

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS 6

École doctorale EDITE

Spécialité

Traitement du signal

présentée par

Etienne CORTEEL

etienne.corteel@ircam.fr

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE PARIS 6

Caractérisation et Extensions de la Wave Field Synthesis en conditions réelles

soutenue le 9 décembre 2004

devant le jury composé de :

Jean-Dominique POLACK	Directeur de thèse
Pierre-Olivier MATTEI	Rapporteur
Diemer DE VRIES	Rapporteur
Xavier MEYNIAL	Examineur
Bernard ROBINET	Examineur
Olivier WARUSFEL	Examineur

Version corrigée : 25 juillet 2006

Résumé

La Wave Field Synthesis est une technique de reproduction sonore qui vise à reproduire les caractéristiques physiques du champ acoustique dans une zone d'écoute étendue. Elle dépasse ainsi les limites des techniques stéréophoniques et permet la restitution d'une véritable perspective sonore à l'aide d'un réseau linéaire de haut-parleurs. Cette technique repose sur un ensemble d'approximations des principes physiques sous-jacents qui n'assure cependant pas une restitution exacte.

Le travail réalisé a consisté à compléter le cadre théorique et, à analyser et à valider de manière systématique la qualité du champ acoustique synthétisé en termes perceptifs (indices de localisation, coloration). Au-delà de cette analyse, nous proposons des techniques d'égalisation multicanal pour compenser les caractéristiques réelles des transducteurs et tenir compte de celles des parois de la pièce de restitution. Enfin, nous présentons une chaîne de production qui permet la gestion de la perspective sonore par le créateur de contenu.

Mots clés : spatialisation, reproduction sonore, Wave Field Synthesis, holophonie, coloration, égalisation, égalisation multicanal, inversion multicanal, contrôle acoustique, compensation de l'effet de salle, indices de localisation, perspective sonore, reproduction d'effet de salle, édition de scène sonore, manipulation de scène sonore.

Abstract

Wave Field Synthesis is a sound reproduction technique that aims at reproducing the physical characteristics of an acoustic field over an extended listening area. It therefore overcomes the limitations of stereophonic based techniques and allows for the synthesis of true sound perspective. The technique is based on a set of approximations of the underlying physical principles, which implies errors in the reproduced sound field.

The work presented in this document consisted in completing the theoretical framework and in analyzing and validating the quality of the synthesized acoustic field in perceptual terms (localization cues, coloration). Beyond this analysis, multichannel equalization techniques are proposed to compensate for the true characteristics of the transducers and account for those of the boundaries of the listening room. Finally, we present a production chain allowing for the control of sound perspective by the content creator.

Key words : spatialization, sound reproduction, Wave Field Synthesis, holophony, coloration, equalization, multichannel equalization, multichannel inversion, acoustic control, listening room compensation, localization cues, sound perspective, room effect synthesis, sound scene manipulation, sound scene edition.

Remerciements

Je voudrais tout d'abord remercier tous ceux que je risque d'omettre ou d'oublier dans les lignes qui viennent... Qu'ils sachent que je leur saurais toujours grès de tout ce qu'ils ont pu m'apporter.

Je voudrais ensuite remercier les membres du jury et spécialement Diemer de Vries et Pierre-Olivier Mattei qui ont bien voulu être les rapporteurs de cette thèse. Les remarques et suggestions qu'ils m'ont faites m'ont été très utiles. J'ai essayé autant que possible de les prendre en compte pour la version finale de ce document. Je remercie Diemer de Vries et toute l'équipe de l'université de Delft d'avoir inventé la Wave Field Synthesis. Leur travail pionnier reste une source immense d'inspiration.

Je remercie les gens de la communauté scientifique et musicale qui m'ont encouragé dans cette "voix". Je pense notamment à l'équipe du Premium Sound Lab de Studer (Ulrich Horbach, Renato Pellegrini, et Attila Karamustafaoglu) qui ont accueilli le petit français que j'étais, ont su le lancer et lui ont donné leur confiance. Je pense aussi aux collaborateurs du projet Carrouso qui ont su communiquer leur enthousiasme et leurs compétences. Je pense en particulier à Rozenn Nicol, Helmut Wittek, Edo Hulsebos, Jan Plogsties, Sascha Spors, Marc Emerit, Gunther Theile, Rinus Boone, et Yannick Mahieux. Ce projet a été un élément clé dans mon travail. Il a fourni un cadre applicatif idéal à mes recherches.

Je tiens également à remercier tous les collaborateurs de l'équipe acoustique des salles pour leur accueil, leurs compétences et leur sympathie (Terence Caulkins, Olivier Delerue, Riitta Vänänen, Sébastien Roux, Alexis Baskind, Cédric Peigne, Emmanuel Rio, Guillaume Vandernoot, Matthieu Noguès, Clemens Kuhn, Suzanne Winsberg). Je remercie aussi les gens dont j'ai pu faire la connaissance à l'IRCAM. Ils m'ont beaucoup apporté tant sur le plan professionnel que sur le plan humain.

Je remercie mes directeurs de thèse, Jean-Dominique Polack et Olivier Warusfel, pour les discussions instructives et productives que nous avons eues au cours de ces années. Je tiens tout spécialement à remercier Olivier Warusfel pour m'avoir accueilli si gentiment au sein de l'équipe acoustique des salles de l'IRCAM. Je le remercie aussi pour m'avoir accordé sa confiance et son soutien tout au long de ces années, ainsi que pour son aide extrêmement précieuse lors de la rédaction de son document.

Enfin, je voudrais remercier une certaine Elise pour son attention et son immense soutien. Son aide fut infiniment précieuse. J'espère pouvoir lui rendre la pareille...

Table des Matières

I	Wave Field Synthesis	11
1	De Huygens à la Wave Field Synthesis	15
1.1	Equations intégrales	16
1.2	Dérivation de la Wave Field Synthesis	17
1.2.1	Approximation 1 : réduction à une ligne de sources secondaires	18
1.2.2	Approximation 2 : passage à un segment	21
1.2.3	Approximation 3 : utilisation de bancs fermés	22
1.2.4	Approximation 4 : échantillonnage	24
1.3	Synthèse de sources virtuelles à directivité arbitraire	24
1.4	Simulations	29
1.4.1	Sources "omnidirectionnelles"	30
1.4.2	Sources de directivité de degré supérieur	35
1.4.3	Discussion : qualité de la synthèse des sources virtuelles en WFS	39
2	Rayonnement synthétisé en champ libre	41
2.1	Évaluation objective de la transparence de reproduction en champ libre	42
2.1.1	Réponse fréquentielle	42
2.1.2	Temps de groupe	43
2.2	Rayonnement d'un banc linéaire continu infini pour la WFS	44
2.2.1	Domaine fréquentiel : calcul de la loi d'atténuation	45
2.2.2	Domaine temporel : caractérisation de la variation de la forme d'onde synthétisée	48
2.2.3	Simulations : évolution spatiale du champ acoustique synthétisé	54
2.3	Rayonnement d'un banc linéaire continu de taille finie : diffraction	60
2.3.1	Domaine temporel : fenêtrage	60
2.3.2	Domaine fréquentiel : contributions des haut-parleurs à une position d'écoute	61

2.3.3	Simulations : évolution spatiale du champ acoustique synthétisé	66
2.4	Aliasing spatial	72
2.5	Discussion : qualité de la synthèse des sources virtuelles en WFS	77
3	Rayonnement en espace clos	79
3.1	Interaction du système WFS avec la salle d'écoute	81
3.1.1	Rayonnement en 3 dimensions	81
3.1.2	Réflexion des fronts d'onde synthétisés	82
3.1.3	Simulations : champ réfléchi sur les parois d'une salle rectangulaire	83
3.1.4	Discussion : impression spatiale et source virtuelle synthétisée	86
3.2	Atténuation des réflexions par la WFS	87
3.2.1	Compensation du champ réfléchi par une paroi	88
3.2.2	Adaptation à des situations réelles	100
3.2.3	Réflexion des haut-parleurs image et compensation sur les parois de la pièce	101
3.2.4	Généralisation à l'ensemble des parois verticales	103
3.2.5	Discussion : validité de la méthode proposé	107
II	Méthodes d'inversion multicanal pour la Wave Field Synthesis	109
4	Techniques d'égalisation pour la reproduction sonore	115
4.1	Méthodes d'égalisation d'un système multicanal	116
4.1.1	Égalisation individuelle	116
4.1.2	Égalisation multicanal	117
4.2	Inversion multicanal	118
4.2.1	Inversion dans le domaine fréquentiel	119
4.2.2	Inversion dans le domaine temporel	120
4.3	Égalisation multicanal pour la captation et la reproduction	122
4.3.1	Application à la captation	122
4.3.2	Application à la reproduction sonore	124
4.3.3	1 haut-parleur	125
4.3.4	Plusieurs haut-parleurs avec inversion multicanal	126
4.3.5	Plusieurs haut-parleurs, méthodes alternatives	128
4.4	Discussion : égalisation multicanal et contexte de reproduction	130
5	Contrôle du rayonnement en champ libre	133

5.1	Égalisation multicanal pour la WFS	133
5.1.1	Identification du système MIMO	134
5.1.2	Champ sonore cible	136
5.1.3	Inversion multicanal	138
5.1.4	Au-dessus de la fréquence d'aliasing	147
5.2	Résultats	149
5.2.1	Résultats par simulation de haut-parleurs idéaux	150
5.2.2	Résultats avec des haut-parleurs MAP	164
5.3	Discussion : validité de la méthode d'égalisation multicanal	178
6	Prise en compte de la pièce de restitution	181
6.1	Méthodes d'inversion multicanal pour la compensation de l'effet de salle	183
6.1.1	Identification du système MIMO	183
6.1.2	Champ sonore cible	186
6.1.3	Inversion multicanal	187
6.1.4	Au-dessus de la fréquence d'aliasing	190
6.2	Résultats	191
6.2.1	Simulations	191
6.2.2	Résultats avec des haut-parleurs MAP	202
6.3	Discussion	206
III	Mise en œuvre et validation	209
7	Validation perceptive de la perspective sonore en WFS	213
7.1	Validation objective de la localisation des sources en WFS	214
7.1.1	Localisation des sources sonores	214
7.1.2	Cohérence des indices binauraux de localisation pour la WFS	217
7.2	Exploration spatiale des scènes sonores	225
7.2.1	Utilisation de la distance holophonique pour la création d'une perspective sonore	226
7.2.2	Gestion de l'effet de distance en Wave Field Synthesis	230
7.2.3	Discussion : gestion de la distance en WFS	235
8	Création/Manipulation/Reproduction de scènes sonores	237
8.1	Description et synthèse d'une scène sonore en WFS	238

8.1.1	Description et synthèse de l'effet de salle	238
8.1.2	Reproduction sur le système WFS	249
8.2	Validité de la scène reproduite	252
8.2.1	Restitution de la perspective spatiale	253
8.2.2	Impression spatiale restituée	257
8.3	Chaîne de production	259
8.3.1	Chaîne de production sonore	259
8.3.2	Structure de la chaîne de production	259
8.3.3	Manipulation de la description de la scène sonore	261
A	Validité de l'approximation de la phase stationnaire	271
A.1	Approximation de la phase stationnaire	272
A.1.1	Formulation	272
A.1.2	Domaine de validité	273
A.2	Phase stationnaire et Wave Field Synthesis	274
A.2.1	Identification des fonctions pour la Wave Field Synthesis	274
A.2.2	Domaine de validité	275
B	Description de l'algorithme MFAP	279
B.1	Algorithme adaptatif MFAP pour l'inversion multicanal	280
B.1.1	Détail de l'algorithme MFAP simple	280
B.1.2	Dérivation de l'algorithme MFAP multicanal	283
	References	287

Introduction

L'audition est un canal sensoriel essentiel dans la perception spatiale de l'environnement. Alors que la vue doit se focaliser dans une direction particulière, l'oreille reçoit à chaque instant des informations provenant de l'ensemble de l'espace environnant. On parle à ce sujet de "sphère auditive". En plus des messages sonores proprement dits, une grande partie de ces informations concerne le ressenti de l'espace. D'où viennent les sources sonores ? Sont-elles proches ? Lointaines ? Quelle est la nature de l'espace dans lequel on se trouve ? Est-il grand ? Petit ? Autant de questions et de sensations auditives qui permettent une immersion effective et totale dans l'environnement spatial.

Partant de ce constat, les recherches sur la reproduction sonore se sont très vite intéressées à la reproduction de la sensation d'espace. Un effort constant a été fourni pour restituer l'ensemble de la sphère perceptive. Les travaux sur le "surround", (quadraphonie dans les années 1970 puis dolby surround et 5.1 depuis les années 1980) témoignent de cette volonté. Le but de ces techniques est d'augmenter la sensation d'immersion de l'auditeur dans un espace différent de celui où il se trouve ; ou bien, de lui donner à entendre des "effets" provenant de l'extérieur du champ visuel (dolby surround pour le cinéma).

Cependant, ces dispositifs ont été conçus pour des situations statiques, où l'auditeur se situe à un endroit précis et reste immobile. Les principes de reproduction mis en œuvre se basent sur une *illusion perceptive* (principes stéréophoniques), un "*trompe l'oreille*" qui ne s'apprécie qu'au centre du dispositif. Cette position d'écoute est communément appelée le "sweet spot". Pour une position d'écoute différente, l'illusion disparaît et on entend à nouveau les sources acoustiques réelles (haut-parleurs).

Au cours des années 70, l'installation et la performance acquièrent le statut d'expression artistique à part entière. Ces courants artistiques font voler en éclat le schéma statique. Le visiteur bouge. Il circule, déambule dans le lieu, se l'approprie aussi par ses mouvements.

Dans un contexte plus technologique, les travaux sur la réalité virtuelle et sur la réalité augmentée au cours des années 80, à leur tour, mettent en avant la nécessité de placer l'individu dans un contexte d'intégration. Ces applications reposent sur la stimulation de plusieurs canaux sensoriels (vision, ouïe, toucher, ...) et sur une interaction de l'individu avec l'environnement. Il/elle peut agir soit directement sur les objets qui composent la scène, soit par ses propres mouvements à l'intérieur de l'espace réel ou virtuel proposé.

En parallèle à cette évolution des pratiques artistiques et technologiques, se sont développés des systèmes ayant l'ambition de reconstruire les propriétés physiques du champ sonore : le système Ambisonique et la Wave Field Synthesis.

Le système Ambisonique proposé par Gerzon au cours des années 70 et 80 (Gerzon, 1985) est le premier dans le genre. L'objectif initial de ce système s'en écartait toutefois. Il visait plutôt la synthèse de sources sonores en dehors du plan horizontal. Ce système permet en effet une capture et une restitution du champ sonore en 3 dimensions mais dans une zone de l'espace qui reste très limitée par sa taille (du même ordre de que celle du sweet spot stéréophonique). Il

faudra attendre les travaux récents sur Ambisonique généralisé pour étendre réellement la zone d'écoute (Daniel, 2000) (Daniel, 2003).

Pendant les années 80 le laboratoire de sismique et d'acoustique de l'université de Delft aux Pays-Bas s'intéresse de son côté à la synthèse de champs sonores, "Wave Field Synthesis" en anglais. L'objectif de la Wave Field Synthesis (WFS) est de synthétiser les caractéristiques physiques du champ sonore issu d'une source omnidirectionnelle située dans le plan horizontal par un banc linéaire de haut-parleurs omnidirectionnels. Ces recherches aboutissent à la fin des années 80 à la publication par Berkhout de l'article fondateur de la Wave Field Synthesis (Berkhout, 1988). Le terme de Wave Field Synthesis n'apparaît cependant qu'en 1993 dans un article écrit par Berkhout, de Vries et Vogel (Berkhout *et al.*, 1993) et dans la thèse de Vogel (Vogel, 1993). C'est à ce moment là que la plupart des principes fondamentaux de la technique sont mis en place. Ses limites aussi : cette synthèse est a priori assurée uniquement dans le plan horizontal en aval du banc de haut-parleurs et seulement jusqu'à une fréquence limite.

La Wave Field Synthesis s'inspire du principe de Huygens. Ce principe, énoncé dès 1678, stipule qu'un front d'onde est équivalent à un ensemble de sources secondaires. Une onde qui se propage à partir d'un certain front, dit primaire, peut être considérée, de manière équivalente, comme émise par la source d'origine (source primaire) ou issue d'une distribution continue de sources secondaires réparties le long du front d'onde primaire. Autrement dit, au delà du front d'onde de référence, le champ d'onde généré par les sources secondaires se substitue de manière parfaitement équivalente à celui émis par la source primaire. Les sources secondaires prennent, en quelque sorte, le *relais* de la source primaire.

La mise en équation de ce principe est réalisée par Kirchhoff, Helmholtz et Rayleigh au cours du XIXème siècle. Ce pas est franchi grâce à l'écriture de la synthèse du champ sous forme de problème aux limites de l'acoustique linéaire. Dans la formulation du problème, ils effectuent une séparation de l'espace à trois dimensions par une surface fermée entourant la source primaire (version de Kirchhoff-Helmholtz) ou par un plan infini (version de Rayleigh) entre un sous-espace, dit "des sources", où sont situées les sources primaires et un sous-espace, dit "de reproduction", dans lequel la synthèse du champ est assurée. Dans ces intégrales, la distribution des sources secondaires à utiliser est continue et se situe à l'interface entre les deux sous-espaces complémentaires.

La mise en pratique de la Wave Field Synthesis s'appuie sur un ensemble d'approximations de ces équations intégrales.

Les objectifs des développements postérieurs aux articles fondateurs s'attachent à garantir une qualité de reproduction sonore indépendante de la position de l'auditeur au sein du dispositif. Il s'agit de s'affranchir, ou du moins d'étendre, le "sweet spot".

Ces travaux, réalisés à l'université de Delft et à France Télécom R & D, ont consisté à :

- préciser les conditions de validité des approximations utilisées (Verheijen, 1997) (Start, 1997) (Nicol, 1999) ;
- les valider d'un point de vue perceptif (Verheijen, 1997) (Start, 1997) ;

- assurer la synthèse d'un effet de salle (Boone & van Tol, 1995) (Sonke & de Vries, 1997) (Boone & Verheijen, 1998) (Sonke, 2000) ;
- développer le concept dual de Wave Field Analysis pour la caractérisation de champ sonore en zone étendue et la mesure des caractéristiques spatiales de la réponse impulsionnelle d'une salle (Berkhout *et al.*, 1997) (Sonke, 2000) (Hulsebos *et al.*, 2001) (Hulsebos & de Vries, 2002) ;
- déterminer la possibilité d'utilisation de technologies de transduction électroacoustiques alternatives aux haut-parleurs électro-dynamiques : électro-statiques (Verheijen, 1997), DML (Boone & de Bruijn, 2000) ;
- s'adapter aux caractéristiques de directivité des haut-parleurs utilisés (de Vries, 1996) ;
- étudier les rapprochements de la WFS avec Ambisonique généralisé (Nicol, 1999) ;
- valider l'utilisation de la WFS dans un contexte de visio-conférence (Nicol, 1999) (de Bruijn & Boone, 2003) ;

Toutes ces études ont mené à la mise en place, dans la deuxième moitié des années 90, de prototypes dans les deux instituts. Les limites technologiques étaient de taille. En plus du nombre important de haut-parleurs et d'amplificateurs utilisés, la puissance de calcul nécessaire s'avérait très importante par rapport aux moyens informatiques disponibles à l'époque. Aujourd'hui, ces contraintes techniques sont relâchées.

