) et la phase du champ mesuré au niveau de la sphère de microphones. **Colonne de Droite :** Module des coefficients b_{lm} de la décomposition, niveau linéaire.

Source virtuelle Y^0  centrée

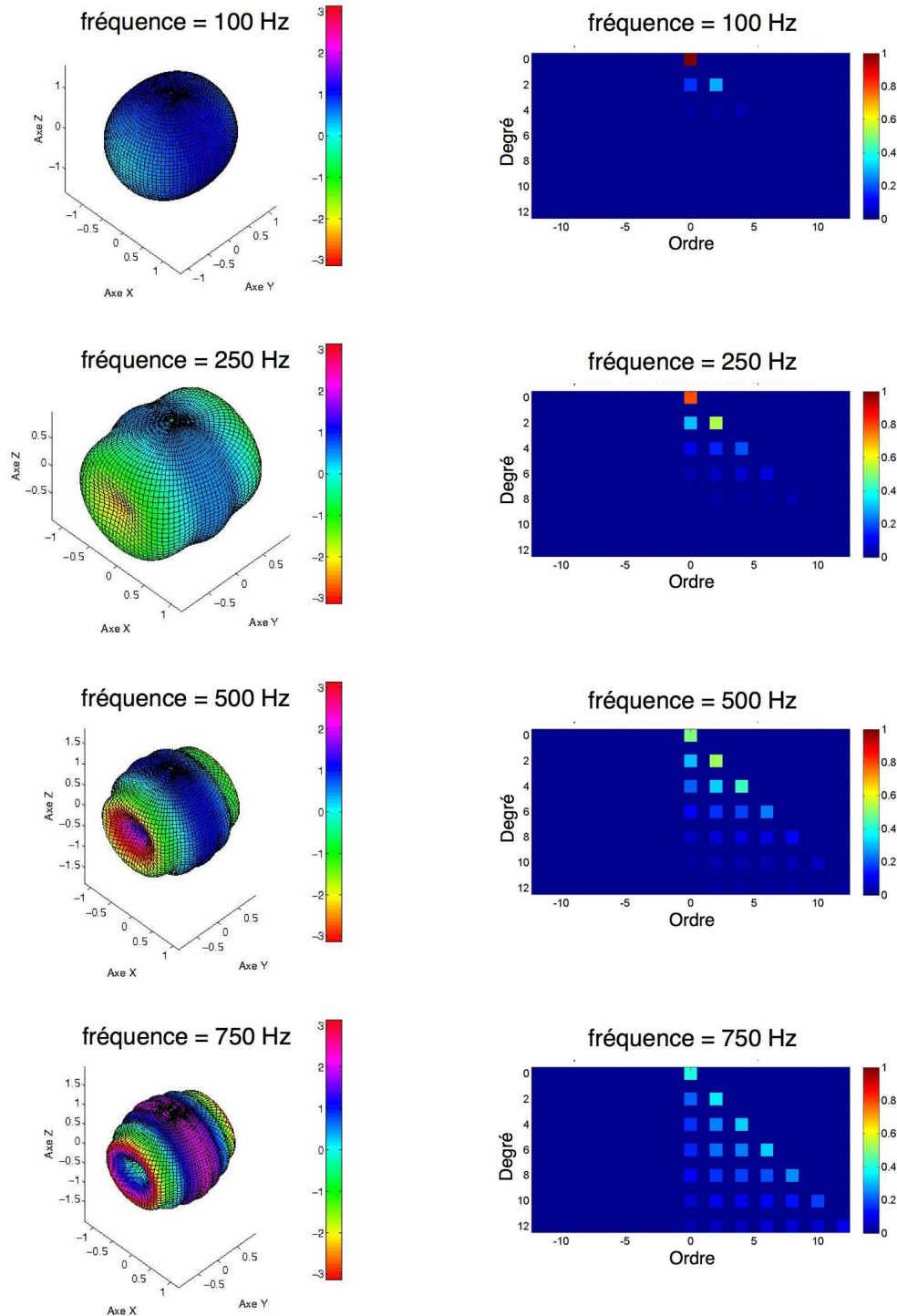



FIG. C.3 – Décomposition en harmoniques sphériques du champ associé à un monopole Y^0 synthétisé à une position centrée, 1m derrière le banc de haut parleurs. **Colonne de Gauche :** Graphiques représentant le module (échelle linéaire) et la phase du champ mesuré au niveau de la sphère de microphones. **Colonne de Droite :** Module des coefficients b_{lm} de la décomposition, niveau linéaire.

Source virtuelle Y^1  centrée

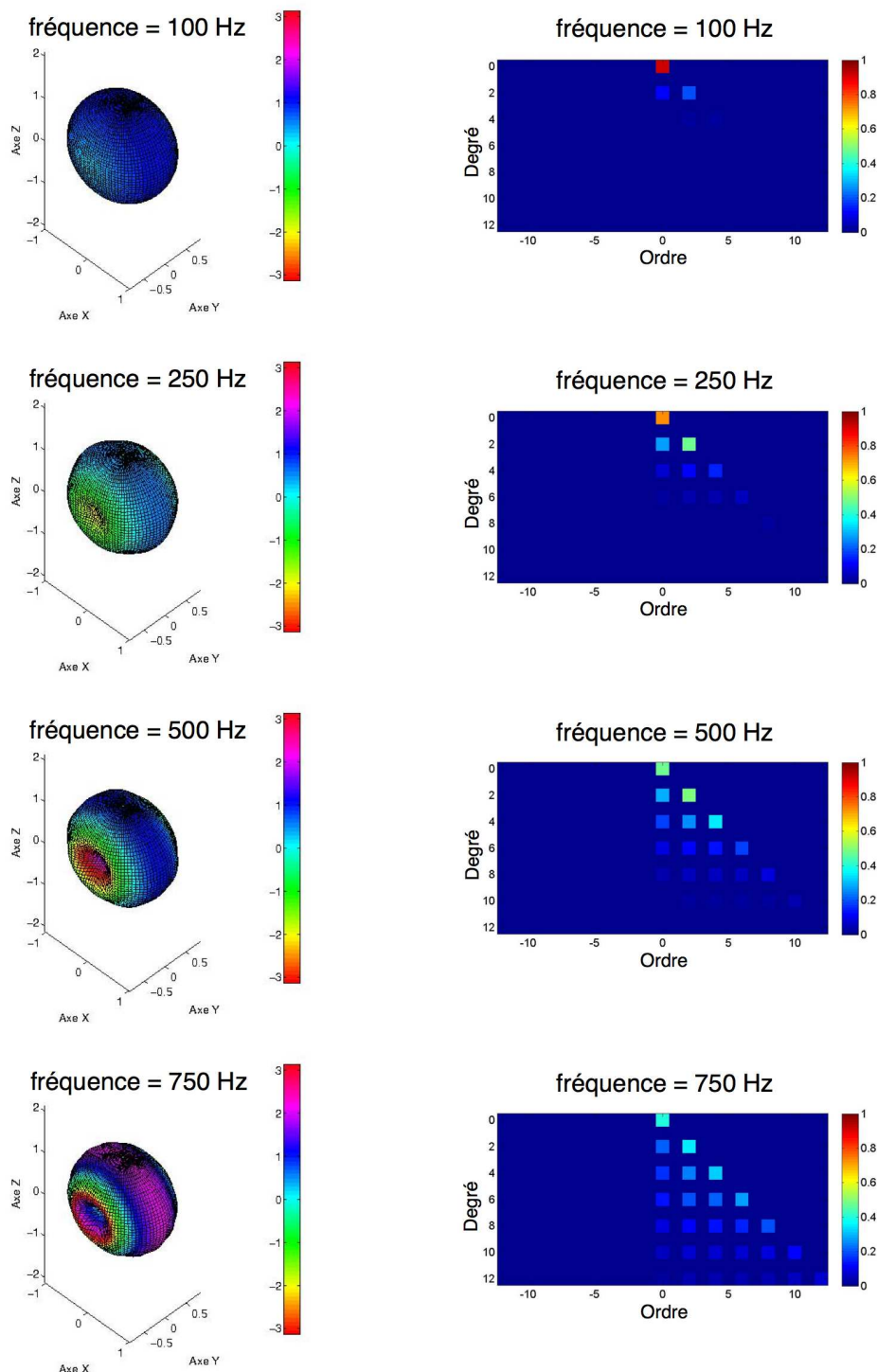



FIG. C.4 – Décomposition en harmoniques sphériques du champ associé à un dipole axial Y^1 synthétisé à une position centrée, 1m derrière le banc de haut parleurs. **Colonne de Gauche :** Graphiques représentant le module (échelle linéaire) et la phase du champ mesuré au niveau de la sphère de microphones. **Colonne de Droite :** Module des coefficients b_{lm} de la décomposition, niveau linéaire.