Ces travaux ont motivé l'idée d'un projet de recherche européen. Le projet Carrouso (Creating, Assessing and Rendering in Real time of high quality audio-visual environments in MPEG-4 context), financé par la communauté européenne (IST 1999-20993) s'est déroulé entre janvier 2001 et juin 2003 (Brix *et al.*, 2001). Il a réuni 10 partenaires universitaires et industriels européens, dont l'université de Delft, France Télécom R&D, l'IRCAM, Studer et le Fraunhofer Institute IIS AEMT, coordinateur du projet.

Son but était de mettre en œuvre et/ou de développer des technologies pour la captation, la description, la transmission et la synthèse de scènes sonores spatialisées, réelles ou virtuelles, en laissant la possibilité à l'utilisateur d'interagir avec le contenu. Il s'est basé sur le standard MPEG-4 (MPEG, 2000), pour la description de la scène, le format de stockage et de transmission, et la Wave Field Synthesis pour la reproduction.

Il a abouti à des réalisations qui sont maintenant commercialisées par le Fraunhofer Institute IIS AEMT * et sonic emotion ag †.

Cette thèse s'est déroulée à l'IRCAM pendant le projet Carrouso. Il s'agissait des premiers travaux réalisés à l'IRCAM sur la WFS. Très tôt, il est apparu important d'explorer les nouvelles possibilités offertes, en terme de création musicale, de cette restitution de scènes sonores spatialisées dans une *zone étendue*. Au delà de l'extension de la zone d'écoute, la question était de savoir quel type d'"effets" accessibles avec cette technique ne l'étaient pas auparavant avec les

*<http://www.iosono-sound.com>

†<http://www.sonicemotion.com>

techniques de reproduction traditionnelles.

Pour préciser cette idée, définissons les trois types de sources virtuelles élémentaires auxquels on peut être confronté.

1. Les sources ponctuelles omnidirectionnelles situées *derrière* le banc de haut-parleurs. Les fronts d'onde synthétisés dans la zone d'écoute sont cohérents avec la position de la source cible. Un auditeur qui se déplace à l'intérieur de la zone d'écoute perçoit des variations des indices acoustiques cohérents avec la *position* de la source cible.
2. Les ondes planes. Elles sont un cas limite des sources mentionnées précédemment. Elles correspondent à des sources situées à une distance "infinie". De telles sources ne sont pas audibles dans un contexte naturel car elles sont aussi infiniment atténuées. Dans le domaine visuel, le soleil donne un exemple de source située à l'"infini". Lors d'un déplacement rectiligne en train ou en voiture, nous avons l'impression que le soleil nous "suit" alors que le paysage semble "défiler". De la même manière, un auditeur se déplaçant à l'intérieur de l'installation sonore aura l'impression que la source le suit et conserve une *direction* de provenance identique quelque soit sa position.
3. Les sources virtuelles ponctuelles situées à l'intérieur de la pièce de reproduction (sources dites "focalisées"). Une extension du principe de la WFS permet de "recréer" des sources sonores à l'intérieur de la pièce, à un endroit où il n'y a pas de source physique. Ce véritable "hologramme sonore" est synthétisé en émettant des ondes acoustiques qui convergent à la position cible de la source virtuelle. Ces ondes acoustiques se propagent alors naturellement dans le reste de l'espace à partir du point de convergence. Le champ sonore synthétisé est incorrect entre la position des haut-parleurs et la position de la source virtuelle mais devient valide au-delà.

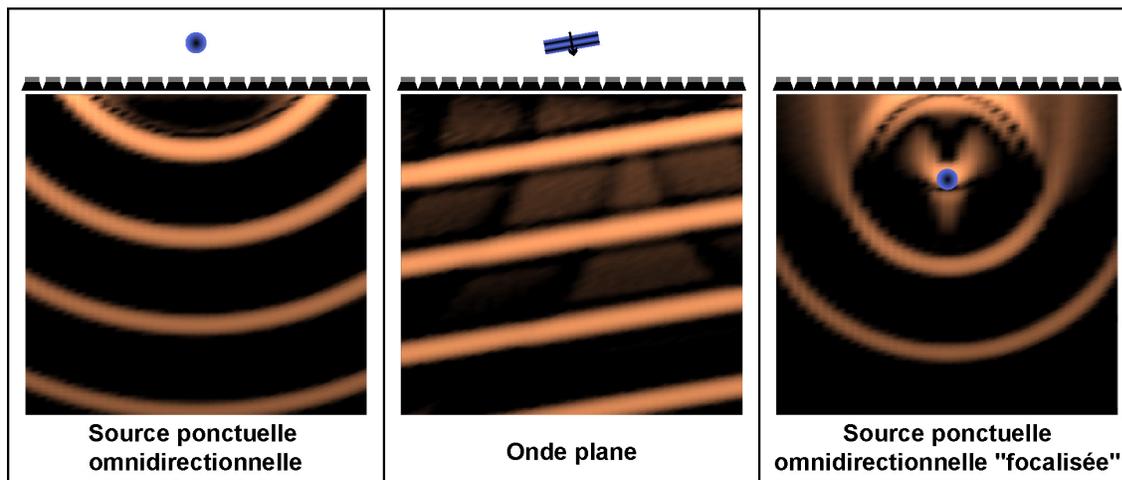


FIG. 1 – Sources virtuelles élémentaires reproduites par la WFS

Ces sources constituent en quelque sorte le *vocabulaire élémentaire* de la WFS.

En plus de la possibilité de s'adresser à un nombre important d'auditeurs, la WFS autoriserait a priori la création et la manipulation d'une véritable *perspective sonore* qui s'apprécierait depuis

différents *points de vue*.

A travers ses mouvements, à l'intérieur de l'installation sonore, l'auditeur pourrait *explorer* la scène, se rapprocher de certaines sources, s'en éloigner d'autres.

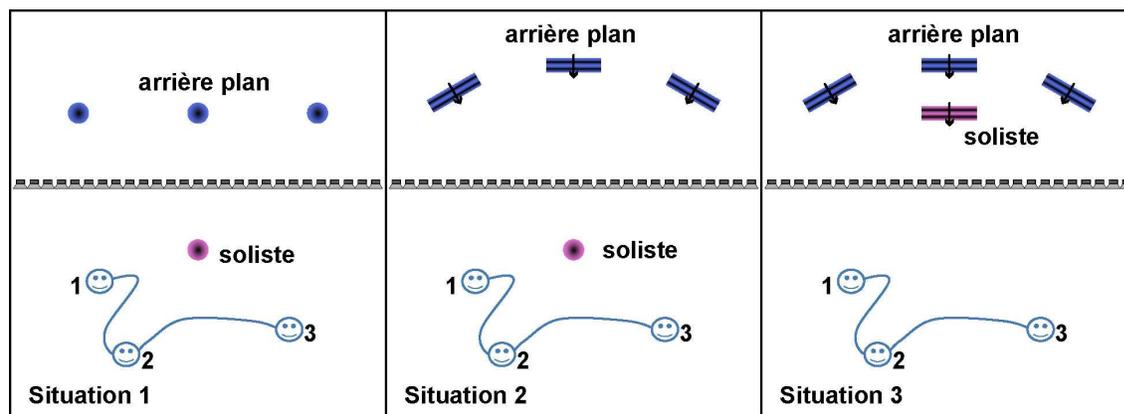


FIG. 2 – Création de perspective sonore pour la WFS

Ces types élémentaires de sources peuvent aussi être combinés pour reproduire des scènes plus complexes. La figure 2 présente une scène composée d'un arrière plan et d'un soliste.

Dans la situation 1, l'arrière plan est synthétisé par trois sources ponctuelles situées *derrière* les haut-parleurs, le soliste par une source ponctuelle située *devant* les haut-parleurs. L'auditeur peut choisir l'attitude qu'il souhaite adopter vis-à-vis de cette scène, soit en se plaçant loin afin de percevoir l'ensemble de la scène, ou bien en se rapprochant physiquement de l'une des sources. Ceci est en particulier le cas pour le soliste avec lequel l'auditeur peut rentrer en contact direct par sa restitution en tant que source focalisée.

Dans la situation 2, l'arrière plan est synthétisé sous forme d'ondes planes. Dans ce cas, l'organisation spatiale de l'arrière plan est *figée*, la direction de provenance des sources *ne dépendant pas* de la position de l'auditeur. Lors d'un déplacement de l'auditeur le long du banc de haut-parleurs, cet arrière plan suit ses mouvements alors que le soliste reste à une position fixe de laquelle il va s'approcher puis s'éloigner.

Dans la situation 3, l'ensemble des sources virtuelles est reproduit en ondes planes. Ceci revient à figer l'organisation spatiale de la scène sonore. Celle-ci reste identique *quelque soit* la position de l'auditeur dans l'installation.

Remarquons que cette notion de perspective sonore s'apprécie principalement lorsque les sources et/ou les auditeurs se déplacent. En effet, les trois situations présentées sont perçues de manière parfaitement identique, en terme de localisation des sources sonores, par un auditeur situé au centre du dispositif.

La WFS semble reproduire les effets de parallaxe naturels liés à la position des sources sonores. En découle un certain "réalisme du rendu". Ceci augmente alors la sensation de "présence" * des sources virtuelles et la sensation d'immersion dans la scène sonore virtuelle (Warusfel & Viaud-

*dans l'acception couramment utilisée en réalité virtuelle, on peut aussi parler de réalisme de rendu

Delmon, 2003).

Plus généralement, cette sensation de présence repose sur la cohérence entre plusieurs dimensions sensorielles (vue/ouïe) et également, sur la congruence des sensations perceptives au cours d'une interaction avec l'objet présenté (préhension, observation dynamique). Ainsi, le fait de se déplacer, se rapprocher, s'éloigner latéralement ou en profondeur dans la scène tout en éprouvant de manière auditive une variation des indices acoustiques cohérente avec ses mouvements contribue à la présence de cette source. A l'inverse, dans un système conventionnel stéréophonique, dès que l'on se déplace, l'illusion sonore disparaît. On ré-entend les haut-parleurs. Avec la WFS pourrait-on espérer une *transparence* de l'outil de diffusion ?

La transparence est une question essentielle pour la reproduction sonore. Cette notion se définit d'un point de vue perceptif par l'absence de modification apportée par la transduction électroacoustique sur la scène sonore spécifiée. Il s'agit en quelque sorte de faire "disparaître" le système de reproduction et de laisser place à l'expérience perceptive souhaitée par le créateur de contenu.

Toutes ces questions ne peuvent se poser que si le champ sonore est effectivement correctement synthétisé dans *l'ensemble de la zone d'écoute*. Or, nous rappelons que la WFS se base sur un ensemble d'approximations sur les principes qui décrivent la reproduction *exacte* des caractéristiques d'un champ sonore par une distribution de haut-parleurs idéaux. Ces approximations n'assurent la restitution que de certaines caractéristiques du champ sonore (forme des fronts d'ondes mais niveau synthétisé incorrect en dehors d'une ligne de référence, champ synthétisé incorrect au dessus d'une fréquence limite) *en champ libre*, et avec des transducteurs idéaux ou bien dont les caractéristiques de directivité sont identiques (de Vries, 1996).

On peut se demander comment mettre en œuvre et assurer, dans les conditions *réelles* (transducteurs hétérogènes placés en espace clos), la cohérence des impressions spatiales auditives dans *l'ensemble de la zone d'écoute*.

Plus précisément, nos questions fondamentales sont les suivantes :

1. Quelles sont les caractéristiques du champ sonore effectivement synthétisé dans *l'ensemble de la zone de d'écoute* d'un point de vue objectif ? Dans quelle mesure celles-ci vont-elles réduire la transparence de la reproduction d'un point de vue perceptif ?
2. Peut-on définir des méthodes qui permettent une prise en compte et une compensation des conditions réelles d'écoute (banc de haut-parleurs, interaction avec la salle de restitution) ? Cette prise en compte peut-elle être efficace dans une zone étendue de l'espace ? Ces méthodes peuvent-elles être compatibles avec une implémentation en temps réel des solutions ?
3. Comment garantir que les impressions sonores spécifiées par le créateur de contenu sont bien conservées sur différents systèmes de reproduction ? Il n'existe pas en effet de système WFS standard. Chaque installation est spécifique et se réfère à une zone d'écoute de taille différente.

4. Quels outils d'édition et de manipulation de la scène sonore doit-on fournir au créateur de contenu ? La création de contenu dans ce contexte ne peut plus être égocentré comme c'est le cas pour les systèmes stéréophoniques pour lesquels il existe une position de référence. Le créateur de contenu doit ainsi pouvoir se déplacer dans la scène sonore tout en manipulant les différents éléments qui la composent.
5. Quel est le format de stockage et de distribution adapté à cette technique ?

Les descriptions des caractéristiques du champ synthétisé fournies dans la littérature se placent uniquement à une distance moyenne du banc de haut-parleurs. Elles observent ou prédisent le champ synthétisé mais sans en préciser la qualité d'un point de vue perceptif. Ces études ne concernent d'autre part qu'un banc de haut-parleurs placé en champ libre mais rarement en espace clos.

La compensation des conditions *réelles* de reproduction est abordée dans (de Vries, 1996) dans le cas restrictif où l'ensemble des haut-parleurs sont identiques. La compensation de l'acoustique de la pièce de reproduction est un thème qui n'a été étudié qu'au cours du projet Carrouso (Corteel & Nicol, 2003) (Spors *et al.*, 2003).

La question de la persistance des impressions spatiales sur différents types de reproduction n'a été abordée que récemment dans le cadre d'applications cinématographiques ou de téléconférence (Melchior *et al.*, 2003) (de Bruijn & Boone, 2003) pour garantir la cohérence des impressions auditives avec une projection d'images 2D.

Des outils d'édition et de manipulation de la scène sonore ont été proposés dans le cadre du projet Carrouso, notamment à l'IRCAM par Riitta Väänänen (Väänänen *et al.*, 2002) autour du format MPEG-4 qui permet le stockage et la distribution de présentations audio-visuelles interactives, a priori, indépendamment de la technique de reproduction utilisée.

Ce document est organisé en trois parties relativement distinctes et nous invitons le lecteur à les lire dans l'ordre qu'il souhaite.

La première partie de ce document (chapitres 1, 2 et 3) est consacrée à l'analyse des caractéristiques du champ acoustique synthétisé dans un cadre idéal (haut-parleurs omnidirectionnels idéaux placés en champ libre puis en espace clos). Nous examinons de façon systématique les caractéristiques du champ sonore synthétisé dans l'ensemble de la zone d'écoute (chapitre 2). Nous mettons en évidence une modification de la qualité des fronts d'onde synthétisés au cours de la propagation. Dans ce cadre théorique, nous proposons deux extensions au cadre initial de la Wave Field Synthesis. La première permet de synthétiser des sources virtuelles dont le rayonnement est de nature plus complexe (chapitre 1), la seconde de compenser de l'acoustique de la salle de restitution (chapitre 3).

La deuxième partie de ce document (chapitres 4, 5 et 6) est consacrée à la prise en compte et à la compensation des conditions réelles d'écoute (directivité des transducteurs, caractéristiques acoustiques des parois de la pièce de restitution) dans l'ensemble de la zone d'écoute. Les méthodes proposées se basent sur des techniques d'inversion multicanal qui permettent de minimiser

l'erreur de reproduction au niveau d'un nombre réduit de points de contrôle. Tout l'enjeu consiste alors à déterminer dans quelles conditions ces techniques assurent la validité du champ synthétisé dans l'ensemble de la zone d'écoute. Après avoir décrit les techniques classiquement utilisées dans le domaine de la captation et de la reproduction sonore (chapitre 4), nous définissons une méthode qui permet de contrôler le champ acoustique synthétisé en champ libre par un banc de haut-parleurs à directivité arbitraire (chapitre 5) et ainsi de compenser une partie des "défauts" de la Wave Field Synthesis et des transducteurs utilisés. Nous proposons ensuite une évolution de cette méthode pour la compensation des réflexions sur les parois verticales de la salle d'écoute (chapitre 6). Les procédures mises en œuvre produisent une base de données de filtres référencée en fonction de la source virtuelle synthétisée et du haut-parleur considéré.

La troisième et dernière partie de ce document est consacrée à la mise en œuvre (chapitre 8) et à la validation de l'utilisation de la perspective sonore en WFS (chapitre 7). Dans un premier temps, nous proposons une procédure qui permet d'objectiver les défauts de localisation des sources virtuelles synthétisées par simulation des indices interauraux perçus par l'auditeur (partie 7.1). Nous présentons ensuite un test d'écoute interactif réalisé auprès d'ingénieurs du son dont le but est de déterminer l'influence respective des paramètres disponibles pour la manipulation de la perspective sonore (partie 7.2). Le chapitre 8 est consacré à la définition d'une chaîne de production et de restitution sonore qui permet de consigner les impressions sonores spécifiées par le créateur de contenu et de garantir que celles-ci sont correctement reproduites quelque soit le système de reproduction considéré. A défaut de répondre totalement à cette question, nous effectuons une analyse critique de la synthèse de scène sonore pour la WFS (son direct + effet de salle) et nous présentons la chaîne de production mise en place à l'IRCAM à la suite du projet Carrouso et en collaboration avec sonic emotion.

Première partie

Wave Field Synthesis

Dans la première partie de la thèse, nous nous plaçons dans un cadre théorique que l'on pourrait qualifier d'"idéal". Il s'agit pour nous de donner une analyse critique des approximations proposées par la WFS, d'en évaluer systématiquement les conséquences sur le champ sonore synthétisé dans la zone d'écoute et, de proposer des extensions pour compléter ce cadre.

Pour cela, nous considérerons un banc linéaire composé de haut-parleurs omnidirectionnels placé dans deux situations :

- en champ libre, afin de donner les caractéristiques du son direct synthétisé par le banc dans une zone étendue de l'espace. Nous verrons alors que le champ synthétisé dans la zone d'écoute présente de nombreuses différences avec le champ associé à la source virtuelle spécifié (champ cible) et nous tâcherons d'en évaluer les conséquences sur la transparence de la reproduction ;
- en espace clos, ce qui nous permettra de décrire l'interaction du banc avec une salle d'écoute "idéale" dont les parois sont parfaitement réfléchissantes. Nous proposerons alors une méthode de compensation des réflexions précoces qui peut se présenter comme une extension à la WFS.

Le chapitre 1 est dédié à la mise en place du cadre théorique. Ce chapitre s'appuie sur une description détaillée des approximations proposées par la Wave Field Synthesis afin de la rendre réalisable en pratique. Ces approximations consistent à effectuer des simplifications sur la géométrie de sources secondaires et à corriger leurs fonctions d'alimentation afin de réduire les erreurs introduites. Nous donnerons une extension à cette formulation pour la synthèse de sources virtuelles à directivité arbitraire. Nous soulignerons néanmoins les limitations de cette méthode inhérentes à la nature des sources secondaires employée et à leur disposition. Enfin, nous présenterons un ensemble de simulations qui mettra en évidence les "défauts" potentiels de la WFS.

Le chapitre 2 effectue un renversement de point de vue. Les approximations et la description de leurs conséquences proposées dans la littérature se rapporte à "ce qui manque" par rapport à la géométrie idéale (distribution planaire continue de sources secondaires). Dans ce chapitre notre analyse se penchera davantage sur "ce qui est". Nous donnerons des méthodes de prédiction qui permettent d'évaluer le rayonnement d'un banc linéaire de sources secondaires auxquelles on applique les signaux d'alimentation donnés par la WFS. Ces méthodes visent à décrire la construction des fronts d'onde par l'ensemble des haut-parleurs en un point de réception donné. Ceci nous permettra d'évaluer, à chaque étape de la simplification, la transparence de reproduction. Nous mettrons ainsi en évidence que la qualité de la restitution *dépend* de la position dans la zone d'écoute.

Enfin, dans le chapitre 3, le banc sera placé dans une salle d'écoute "idéale" dont les parois sont parfaitement réfléchissantes. Dans ce cas simple nous proposerons des méthodes de description de l'interaction du banc avec la salle d'écoute. Nous verrons que cette interaction dépend largement de la source virtuelle considérée. Ceci nous amènera alors à définir une méthode qui utilise la WFS pour compenser partiellement les réflexions de premier ordre. Nous analyserons le comportement

de cette méthode et nous en donnerons les limites. Cette méthode sera enfin présentée en temps qu'extension de la WFS dans le cas où l'ensemble des haut-parleurs de l'installation sont intégrés dans les parois de la salle d'écoute.

Chapitre 1

De Huygens à la Wave Field Synthesis

Le principe de Huygens et ces formulations mathématiques (intégrales de Kirchhoff-Helmholtz et de Rayleigh) permettent la synthèse des caractéristiques physiques d'un champ d'ondes. Ces formulations sont très générales et ne dépendent pas du type d'onde considéré (électromagnétiques, optiques, mécaniques, sonores, etc.).

L'application de ce principe à la reproduction sonore a été proposée pour la première par Jessel en 1973 sous le nom d'holophonie (Jessel, 1973). Il décrit, dans son livre "Acoustique théorique : propagation et holophonie" un système de captation/reproduction sonore qui consiste à enregistrer le champ acoustique sur une surface fermée pour le reproduire à l'intérieur d'un volume à l'aide de haut-parleurs situés à la même position que les microphones utilisés pour la captation. Il s'agit en quelque sorte d'un équivalent de l'holographie dans le domaine acoustique.

Le système de reproduction se calque complètement sur la géométrie du système de captation ce qui rend cette technique inutilisable en pratique. Elle requiert d'autre part un nombre incroyablement élevé de haut-parleurs et de microphones.

Les travaux sur la Wave Field Synthesis partent de ce constat et proposent un ensemble de simplifications visant à séparer le système de captation du système de reproduction et à réduire le nombre de microphones et de haut-parleurs nécessaires. L'article fondateur de cette technique a été publié en 1988 par Berkhout (Berkhout, 1988). Il y décrit un système holophonique dans lequel il ajoute une étape supplémentaire d'extrapolation du champ sonore capté qui permet de dissocier le dispositif de captation du dispositif de restitution. Tous deux sont constitués de bancs linéaires de taille finie.

Le nom de Wave Field Synthesis apparaît pour la première fois en 1993 dans un article de Berkhout, de Vries, et Vogel (Berkhout *et al.*, 1993) et, dans la thèse de Vogel (Vogel, 1993). Apparaissent aussi les deux concepts fondamentaux qui définissent à ce jour la Wave Field Synthesis :

1. le concept de *source notionnelle* qui permet d'effectuer une séparation complète entre le dispositif de captation et le dispositif de restitution. Pour décrire la scène sonore, on la

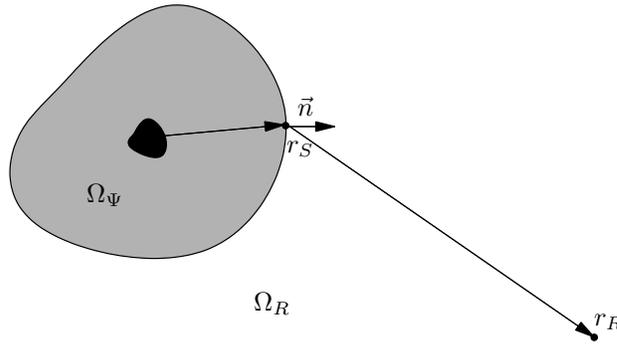


FIG. 1.1 – Géométrie pour l'intégrale de Kirchhoff-Helmholtz

décompose en un ensemble de sources dites notionelles. Chaque source notionelle est modélisée comme une source à directivité omnidirectionnelle. Ceci permet de simuler son rayonnement et d'en déduire les signaux d'alimentation des sources secondaires. Dans la suite on préférera la notion synonyme de source virtuelle, plus couramment utilisée par ailleurs ;

2. un ensemble de simplifications de la géométrie de sources secondaires avec, à chaque étape, une quantification de l'erreur introduite et une correction apportée. Ces approximations permettent de passer d'une distribution théorique plane, continue de sources secondaires à un nombre fini réparties sur une ligne.

Les travaux postérieurs de l'université de Delft puis de France Télécom R&D au cours des années 90 ont permis d'améliorer la compréhension des phénomènes mis en jeu d'un point de vue objectif et subjectif afin de valider l'approche (Verheijen, 1997) (Start, 1997) (Nicol, 1999) (Sonke, 2000).

Dans ce chapitre, nous rappellerons dans un premier temps les équations intégrales de Kirchhoff-Helmholtz et Rayleigh qui fournissent un socle mathématique exact à partir duquel la formulation de la Wave Field Synthesis est dérivée par un ensemble d'approximations. Nous proposerons une extension de ce principe à la synthèse de sources virtuelles à directivité arbitraire en utilisant le même type d'approximations. Nous soulignerons alors les limitations induites par la géométrie linéaire de sources secondaires. Enfin, nous recourrons à des simulations. Celles-ci nous permettront de mettre en évidence les caractéristiques typiques du champ sonore synthétisé.

1.1 Equations intégrales

Pour obtenir l'intégrale de Kirchhoff-Helmholtz, on considère une surface fermée quelconque $\partial\Omega$ séparant l'espace à trois dimensions Ω en deux sous-espaces complémentaires Ω_Ψ (contenant l'ensemble des sources primaires) et Ω_R (sous-espace de reproduction). A partir de la formulation du problème aux limites sur $\partial\Omega$, se déduit l'intégrale de Kirchhoff-Helmholtz :

$$p_R(\vec{r}_R) = \int_{\partial\Omega} p_\Psi(\vec{r}_S) \vec{\nabla}_S G(\vec{r}_R, \vec{r}_S) - G(\vec{r}_R, \vec{r}_S) \vec{\nabla}_S p_\Psi(\vec{r}_S) \vec{n} dS, \quad (1.1)$$

où $p_{\Psi}(r_{\vec{S}})$ est le champ de pression rayonné par l'ensemble des sources primaires sur $\partial\Omega$; \vec{n} , le vecteur normal à $\partial\Omega$ au point r_S ; r , un point de l'espace et G , la fonction de Green de l'espace libre en 3 dimensions :

$$G(r_{\vec{R}}, r_{\vec{S}}) = \frac{e^{-jk(\|r_{\vec{R}} - r_{\vec{S}}\|)}}{4\pi \|r_{\vec{R}} - r_{\vec{S}}\|}. \quad (1.2)$$

L'intégrale de Kirchhoff-Helmholtz vaut alors :

$$p_R(r_{\vec{R}}) = \begin{cases} 0 & \text{si } r \in \Omega_{\Psi} \\ p_{\Psi}(r_{\vec{R}}) & \text{si } r \in \Omega_R \end{cases}. \quad (1.3)$$

Le champ sonore synthétisé par l'ensemble des sources secondaires est donc *nul* dans Ω_{Ψ} et correspond exactement au champ sonore produit par l'ensemble des sources primaires dans Ω_R . L'intégrale de Kirchhoff-Helmholtz assure ainsi une "annulation de l'onde arrière". Les fronts d'onde synthétisés en *aval* de la distribution de sources secondaires correspondent aux fronts d'onde associés aux sources primaires. Le champ synthétisé en *amont* est lui, par contre, nul. De plus, cette formulation ne dépend pas du nombre de sources primaires ni de leur position dans Ω_{Ψ} .

Dans le cadre de cette formulation, la distribution de sources secondaires est la résultante de la superposition d'une distribution continue de sources monopolaires alimentées par le gradient de pression du champ rayonné par les sources primaires sur $\partial\Omega$ et d'une distribution continue de sources dipolaires alimentées par le signal de pression.

Dans le cas où $\partial\Omega$ correspond à un plan infini (cas de la géométrie de Rayleigh), les contributions des sources monopolaires et dipolaires sont identiques dans Ω_R et, en opposition de phase dans Ω_{Ψ} . Si on ne considère que le sous-espace de reproduction, on peut ne conserver qu'un seul type de source.

Dans le cas où l'on garde les sources monopolaires (équation de Rayleigh 1), le champ émis par les sources secondaires n'est plus nul dans Ω_{Ψ} et devient symétrique à celui émis dans Ω_R par rapport à $\partial\Omega$. Lorsque l'on garde les sources dipolaires (équation de Rayleigh 2), celui-ci est symétrique mais en opposition de phase. Les équation ainsi obtenues s'expriment sous la forme suivante :

$$p_R(r_{\vec{R}}) = -2 \int_{\partial\Omega} G(r_{\vec{R}}, r_{\vec{S}}) \vec{\nabla} p_{\Psi}(r_{\vec{S}}) \cdot \vec{n} dS \quad \text{Rayleigh 1;} \quad (1.4)$$

$$p_R(r_{\vec{R}}) = 2 \int_{\partial\Omega} (p_{\Psi}(r_{\vec{S}}) \vec{\nabla} G(r_{\vec{R}}, r_{\vec{S}})) \cdot \vec{n} dS \quad \text{Rayleigh 2.} \quad (1.5)$$

1.2 Dérivation de la Wave Field Synthesis

Dans sa formulation initiale, la Wave Field Synthesis vise la reproduction d'une source ponctuelle omnidirectionnelle Ψ située en r_{Ψ} . Elle s'appuie sur une formulation de type Rayleigh 1. Afin de

simplifier les notations, posons :

$$\begin{aligned} r &= |r_S - r_\Psi|, \\ \Delta r &= |r_R - r_S|. \end{aligned} \quad (1.6)$$

r correspond ainsi à la distance de propagation dans Ω_Ψ , et Δr , à la distance de propagation dans Ω_R . La composante normale à la surface du gradient du champ de pression rayonné par ce type de source peut être simulée au niveau de la surface $\partial\Omega$ sous la forme,

$$\vec{\nabla} p_\Psi(r_S) \cdot \vec{n} = S(\omega) \frac{\vec{r}}{r} \cdot \vec{n} \frac{1 + jkr}{r} \frac{e^{-jkr}}{4\pi r}, \quad (1.7)$$

où $S(\omega)$ est le signal d'alimentation de la source Ψ . Si on place la source en champ lointain ($kr \gg 1$, hypothèse classique), cette expression devient :

$$\vec{\nabla} p_\Psi(r_S) \cdot \vec{n} = S(\omega) \cos \theta_\Psi jk \frac{e^{-jkr}}{4\pi r}, \quad (1.8)$$

où,

$$\cos \theta_\Psi = \frac{\vec{r}}{r} \cdot \vec{n}. \quad (1.9)$$

La Wave Field Synthesis s'appuie sur un ensemble d'approximations qui simplifient la géométrie des sources secondaires pour aboutir à un nombre fini de sources secondaires.

La première approximation se base sur le fait que les sources et les auditeurs sont situées dans le même plan horizontal. Dans ce cas, on peut réduire le plan infini à une ligne. La seconde approximation vise à réduire cette ligne à un segment. Quant à la troisième, elle consiste à utiliser plusieurs segments pour former un banc fermé. La quatrième et dernière approximation réalise un échantillonnage spatial de la distribution des sources secondaires.

Ceci s'effectue au prix de compromis sur la qualité du champ sonore effectivement synthétisé. Ces compromis sont soulignés dans cette partie puis analysés de façon systématique dans le chapitre 2.

1.2.1 Approximation 1 : réduction à une ligne de sources secondaires

Dans la plupart des situations (ex : salle de concert, musées, écoute domestique), la plupart des sources sonores * et les auditeurs sont situés dans un même plan horizontal. On peut alors montrer que les sources secondaires situées dans le plan horizontal produisent la contribution la plus importante du champ sonore perçu par les auditeurs. La géométrie de répartition des sources secondaires peut alors être simplifiée à une ligne infinie, intersection du plan vertical défini pour l'intégrale de Rayleigh 1 et du plan horizontal dans lequel sont situés les sources et les auditeurs.

De par l'utilisation d'une distribution linéaire de sources omnidirectionnelles, le champ sonore synthétisé dans l'espace à 3 dimensions présente une *symétrie cylindrique* dont l'axe principal

*en ne considérant que leur contribution directe et non pas les réflexions qu'elles engendrent

correspond à la ligne de sources secondaires.

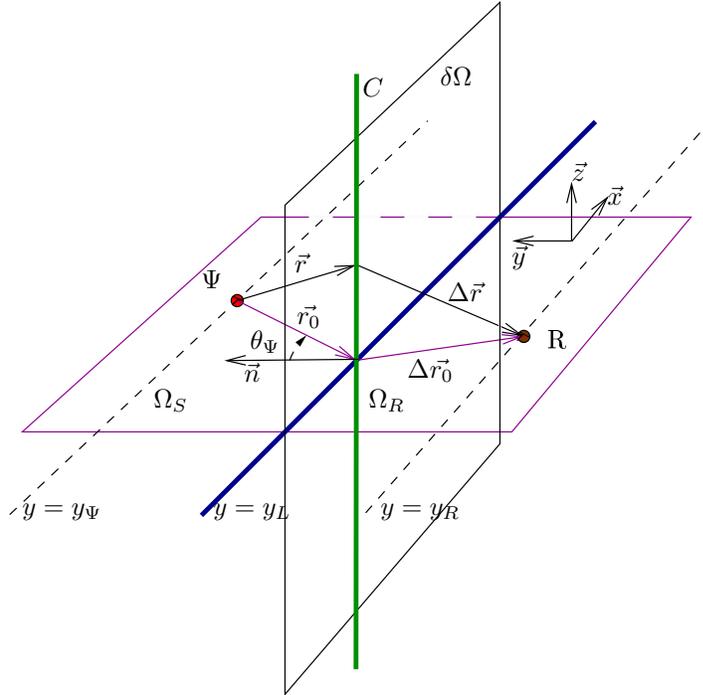


FIG. 1.2 – Réduction de la géométrie de sources secondaires à une ligne infinie par l'approximation de la phase stationnaire

D'un point de vue mathématique, la réduction de la géométrie planaire à la ligne L s'effectue en exprimant le champ sonore synthétisé par chacune des colonnes verticales $C(x_L)$ (cf. figure 1.2) :

$$P_{C(x_L)}(r_R, k) = jkS(\omega) \int_{-\infty}^{+\infty} \cos \theta_\Psi \frac{e^{-jk(r+\Delta r)}}{16\pi^2 r \Delta r} dz. \quad (1.10)$$

On peut utiliser l'approximation de la phase stationnaire (cf. Annexe A.1) afin de simplifier l'expression 1.10,

$$P_{C(x_L)}(r_R, k) = \sqrt{2\pi} \sqrt{jk} S(\omega) \cos \theta_\Psi \sqrt{\frac{r_0 \Delta r_0}{r_0 + \Delta r_0}} \frac{e^{-jk(r_0 + \Delta r_0)}}{16\pi^2 r_0 \Delta r_0}. \quad (1.11)$$

Ceci permet d'évaluer la contribution de la ligne verticale. On approche l'intégrale de Rayleigh 1 par

$$P(r_R, k) = \sqrt{\frac{jk}{2\pi}} \int_{-\infty}^{+\infty} g_\Psi(y_R) S(\omega) \cos \theta_0 \left(\frac{e^{-jkr_0}}{\sqrt{r_0}} \right) \left(\frac{e^{-jk\Delta r_0}}{4\pi \Delta r_0} \right) dx_L, \quad (1.12)$$

où $g_\Psi(y_R)$ est un facteur de compensation de niveau, dû au passage d'une géométrie planaire à une géométrie linéaire,

$$g_\Psi(y_R) = \sqrt{\frac{\Delta r_0}{r_0 + \Delta r_0}} = \sqrt{\frac{|y_R - y_L|}{|y_R - y_\Psi|}}. \quad (1.13)$$

Ce facteur dépend de la profondeur d'écoute y_R . Ceci est équivalent à une modification de la

loi d'atténuation des sources secondaires. Afin de conserver des sources secondaires omnidirectionnelles idéales, on néglige la dépendance entre g_Ψ et y_R et on choisit une profondeur d'écoute moyenne $y_{R_{av}}$. Les fonctions d'alimentation des sources secondaires $U_\Psi(x_L, k)$ situées sur la ligne L sont alors définies par

$$U_\Psi(x_L, k) = S(\omega) \sqrt{\frac{k}{2\pi}} g_\Psi(y_{R_{av}}) \cos \theta_0 \frac{e^{-j(kr_0 - \frac{\pi}{4})}}{\sqrt{r_0}}. \quad (1.14)$$

Le niveau du champ synthétisé n'est parfaitement égal que pour la profondeur d'écoute moyenne $y_{R_{av}}$. En dehors de cette profondeur, la loi d'atténuation du champ synthétisé n'est plus correcte. Nous en donnons une formulation dans la partie 2.2.1.

En plus des facteurs g_Ψ et $\cos \theta_0$ (composante normale du gradient de pression), un filtrage global *indépendant* de la source virtuelle synthétisée est appliqué sur le signal sonore d'origine $S(\omega)$. Ce filtrage présente une atténuation en \sqrt{k} qui correspond à une pente de +3 dB par octave ainsi qu'une phase constante. Traduit en terme de retard, ce déphasage correspond à une avance d'un huitième de période. Les basses fréquences sont ainsi émises en *avance* et avec un niveau plus faible que les hautes fréquences pour l'ensemble des sources secondaires.

Une extension de cette formulation permet de synthétiser des sources dites focalisées ("focused sources" en anglais) situées à l'intérieur de l'espace de reproduction Ω_R . Ces sources sont synthétisées en créant des fronts d'onde qui convergent vers le point de l'espace où est située la source virtuelle. Les ondes se propagent ainsi naturellement à partir de cette position dans le reste de l'espace de reproduction.

Les fonctions d'alimentations $U_\Psi^{foc}(x_L, k)$ sont obtenues par retournement temporel (propagation rétrograde) des fronts d'onde depuis la source virtuelle, vers les sources secondaires. On obtient alors,

$$U_\Psi^{foc}(x_L, k) = S(\omega) \sqrt{\frac{k}{2\pi}} g_\Psi(y_{R_{av}}) \cos \theta_0 \frac{e^{-j(\frac{\pi}{4} - kr_0)}}{\sqrt{r_0}}. \quad (1.15)$$

La principale différence entre U_Ψ^{foc} et U_Ψ , pour des sources virtuelles symétriques par rapport à $\partial\Omega$, est l'inversion de phase due au retournement temporel. Autrement dit, les retards à appliquer aux haut-parleurs sont simplement opposés. Le filtrage global pour une source focalisée présente un *retard* d'un huitième de période. Les basses fréquences doivent ainsi être émises en *retard* par rapport aux hautes fréquences.