Source virtuelle Y^2  centrée

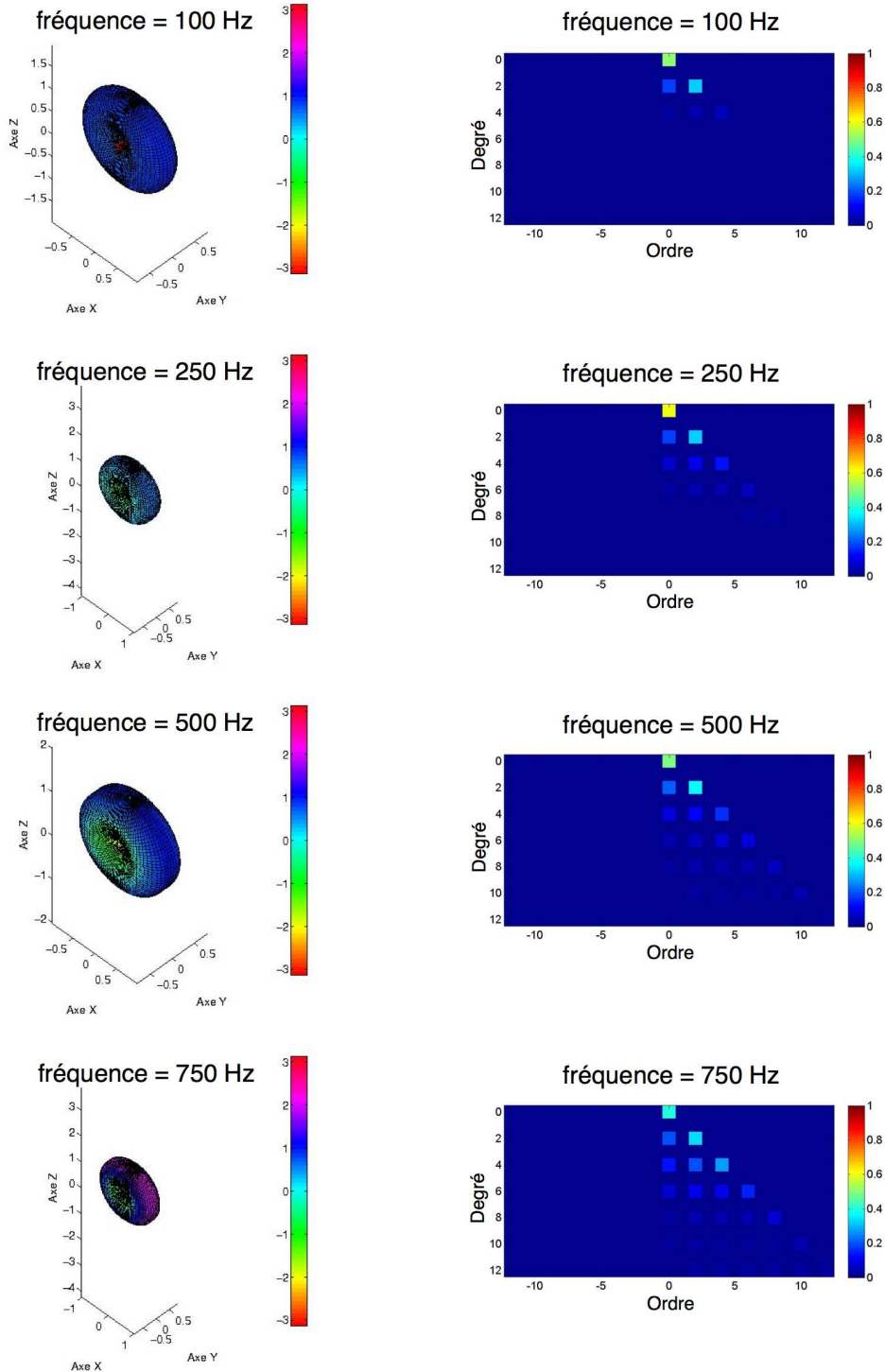


FIG. C.5 – Décomposition en harmoniques sphériques du champ associé à un quadrupole axial Y^2 synthétisé à une position centrée, 1m derrière le banc de haut parleurs. **Colonne de Gauche :** Graphiques représentant le module (échelle linéaire) et la phase du champ mesuré au niveau de la sphère de microphones. **Colonne de Droite :** Module des coefficients b_{lm} de la décomposition, niveau linéaire.

Calcul de Compensation de Contexte

La figure D.1 (tirée du brevet (Jot *et al.*, 1998)) illustre en détail les opérations d'addition et de convolution énergétique qui entrent en jeu dans la synthèse d'un effet de salle utilisant le Spatialisateur. Le bilan énergétique illustré par la figure D.1 peut s'écrire sous la forme d'un système d'équations faisant intervenir "cible", "contexte", et "live".

Dans le cadre qui a été fixé pour le travail du chapitre 4, le contexte central (i.e. $OD_{central}$, $R1_{central}$, $R2_{central}$, $R3_{central}$) possède une dépendance vis-à-vis de la position et de l'orientation de l'instrument virtuel WFS $\Psi(\vec{r}_S, \alpha_S)$. Le système d'équations issu du bilan énergétique s'exprime donc de la manière suivante :

$$\left\{ \begin{array}{l} OD_{cible} = OD_{live} + OD_{central}(\vec{r}_S, \alpha_S) * OD \\ R1_{cible} = R1_{live} + OD * R1_{central}(\vec{r}_S, \alpha_S) * R1 + \frac{7}{8}R1 * OD_{lateral} \\ R2_{cible} = R2_{live} + OD * R2_{central}(\vec{r}_S, \alpha_S) + R1 * (\frac{1}{8}OD_{lateral} + R1_{lateral} + \frac{1}{2}R2_{lateral}) \\ \quad + R2 * (\frac{23}{24}OD_{diffus} + \frac{1}{2}R1_{diffus} + \frac{1}{18}R2_{diffus}) \\ R3_{cible} = R3_{live} + OD * R3_{central}(\vec{r}_S, \alpha_S) + R1 * (\frac{1}{2}R2_{lateral} + R3_{lateral}) \\ \quad + R2 * (\frac{1}{24}OD_{diffus} + \frac{1}{2}R1_{diffus} + \frac{17}{18}R2_{diffus} + R3_{diffus}) \\ \quad + R3 * (OD_{diffus} + R1_{diffus} + R2_{diffus} + R3_{diffus}), \end{array} \right.$$

où * représente l'opération de convolution et OD, R1, R2, et R3 représentent les sorties du processeur d'effet de salle. Les relations D.1 sont données pour des limites temporelles 20, 40