L'approximation de la géométrie des sources secondaires par une ligne L modifie donc la nature du champ sonore synthétisé. Celui-ci n'est conforme à la cible qu'au niveau d'une ligne du plan horizontal, parallèle à L , pour une profondeur moyenne d'écoute $y_{R_{av}}$. A cause de la géométrie et de la nature des sources secondaires, le champ synthétisé présente une symétrie de révolution autour de l'axe de L et ne correspond donc plus à celui de la source primaire. Dans le plan

horizontal et en dehors de la profondeur d'écoute moyenne, l'amplitude du champ synthétisé n'est pas correctement reproduite. De plus, on a dérivé les fonctions d'alimentation des sources secondaires à partir de l'*approximation* de la phase stationnaire. Celle-ci n'est a priori valable que pour les hautes fréquences.

Dans le chapitre 2, nous analyserons la perte de qualité effective liée à ces approximations. Nous donnerons aussi la loi d'atténuation du champ effectivement synthétisé.

1.2.2 Approximation 2 : passage à un segment

En pratique, il est impossible de réaliser un banc de haut-parleurs de taille infinie. Il est donc nécessaire d'effectuer une troncature de la ligne.

Un banc de haut-parleurs de taille finie peut être assimilé à une fenêtre acoustique à travers laquelle on "voit" un certain espace extérieur. De la même manière, on peut définir une zone de "visibilité", Ω_R^Ψ , de la source primaire à travers la distribution de sources secondaires (cf fig. 1.3).

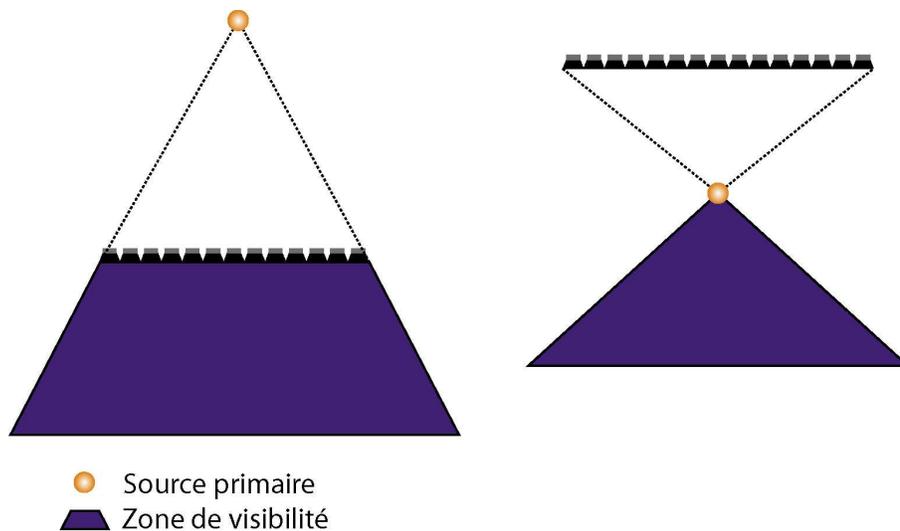


FIG. 1.3 – Mise en évidence des problèmes de visibilité

Ce fenêtrage introduit des effets de diffraction des ondes acoustiques qui vont modifier le champ synthétisé dans Ω_R^Ψ . A l'extérieur de Ω_R^Ψ , seules les contributions diffractées par l'ouverture constituée par la distribution de sources secondaires seront émises.

Il est néanmoins possible de réduire la diffraction en atténuant les contributions des haut-parleurs situés aux extrémités du banc (Start, 1997).

Dans le chapitre 2, nous montrons que l'atténuation des haut-parleurs situés aux extrémités a cependant des limites et nous apportons une analyse qualitative de l'effet de la diffraction sur le champ synthétisé dans l'ensemble de la zone de visibilité.

1.2.3 Approximation 3 : utilisation de bancs fermés

Une distribution de sources secondaires disposées le long d'un segment ne permet à un auditeur que de "voir" un espace acoustique réduit, principalement frontal. Afin d'augmenter les possibilités de disposition des sources sonores, on utilise un dispositif constitué de plusieurs segments linéaires qui entourent la zone d'écoute privilégiée.

Une analyse du champ synthétisé par chacun des segments est effectuée dans (Nicol, 1999) et (Start, 1997). Cette analyse utilise un formalisme de type Kirchhoff-Helmholtz (sources monopolaires *et* dipolaires) pour lequel le champ sonore est, a priori, parfaitement synthétisé dans la zone d'écoute. Il apparaît que, pour les sources secondaires situées entre la zone d'écoute et la source primaire, les composantes du champ synthétisé dans la zone d'écoute par les sources monopolaires et dipolaires sont pratiquement identiques. Pour les autres sources secondaires, les composantes issues des sources monopolaires et dipolaires sont pratiquement en opposition de phase. Pour la synthèse d'une source primaire donnée, il est alors suffisant d'utiliser les segments de sources monopolaires situées entre la source primaire et la zone d'écoute.

Pour chaque segment i , on peut définir un sous-espace de reproduction Ω_R^i et un sous-espace des sources Ω_Ψ^i . On considère que la zone d'écoute est située dans *tous* les sous-espaces de reproduction Ω_R^i . Le critère énoncé précédemment équivaut à utiliser le segment i pour la synthèse de la source Ψ si celle-ci est située dans le sous-espace des sources Ω_Ψ^i associé au segment i .

Dans ce cas, on remarque que les zones de visibilité de la source primaire à travers chacun des segments utilisés sont disjointes, juxtaposées dans la zone d'écoute (cf. figure 1.4). Chacun des segments permet, en quelque sorte, "d'éclairer" une partie de la zone d'écoute plus ou moins indépendamment les uns des autres (à la diffraction près).

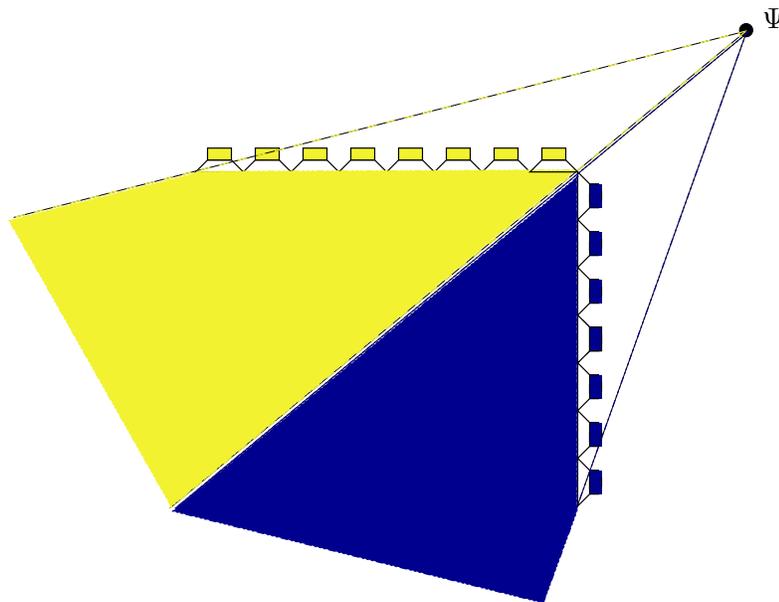


FIG. 1.4 – Zones de visibilité associées à deux bancs de haut-parleur non colinéaires

Ce raisonnement n'est cependant plus valable si Ψ est une source focalisée. Par définition, elle n'est alors située dans aucun des sous-espaces des sources Ω_{Ψ}^i . D'autre part, pour la synthèse d'une telle source, le champ synthétisé entre les haut-parleurs utilisés et la position de la source n'est pas correct. Nous proposons de considérer une zone d'écoute réduite (plus petite que l'intersection de l'ensemble des sous-espaces de reproduction Ω_R^i) et de contraindre la position des source virtuelles intérieures. Pour la synthèse d'une source donnée, on choisit alors les haut-parleurs qui ne sont pas situés entre la zone d'écoute et la position de la source virtuelle.

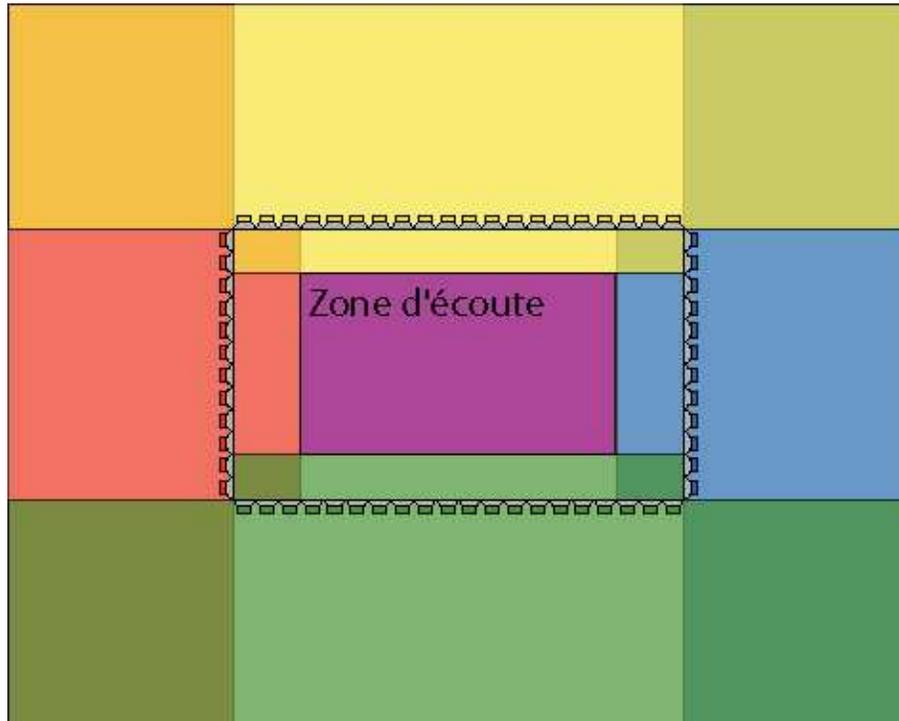


FIG. 1.5 – Bancs de haut-parleurs utilisés en fonction de la source virtuelle synthétisée

La figure 1.5 illustre le choix des haut-parleurs dans le cas d'une distribution rectangulaire de sources secondaires. On définit quatre bancs de haut-parleurs linéaires auxquels on associe une couleur particulière et une zone d'écoute réduite.

Pour les sources situées à l'extérieures de la ceinture de haut-parleurs, on a représenté les sous-espaces des sources Ω_{Ψ}^i associés à chacun des bancs linéaires. Ceux-ci définissent des zones d'utilisation de chacun des bancs de haut-parleurs en fonction de la zone d'écoute utilisée.

Pour les sources focalisées, on a représenté des zones plus réduites entre la zone d'écoute et la ceinture de haut-parleurs. Lorsqu'il n'y a pas de recouvrement entre les zones, on utilise le banc de haut-parleurs associé à la couleur de la zone *et* les haut-parleurs des bancs de haut-parleurs des bancs adjacents situés dans la zone considérée. Lorsque les zones se recouvrent, on peut choisir d'utiliser l'un ou l'autre ensemble de haut-parleurs. Le résultat est, a priori, équivalent.

L'analyse systématique des conséquences de l'utilisation de plusieurs bancs linéaires dépasse le cadre de cette étude.

1.2.4 Approximation 4 : échantillonnage

Le contrôle d'une distribution continue de sources omnidirectionnelles n'est pas possible dans l'état actuel de la technologie. La solution retenue consiste à utiliser un nombre fini de sources omnidirectionnelles régulièrement espacées, ce qui équivaut à effectuer un échantillonnage spatial régulier de la distribution de sources secondaires. La préservation du champ synthétisé est, en partie, garantie par l'écriture de la Wave Field Synthesis dans le domaine de Fourier spatial (Nicol, 1999) (Start, 1997).

La transformée de Fourier spatiale effectue une décomposition du champ sonore dans le domaine spatial des nombres d'onde k_x/k . On met alors en évidence un phénomène de recouvrement des composantes spatiales du champ sonore synthétisé à partir d'une fréquence limite, la fréquence d'aliasing spatial. En dessous de cette fréquence, le champ sonore émis par la distribution "échantillonnée" de sources secondaires est identique à celui émis par la distribution continue. Au-dessus, un phénomène de répétition des composantes spatiales du champ cible (celui de la source primaire) apparaît et le champ sonore synthétisé n'est plus correct.

Dans le cas d'une distribution de sources secondaires linéaire de taille infinie, la fréquence d'aliasing est donnée par

$$f_{al}^{\Psi} = \frac{c}{\Delta x(1 + \sin \theta_{\Psi, max})}, \quad (1.16)$$

où, Δx correspond à l'écartement entre les haut-parleurs, c à la vitesse du son dans l'air et $\theta_{\Psi, max}$ à l'angle maximal d'incidence du champ rayonné par la source primaire par rapport à la normale à la ligne des sources secondaires.

La fréquence d'aliasing dépend à la fois de la source primaire considérée et de l'écartement des sources secondaires. Des études menées à l'université de Delft ont permis de montrer qu'un écartement de 15 à 20 cm est suffisant pour assurer une localisation efficace des sources virtuelles ainsi synthétisées. Un tel écartement correspond en effet à une fréquence d'aliasing comprise entre 1500 et 2000 Hz pour la plupart des sources primaires considérées. Ceci permet d'assurer une reproduction idéale des différences interaurales de phase (IPD) qui constituent un indice important dans la localisation des sources sonores. Nous reviendrons sur ce point dans la partie 7.1.2 dans laquelle on estimera les indices de localisation auditive pour la WFS.

Cette fréquence limite est a priori indépendante de la position d'écoute dans le sous-espace de reproduction. Dans la suite de ce chapitre, nous verrons que ceci n'est plus valable dans le cas de bancs de haut-parleurs de taille finie et nous donnerons une formulation exacte de la fréquence d'aliasing en fonction de la position d'écoute dans la partie 2.4.

1.3 Synthèse de sources virtuelles à directivité arbitraire

La formulation de la WFS proposée par l'université de Delft et France Télécom R&D n'assure "que" la synthèse de sources virtuelles à directivité omnidirectionnelles.

Dans cette partie, nous proposons une extension de la formulation de la WFS pour la synthèse

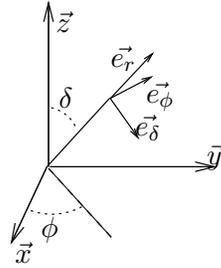


FIG. 1.6 – Coordonnées sphériques

de sources virtuelles à directivité arbitraire.

Rappelons pour cela que le champ de pression synthétisé par une source quelconque dans l'espace à trois dimensions en dehors de sa position se décompose, de manière unique, sur un ensemble de fonctions appelées harmoniques sphériques. Les harmoniques sphériques constituent les solutions générales de l'équation des ondes en coordonnées sphériques (cf. figure 1.6) en prenant l'hypothèse, classique, de séparabilité des variables (Bruneau, 1998).

On définit les harmoniques sphériques $Y_{mn}^\sigma(\delta, \phi)$ de degré $m \geq 0$ et d'ordre $0 \leq n \leq m$,

$$Y_{mn}^\sigma(\delta, \phi) = P_{mn}(\cos \delta) \times \begin{cases} \cos m\phi & \text{si } \sigma = 1 \\ \sin m\phi & \text{si } \sigma = -1 \end{cases}, \quad (1.17)$$

où, les P_{mn} sont les fonctions de Legendre. Celles-ci traduisent ainsi la dépendance angulaire du champ sonore. À ces harmoniques sphériques sont associées des fonctions de propagation, h_m^\pm , fonctions de Hankel sphériques.

L'ensemble des fonctions $\{h_m^- Y_{mn}^\sigma, 0 \leq n \leq m, m, n \in \mathbb{N}\}$ et $\{h_m^+ Y_{mn}^\sigma, 0 \leq n \leq m, m, n \in \mathbb{N}\}$ forme une base orthogonale de description d'un champ sonore quelconque dans une portion de l'espace à trois dimensions ne comportant pas de source. La normalisation de la base est réalisée en ajoutant des facteurs correctifs aux fonctions de Legendre (Laborie, 2000) (Daniel, 2000). Le sous-ensemble $\{h_m^- Y_{mn}^\sigma, 0 \leq n \leq m, m, n \in \mathbb{N}\}$ représente les ondes *divergentes* ; et $\{h_m^+ Y_{mn}^\sigma, 0 \leq n \leq m, m, n \in \mathbb{N}\}$, les ondes *convergentes*.

Dans Ω_R , en omettant provisoirement les focalisées, le champ est uniquement divergent. Le rayonnement d'une source sonore quelconque dans Ω_R s'exprime alors comme une somme pondérée des fonctions des bases $\{W_{mn}^\sigma = h_m^- Y_{mn}^\sigma, 0 \leq n \leq m, m, n \in \mathbb{N}\}$,

$$P(\vec{r}, k) = \sum_{m=0}^{+\infty} \sum_{0 \leq n \leq m, \sigma=\pm 1} A_{mn}^\sigma(k) W_{mn}^\sigma. \quad (1.18)$$

La synthèse du champ sonore associé à une source virtuelle Ψ_{mn} située en r_S , et dont les caractéristiques de rayonnement sont données par une harmonique sphérique W_{mn}^σ de degré m et

d'ordre n , dans le sous-espace de reproduction Ω_R , peut se décrire par l'intégrale de Rayleigh 1,

$$p_R(\vec{r}_R) = -2 \int_{\partial\Omega} G(\vec{r}_R, \vec{r}_S) \vec{\nabla} W_{mn}^\sigma(\vec{r}_S) \cdot \vec{n} dS \quad \text{Rayleigh 1.} \quad (1.19)$$

Le repère en coordonnées sphériques sur lequel on décrit la dépendance angulaire du rayonnement de Ψ_{mn} est ici centré sur la position de la source.

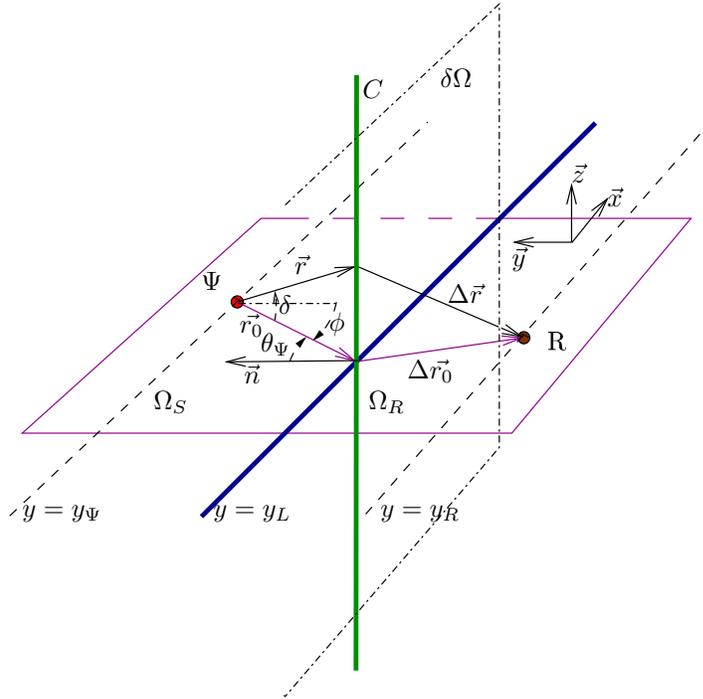


FIG. 1.7 – Réduction de la géométrie de sources secondaires à une ligne infinie par l'approximation de la phase stationnaire

On rappelle que pour la WFS, la distribution de sources secondaires est réduite à une ligne infinie dans le plan horizontal. La synthèse du champ n'est alors assurée correctement que dans le plan horizontal ($\delta = \frac{\pi}{2}$ dans le système de coordonnées sphériques). On s'intéresse au sous-ensemble des W_{mn}^σ pour lesquelles $P_{mn}(0) \neq 0$.

Dans le but de simplifier les écritures, on redéfinit $m = \sigma m$ et $0 \leq n \leq |m|$. Afin d'assurer la cohérence avec les sources ponctuelles omnidirectionnelles ($m = 0, n = 0$), on propose de considérer le sous-ensemble $\{P_{mn} \mid P_{mn}(0) \neq 0\}$. Lorsqu'on utilise l'approximation de champ lointain ($kr \gg 1$),

$$h_n^-(r) \approx j^{n+1} \frac{e^{-jkr}}{kr}. \quad (1.20)$$

On définit à un facteur de normalisation près,

$$\widetilde{W}_{mn}(r, \delta, \phi) = \frac{P_{mn}(\cos(\delta))}{P_{mn}(0)} \times \frac{e^{-jkr}}{4\pi r} \begin{cases} \cos(m\phi) & \text{pour } m \geq 0 \\ \sin(m\phi) & \text{pour } m < 0 \end{cases}, \quad (1.21)$$

On introduit alors le sous-ensemble des harmoniques sphériques qui peut a priori être reproduit en WFS,

$$\widetilde{Y}_{mn}(\delta, \phi) = \frac{P_{mn}(\cos(\delta))}{P_{mn}(0)} \begin{cases} \cos m\phi & \text{pour } m \geq 0 \\ \sin m\phi & \text{pour } m < 0 \end{cases}. \quad (1.22)$$

Comme dans le cas de sources ponctuelles omnidirectionnelles, on considère le champ sonore synthétisé par chacune des colonnes verticales $C(x_L)$ (cf. figure 1.2)

$$P_{C(x_L)}(r_R, k) = S(\omega) \int_{-\infty}^{+\infty} \vec{\nabla} \widetilde{W}_{mn}(\vec{r}) \cdot \vec{n} \frac{e^{-jk\Delta r}}{4\pi\Delta r} dz. \quad (1.23)$$

Rappelons la formule du gradient dans le système de coordonnées sphériques,

$$\vec{\nabla} f(r, \delta, \phi) = \frac{\partial f}{\partial r} \vec{e}_r + \frac{1}{r} \frac{\partial f}{\partial \delta} \vec{e}_\delta + \frac{1}{r \sin \delta} \frac{\partial f}{\partial \phi} \vec{e}_\phi. \quad (1.24)$$

On a alors,

$$\vec{\nabla} \widetilde{W}_{mn}(r, \delta, \phi) = \left(\frac{1 + jkr}{r} \widetilde{Y}_{mn}(\delta, \phi) \vec{e}_r + \frac{1}{r} \frac{\partial \widetilde{Y}_{mn}}{\partial \delta}(\delta, \phi) \vec{e}_\delta + \frac{1}{r \sin \delta} \frac{\partial \widetilde{Y}_{mn}}{\partial \phi}(\delta, \phi) \vec{e}_\phi \right) \frac{e^{-jkr}}{4\pi r}. \quad (1.25)$$

En champ lointain, on peut écrire

$$\vec{\nabla} \widetilde{W}_{mn}(r, \delta, \phi) \approx \left(jk \widetilde{Y}_{mn}(\delta, \phi) \vec{e}_r + \frac{1}{r} \frac{\partial \widetilde{Y}_{mn}}{\partial \delta}(\delta, \phi) \vec{e}_\delta + \frac{1}{r \sin \delta} \frac{\partial \widetilde{Y}_{mn}}{\partial \phi}(\delta, \phi) \vec{e}_\phi \right) \frac{e^{-jkr}}{4\pi r}. \quad (1.26)$$

Ainsi, la contribution de la colonne $C(x_L)$ s'écrit

$$P_{C(x_L)}(r_R, k) = S(\omega) \left[jk \int_{-\infty}^{+\infty} \widetilde{Y}_{mn}(\delta, \phi) \vec{e}_r \cdot \vec{n} \frac{e^{-jk(r+\Delta r)}}{16\pi^2 r \Delta r} dz + \int_{-\infty}^{+\infty} \left(\frac{1}{r} \frac{\partial \widetilde{Y}_{mn}}{\partial \delta}(\delta, \phi) \vec{e}_\delta + \frac{1}{r \sin \delta} \frac{\partial \widetilde{Y}_{mn}}{\partial \phi}(\delta, \phi) \vec{e}_\phi \right) \cdot \vec{n} \frac{e^{-jk(r+\Delta r)}}{16\pi^2 r \Delta r} dz \right]. \quad (1.27)$$

Chacune de ces deux intégrales se prête à l'utilisation de l'approximation de la phase stationnaire. La fonction de phase est la même que dans le cas de sources omnidirectionnelles, seule change la fonction d'enveloppe qui ne doit être correctement évaluée qu'au point de phase stationnaire (intersection de $C(x_L)$ avec le plan horizontal). En remarquant qu'en ce point $\delta = \frac{\pi}{2}$, $\vec{e}_\delta \cdot \vec{n} = 0$ et $\vec{e}_\phi \cdot \vec{n} = \sin \theta_\Psi$, on écrit

$$P_{C(x_L)}(r_R, k) = \sqrt{2\pi} S(\omega) \left(\sqrt{jk} \cos \theta_\Psi \widetilde{Y}_{mn}\left(\frac{\pi}{2}, \phi_0\right) + \frac{1}{r_0 \sqrt{jk}} \sin \theta_\Psi \frac{\partial \widetilde{Y}_{mn}}{\partial \phi}\left(\frac{\pi}{2}, \phi_0\right) \right) \times \sqrt{\frac{r_0 \Delta r_0}{r_0 + \Delta r_0}} \frac{e^{-jk(r_0 + \Delta r_0)}}{16\pi^2 r_0 \Delta r_0}. \quad (1.28)$$

Pour $m \geq 0$, on a

$$P_{C(x_L)}(r_R, k) = \sqrt{2\pi}S(\omega) \left(\sqrt{jk} \cos \theta_\Psi \cos m\phi_0 - \frac{m}{r_0\sqrt{jk}} \sin \theta_\Psi \sin m\phi_0 \right) \\ \times \sqrt{\frac{r_0\Delta r_0}{r_0 + \Delta r_0}} \frac{e^{-jk(r_0+\Delta r_0)}}{16\pi^2 r_0\Delta r_0}, \quad (1.29)$$

et pour $m < 0$:

$$P_{C(x_L)}(r_R, k) = \sqrt{2\pi}S(\omega) \left(\sqrt{jk} \cos \theta_\Psi \sin m\phi_0 + \frac{m}{r_0\sqrt{jk}} \sin \theta_\Psi \cos m\phi_0 \right) \\ \times \sqrt{\frac{r_0\Delta r_0}{r_0 + \Delta r_0}} \frac{e^{-jk(r_0+\Delta r_0)}}{16\pi^2 r_0\Delta r_0}. \quad (1.30)$$

Ces approximations permettent d'évaluer la contribution de la ligne verticale et de simplifier l'intégrale de Rayleigh 1, pour $m \geq 0$,

$$P_{mn}^1(r_R, k) = S(\omega)\sqrt{jk} \int_{-\infty}^{+\infty} g_\Psi(y_R) \left(\cos \theta_\Psi \cos m\phi_0 - \frac{m}{jkr_0} \sin \theta_\Psi \sin m\phi_0 \right) \frac{e^{-jkr_0}}{\sqrt{2\pi}\sqrt{r_0}} \frac{e^{-jk\Delta r_0}}{4\pi\Delta r_0} dx_L, \quad (1.31)$$

et pour $m < 0$,

$$P_{mn}^{-1}(r_R, k) = S(\omega)\sqrt{jk} \int_{-\infty}^{+\infty} g_\Psi(y_R) \left(\cos \theta_\Psi \sin m\phi_0 + \frac{m}{jkr_0} \sin \theta_\Psi \cos m\phi_0 \right) \frac{e^{-jkr_0}}{\sqrt{2\pi}\sqrt{r_0}} \frac{e^{-jk\Delta r_0}}{4\pi\Delta r_0} dx_L. \quad (1.32)$$

On remarque que dans les deux cas, l'expression ne dépend plus de l'ordre n . En effet, l'ordre traduit les variations de directivité dans le plan vertical. Or, la WFS ne permet la synthèse d'un champ cible que dans le plan horizontal. De plus, le champ sonore synthétisé présente une symétrie autour de l'axe de la distribution de sources secondaires. On a alors une indétermination sur l'ordre n . On nommera dans la suite la source Ψ_m . On rappelle que les degrés négatifs correspondent à $\sigma = -1$.

Afin de simplifier les expressions, on choisit de placer la source virtuelle Ψ_m en champ lointain par rapport à la distribution de sources secondaires ($kr_0 \gg 1$) et de considérer des degrés m petits. On définit

$$\Phi_m(\phi) = \begin{cases} \cos m\phi & \text{si } m \geq 0 \\ \sin m\phi & \text{pour } m < 0 \end{cases}. \quad (1.33)$$

On réécrit la fonction d'alimentation $U_{\Psi_m}(x_L, k)$ à considérer pour la reproduction d'une source virtuelle Ψ_m

$$U_{\Psi_m}(x_L, k) = S(\omega)\sqrt{\frac{k}{2\pi}} g_\Psi(y_{R_{av}}) \cos \theta_0 \frac{e^{-j(kr_0 - \frac{\pi}{4})}}{\sqrt{r_0}} \Phi_m(\phi_0). \quad (1.34)$$

Rappelons que $y_{R_{av}}$ est la profondeur moyenne d'écoute pour laquelle le niveau de la source est correctement synthétisé.

De manière similaire, on peut obtenir une fonction d'alimentation pour la synthèse de sources

virtuelles focalisées de degré $m < 0$:

$$U_{\Psi_m}^{foc}(x_L, k) = S(\omega) \sqrt{\frac{k}{2\pi}} g_{\Psi}(y_{R_{av}}) \cos \theta_0 \frac{e^{-j(\frac{\pi}{4} - kr_0)}}{\sqrt{r_0}} \Phi_m(\phi_0 + \pi). \quad (1.35)$$

Les équations que nous avons dérivées constituent une extension à la Wave Field Synthesis pour la synthèse de sources virtuelles dont les fonctions de directivité décrivent une base incomplète des harmoniques sphériques.

On a vu cependant que la synthèse du rayonnement est limitée au plan horizontal. En dehors du plan horizontal, le champ émis présente une symétrie de révolution autour de l'axe de la distribution de sources secondaires.

La définition des fonctions d'alimentation repose sur de nombreuses approximations (source située en champ lointain du banc de haut-parleurs, degré m petit, phase stationnaire pour dériver les fonctions d'alimentation d'une ligne de sources secondaires). L'analyse systématique des conséquences de ces approximations dépasse le cadre de notre étude. Nous donnerons cependant, dans la partie suivante, quelques indications sur la nature du champ effectivement synthétisé.

Enfin, la synthèse d'une figure de directivité quelconque est réalisée par recombinaison sur la base incomplète d'harmoniques sphériques. La prise en compte des "défauts" de reproduction des harmoniques sphériques ainsi que de l'impossibilité de synthétiser l'ordre n dépasse le cadre de cette étude.

1.4 Simulations

Dans cette partie, nous présentons les résultats de simulations du champ acoustique synthétisé par un banc linéaire de 60 haut-parleurs idéaux omnidirectionnels espacés de 10 cm suivant une répartition régulière. Pour cela, on simule les réponses impulsionnelles de ce banc, captées par un ensemble de 60 microphones omnidirectionnels, répartis sur une ligne située à 2 m en face du banc de haut parleurs. Les microphones sont régulièrement répartis tous les 30 cm (cf. figure 1.8). L'ensemble de positions de sources virtuelles à directivité omnidirectionnelle servira de cadre d'analyse dans toute la suite du document :

- source focalisée située en (1, 1.2), soit 80 cm *devant* le banc de haut-parleur ;
- source omnidirectionnelle située en (3, 5), soit 3 m *derrière* le banc de haut-parleur et en face de l'extrémité droite du banc de haut-parleurs ;
- une onde plane * se propageant perpendiculairement au banc de haut-parleurs (onde plane à 0 degré).

*L'utilisation de bancs de haut-parleurs linéaire ne permet pas à proprement parler la synthèse d'une onde plane mais d'un front d'onde à symétrie cylindrique. Il s'agit donc d'un abus de langage que l'on retrouve dans l'ensemble de la littérature concernant la Wave Field Synthesis. On conserve ainsi le terme dans le reste du document pour en faciliter la compréhension et la cohérence au reste de la littérature.

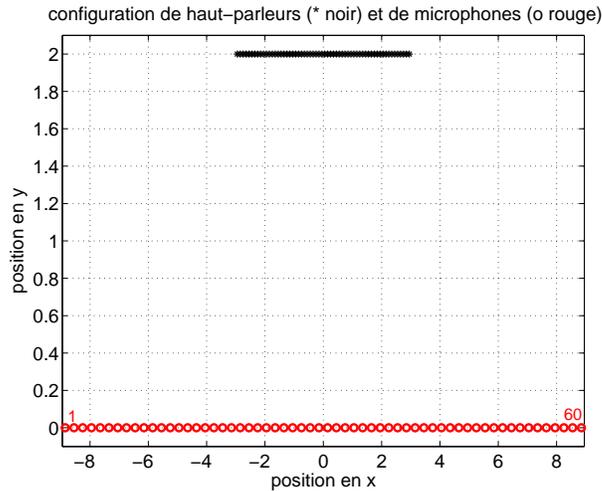


FIG. 1.8 – Haut-parleurs et Microphones utilisés

Dans un premier temps, nous nous attachons à mettre en évidence les caractéristiques du champ acoustique pour la synthèse de ces sources à directivité omnidirectionnelle (directivité de degré 0) puis on montre des résultats pour des sources de directivité de degré -1 (dipôles orientés parallèlement au banc de haut-parleurs) puis de degré 4.

La directivité d'une onde plane n'ayant aucun sens, on la remplace par une source située en $(0, 9)$, soit 7 m derrière le banc de haut-parleurs. La source virtuelle focalisée en $(1, 1.2)$ est remplacée par une source en $(0, 1)$, soit 1 m devant le banc de haut-parleurs et centrée.

Pour la clarté de la représentation des effets de directivité des sources de degré supérieur, on les représente sur une répartition de microphones de même longueur mais avec un espacement de 10 cm.

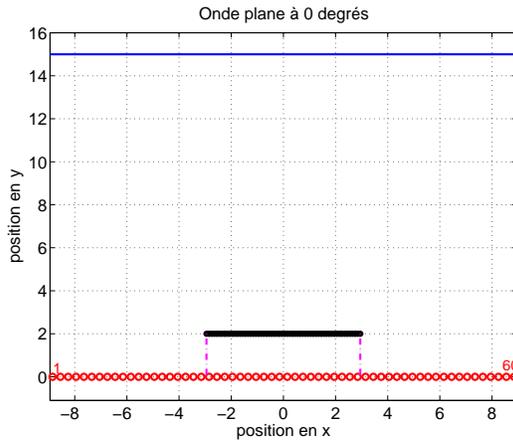
L'ensemble de ces sources sont normalisées de telle manière que leur niveau, en dehors de leurs caractéristiques de directivité, soit égale à 1 en $(0, 0)$ et qu'elles arrivent avec un retard fixe en ce point.

1.4.1 Sources "omnidirectionnelles"

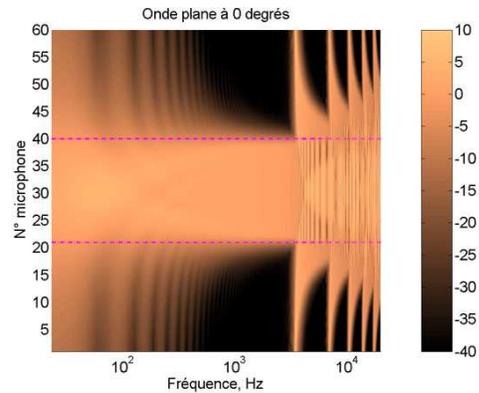
Pour chacune des sources virtuelles omnidirectionnelles, un ensemble de quatre figures est présenté. Il comprend :

- la position de la source et la zone de visibilité associée.
- les modules des réponses fréquentielles pour l'ensemble des positions des microphones.
- les réponses impulsionnelles auxquelles on applique un filtrage passe-bas (fréquence limite de 2000 Hz qui est proche de la fréquence d'aliasing spatial). Le niveau est donné en échelle logarithmique.
- les réponses impulsionnelles auxquelles on applique un filtrage passe-haut (fréquence limite de 2000 Hz). Le niveau est donné en échelle logarithmique.

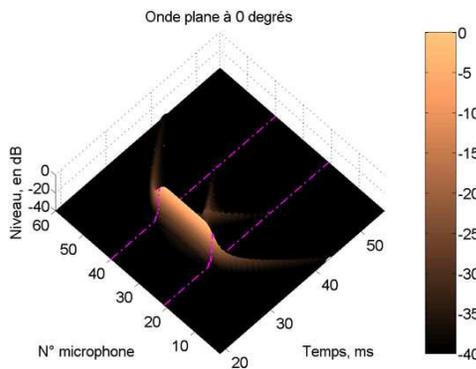
Pour chacune de ces figures, la zone de visibilité de la source virtuelle à travers le banc de haut-parleurs est indiquée par une ligne discontinue magenta. Ceci permet de distinguer le champ "cible" dans la zone de visibilité et le champ "diffracté" en dehors.



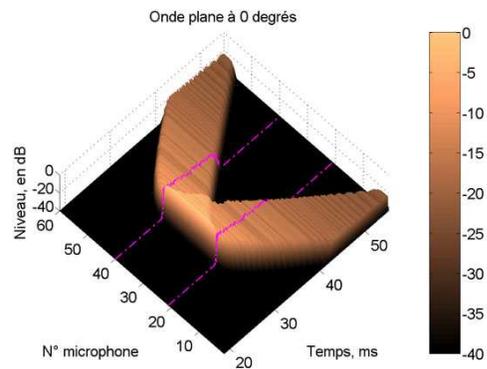
(a) Source virtuelle et zone de visibilité associée



(b) Réponses fréquentielles



(c) Réponses impulsionnelles filtrées passe bas, fréquence de coupure correspond à la fréquence d'aliasing



(d) Réponses impulsionnelles filtrées passe haut, fréquence de coupure correspond à la fréquence d'aliasing

FIG. 1.9 – Champ synthétisé pour une onde plane se propageant perpendiculairement au banc de haut-parleurs

La synthèse de l'onde plane à 0 degré laisse apparaître de nombreuses caractéristiques typiques de la WFS (cf. figure 1.9). On remarque notamment la synthèse précise du front d'onde en dessous de la fréquence d'aliasing. On note toutefois la présence de deux fronts d'onde de faible amplitude semblant provenir des extrémités du banc de haut-parleurs (cf. figure 1.9(c)). Ceux-ci sont dus à la diffraction du champ synthétisé à travers le banc de haut-parleurs de taille finie. Ces fronts d'onde subsistent même lorsque l'on réduit le niveau appliquée sur les haut-parleurs situés aux extrémités. Ils viennent légèrement perturber la réponse fréquentielle (cf. figure 1.9(b)). Ceci se produit principalement en basses fréquences.