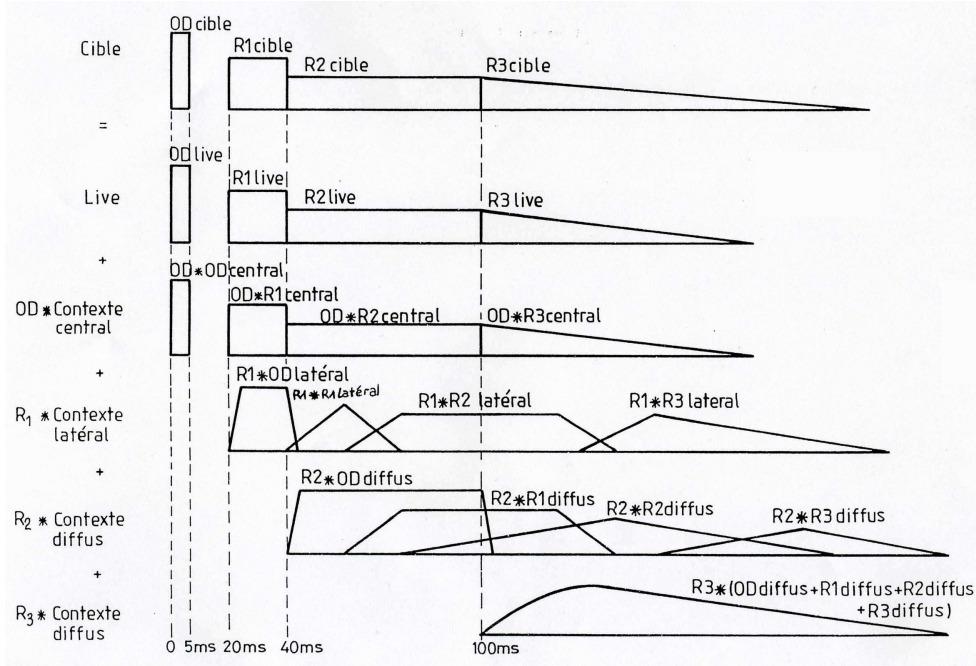


FIG. D.1 – Bilan énergétique pour la synthèse d'un effet de salle cible utilisant le processeur d'effet de salle du Spatialisateur (image tirée directement du Brevet du Spatialisateur). Les composantes OD, R1, R2, et R3 représentent les sorties du processeur d'effet de salle. Les composantes affectées de l'apposition "central", "latéral" et "diffus" représentent le contexte mesuré dans la salle d'écoute.

et 100 ms (fig. D.1) mais demeurent valables lorsque ces limites sont modifiées de manière homothétique (ex : 10, 20 et 50 ms).

La résolution du système d'équations D.1 équivaut à l'inversion d'une matrice triangulaire inférieure. On obtient après inversion de cette matrice les expressions suivantes pour les énergies OD, R1, R2, et R3 que doit produire le Spatialisateur de manière à reproduire l'acoustique cible en minimisant l'influence de la salle d'écoute au point de référence :

$$\left\{ \begin{array}{l} OD = \frac{(OD_{cible} - OD_{live})}{OD_{central}(\vec{r}_S, \alpha_S)} \\ R1 = \frac{R1_{cible} - [R1_{live} + \frac{7}{8} R1_{central}(\vec{r}_S, \alpha_S) * OD]}{OD_{latéral}} \\ R2 = \frac{R2_{cible} - [R2_{live} + R2_{central}(\vec{r}_S, \alpha_S) * OD + R1 * (\frac{1}{8} OD_{latéral} + R1_{latéral} + \frac{1}{2} R2_{latéral})]}{\frac{23}{24} OD_{diffus} + \frac{1}{2} R1_{diffus} + \frac{1}{18} R2_{diffus}} \\ R3 = \frac{R3_{cible} - R3_{live} - OD * R3_{central}(\vec{r}_S, \alpha_S) - R1 * (\frac{1}{2} R2_{latéral} + R3_{latéral})}{OD_{diffus} + R1_{diffus} + R2_{diffus} + R3_{diffus}} \\ \quad - \frac{R2 * (\frac{1}{24} OD_{diffus} + \frac{1}{2} R1_{diffus} + \frac{17}{18} R2_{diffus} + R3_{diffus})}{OD_{diffus} + R1_{diffus} + R2_{diffus} + R3_{diffus}}, \end{array} \right.$$

Ces expressions sont valables quelle que soit la bande de fréquences considérée. Si la qualité acoustique cible est globalement moins réverbérante que les qualités "live" et "contexte" il peut arriver que les valeurs calculées pour OD, R1, R2, ou R3 soient négatives. Dans ce cas ces valeurs sont seuillées à zéro de manière à rester cohérent avec le fait qu'elles représentent des énergies. En d'autres termes, ce procédé ne peut avoir d'influence que si l'acoustique "cible" est globalement plus réverbérante que la pièce d'écoute.

Apparent Source Width (ASW) - Largeur Apparente de Source

l'ASW est un des deux facteurs (avec le LEV -Listener Envelopment) contribuant à l'impression spatiale (Spatial Impression) perçue dans une salle. Comme son nom l'indique, l'ASW correspond à la largeur apparente de la source perçue par l'auditeur. On montre que l'ASW est déterminé par le niveau relatif et l'angle d'arrivée des premières réflexions dont les retards n'excèdent pas les 80 premières millisecondes de la réponse impulsionnelle (Bradley & Soulodre, 1995a). On utilise classiquement l'IACC (Interaural Cross Correlation) et le LEF (Lateral Energy Fraction) pour prédire l'ASW. Notons que certaines études tendent à préférer le LEF en tant que prédicteur de l'impression spatiale du fait que des variations de l'ASW peuvent être masquées lorsque le niveau d'énergie tardif est élevé (Bradley & Soulodre, 1995b).

Auralisation

Technique utilisant une modélisation informatique de l'environnement acoustique et de méthodes de traitement de signal 3D pour rendre audible le champ sonore rayonné par une source dans l'environnement modélisé. De manière analogue à la fabrication d'un modèle réduit d'immeuble avant construction, l'auralisation permet à l'acousticien ou le designer sonore de construire un modèle informatique du lieu et de pouvoir "écouter" le son de la salle sur casque (Yates, 1995).

"Auralization is the process of rendering audible, by physical or mathematical modeling, the

soundfield, in such a way as to simulate the binaural listening experience at a given position in the modeled space.” (Kleiner et al., 1993).

Champ libre

Lorsque l'on cherche à caractériser le rayonnement d'une source sonore il est utile de pouvoir isoler le son direct de cette source en créant des conditions dites de *champ libre* dans lesquelles aucun écho ne vient perturber la mesure. La création de conditions de champ libre à l'intérieur d'un espace fermé réverbérant nécessite que l'on place la source sonore et le récepteur loin de toute surface réfléchissante. Si la salle est grande et le support temporel de la réponse étudiée est suffisamment compacte, on est alors en mesure d'isoler le son direct de la source en tronquant sa réponse impulsionnelle avant l'arrivée des premières réflexions. Ainsi par exemple le site de recherche et développement de Bang et Olufsen (B&O, 2007) contient une enceinte cubique dont les arêtes mesurent 40m destinée précisément à ce type de mesure. L'enceinte est équipée d'un bras robotique permettant de déplacer et tourner les haut-parleurs mesurés de manière à pouvoir caractériser leur rayonnement en champ libre dans toutes les directions.

Une manière alternative qui permet de recréer des conditions de champ libre consiste à se placer dans une *chambre anéchoïque*. Une telle chambre, appelée également *chambre sourde*, est une chambre entièrement recouverte de dièdres absorbants (souvent fait de mélamine ou fibre de verre) permettant d'absorber l'énergie acoustique arrivant sur ses murs au dessus d'une certaine fréquence de coupure. A titre d'exemple, la chambre anéchoïque de l'IRCAM (fig. E.1) possède une fréquence de coupure de 75Hz.

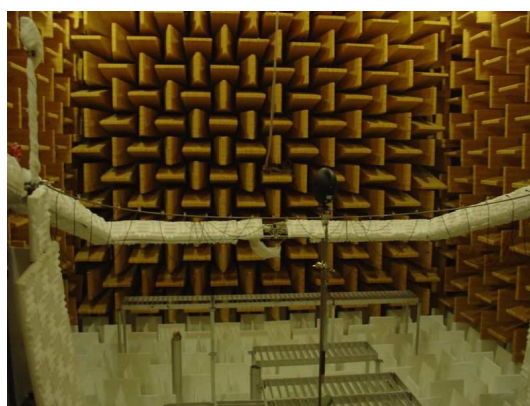


FIG. E.1 – Photographie de la chambre anéchoïque de l'IRCAM

Champ diffus

Champ sonore uniforme et isotrope, pour lequel l'énergie est répartie de manière égale selon toutes les directions et à toutes les positions. On peut modéliser un champ diffus comme étant composé d'ondes planes décorréliées provenant d'incidences uniformément distribuées dans l'espace, dans lequel il existe de l'énergie acoustique à toutes les fréquences et dans toutes les directions (Larcher, 2001). On considère de plus que les phases sont distribuées de façon aléatoire, de sorte que les ondes incidentes en un point soient indépendantes d'une direction à l'autre et que leurs énergies puissent être ajoutées (Baskind, 2003).