A l'extérieur de la zone de visibilité, on remarque l'émission de front d'onde semblant également

provenir des extrémités des haut-parleurs dont le contenu fréquentiel se concentre principalement en basses fréquences et présente de légères oscillations (cf. figure 1.9(b)). La concentration de l'énergie dans la fenêtre de visibilité est ainsi plus efficace en moyennes/hautes fréquences.

Au dessus de la fréquence d'aliasing, on note des contributions successives qui proviennent de chacun des haut-parleurs. Contrairement aux basses fréquences, ceux-ci ne fusionnent pas en un unique front d'onde. Néanmoins, on remarque que le premier front d'onde reçu est synchrone avec celui reçu en dessous de la fréquence d'aliasing. L'attaque de l'enveloppe de la réponse impulsionnelle reste ainsi cohérente avec le front d'onde en basses fréquences. Ceci assure une restitution cohérente des différences interaurales d'enveloppe (IED). Rappelons que l'IED est un indice de localisation utilisé par le système auditif au-dessus de 1500 Hz (cf. partie 7.1.1).

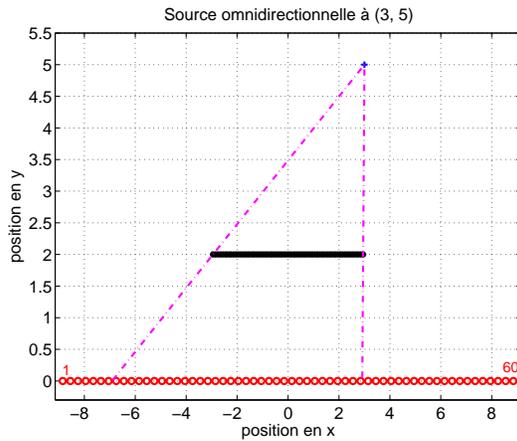
Le phénomène d'aliasing spatial se manifeste par des irrégularités dans la réponse fréquentielle. Ces irrégularités forment des motifs qui dépendent de la position. On note aussi que la structure temporelle du champ d'aliasing spatial change en fonction de la position de réception. Ainsi, pour les positions centrales (microphones 29-30), l'étalement temporel du champ d'aliasing est réduit par rapport aux positions situées en face des extrémités du banc de haut-parleurs (microphones 21-22 ou 39-40). En dehors, de la zone de visibilité, on note une émergence très nette de l'énergie émise à partir de la fréquence d'aliasing spatial qui s'accompagne, pour cette source virtuelle particulière, de phénomènes de lobes et de résonances.

Pour la source omnidirectionnelle à (3, 5), on remarque que le front d'onde associé à la source virtuelle cible est, là aussi, correctement synthétisé en basses fréquences (cf. figure 1.10(c)). Le premier front d'onde émis au-dessus de la fréquence d'aliasing reste cohérent avec le front d'onde cible (cf. figure 1.10(d)).

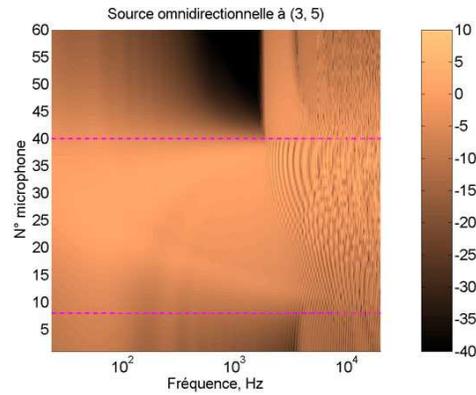
On note que la diffraction est globalement plus importante que pour l'onde plane, notamment pour les microphones 1 à 5 pour lesquels la réponse fréquentielle est pratiquement indépendante de la fréquence (cf. figure 1.10(b)). De l'autre côté de la zone de visibilité (microphones 41 à 60), la diffraction est plus concentrée en basses fréquences. On remarque une rupture très nette à partir de la fréquence d'aliasing spatial. On retrouve un niveau similaire à celui synthétisé à l'intérieur de la zone de visibilité.

On remarque une hausse sensible de la fréquence d'aliasing spatial entre les microphones 41 à 60 et les microphones 1 à 8 qui se traduit par une variation importante du support temporel du champ d'aliasing spatial (5 à 20 ms) en fonction de la position d'écoute considérée (cf. figure 1.10(d)).

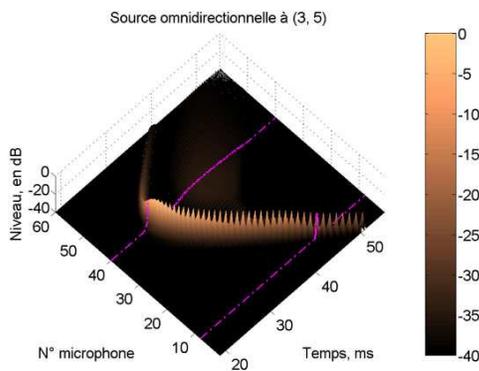
Pour la source omnidirectionnelle focalisée à (1, 1.2), la principale différence, par rapport aux deux autres sources considérées, concerne le retournement temporel. Ceci se traduit dans le domaine temporel par un résidu de composantes diffractées (cf. figure 1.11(c)) et un champ d'aliasing spatial (cf. figure 1.11(d)) émis *avant* le front d'onde principal. Ceci est ainsi susceptible d'introduire des problèmes pour la localisation des sources virtuelles. Remarquons que le support temporel du champ d'aliasing spatial dépend de la *longueur* du banc de haut-parleurs. On peut artificiellement réduire cette longueur en atténuant le niveau des haut-parleurs situés sur les côtés au-dessus de la fréquence d'aliasing. Ceci sera appliqué lors de l'égalisation multicanal



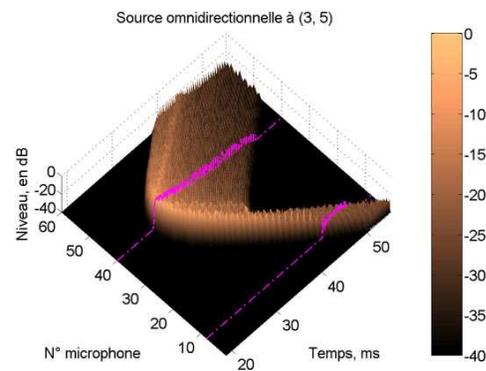
(a) Source virtuelle et zone de visibilité associée



(b) Réponses fréquentielles



(c) Réponses impulsionnelles filtrées passe bas, fréquence de coupure correspond à la fréquence d'aliasing



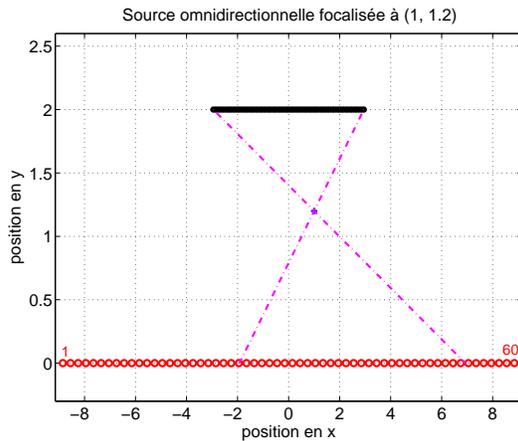
(d) Réponses impulsionnelles filtrées passe haut, fréquence de coupure correspond à la fréquence d'aliasing

FIG. 1.10 – Champ synthétisé pour une source omnidirectionnelle situé 3m derrière le banc de haut-parleurs et décentrée de 3m sur la droite

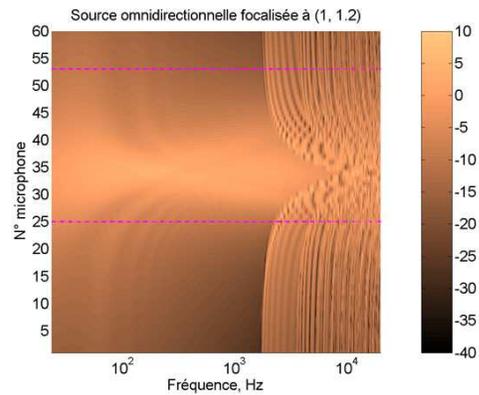
que nous proposons dans le chapitre 5.

En dehors de la zone de visibilité, le champ diffracté est globalement plus important en basses fréquences (cf. figure 1.11(b)). Néanmoins, il reste relativement important aux fréquences moyennes. Dans cette zone, on note, d'autre part, une émergence très forte à partir de la fréquence d'aliasing spatial.

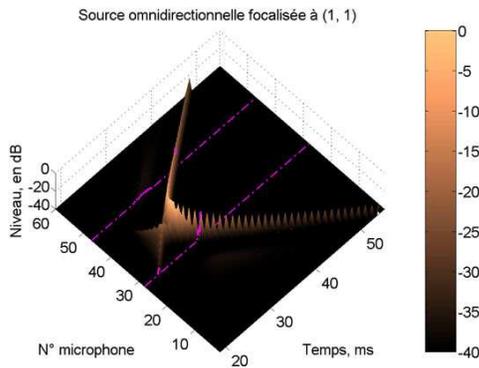
La fréquence d'aliasing spatial varie très sensiblement entre 2000 Hz pour les positions de micros aux extrémités et environ 10000 Hz pour la position 34 située en face de la source virtuelle. On peut mettre ceci en rapport avec l'étalement temporel du champ d'aliasing dont la durée est inversement proportionnelle à la fréquence d'aliasing observée, (cf. figure 1.11(d)).



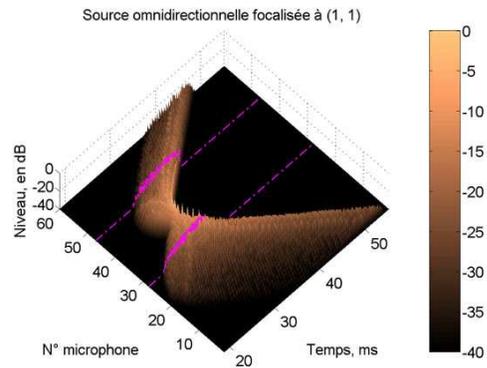
(a) Source virtuelle et zone de visibilité associée



(b) Réponses fréquentielles



(c) Réponses impulsionnelles filtrées passe bas, fréquence de coupure correspond à la fréquence d'aliasing



(d) Réponses impulsionnelles filtrées passe haut, fréquence de coupure correspond à la fréquence d'aliasing

FIG. 1.11 – Champ synthétisé pour une source omnidirectionnelle focalisée situé 80cm devant le banc de haut-parleurs et décentrée de 1m sur la droite

1.4.2 Sources de directivité de degré supérieur

Dans cette partie, nous présentons la synthèse de sources virtuelles similaires à celles de la partie précédente, mais avec des caractéristiques de directivité plus complexes. L'onde plane est remplacée par une source située en $(0, 9)$ et la source focalisée est centrée par rapport au banc de haut-parleur et située 1 m devant, soit en $(0, 1)$. Nous considérons des sources de degré -1 (dipôle orienté parallèlement au banc de haut-parleur) et de degré 4. Pour chacune des sources virtuelles, on présente uniquement deux figures :

- les modules des réponses fréquentielles pour l'ensemble des positions de microphone.
- les réponses impulsionnelles filtrées passe-bas (fréquence limite de 2000 Hz), contenu des réponses impulsionnelles "sous" la fréquence d'aliasing, niveau en échelle linéaire.

Ces figures sont regroupées par six et par degré de directivité. Rappelons que les 180 microphones sont disposés sur une ligne à 2 m du banc de haut-parleurs et sont espacés de 10 cm.

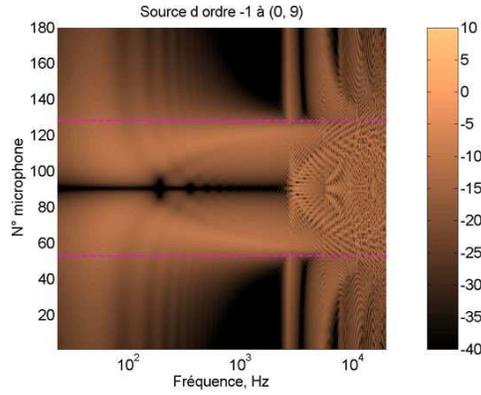
Pour la source de degré -1 située à $(0, 9)$, dans le domaine fréquentiel, le zéro du dipôle est bien reproduit jusqu'à la fréquence d'aliasing. Les réponses fréquentielles présentent cependant des irrégularités dans la synthèse des lobes(cf. figure 1.12(a)). Dans le domaine temporel, les deux lobes attendus sont bien de signes opposés (cf. figure 1.12(b)). Chacun d'entre présente cependant des composantes en opposition de phase. Celles-ci sont probablement inaudibles mais risquent d'introduire des erreurs lors de la synthèse de figures de directivité plus complexes par combinaison des harmoniques sphériques. Enfin, on remarque que l'on ne synthétise qu'une portion de la figure de directivité : celle qui est contenue dans la zone de visibilité.

Cet effet est particulièrement notable pour la source située à $(3, 5)$ pour laquelle un seul lobe est situé dans la zone de visibilité, dont la synthèse se dégrade légèrement en basses fréquences (cf. figures 1.12(c) et 1.12(d)).

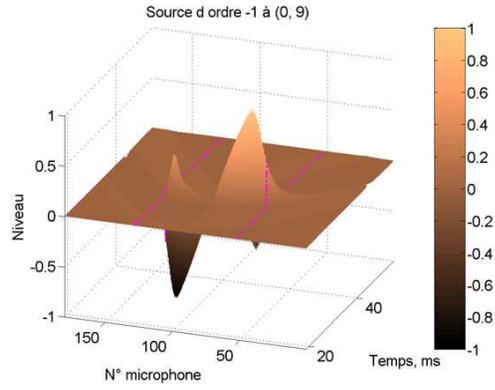
La source focalisée de degré -1 fait apparaître un zéro correctement reproduit jusqu'à la fréquence d'aliasing pour cette position, soit environ 10000 Hz (cf. figure 1.12(e)). On remarque néanmoins que la synthèse est moins précise en basses fréquences, les deux lobes semblant fusionner en dessous de 100 Hz. D'autre part, on note la présence de composantes en opposition de phase d'un niveau relativement élevé pour chacun des lobes, (cf. figure 1.12(f)).

Pour les sources de degré 4, on remarque des comportements similaires à ceux observés pour les sources virtuelles de directivité dipolaire (cf. ensemble des figures 1.13). Ainsi, la source à $(0, 9)$ ne laisse apparaître qu'un seul lobe à l'intérieur de la zone de visibilité. La source à $(3, 5)$ synthétise deux lobes. Celui situé entre les positions des microphones 100 et 120 présente de nombreux défauts aussi bien dans le domaine temporel que dans le domaine fréquentiel (variation de position du zéro séparant les lobes, niveau important de la composante en opposition de phase associée). On remarque, d'autre part, une fusion des deux lobes en basses fréquences (20-200 Hz). La source focalisée a le même comportement avec des composantes en opposition de phase pour chacun des lobes d'un niveau très élevé et une fusion des trois lobes en basses et moyennes fréquences (20-400 Hz).

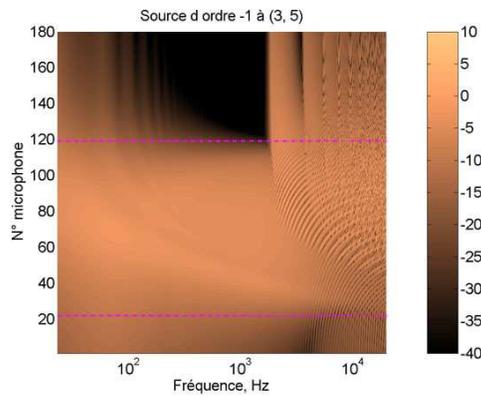
De manière générale, la synthèse de la figure de directivité est correcte en dessous de la fréquence d'aliasing et se dégrade légèrement en basses fréquences quelque soit la source considérée. Cette dégradation est particulièrement remarquable pour les sources de degré plus élevé. D'autre part, la synthèse de la figure de directivité n'est effectuée que dans la fenêtre de visibilité déterminée par la position de la source et l'ouverture du banc de haut-parleur. On ne synthétise en fait que le rayonnement de la source dans un angle solide limité qui se réduit lorsque la source s'éloigne du banc de haut-parleur.



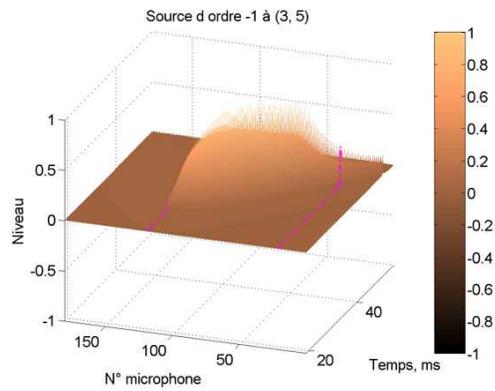
(a) Source de degré -1 située à (0, 9), réponses fréquentielles



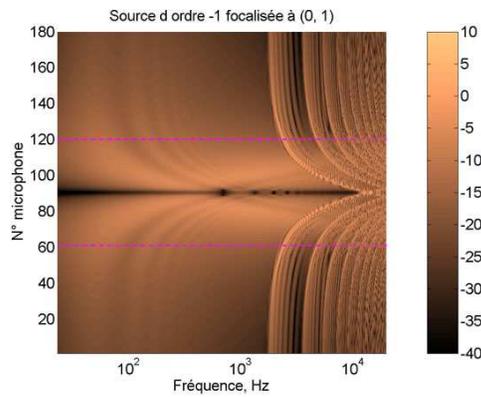
(b) Source de degré -1 située à (0, 9), réponses impulsionnelles filtrées passe bas



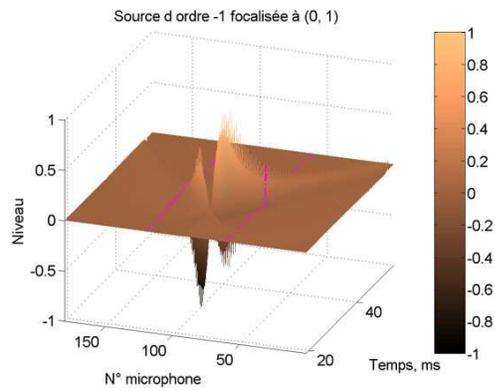
(c) Source de degré -1 située à (3, 5), réponses fréquentielles



(d) Source de degré -1 située à (3, 5), réponses impulsionnelles filtrées passe bas

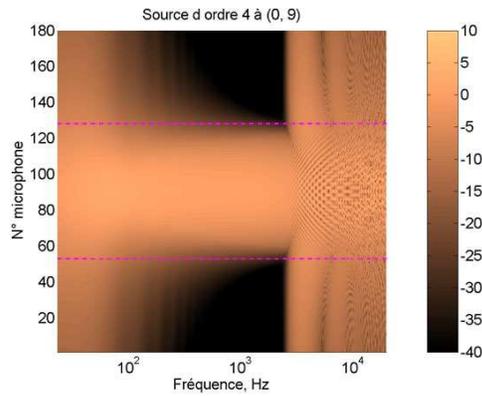


(e) Source de degré -1 située à (0, 1), réponses fréquentielles

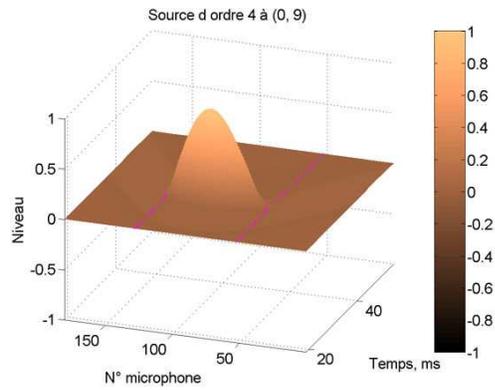


(f) Source de degré -1 située à (0, 1), réponses impulsionnelles filtrées passe bas