Une vision alternative de la diffusion consiste à prendre en compte les différents modes d'une salle : dans un champ diffus, tous les modes sont excités de manière égale. Ceci n'est vrai que si le recouvrement des modes à une fréquence donnée est suffisant, i.e. au delà de la *fréquence de Schroeder* de la salle.

Notons qu'en réalité un champ n'est jamais complètement diffus, bien qu'il puisse s'en rapprocher dans certains environnements comme celui d'une chambre réverbérante. Les chambres réverbérantes sont beaucoup utilisées dans le domaine de l'acoustique notamment pour mesurer des coefficients d'absorption de matériaux acoustiques, mesurer des niveaux de puissance de sources ou encore caractériser/égaliser des transducteurs (Cook *et al.*, 1955).

(Voir également : *Diffusivité*).

Coefficient de corrélation

Le coefficient de corrélation entre les pressions mesurées en deux points A et B est défini par la formule suivante (Cook *et al.*, 1955) :

$$C_{corr} = \frac{\langle p_1 p_2 \rangle_{moy}}{(\langle p_1^2 \rangle_{moy} \langle p_2^2 \rangle_{moy})^{1/2}} \quad (\text{E.1})$$

où $p_1(t)$ et $p_2(t)$ représentent les pressions instantanées en A et B, et $\langle . \rangle_{moy}$ représente la moyenne temporelle, d'où :

$$C_{corr} = \frac{\int_0^T p_1(t) p_2(t) dt}{\left[\int_0^T p_1^2(t) dt \int_0^T p_2^2(t) dt \right]^{1/2}} \quad (\text{E.2})$$

On pourra noter que le coefficient de corrélation représente la valeur maximale de la corrélation normalisée, qui est le rapport de la corrélation par le produit des puissances.

Cohérence

1. Rapport de l'interspectre par la racine du produit des autospectres.
2. Ensemble des propriétés de corrélation d'un système ondulatoire. Son sens initial était la mesure de la capacité d'onde(s) à donner naissances à des interférences -du fait de l'existence d'une relation de phase définie- mais il s'est élargi. On peut parler de cohérence entre 2 ondes, entre les valeurs d'une même onde à deux instants différents (cohérence temporelle) ou entre les valeurs d'une même onde à deux endroits différents (cohérence spatiale). Le terme cohérent s'emploie à la fois pour une radiation électromagnétique ou pour des particules, dont le comportement ondulatoire a été postulé en 1924 par Louis de Broglie (Wikipedia, 2007).

Coloration

Attribut de la sensation cochléaire par lequel un auditeur peut juger de la dissimilarité entre deux sons ayant la même sonie. La coloration comprend donc le timbre, la sensation de rythme ainsi que la hauteur du signal. On définit souvent la coloration comme étant une distorsion audible du signal qui altère sa couleur (naturelle).

Densité modale

La densité modale représente le nombre moyen de modes par Hz. Dans un cas de figure classique de salle parallépipédique, on peut montrer que la densité modale est environ proportionnelle au carré de la fréquence :

$$\frac{dN_{modes}}{df} = \frac{4\pi V}{c^3} \cdot f^2 \quad (s^{-1}) \quad (E.3)$$

ou N_{modes} représente le nombre de modes présents et V représente le volume de la salle. Ce résultat est généralisable à des salles de géométrie quelconque (Cremer *et al.*, 1978), (Kuttruff, 1991).

Densité de réflexions

Dans un cas de figure classique de salle parallépipédique, la densité temporelle de réflexions spéculaires peut être relié au temps écoulé de la manière suivante :

$$\frac{dN_{ref}}{dt} = \frac{4\pi c^3}{V} \cdot t^2 \quad (s^{-1}) \quad (E.4)$$

avec N_{ref} le nombre de réflexions présentes dans la salle à un instant donnée et V le volume de la salle. Ce résultat est généralisable à des salles de géométrie quelconque (Cremer *et al.*, 1978), (Kuttruff, 1991).

Directivité, Indice et Facteur

Le facteur de directivité $Q(f)$ (Directivity Factor) est défini de la manière suivante (Beranek, 1954) :

$$Q(f) = \frac{4\pi |p_{ax}|^2}{\int_0^{2\pi} \int_0^\pi |p(\theta, \phi)|^2 \sin \theta d\theta d\phi} \quad (E.5)$$

où p_{ax} représente la magnitude de la pression quadratique selon l'axe de rayonnement principal de la source à une distance donnée r . Ce rapport de puissances indique à quel point une source directive parvient à concentrer l'énergie dans une direction donnée. L'indice de directivité (Directivity Index) $DI(f)$ se calcule à partir du facteur de directivité de la manière suivante :

$$DI(f) = 10 \log(Q(f)). \quad (E.6)$$

Une source monopolaire présente un indice de directivité $DI_{monop} = 0$. Un dipole possède quand à lui un indice de directivité $DI_{dipole} = 10 \log(3) \approx 4.8 dB$

Diffusivité

La mesure de diffusivité sert à vérifier qu'un champ acoustique est diffus selon une ou plusieurs de ses dimensions. Dans un champ diffus 3D, si on mesure le coefficient de corrélation C_{corr} entre deux points A et B, on trouve :

$$C_{corr} = \frac{\sin kd}{kd} \quad (E.7)$$

avec d la distance entre A et B. Dans un champ diffus seulement dans le plan contenant A et B, on trouve :

$$C_{corr} = J_0(kd) \quad (E.8)$$

Ces propriétés permettent de détecter des anisotropies dans des champ diffus. Notons que les formules E.7 et E.8 sont valables uniquement pour des champ diffus sinusoidaux purs. Rafaely a récemment proposé une généralisation de ces résultats à des champ diffus large-bande ??, ??. (Voir également : *Champ diffus, Coefficient de corrélation*)

Energy Decay Relief (*EDR*)

: “Relief de Décroissance Energétique”, proposé par Jot et. al. (Jot *et al.*, 1997) comme extension de l’Energy Decay Curve de Schroeder (Schroeder, 1965).

Fréquence de Schroeder

La fréquence de Schroeder F_s donne un ordre de grandeur de la fréquence à partir de laquelle on passe d’un modèle déterministe (théorie modale) à un modèle stochastique pour décrire l’effet de salle.

$$F_s \simeq 2000 \cdot \sqrt{\frac{Tr_{60}}{V}} \quad (\text{E.9})$$

HRIR (Head Related Impulse Response)

Réponse impulsionnelle d’une source située en champ libre mesurée à l’entrée du canal auditif (voir HRTF).

HRTF (Head Related Transfer Function)

Transformée de Fourier d’une HRIR. En pratique on déconvolue la HRTF mesurée à l’entrée du canal auditif par la réponse d’un microphone de référence placé au centre de la tête du participant en absence de celui-ci, de manière à éliminer la coloration liée à la chaîne de mesure employée (amplificateur, haut-parleur, etc..). Une HRTF contient donc uniquement l’information liée à la diffraction par le torse, la tête et l’oreille externe d’une personne pour une position de source donnée (distance et angle d’incidence). En d’autres termes, une HRTF contient tous les indices (binauraux et monauraux) nécessaires à l’écoute spatialisée pour une personne donnée. Une base de donnée de HRTF contient les HRTF mesurées sur un même sujet pour un ensemble de positions de source (Page web de l’équipe Acoustique des Salles, 2007) (Larcher, 2001).

Incohérence

Qualité d'un faisceau d'ondes ne disposant pas de caractéristiques cohérentes, i.e. pour lequel il n'y a en moyenne ni interférence destructives, ni interférences constructives (basé sur (TechnoScience, 2007)).

Listener Envelopment (LEV)

Le LEV (enveloppement de l'auditeur) est un des deux composantes (avec l'ASW) de "l'impression spatiale" dans une salle de concert. On montre que le LEV est influencé principalement par les réflexions latérales arrivant plus de 80ms après le son direct. Une étude menée dans (Bradley & Soulodre, 1995b) a montré qu'un descripteur simple et efficace du LEV était le LG_{80} défini de la manière suivante :

$$LG_{80}^{\infty} = 10 \log \left[\frac{\int_{80ms}^{\infty} (g(t) \cos\theta)^2 dt}{\int_0^{\infty} g_a^2(t) dt} \right] dB \quad (E.10)$$

et restreint aux bandes de fréquences moyennes (125, 250, 500, and 1000 Hz) de manière à tenir compte du fait que les basses fréquences ne contribuent pas à l'enveloppement.

Parallaxe de mouvement

L'effet de parallaxe, ou plus précisément la "parallaxe de mouvement", correspond à un changement des directions d'incidences respectives de deux points immobiles, perçu par un observateur lors d'un déplacement de celui-ci. Dit autrement, la parallaxe correspond au glissement d'un objet proche par rapport à l'arrière plan lors d'un déplacement de l'observateur. Il ne peut y avoir de parallaxe entre deux objets s'ils se situent côte à côte à une même profondeur (Wikipedia, 2007). La parallaxe de mouvement n'est pas à confondre avec la "parallaxe acoustique" qui correspond aux variations de différences intéraurales perçues lorsqu'une source sonore entre dans le champ proche de la tête de l'auditeur, et qui contribuent à renseigner celui-ci sur la distance de la source (Brungart & Rabinowitz, 1999; Brungart *et al.*, 1999).

Réverbération

La persistance plus ou moins longue de son entendue lors de l'arrêt soudain d'une source sonore (Cremer *et al.*, 1978) (voir *Temps de Réverbération*).

Son Direct

Partie du champ acoustique émise par la source arrivant à l'auditeur suivant le trajet le plus court, sans subir l'influence des parois de la salle d'écoute (Kuttruff, 1991).

Source focalisée

Source virtuelle synthétisée à l'intérieur de l'espace de reproduction par le biais de fronts d'ondes convergents émis par le banc WFS (sec 1.3.3).

Source virtuelle

Source acoustique dont on vise à reproduire le rayonnement via à un système électroacoustique. On trouve également le terme équivalent de *source notionnelle* dans la littérature.

Spaciousness/Spatial Impression (Impression d'espace)

L'impression d'espace est une quantité multidimensionnelle dont on peut dire qu'elle dépend principalement des réflexions latérales arrivant sur l'auditeur. On peut diviser cette impression en deux dimensions principales : d'une part, la dimension d'enveloppement ou *Listener Envelopment* (LEV) et d'autre part la largeur apparente de source ou *Apparent Source Width* (ASW). Cependant, on montre que l'impression spatiale dépend également du niveau, et qu'elle croît de manière monotone avec le rapport

$$H = 10 \log \left[\frac{E_r}{E_0} \right] \text{ dB} \quad (\text{E.11})$$

entre l'énergie réverbérée E_r et l'énergie du son direct E_0 (Kuttruff, 2000).

(voir également **Apparent Source Width** et **Listener Envelopment**)

Temps de mélange

Dans une réponse impulsionnelle de salle, le temps de mélange T_{mix} correspond à la durée au bout de laquelle on peut appliquer un modèle statistique pour décrire l'effet de salle, du fait que l'audition n'est plus sensible à l'effet individuel de chaque réflexion (Polack, 1988). On peut estimer que lorsque le nombre d'échos arrivant à l'oreille dépasse 10 dans un intervalle de temps Δt égal au seuil d'écho de l'audition on obtient un mélange perceptivement convaincant

quel que soit le point d'écoute dans une salle ergodique. On trouve :

$$T_{mix} \simeq \sqrt{\frac{10V}{4\pi c^3 \Delta t}} \quad (\text{E.12})$$

En prenant un seuil d'écho de l'ordre de 20ms (ce qui est cohérent avec les résultats d'expérience sur la précédence) on obtient un expression simplifiée de l'équation E.12 traduisant que le temps de mélange dans une salle ergodique est proportionnel à la racine carrée du volume V de la salle :

$$T_{mix} \simeq \sqrt{V} \quad (\text{E.13})$$

Temps de réverbération (Tr_{60})

Temps mis pour que le niveau d'énergie acoustique présente dans une salle chute de 60dB après interruption soudaine de la source (Sabine, 1900). Aujourd'hui, dans la plupart des standards, le temps de réverbération est calculé par régression linéaire sur la pente de décroissance intégrée d'une réponse impulsionnelle de la salle (Schroeder, 1965), dans la plage énergétique de -5 à -35dB par rapport au niveau maximal (ISO3382, 1993).

Timbre

Attribut de la sensation cochléaire par lequel un auditeur peut juger de la dissimilarité entre deux sons ayant même sonie et même hauteur.

Transparence

Qualité attribuée à des haut-parleurs dont les propriétés acoustiques permettent de faire disparaître leur présence individuelle au profit de la source virtuelle qu'ils synthétisent ; on dira qu'ils permettent d'assurer une "illusion de non-médiation" (Lombard & Ditton, 1997).

References

- Abhayapala, T. D., & Ward, D. B. 2002. Theory et Design of High Order Sound Field Microphones Using Spherical Microphone Array. *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Mai.
- Baskind, A. 1999 (Juillet). *Etude de la puissance acoustique rayonnee en champ diffus*. M.Phil. thesis, DEA ATIAM.
- Baskind, A. 2003. *Modeles et Methodes de Description Spatiale des Scenes Sonores, application aux enregistrements binauraux*. Ph.D. thesis, Universite de Paris VI, Paris, France.
- Beer, Daniel, Brix, Sandra, de Vries, Diemer, & Kuster, Martin. 2005. Structural and Acoustic Analysis of Multi-Actuator Panels. *In : AES 118th Convention*.
- Beranek, L. 1954. *Acoustics*. McGraw Hill.
- Berkhout, A. J. 1988. A holographic approach to acoustic control. *Journal of the Audio Engineering Society*, **36**(12), 977–995.
- Berkhout, A. J., de Vries, D., & Vogel, P. 1993. Acoustic Control by Wave Field Synthesis. *Journal of the Acoustical Society of America*, **93**(5), 2764–2778.
- Berkhout, A. J., de Vries, D., Baan, J., & van den Oetelaar, B. W. 1998. A wave field extrapolation approach to acoustical modeling in enclosed spaces. *Journal of the Acoustical Society of America*, **105**(3), 1725–1733.
- Betlehem, T., & Abhayapala, T.D. 2005. Theory and design of sound field reproduction in reverberant rooms. *Journal of the Acoustical Society of America*, **117**(4), 2100–2111.
- B&O. 2007. *Virtual Factory Tour*. Website.
- Bortfeldt, J., & Kramer, B. 1991. *Units in Physics and Chemistry - SubVol A*. Vol. 16. Springer-Verlag.
- Boulez, P. 1983 (Juillet). Intervention Table Ronde. *Pages 213–216 of : International Congress on Acoustics*, vol. 11.
- Bradley, J., & Soulodre, G. 1995a. The Influence of late arriving energy on spatial impression. *Journal of the Acoustical Society of America*, **97**(4), 2263–2271.