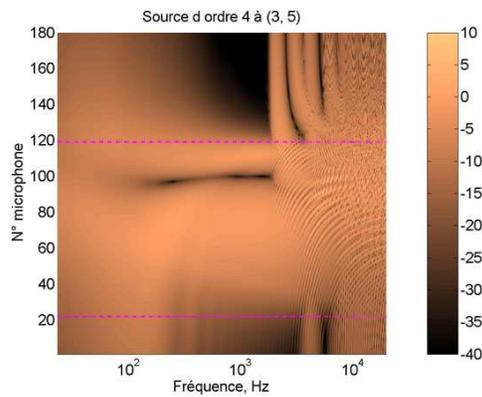
FIG. 1.12 – Champ synthétisé pour un ensemble de sources virtuelles de directivité dipolaire orientées parallèlement au banc de haut-parleurs



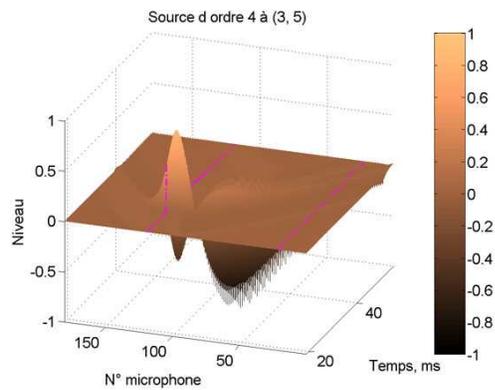
(a) Source de degré 4 située à $(0, 9)$, réponses fréquentielles



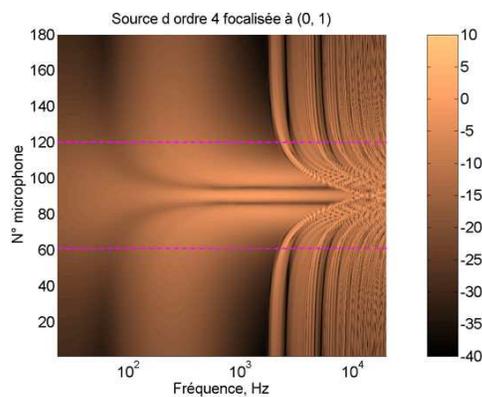
(b) Source de degré 4 située à $(0, 9)$, réponses impulsionnelles filtrées passe bas



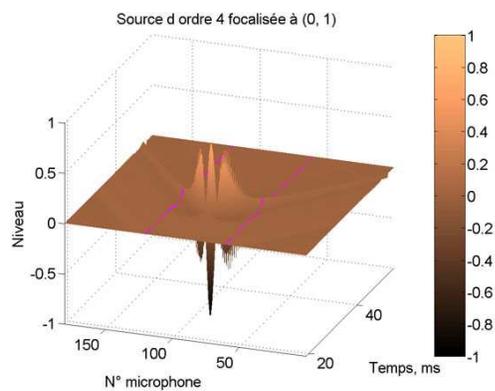
(c) Source de degré 4 située à $(3, 5)$, réponses fréquentielles



(d) Source de degré 4 située à $(3, 5)$, réponses impulsionnelles filtrées passe bas



(e) Source de degré 4 située à $(0, 1)$, réponses fréquentielles



(f) Source de degré 4 située à $(0, 1)$, réponses impulsionnelles filtrées passe bas

FIG. 1.13 – Champ synthétisé pour un ensemble de sources virtuelles de directivité de degré 4

1.4.3 Discussion : qualité de la synthèse des sources virtuelles en WFS

Au cours de ce chapitre, nous avons mis en évidence un certain nombre de caractéristiques du champ sonore rayonné par un banc linéaire de haut-parleurs omnidirectionnels pour la synthèse d'une source virtuelle en WFS.

1. Le champ acoustique rayonné possède une symétrie de révolution autour de l'axe du banc de haut-parleurs.
2. Le niveau synthétisé n'est conforme à celui de la source cible que sur une ligne parallèle au banc de haut-parleurs dont on peut choisir la position. La loi d'atténuation du champ synthétisé n'est donc pas celle de la source cible.
3. Le champ synthétisé est émis à travers une "ouverture" de taille finie et n'est correctement synthétisé que dans une certaine zone de visibilité de la source à travers cette ouverture. A l'intérieur de cette zone, il présente des composantes diffractées sur les bords de l'ouverture qui créent de légères perturbations dont les caractéristiques dépendent de la source considérée et de la taille du banc de haut-parleurs (Start, 1997) (Nicol, 1999).
4. En basses fréquences, la directivité des sources n'est pas convenablement reproduite. Pour les sources de degré supérieur, les lobes tendent à fusionner en basses fréquences.
5. Le champ est "incorrect" à partir d'une fréquence limite, qui dépend de la position de la source virtuelle, de l'espacement des haut-parleurs mais aussi, on l'a vu, de la position d'écoute considérée. Le champ synthétisé laisse apparaître séparément les contributions des haut-parleurs du banc linéaire qui ne fusionnent plus en un unique front d'onde. Pour une source située dans Ω_Ψ , le premier front d'onde synthétisé est cohérent avec le front d'onde cible. Pour les sources focalisées, le champ synthétisé s'étend *devant* le front d'onde cible et est ainsi susceptible d'introduire des problèmes de localisation en hautes fréquences.

Ces caractéristiques peuvent être considérés comme autant de "défauts" de la synthèse de sources virtuelles par la WFS. Dans le chapitre suivant, nous nous attacherons à évaluer l'importance de ces défauts sur la transparence de la reproduction en fonction de la position d'écoute.

Dans les chapitres 2 et 3, notre étude se limitera à des sources virtuelles *omnidirectionnelles*.

Chapitre 2

Rayonnement synthétisé en champ libre

Dans ce chapitre, nous définissons des méthodes de prédiction du champ sonore synthétisé par un banc linéaire de haut-parleurs omnidirectionnels alimentés par des signaux donnés par la Wave Field Synthesis. Ces méthodes permettent de décrire la synthèse du front d'onde pour n'importe quelle position de la zone d'écoute.

Nous mettrons en évidence une propagation des fronts d'onde synthétisés de type *dispersive*. C'est à dire pour laquelle la loi d'atténuation et la vitesse de propagation ne sont pas indépendantes de la fréquence. Cette propagation dispersive affecte la transparence de la reproduction en fonction de la position d'écoute principalement dans le cas de bancs de taille finie. Nous tenterons de quantifier ce phénomène et d'évaluer son importance d'un point de vue perceptif.

Dans un premier temps, nous donnerons des critères objectifs issus de résultats psychacoustiques qui permettent de juger de la transparence d'un système de restitution. Nous définirons une *fonction de qualité* à partir de la réponse impulsionnelle du système mesurée en champ libre.

Dans le cadre de la description du rayonnement d'une distribution continue infinie de sources secondaires, nous proposerons des méthodes de calcul ou d'estimation du front d'onde synthétisé en un point de l'espace, à la fois dans le domaine fréquentiel, et dans le domaine temporel. Un ensemble de simulations permettra de valider les méthodes proposées et d'évaluer la fonction de qualité dans ce cas.

Nous analyserons ensuite le champ rayonné par un banc de taille finie dans le domaine temporel et dans le domaine fréquentiel. Nous montrerons ainsi l'influence de la diffraction sur la fonction de qualité ainsi que son évolution spatiale.

Enfin, nous présenterons une définition originale de la fréquence d'aliasing spatial par un calcul du front d'onde synthétisé dans le domaine temporel. Cette définition permettra de tenir compte de sa dépendance envers la position d'écoute.

2.1 Évaluation objective de la transparence de reproduction en champ libre

La notion de transparence du système de reproduction peut être définie comme la conservation des caractéristiques perceptives d'un signal transmis par un système électroacoustique à un auditeur situé à une position de l'espace.

Une évaluation objective de la transparence de la reproduction sonore en champ libre consiste à analyser les caractéristiques de la fonction de transfert électroacoustique du système considéré *. On effectue pour cela une mesure de la réponse impulsionnelle $h(t)$ du système de transduction électroacoustique en champ libre (par exemple, dans une chambre anéchoïque).

A partir de ce type de mesure, on extrait deux informations essentielles pour juger de la transparence du système considéré :

1. le module de la réponse fréquentielle $H(\omega)$ donne le niveau synthétisé par le système en fonction de la fréquence,
2. le temps de groupe, dérivée de la phase de la réponse fréquentielle par rapport à la fréquence $\frac{\partial \arg H(\omega)}{\partial \omega}$, détermine d'arrivée de l'énergie en fonction de la fréquence.

Idéalement, cette réponse impulsionnelle doit avoir les caractéristiques d'une distribution de Dirac. Ceci garantit une reproduction identique et synchrone de l'ensemble des fréquences.

En pratique, on réduit l'étendue des fréquences considérées au spectre des fréquences audibles (20-20000 Hz). On ne considère pas les niveaux absolus de ces deux paramètres mais leurs variations par rapport à une valeur moyenne.

On définit une fonction de qualité $q(t)$,

$$q(t) = \frac{h(t - t_0)}{\sqrt{\int_{t_1}^{t_2} h(t)^2 dt}}, \quad (2.1)$$

où, t_0 est le temps de propagation théorique de l'énergie. t_1 et t_2 déterminent le support temporel de la réponse impulsionnelle. On s'intéresse au module de la transformée de Fourier $Q(\omega)$ de $q(t)$ et au temps de groupe $TG(\omega)$,

$$TG(\omega) = -\frac{\partial \arg Q(\omega)}{\partial \omega}. \quad (2.2)$$

L'évaluation de la transparence du système vis-à-vis de la perception passe ainsi par la détermination des seuils différentiels d'audition pour ces deux quantités.

2.1.1 Réponse fréquentielle

Les variations de la réponse fréquentielle autour de la valeur moyenne peuvent se modéliser comme un ensemble de résonances ("pics") et d'anti-résonances ("zéros"). Une résonance se

*On ne considère ici que les caractéristiques linéaires du système. Les non-linéarités, même si elles influent de manière très significatives sur la transparence du système étudié, ne sont pas prises en compte.

caractérise par son niveau, sa fréquence centrale, située au maximum de niveau dans la réponse fréquentielle, et son facteur de qualité (inverse de largeur de bande à $-3dB$ par rapport au niveau maximum multiplié par la fréquence centrale). De manière équivalente, une anti-résonance se caractérise par son niveau, sa fréquence centrale, située au minimum de niveau dans la réponse fréquentielle, et son facteur de qualité (inverse de largeur de bande à $+3dB$ par rapport au niveau minimum multiplié par la fréquence centrale).

Toole et Olive (Toole & Olive, 1988) puis Olive *et al.* (Olive *et al.*, 1997) poursuivant les travaux de Bucklein (Bucklein, 1985) se sont attachés à mesurer des seuils d'audition. Le seuil d'audition est défini par la valeur minimale du niveau au-dessus duquel on perçoit la résonance (resp. l'anti-résonance). On donne ces seuils d'audition en fonction de la fréquence centrale et du facteur de qualité.

Leurs études consistent à ajouter, par filtrage, une résonance (ou une anti-résonance) dans un signal de référence (bruit rose, voix, impulsion). Les signaux sont présentés au casque en écoute diotique (même signaux pour les deux oreilles). Ils demandent au sujet de juger de la dissemblance entre le signal de référence et le signal filtré.

Olive *et al.* (Olive *et al.*, 1997) montrent qu'en fonction du signal test, les seuils de détection varient. Tous signaux confondus, le seuil de détection des résonances ou des anti-résonances se situe autour de $\pm 1dB$ ce qui correspond à la valeur communément admise. On accepte généralement une variation autour de la valeur moyenne de l'ordre $\pm 2dB$ sans coloration excessive.

Contrairement à ce qui est couramment admis, Olive *et al.* montrent que les anti-résonances sont détectées de manière équivalente aux résonances pour des signaux de type impulsif. En effet, la présence d'une résonance ou une anti-résonance se traduit par une modification de l'allure temporelle du signal qui est plus facilement détectée sur des signaux de type impulsif. Sur ce type de signaux, les anti-résonances sont d'autant plus détectées que le facteur de qualité est important.

2.1.2 Temps de groupe

Les études portant sur la détection des variations du temps de groupe sont relativement anciennes (entre 1982 et 1985) (Preis, 1982) (Preis & Bloom, 1984) (Deer *et al.*, 1985). Elles se basent sur l'utilisation de filtres passe-tout analogiques simples. Ceux-ci permettent de manipuler la phase du signal indépendamment du niveau de la réponse fréquentielle. De la même manière que pour les filtres résonants, on peut définir un facteur de qualité (ratio de la fréquence centrale du filtre sur la largeur de la bande de fréquence dans laquelle le temps de groupe additionnel est supérieur à la moitié du temps de groupe additionnel maximal). Ce type de filtre n'autorise malheureusement pas un réglage indépendant du temps de groupe et du facteur de qualité. A un facteur de qualité élevé correspond un temps de groupe élevé et inversement. Ces études ne permettent donc pas de déterminer le seuil de détection du temps de groupe indépendamment du facteur de qualité (largeur de bande sur laquelle le temps de groupe est modifié).

Une étude récente, menée par Koya en 2000 (Koya, 2000), utilise le même type de structure mais

avec des signaux différents et ne résout pas les problèmes que l'on vient de citer.

Pour toutes ces études, les signaux de test sont présentés en écoute diotique sur casque ou sur un haut-parleur dans une pièce d'écoute semi-réverbérante.

A la lecture de leurs résultats, il semble que les variations de temps de groupe soient pas ou peu audibles. Il apparaît cependant que le paramètre important pour la détection de variation du temps de groupe soit le facteur de qualité. Plus la largeur de bande est importante (facteur de qualité faible), plus la variation de temps de groupe semble être perçue. La littérature n'apporte pas, à notre connaissance, de réponse définitive sur ce point.

Dans la suite, nous considérerons que seule des variations de l'ordre de 2 à 4 ms sur une large bande de fréquences (octave) sont susceptibles d'être audibles.

2.2 Rayonnement d'un banc linéaire continu infini pour la WFS

Dans cette partie, nous proposons d'estimer les caractéristiques de rayonnement du champ sonore émis par une ligne infinie avec les fonctions d'alimentation définies par la WFS. Une estimation dans le domaine fréquentiel donne la loi d'atténuation du champ sonore synthétisé. Un calcul dans le domaine temporel donne la forme du front d'onde synthétisé et son évolution lors de la propagation. Enfin nous proposons une définition de la fonction de qualité dans le cas de la WFS ainsi qu'un ensemble de simulations. Celles-ci permettent d'estimer l'audibilité des "défauts" mis en évidence dans ce cas.

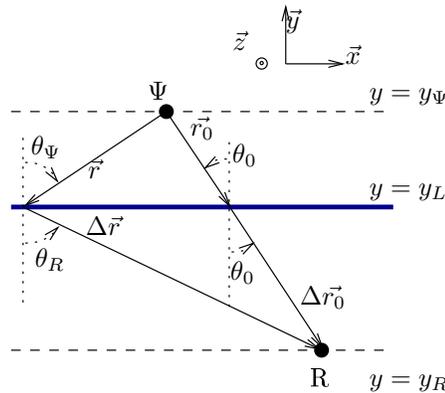


FIG. 2.1 – description du rayonnement d'un banc de haut-parleurs linéaire

On considère une distribution linéaire L de sources monopolaires idéales de taille infinie disposée le long de l'axe x ($y_L = 0, z_L = 0$). Dans le cadre de la WFS (cf. équations 1.34 et 1.35), le champ sonore associé à une source virtuelle Ψ_m est synthétisé en appliquant à la source secondaire située en x_L un gain $G_{\Psi_m}(x_L)$, un retard $\tau_{\Psi_m}(x_L)$ et un filtre global $f(t)$ dont la réponse fréquentielle

$F(k)$ s'exprime sous la forme :

$$F(k) = \begin{cases} \sqrt{\frac{k}{2\pi}} e^{j\frac{\pi}{4}} & \text{si } \Psi_m \in \Omega_R \\ \sqrt{\frac{k}{2\pi}} e^{-j\frac{\pi}{4}} & \text{si } \Psi_m \in \Omega_\Psi \end{cases}. \quad (2.3)$$

$F(k)$ est la transformée de Fourier de $f(t)$ et k le nombre d'onde. $\tau_{\Psi_m}(x_L)$ est donnée par :

$$\tau_{\Psi_m}(x_L) = \begin{cases} \frac{r_0}{c} & \text{if } \Psi_m \in \Omega_R \\ \frac{-r_0}{c} & \text{if } \Psi_m \in \Omega_\Psi \end{cases}, \quad (2.4)$$

et $G_{\Psi_m}(x_L)$,

$$G_{\Psi_m}(x_L) = \cos(\theta_{inc}) \sqrt{\frac{|y_L - y_{R_{av}}|}{|y_{R_{av}} - y_{\Psi_m}| r_0}} \Phi_m(\phi_0). \quad (2.5)$$

On rappelle que $y_{R_{av}}$ est la profondeur d'écoute moyenne pour laquelle le niveau de la source virtuelle synthétisée est correct. Le champ de pression émis par L lors de la synthèse de la source Ψ_m est donnée par :

$$p_{\Psi_m}(r_R, t) = s(t) * f(t) * \int_{-\infty}^{+\infty} G_{\Psi_m}(x_L) \frac{\delta(t - \frac{\Delta r}{c} - \tau_{\Psi_m}(x_L))}{4\pi\Delta r} dx_L, \quad (2.6)$$

où $s(t)$ est le signal d'alimentation de la source virtuelle Ψ_m .