- Bradley, J., & Soulodre, G. 1995b. Objective measures of listener envelopment. *Journal of the Acoustical Society of America*, **98**(5), 2590–2597.
- BruelKjaer. 1984. *Measuring Sound*. Tech. rept. Bruel and Kjaer.
- BruelKjaer. 1993. *Sound Intensity*. Tech. rept. Bruel and Kjaer.
- Bruneau, M. 1998. *Manuel d'Acoustique Fondamentale*. Hermes.
- Brungart, D., & Rabinowitz, W. 1999. Auditory localization of nearby sources. Head-related transfer functions. *Journal of the Acoustical Society of America*, **106**(3), 1465–1479.
- Brungart, D., Durlach, N., & Rabinowitz, W. 1999. Auditory localization of nearby sources. II. Localisation of a broadband source. *Journal of the Acoustical Society of America*, **106**(4), 1956–1968.
- Caulkins, T. 2003 (Juillet). *Restitution du effet de salle coherent avec le lieu d'ecoute dans le cadre de la Wave Field Synthesis*. M.Phil. thesis, Dea Atiam, Universite Paris 6.
- Caulkins, T., & Girier, A. 2007 (August). *Loudspeaker radiation prediction method based on a spherical harmonic decomposition*. submitted to Journal of the Acoustical Society of America.
- Caulkins, T., & Warusfel, O. 2006 (May). Characterization of the Reverberant Sound Field Emitted by a Wave Field Synthesis Driven Loudspeaker Array. *In : AES 120th Convention*.
- Caulkins, T., Corteel, E., & Warusfel, O. 2003 (Septembre). Wave Field Synthesis Interaction with the Listening Environment, Improvements in the Reproduction of Virtual Sources Located Inside the Listening Room. *In : DAFX03*. poster.
- Caulkins, T., Corteel, E., & Warusfel, O. 2004 (Novembre). Analysis of Certain Challenges for the use of Wave Field Synthesis in Concert-Based Applications. *In : DAFX04*.
- Caulkins, T., Laborie, A., Corteel, E., Bruno, R., Montoya, S., & Warusfel, O. 2006 (July). Use of a high spatial resolution microphone to characterize the early reflections generated by a WFS loudspeaker array. *In : AES 28th Conference*.
- Chiron, G., Poletti, M., Corteel, E., & Caulkins, T. 2007 (Mars - Juin). *Sound Installation*. Samuel Beckett exhibition.
- Clement, G. T., Liu, R., Letcher, S. V., & Stepanishen, P. R. 1997. Temporal backward planar projection of acoustic transients. *Journal of the Acoustical Society of America*, **103**(4), 1723–1726.
- Cook, R. K., Waterhouse, R. V., Berendt, R. D., Edelman, S., & Thompson, M. C. 1955. Correlation Coefficients in Reverberant Sound Fields. *Journal of the Acoustical Society of America*, **27**(6), 1072–1077.
- Copley, L. G. 1968. Fundamental Results Concerning Integral Representations in Acoustic Radiation. *Journal of the Acoustical Society of America*, **44**(1), 28–32.
- Corsino, N+N, & Corteel, E. 2006. *Installation sonore*. Exposition Seule avec Loup.
- Corteel, E. 2004. *Caracterisation et Extensions de la Wave Field Synthesis en conditions reelles*. Ph.D. thesis, Universite de Paris VI, Paris, France.
- Corteel, E. 2006a. Equalization in an extended area using multichannel inversion and Wave Field Synthesis. *Journal of Audio Engineering Society*, **54**(12), 1140–1161.

REFERENCES

- Corteel, E. 2006b (September). On the Use of Irregularly Spaced Loudspeaker arrays for Wave Field Synthesis, potential impact on spatial aliasing frequency. *In : DAFX-06*.
- Corteel, E. 2007. Synthesis of Directional Sources Using Wave Field Synthesis, Possibilities, and Limitations. *EURASIP Journal on Advances in Signal Processing*.
- Corteel, E., & Nicol, R. 2003 (Juin). Listening Room Compensation for Wave Field Synthesis. what can be done ? *In : 23th International Conference of the Audio Engineering Society*.
- Corteel, E., Horbach, U., & Pellegrini, R. S. 2002 (Mai). Multichannel Inverse Filtering of Multiexciter Distributed Mode Loudspeaker for Wave Field Synthesis. *In : 112th Convention of the Audio Engineering Society*. Preprint Number 5611.
- Corteel, E., NGuyen, K.V., Warusfel, O., Caulkins, T., & Pellegrini, R. 2007 (March). Objective and subjective comparison of electrodynamic and MAP loudspeakers for Wave Field Synthesis. *In : AES 30th Conference*.
- Cremer, L., Muller, H., & Schultz, T. 1978. *Principles and Applications of Room Acoustics*. Applied Science Publisher Ltd.
- Daniel, J. 2000. *Representation de Champs Acoustiques, Application a la Transmission et a la Reproduction de Scenes Sonores Complexes dans un Contexte Multimedia*. Ph.D. thesis, Universite de Paris VI, Paris, France.
- Daniel, J., & Moreau, S. 2004 (Mars). Design refinement of high order ambisonics microphones - Experiments with a 4th order prototype. *In : CFA/DAGA 04*.
- Daniel, J., Nicol, R., & Moreau, S. 2003 (Mars). Further Investigations of High Order Ambisonics and Wave Field Synthesis for Holophonic Sound Imaging. *In : 114th Convention of the Audio Engineering Society*.
- DeBruijn, W. 2004. *Applications of Wave Field Synthesis in Videoconferencing*. Ph.D. thesis, TU Delft, Delft, Netherlands.
- Derogis, P. 1999 (Dec.). *Analyse des vibrations et du rayonnement de la table d'harmonie d'un piano droit et conception d'un systeme de reproduction du champ acoustique*. Ph.D. thesis, Université Paris 6.
- Desnoux, J.C., & Caulkins, T. 2005 (octobre). *Installation sonore*. Exposition Permanente de la Cinematheque Francaise.
- Evans, M. J., Angus, J. A. S., & Tew, A. I. 1998. Analyzing head-related transfer function measurements using surface spherical harmonics. *Journal of the Acoustical Society of America*, **104**(4), 2400–2411.
- Farina, A. 2000. Simultaneous Measurement of Impulse Response and Distorsion with a swept-sine technique. *In : AES 108th convention*.
- Farina, A., Martignon, P., Capra, A., & Fontana, S. 2006. Measuring impulse responses containing complete spatial information. *Journal of the Acoustical Society of America*, **120**(5).
- Forbes, M., Letcher, S., & Stepanishen, P. R. 1991. A wave vector, time-domain method of forward projecting time-dependent pressure fields. *Journal of the Acoustical Society of America*, **90**(5), 2782–2793.
- Gardner, M. 1968. Historical Background of the Haas and/or Precedence Effect. *Journal of the Acoustical Society of America*, **43**(6), 1243–1248.

- Gardner, William S. 1995. *Transaural 3D audio*. Tech. rept. MIT Media Lab.
- Gauthier, P. A., & Berry, A. 2005. Sound-field reproduction in-room using optimal control techniques : Simulations in the frequency domain. *Journal of the Acoustical Society of America*, **117**(2), 662–678.
- Gauthier, P. A., & Berry, A. 2006. Adaptive wave field synthesis with independent radiation mode control for active sound field reproduction : Theory. *Journal of the Acoustical Society of America*, **119**(5), 2721–2737.
- Gerzon, M. A. 1973. Periphony : With-Height Sound Reproduction. *Journal of the Audio Engineering Society*, **21**(1), 2–10.
- Gerzon, M. A. 1975. Recording Concert Hall Acoustics for Posterity. *Journal of the Audio Engineering Society*, **23**(7), 569.
- Gerzon, M. A. 1985. Ambisonics in Multichannel Broadcasting et Video. *Journal of the Audio Engineering Society*, **33**(11), 859–871.
- Giron, Franck. 1996. *Investigations about the directivity of sound sources*. Ph.D. thesis, Ruhr-Universität, Bochum.
- Gontcharov, V., Hill, N., & Taylor, V. 1999 (April). Measurement Aspects of Distributed Mode Loudspeakers. In : *AES 106th Convention*.
- Grace, O. 1981. Instantaneous power spectra. *Journal of the Acoustical Society of America*, **69**(1), 191–198.
- Grand, G., Corteel, E., & Kronenberg, R. 2005 (October 2005 - January 2006). *Installation sonore*. DADA exhibition.
- Guillemin, V., & Sternberg, S. 1977. *Geometric Asymptotics*. AMS.
- Hartmann, W.M. 1983. Localization of sound in rooms. *Journal of the Acoustical Society of America*, **74**(5), 1380–1391.
- HDDigest. 2007. Disc Sales : Hard Numbers in for First Half of 2007. *High-Def Digest*.
- Horton, C.W., & Innis, G.S. 1961. The Computation of Far-Field Radiation Patterns from Measurements Made near the Source The Computation of Far-Field Radiation Patterns from Measurements The computation of far-field radiation patterns from measurements made near the source. *Journal of the Acoustical Society of America*, **33**(7), 877–880.
- Hulsebos, E. 2004. *Auralization using Wave Field Synthesis*. Ph.D. thesis, TU Delft, Delft, Netherlands.
- Hulsebos, E., & de Vries, D. 2002 (Mai). Parameterization et reproduction of concert hall acoustics measured with a circular microphone array. In : *112th Convention of the Audio Engineering Society*. Preprint Number 5579.
- Hulsebos, E., de Vries, D., & Bourdillat, E. 2001 (Mai). Improved microphone array configurations for auralization of sound fields by wave field synthesis. In : *110th Convention of the Audio Engineering Society*. Preprint Number 5337.
- Huygens, C. 1690. *Traite de la Lumiere*. Pierre van der Aa, Leiden.
- ISO3382. 1993. *Measurement of the reverberation time of rooms with reference to other acoustical parameters*.

REFERENCES

- Jacob, K., & Birkle, T. 1990. Prediction of the Full-Space Directivity Characteristics of Loudspeaker Arrays. *Journal of Audio Engineering Society*, **38**(4), 250–258.
- Jacobsen, F. 1997. An Overview of the Sources of Error in Sound Power Determination Using the Intensity Technique. *Applied Acoustics*, **50**(2), 155–166.
- Jacobsen, F., & DeBree, Hans. 2004. A comparison of p-p and p-u sound intensity measurement systems. In : *Eleventh International Congress on Sound and Vibration*.
- Jacobsen, F., Cutanda, V., & Juhl, P. M. 1998. A numerical and experimental investigation of the performance of sound intensity probes at high frequencies. *Journal of the Acoustical Society of America*, **103**(2), 953–961.
- Jacques, R., Albrecht, B., Melchior, F., & De Vries, D. 2005 (October). An approach for multi-channel Recording and Reproduction of Sound Source Directivity. In : *Proceedings of the AES 119th convention*. AES, NY, NY USA.
- Johansen, T. F. 1994. On the Directivity of Horn Loudspeakers. *Journal of Audio Engineering Society*, **42**(12), 1008–1019.
- Jot, J. M., & Warusfel, O. 1995. Spat : A Spatial Processor for Musicians and Sound Engineers. In : *CIARM 95*.
- Jot, J. M., Jullien, J. P., & Warusfel, O. 1998. *Method to simulate the acoustical quality of a room and associated audio-digital processor*. U.S. Patent Number 5812674.
- Jot, J.M. 1992 (Sept.). *Etude et réalisation d'un spatialisateur de sons par modèles physiques et perceptifs*. Ph.D. thesis, ENST.
- Jot, J.M., Cerveau, L., & Warusfel, O. 1997 (Sept.). Analysis and synthesis of room reverberation based on a time-frequency model. In : *AES 103rd convention preprint*. AES. preprint 4629.
- Jullien, J.-P. 1995. Structured model for the representation and the control of room acoustical quality. *Pages 517–520 of : Proceedings of the 15th International Conference on Acoustics*.
- Jullien, J.-P., Kahle, E., Winsberg, S., & Warusfel, O. 1992. Some results on the objective and perceptual characterization of room acoustical quality in both laboratory and real environments. *Proc. Institute of Acoustics*, **XIV**(3).
- Kahle, Eckhard. 1995 (June). *Validation d'un modèle objectif de la perception de la qualité acoustique dans un ensemble de salles de concerts et d'opéras*. Ph.D. thesis, Université du Maine, Le Mans.
- Karjalainen, M., Ikonen, V., Antsallo, P., Maijala, P., Savioja, L., Suutala, A., & Pohjolainen, S. 2001. Comparison of Numerical Simulation Models and Measured Low-Frequency Behavior of Loudspeaker Enclosures. *Journal of Audio Engineering Society*, **49**(12), 1148–1166.
- Kassakian, P., & Wessel, D. 2004 (Oct). Characterization of Spherical Loudspeaker Arrays. In : *117th Convention of the Audio Engineering Society*.
- Katz, B., & d'Alessandro C. 2007 (September). Directivity Measurements of the Singing voice. In : *International Congress on Acoustics*, vol. 19.
- Katz, B., Prezat, F., & d'Alessandro C. 2006. Human voice phoneme directivity pattern measurements. In : *4th Joint Meeting Acoustical Society of America And Acoustical Society of Japan*.

- Kawakami, F., & Yamaguchi, K. 1981. Space-ensemble average of reverberation decay curves. *JASA*, **70**(4), 1071–1082.
- Kinsler, L., Frey, A., Coppens, A., & J., Sanders. 2000. *Fundamentals of Acoustics*. 4th edition edn. John Wiley and sons.
- Kleiner, M., Dalenback, B.-I., & Svensson, P. 1993. Auralization - an overview. *Journal of the Audio Engineering Society*, **41**(11), 861–875.
- Kuttruff, H. 1991. *Room Acoustics*. Applied Science Publisher Ltd.
- Kuttruff, H. 2000. *Room Acoustics*. 4th edn. SPON Press.
- Laborie, A. 2000 (Octobre). *Capture échantillonnage et manipulations de fonctions de directivité. Application à l'enregistrement de scènes sonores 3D avec un microphone SoundField*. M.Phil. thesis, Ecole Nationale Supérieure des Télécommunications.
- Laborie, A., Bruno, R., & Montoya, S. 2003 (Mars). A new comprehensive approach of surround sound recording. In : *114th Convention of the Audio Engineering Society*. Preprint Number 5717.
- Laborie, A., Bruno, R., & Montoya, S. 2004a (Oct). Designing High Spatial Resolution Microphones. In : *117th Convention of the Audio Engineering Society*.
- Laborie, A., Bruno, R., & Montoya, S. 2004b (Mars). High Spatial Resolution Multichannel Recording. In : *116th Convention of the Audio Engineering Society*. Poster.
- Larcher, V. 2001. *Techniques de spatialisation de sons pour la realite virtuelle*. Ph.D. thesis, Universite de Paris VI, Paris, France.
- Lavandier, C. 1989. *Validation perceptive d'un modèle objectif de caracterisation de la qualité en acoustique des salles*. Ph.D. thesis, Universite du Maine, Le Mans, France.
- Lombard, M., & Ditton, T. 1997. At the heart of it all : The concept of presence. *Journal of Computer-Mediated Communication*, **3**(2).
- Lopez, J., Gonzales, A., & Fuster, L. 2005 (October). Room compensation in wave field synthesis by means of multichannel inversion. *Pages 146–149 of : IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*.
- Malham, David. 2001. Spherical Harmonic Encoding of Sound Objects - the Ambisonic 'O' Format. In : *AES 19th International Conference*.
- Mann, J.A., Tichy, J., & Romano, A.J. 1987. Instantaneous and time-averaged energy transfer in acoustic fields. *Journal of the Acoustical Society of America*, **82**(1), 17–30.
- Maresz, Y., Naves S., & Corteel, E. 2004. *Sul Segno : Adaptation pour WFS*.
- Mathworks, The. 2006. *MATLAB version 7.2.0*. computer software.
- Maynard, J. D., Williams, E. G., & Lee, Y. 1985. Nearfield acoustic holography : I. Theory of generalized holography and the development of NAH. *Journal of the Acoustical Society of America*, **78**(4), 1395–1413.
- Menzies, D. 1999. *New electronic performance instruments for electroacoustic music*. Ph.D. thesis, University of York.
- Meyer, J. 1972. Directivity of the Bowed Stringed Instruments and Its Effects on Orchestral Sound in Concert Halls. *Journal of the Acoustical Society of America*, **51**(6), 1994–2009.

REFERENCES

- Meyer, J., & Elko, G. W. 2002. A Highly Scalable Spherical Microphone Array Based on an Orthonormal Decomposition of the Soundfield. *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 2(May), 1781–1784.
- Moore, B. C. J., & Tan, A. 2004. Development and Validation of a Method for Predicting the Perceived Naturalness of Sounds Subjected to Spectral Distorsion. *Journal of Audio Engineering Society*, 52(9), 900–914.
- Moorer, J. 1978. About this reverberation business. *Computer Music Journal*.
- Morse, P., & Ingard, U. 1968. *Theoretical Acoustics*. Princeton University Press.
- Nicol, R. 1999. *Restitution sonore spatialisée sur une zone étendue : Application à la téléprésence*. Ph.D. thesis, Université du Maine, Le Mans, France, Le Mans, France.
- Nicol, R., & Emerit, M. 1999. 3D-Sound Reproduction over an Extensive Listening Area : A Hybrid Method Derived From Holophony and Ambisonic. *In : AES 16th*.
- Nogues, M., Corteel, E., & Warusfel, O. 2003 (Septembre). Monitoring Distance Effect with Wave Field Synthesis. *In : DAFX03*. poster.
- Norcross, S. G., Soulodre, G. A., & Lavoie, M. C. 2004. Subjective Effects of Inverse Filtering. *Journal of the Audio Engineering Society*, 52(10), 1003–1028.
- Page web de l'équipe Acoustique des Salles, IRCAM. 2007. <http://recherche.ircam.fr/equipes/salles/>.
- Pellegrini, R. S. 2001. *A Virtual Listening Room as an Application of Auditory Virtual Environments*. Ph.D. thesis, Ruhr-Universität, Bochum, Allemagne.
- Pellerin, G., Polack, J. D., & Morkerken, J. P. 2003. Finite element methods and equivalent electrical models for loudspeaker characterization. *In : AES 114th Convention*.
- Penna, P. 1993. *Some statistical properties of reverberation process in diffuse sound fields*. M.Phil. thesis, Université du Maine and Laboratoire for Akustik(Danemark).
- Polack, J. D. 1988. *La transmission de l'énergie sonore dans les salles*. Ph.D. thesis, Université du Maine.
- Pulkki, V. 1997. Virtual Sound Source Positioning Using Vector Based Amplitude Panning. *Journal of the Audio Engineering Society*, 45(6).
- Rafaely, B. 2005. Analysis and Design of Spherical Microphone Arrays. *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 13(1), 135–143.
- Ritchie, D., & Kemp, G. J. L. 1999. Fast computation, rotation, and comparison of low resolution spherical harmonic molecular surfaces. *Journal of Computational Chemistry*, 20(4), 383–395.
- Rochefoucauld, O., Melon, M., & Garcia, A. 2004. Time domain holography : forward projection of simulated and measured sound pressure fields. *Journal of the Acoustical Society of America*, 116(1), 142–153.
- Russell, D. A., Titlow, J., & Bemmen, Y. 1999. Acoustic Monopoles, dipoles and quadrupoles : An experiment revisited. *American Journal of Physics*, 67(8), 660–664.
- Sabine, W. C. 1900. Reverberation. *The American Architect*.

- Sanson, J. 2007. *Evaluation objective et subjective des erreurs de localisation en Wave Field Synthesis*. M.Phil. thesis, Universite Paris 6.
- Schenck, H. A. 1968. Improved integral formulation for acoustic radiation problems. *Journal of the Acoustical Society of America*, **44**(July), 41–58.
- Schoeller, P., & T., Caulkins. 2004. *Vertigo Apocalypse : Adaptation pour WFS*.
- Schroeder, M. R. 1965. New method of measuring reverberation time. *Journal of the Acoustical Society of America*, **37**(3), 409–412.
- Schroeder, M. R. 1979. Integrated-impulse method measuring sound decay without using impulses. *Journal of the Acoustical Society of America*, **66**(2), 497–500.
- Schroeder, M. R. 1987. Statistical parameters of the frequency response curves of large rooms. *JAES*, **35**(5), 299–305.
- Shindo, T., Yoshioka, T., & Fukuyama, K. 1990. Calculation of sound radiation from an un baffled, rectangular-cross-section Horn Loudspeaker Using Combined analytical and boundary-element methods. *Journal of Audio Engineering Society*, **38**(5), 340–349.
- Skudrzyk, E. 1971. *The Foundations of Acoustics : Basic Mathematics and Basic Acoustics*. Springer-Verlag.
- Somerville, T., Gilford, C. L. S., Spring, N. F., & Negus, R. D. M. 1966. Recent work on the effects of reflectors in concert halls and music studios. *Journal of Sound and Vibration*, **3**(2), 127–134.
- Sonke, J. J. 2000. *Variable Acoustics by Wave Field Synthesis*. Ph.D. thesis, TU Delft, Delft, Netherlands.
- Sonke, J. J., & de Vries, D. 1997 (Mars). Generation of diffuse reverberation by plane wave synthesis. In : *102nd Convention of the Audio Engineering Society*. Preprint Number 4455.
- Spors, S., Kuntz, A., & Rabenstein, R. 2003 (Juin). An Approach to Listening Room Compensation with Wave Field Synthesis. In : *24th International Conference of the Audio Engineering Society*.
- Spors, S., Buchner, H., & Rabenstein, R. 2004a (Mars). Efficient Active Listening Room Compensation for Wave Field Synthesis. In : *116th Convention of the Audio Engineering Society*.
- Spors, S., Buchner, H., & Rabenstein, R. 2004b (Mars). A Novel Approach to active listening room compensation for Wave Field Synthesis using Wave-Domain Adaptive Filtering. In : *116th Convention of the Audio Engineering Society*.
- Spors, S., Buchner, H., Rabenstein, R., & Herbordt, W. 2007. Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering. *Journal of the Acoustical Society of America*, **122**(1), 354–369.
- Start, E. W. 1997. *Direct Sound Enhancement by Wave Field Synthesis*. Ph.D. thesis, TU Delft, Delft, Netherlands.
- TechnoScience. 2007. *Encyclopédie scientifique en ligne* (<http://www.techno-science.net/>).
- Theile, G., Wittek, H., & Reisinger, M. 2003 (Juin). Potential Wave Field Synthesis Application in the Multichannel Stereophonic World. In : *24th International Conference of the Audio Engineering Society*.

REFERENCES

- Tsuchiya, T., Kagawa, Y., Doi, M., & Tsuji, T. 2003. Finite element simulation of non-linear acoustic generation in a horn loudspeaker. *Journal of Sound and Vibration*, **266**(5), 993–1008.
- Vaananen, R., Warusfel, O., & Emerit, M. 2002 (Juin). Encoding and Rendering of perceptual sound scenes in the Carrouso Project. *In : 22nd International Conference of the Audio Engineering Society*.
- van Zon, R. 2003 (Octobre). *Room compensation for Wave Field Synthesis using MAP loudspeakers*. M.Phil. thesis, TU Delft.
- Verheijen, E. N. G. 1997. *Sound Reproduction by Wave Field Synthesis*. Ph.D. thesis, TU Delft, Delft, Netherlands.
- Verin, N., J., Perraudou, & Corteel, E. 2003. *21 rue du Midi : Adaptation pour WFS*.
- Veronesi, W. A., & Maynard, J. D. 1987. Nearfield acoustic holography (NAH) II. Holographic reconstruction algorithms and computer simulations. *Journal of the Acoustical Society of America*, **81**(5), 1307–1322.
- Vogel, P. 1993. *Application of Wave Field Synthesis in room acoustics*. Ph.D. thesis, TU Delft, Delft, Netherlands.
- Wapenaar, C.P.A., & Berkhout, A.J. 1989. *Elastic Wave Field Extrapolation*. Elsevier.
- Warusfel, O. 1990. Etude des paramètres liés à la prise de son pour les applications d'acoustique virtuelle. *Pages 877–880 of : Congrès Français d'Acoustique*, vol. 2.
- Warusfel, O., & Misdariis, N. 2001 (Dec). Directivity Synthesis with a 3D array of Loudspeakers Application for Stage Performance. *In : DAFX-01*.
- Warusfel, O., Corteel, E., & Caulkins, T. 2004a (Mars). Perceptual Spatial Exploration of WFS Soundscapes. *In : CFA/DAGA 04*.
- Warusfel, O., Corteel, E., Misdariis, N., & Caulkins, T. 2004b (Avril). Reproduction of sound source directivity for future audio applications. *In : ICA04*.
- Weinreich, G. 1997. Directional Tone Color. *Journal of the Acoustical Society of America*, **101**(4), 2338–2346.
- Weinreich, G., & Arnold, E. 1980. Method for measuring acoustic radiation fields. *Journal of the Acoustical Society of America*, **68**(2), 404–411.
- Wightman, F. L., & Kistler, D. J. 1992. The dominant role of low-frequency interaural time differences in sound localization. *Journal of the Acoustical Society of America*, **91**(3), 1648–1661.
- Wikipedia. 2007. *Wikipedia : The free encyclopedia (www.wikipedia.org)*.
- Williams, E. 1999. *Fourier Acoustics*. Academic Press.
- Witteck, H. 2004 (Mars). Spatial perception in Wave Field Synthesis rendered sound fields : Distance of real and virtual nearby sources. *In : CFA/DAGA 04*.
- Yates, K. 1995. *Virtual Acoustic Reality*.