On définit la fonction de transfert électroacoustique de la distribution de sources secondaires L pour la reproduction de la source virtuelle Ψ_m au point d'écoute R . On la note $h_{\Psi_m}^{wfs}(R, t)$ et sa transformée de Fourier $H_{\Psi_m}^{wfs}(R, k)$,

$$h_{\Psi_m}^{wfs}(r_R, t) = f(t) * \int_{-\infty}^{+\infty} G_{\Psi_m}(x_L) \frac{\delta(t - \frac{\Delta r}{c} - \tau_{\Psi_m}(x_L))}{4\pi\Delta r} dx_L, \quad (2.7)$$

$$H_{\Psi_m}^{wfs}(r_R, k) = F(k) \int_{-\infty}^{+\infty} G_{\Psi_m}(x_L) \frac{e^{-jk(\Delta r + c\tau_{\Psi_m}(x_L))}}{4\pi\Delta r} dx_L. \quad (2.8)$$

2.2.1 Domaine fréquentiel : calcul de la loi d'atténuation

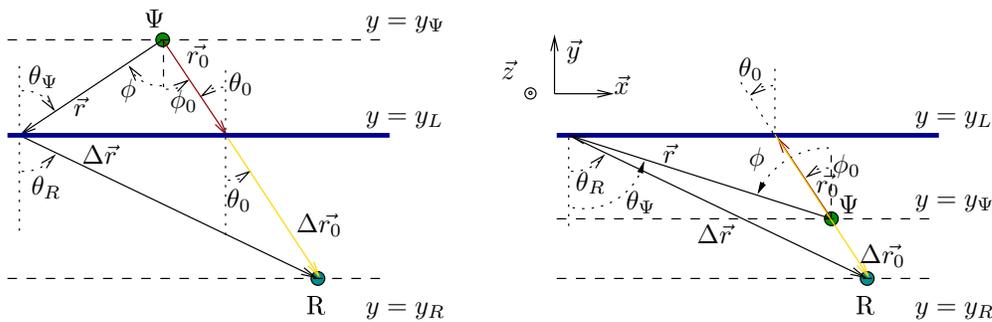


FIG. 2.2 – utilisation de la phase stationnaire pour un banc linéaire infini

L'équation 2.8 peut se mettre sous la forme,

$$I = \int_{-\infty}^{+\infty} A(x) e^{-jk\chi(x)} dx, \quad (2.9)$$

en définissant une fonction d'enveloppe $A(x)$

$$A(x) = F(k) \frac{G_{\Psi_m}(x_L)}{4\pi\Delta r}, \quad (2.10)$$

et une fonction de phase

$$\chi(x) = (\Delta r + c\tau_{\Psi_m}(x_L)), \quad (2.11)$$

ce qui donne

$$\chi'(x) = (\sin\theta_R - \sin\theta_{\Psi_m}), \quad (2.12)$$

et

$$\chi''(x) = \left(\frac{\cos^2\theta_R}{\Delta r} + \frac{\cos^2\theta_{\Psi_m}}{c\Delta t_{\Psi_m}(x_L)} \right). \quad (2.13)$$

On appelle point de phase stationnaire pour la fonction de phase $\chi(x)$ tout extremum de $\chi(x)$. Si x_0 est un point de phase stationnaire pour $\chi(x)$ il satisfait,

$$\begin{cases} \chi'(x_0) = 0 \\ \chi''(x_0) \neq 0 \end{cases}. \quad (2.14)$$

Dans notre cas, ce point est unique et donné par,

$$\theta_R = \theta_{\Psi_m} = \theta_0. \quad (2.15)$$

On a alors,

$$\chi''(x_0) = \begin{cases} \cos^2\theta_0 \left(\frac{1}{r_0} - \frac{1}{\Delta r_0} \right) & \text{if } \Psi_m \in \Omega_R \\ \cos^2\theta_0 \left(\frac{1}{r_0} + \frac{1}{\Delta r_0} \right) & \text{if } \Psi_m \in \Omega_{\Psi} \end{cases}. \quad (2.16)$$

Les conditions 2.14 sont toujours respectées si $\Psi_m \in \Omega_{\Psi}$ sauf pour une source *et* une position d'écoute lointaines ($\Delta r_0 \rightarrow +\infty$ et $r_0 \rightarrow +\infty$). Si $\Psi_m \in \Omega_R$ (source focalisée), cette condition est toujours remplie sauf si le point d'écoute coïncide avec la position de Ψ_m (point de focalisation des ondes sonores).

Dans le cas où Ψ_m est dans Ω_{Ψ} , l'approximation de la phase stationnaire appliquée à l'équation 2.8 donne,

$$H_{\Psi_m}(r_R, k) = F(k) \frac{G_{\Psi_m}(x_0)}{\cos\theta_0} \sqrt{\frac{2\pi}{jk}} \sqrt{\frac{r_0}{\Delta r_0}} \frac{e^{-j(k(\Delta r_0 + r_0))}}{4\pi\sqrt{\Delta r_0 + r_0}}. \quad (2.17)$$

En intégrant les expressions respectives de $F(k)$ et $G_{\Psi_m}(x_0)$ et en considérant $y_L = 0$, on obtient

$$H_{\Psi_m}(r_R, k) = \sqrt{\frac{|y_{R_{av}}|}{|y_{R_{av}} - y_{\Psi_m}| \Delta r_0}} \Phi_m(\phi_0) \frac{e^{-j(k(\Delta r_0 + r_0))}}{4\pi\sqrt{\Delta r_0 + r_0}}. \quad (2.18)$$

Après quelques manipulations,

$$H_{\Psi_m}(r_R, k) = \sqrt{\frac{|y_{R_{av}}|}{|y_R|}} \Phi_m(\phi_0) \frac{e^{-j(k(\Delta r_0 + r_0))}}{4\pi\sqrt{\Delta r_0 + r_0}\sqrt{\cos\theta_0|y_{R_{av}}| + r_0}}, \quad (2.19)$$

ou encore,

$$H_{\Psi_m}(r_R, k) = \sqrt{\frac{|y_{R_{av}}|}{|y_R|}} \sqrt{\frac{|y_R - y_{\Psi_m}|}{|y_{R_{av}} - y_{\Psi_m}|}} \Phi_m(\phi_0) \frac{e^{-j(k(\Delta r_0 + r_0))}}{4\pi(\Delta r_0 + r_0)}. \quad (2.20)$$

Si Ψ_m est dans Ω_R (source focalisée), on obtient de la même manière,

$$H_{\Psi_m}(r_R, k) = \sqrt{\frac{|y_{R_{av}}|}{|y_R|}} \Phi_m(\phi_0) \frac{e^{-j(k(\Delta r_0 - r_0))}}{4\pi\sqrt{\Delta r_0 - r_0}\sqrt{\cos\theta_0|y_{R_{av}}| - r_0}}, \quad (2.21)$$

ou encore,

$$H_{\Psi_m}(r_R, k) = \sqrt{\frac{|y_{R_{av}}|}{|y_R|}} \sqrt{\frac{|y_R - y_{\Psi_m}|}{|y_{R_{av}} - y_{\Psi_m}|}} \Phi_m(\phi_0) \frac{e^{-j(k(\Delta r_0 - r_0))}}{4\pi\sqrt{\Delta r_0 - r_0}}. \quad (2.22)$$

On définit alors la loi d'atténuation $Att_{\Psi_m}(r_R)$, pour une source Ψ_m située dans Ω_R

$$Att_{\Psi_m}(r_R) = \sqrt{\frac{d_{R_{av}}}{d_R}} \sqrt{\frac{d_R + d_{\Psi_m}}{d_{R_{av}} + d_{\Psi_m}}} \frac{1}{4\pi(\Delta r_0 + r_0)}. \quad (2.23)$$

Pour une source Ψ_m (focalisée) située dans Ω_{Ψ} , on aura

$$Att_{\Psi}(r_R) = \sqrt{\frac{d_{R_{av}}}{d_R}} \sqrt{\frac{d_R - d_{\Psi_m}}{d_{R_{av}} - d_{\Psi_m}}} \frac{1}{4\pi(\Delta r_0 - r_0)}, \quad (2.24)$$

où, d_R est la distance du point R à la ligne axe du banc de haut-parleurs et, $d_{R_{av}}$ et d_{Ψ} sont définies de la même manière. Ces notations permettent de généraliser les résultats au cas où le banc de haut-parleurs n'est pas dirigé suivant l'axe x .

On retrouve alors que le niveau de la source n'est bien reproduit que sur la ligne $d_R = d_{R_{av}}$. La loi d'atténuation que l'on obtient correspond à la combinaison de la loi d'atténuation typique de la source linéaire constituée par le banc de haut-parleurs (terme en $\frac{1}{\sqrt{d_R}}$) et de celle de la source ponctuelle Ψ_m (terme en $\frac{1}{\Delta r_0 \pm r_0}$) contrebalancée par un terme dépendant de la source (terme en $\sqrt{\frac{d_R + d_{\Psi_m}}{d_{R_{av}} + d_{\Psi_m}}}$). Ce dernier terme annule celui lié à l'atténuation de la ligne lorsque la source Ψ_m est proche du banc de haut-parleurs et devient négligeable lorsque la source s'éloigne quand d_R est de l'ordre de $d_{R_{av}}$.

Ceci n'est valable que si on peut appliquer l'approximation de la phase stationnaire. On renvoie le lecteur à l'annexe A.1 pour plus de détails sur l'examen de la validité cette approximation dans le cadre de la WFS.

Dans le cas où la source est une "onde plane" se propageant perpendiculairement au banc de haut-parleurs (onde plane à 0 degré, Ψ_{pw0}), la fonction d'alimentation des haut-parleurs devient constante et tous les haut-parleurs sont alimentés en phase.

Le champ de pression émis par cette source linéaire infinie ("line source" en anglais) n'est autre que la fonction de Green de l'espace en deux dimensions (Bruneau, 1998). On a alors

$$H_{\Psi_{pw0}}(r_R, k) = -\frac{j}{4}H_0(kd_R), \quad (2.25)$$

où H_0 est la fonction de Hankel cylindrique de degré 0. En champ lointain ($kd_R \gg 1$), celle-ci peut s'approcher par

$$H_{\Psi_{pw0}}(r_R, k) \simeq \frac{1}{\sqrt{8\pi j k d_R}} e^{-jkd_R}. \quad (2.26)$$

On retrouve ainsi la dépendance en $\sqrt{jk d_R}$ suggérée par l'approximation de la phase stationnaire. Afin de vérifier la qualité de cette approximation au cours de la propagation, on définit le rapport

$$\tilde{H}_{\Psi_{pw0}}(r_R, k) = H_{\Psi_{pw0}}(r_R, k) \times \sqrt{8\pi j k d_R}. \quad (2.27)$$

La fonction $\tilde{H}_{\Psi_{pw0}}(r_R, k)$ rendra donc compte de la validité de l'approximation. Lorsqu'elle est proche de 1, l'approximation est correcte ; lorsqu'elle s'en éloigne, celle-ci n'est plus justifiée.

La figure 2.3 représente le module de $\tilde{H}_{\Psi_{pw0}}(r_R, k)$ entre 20 et 20000 Hz pour des distances d_R entre 1 cm et 1000 m.

L'approximation reste valable dans une partie de l'espace de reproduction. On note tout de même une chute du niveau de la réponse fréquentielle à proximité de la ligne, principalement aux basses fréquences (entre 20 et 100 Hz à 10 cm). Cette chute est de l'ordre de -10 dB pour les positions les plus proches de la ligne. Ce comportement, remarquable sur la figure, est très probablement inaudible pour la plupart des positions d'écoute.

On verra dans la partie 2.2.3 que l'on observe le même type de phénomène pour les autres sources. La propagation du champ synthétisé est ainsi faiblement "dispersive". Ce comportement est caractéristique du rayonnement des sources acoustiques linéaires dans l'espace à trois dimensions (Nicol & Depollier, 1998) (Caulkins *et al.*, 2004).

Le calcul de la fonction de transfert électroacoustique dans le domaine temporel que nous proposons dans la partie suivante nous permet de mettre en évidence la variation de la forme temporelle du front d'onde synthétisé. Ceci confirme l'existence de ce phénomène de dispersion pour la synthèse de n'importe quelle source virtuelle.

2.2.2 Domaine temporel : caractérisation de la variation de la forme d'onde synthétisée

On définit L_{Ψ_m} , la distribution de sources secondaires L auxquelles on applique la fonction d'alimentation données par la WFS pour la synthèse de la source Ψ_m . Le front d'onde synthétisé est donné par la somme de l'ensemble des contributions des sources secondaires de la ligne L_{Ψ_m} .

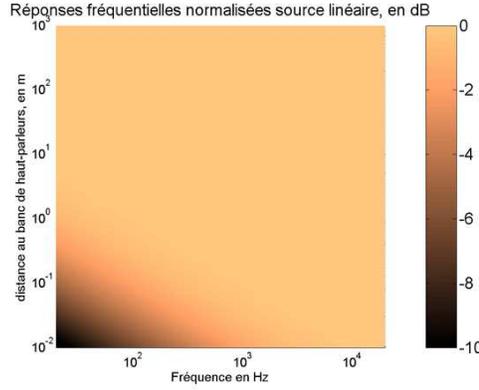


FIG. 2.3 – Réponses fréquentielles normalisées d'une source linéaire à différentes distances de la ligne

Le temps d'arrivée $\Delta t_{\Psi_m}(r_R, x_L)$ en un point d'écoute R de Ω_R , de la contribution émise par un haut-parleur de la ligne L situé en x_L , et retardé de $\Delta t_{\Psi_m}(x_L)$ est lié à la fonction de phase $\chi(x)$ (cf. équation 2.11) par,

$$\Delta t_{\Psi_m}(r_R, x) = \frac{\chi(x)}{c}. \quad (2.28)$$

$\Delta t_{\Psi_m}(r_R, x_L)$ a un extremum unique en $x_L = x_0$. Il s'agit d'un minimum si Ψ_m est dans Ω_Ψ et d'un maximum si Ψ_m est dans Ω_R (source focalisée).

Considérons tout d'abord le cas où Ψ_m est dans Ω_Ψ . La fonction $\chi'(x)$ ne s'annulant qu'en $x_L = x_0$, pour $x_L < x_0$ (resp. for $x_L > x_0$), Δt_{Ψ_m} est une fonction strictement croissante (resp. une fonction strictement décroissante). La réponse impulsionnelle de L_{Ψ_m} correspond alors à la somme de la réponse impulsionnelle $h_{\Psi_m, R}^-$ de $L_{\Psi_m, R}^-$ (définie par $x_L^- < x_0$) et de la réponse impulsionnelle $h_{\Psi_m, R}^+$ de $L_{\Psi_m, R}^+$ (définie par $x_L^+ \geq x_0$). Ces deux réponses impulsionnelles ont des valeurs nulles jusqu'à $t_0 = \Delta t_{\Psi_m}(r_R, x_0)$. A partir de ce temps initial, chacune des réponses impulsionnelles se construit à partir des contributions successives des sources secondaires de $L_{\Psi_m, R}^-$ et $L_{\Psi_m, R}^+$. Celles-ci sont données par les sources secondaires situées en x_L^+ et x_L^- en "s'éloignant" de x_0 .

$h_{\Psi_m, R}^-$ et $h_{\Psi_m, R}^+$ s'expriment sous la forme,

$$\begin{aligned} h_{\Psi_m, R}^-(t) &= \int_{-\infty}^{x_0} G_{\Psi_m}(x_L) \frac{\delta(t - \Delta t_{\Psi_m}(r_R, x_L))}{4\pi\Delta r(x_L)} dx_L, \\ h_{\Psi_m, R}^+(t) &= \int_{x_0}^{+\infty} G_{\Psi_m}(x_L) \frac{\delta(t - \Delta t_{\Psi_m}(r_R, x_L))}{4\pi\Delta r(x_L)} dx_L. \end{aligned} \quad (2.29)$$

On utilise un changement de variable afin de simplifier ces expressions,

$$\begin{aligned} h_{\Psi_m, R}^-(t) &= - \int_{t_0}^{+\infty} G_{\Psi_m}(x^-(t^-)) \frac{\delta(t - t^-)}{4\pi\Delta r(x^-(t^-))} \frac{dx^-(t^-)}{dt^-} dt^-, \\ h_{\Psi_m, R}^+(t) &= \int_{t_0}^{+\infty} G_{\Psi_m}(x^+(t^+)) \frac{\delta(t - t^+)}{4\pi\Delta r(x^+(t^+))} \frac{dx^+(t^+)}{dt^+} dt^+, \end{aligned} \quad (2.30)$$

où, $t^-(x_L)$ et $t^+(x_L)$ sont donnés par

$$\begin{aligned} t^-(x_L) &= \Delta t_{\Psi_m}(r_R, x_L) \quad \forall x_L \in]-\infty, x_0], \\ t^+(x_L) &= \Delta t_{\Psi_m}(r_R, x_L) \quad \forall x_L \in]x_0, +\infty[, \end{aligned} \quad (2.31)$$

et, $x^-(t^-(x_L))$ et $x^+(t^+(x_L))$,

$$\begin{aligned} x^-(t^-(x_L)) &= x_L \quad \forall x_L \in]-\infty, x_0], \\ x^+(t^+(x_L)) &= x_L \quad \forall x_L \in]x_0, +\infty[. \end{aligned} \quad (2.32)$$

Ce changement de variable est possible car $t^-(x_L)$ et $t^+(x_L)$ sont des fonctions strictement monotones et donc bijectives. La propriété fondamentale de la distribution de Dirac permet de simplifier les expressions,

$$\int_{t_a}^{t_b} \delta(t - t_c) f(t) dt = \begin{cases} f(t_c) & \text{si } t_c \in [t_a, t_b] \\ 0 & \text{sinon.} \end{cases} \quad (2.33)$$

On obtient alors

$$\begin{aligned} h_{\Psi_m, R}^-(t) &= -Y(t - t_0) \frac{G_{\Psi_m}(x^-(t))}{4\pi\Delta r(x^-(t))} \frac{dx^-(t)}{dt}, \\ h_{\Psi_m, R}^+(t) &= Y(t - t_0) \frac{G_{\Psi_m}(x^+(t))}{4\pi\Delta r(x^+(t))} \frac{dx^+(t)}{dt}, \end{aligned} \quad (2.34)$$

où, $Y(t)$ est la fonction de Heaviside.

Les réponses impulsionnelles $h_{\Psi_m, R}^-(t)$ et $h_{\Psi_m, R}^+(t)$ se construisent ainsi en considérant le niveau de la contribution de la source secondaire située en $x^-(t)$ (resp. $x^+(t)$) pondéré par la densité de sources secondaires dont la contribution est reçue au temps t en R .

La réponse impulsionnelle complète h_{Ψ_m} de L_{Ψ_m} est alors

$$h_{\Psi_m}(r_R, t) = Y(t - t_0) \left(\frac{G_{\Psi_m}(x^+(t))}{4\pi\Delta r(x^+(t))} \frac{dx^+(t)}{dt} - \frac{G_{\Psi_m}(x^-(t))}{4\pi\Delta r(x^-(t))} \frac{dx^-(t)}{dt} \right). \quad (2.35)$$

On peut ainsi obtenir une expression exacte du champ de pression rayonné dans l'espace par la ligne à condition d'exprimer les fonctions x^- et x^+ . Il s'agit des inverses de la fonction Δt_{Ψ_m} pour chaque demi-distribution de sources secondaires $L_{\Psi_m, R}^-$ et $L_{\Psi_m, R}^+$, ce qui revient à déterminer la position x^- (resp. x^+) de la source secondaire dont la contribution est reçue à l'instant t en R . Ceci équivaut à trouver les coordonnées des points M ($x_M(t)$, $y_M(t)$) solutions du système d'équation suivant

$$\begin{cases} \sqrt{(x_{\Psi_m} - x_M)^2 + (y_{\Psi_m} - y_M)^2} + \sqrt{(x_R - x_M)^2 + (y_R - y_M)^2} = \frac{t}{c} \\ y_M = 0 \end{cases} \quad (2.36)$$

Cette formulation du problème fait apparaître une interprétation géométrique simple. La pre-

mière équation correspond à l'équation d'une ellipse dont les foyers sont Ψ_m et R et de grand axe $\frac{t}{c}$. Les points M sont donc donnés par l'intersection de l'ellipse et de la ligne L , support des sources secondaires (cf. figure 2.4). Ce raisonnement est équivalent à celui donné par Sommerfeld (Sommerfeld, 1964) pour le calcul du rayonnement des ondes lumineuses à travers une ouverture mais nos objectifs sont différents. Les approximations de champ lointain utilisées par Sommerfeld ne seraient valables qu'en très hautes fréquences dans le domaine acoustique et fausseraient complètement la description des effets de dispersion que l'on cherche à mettre en évidence.

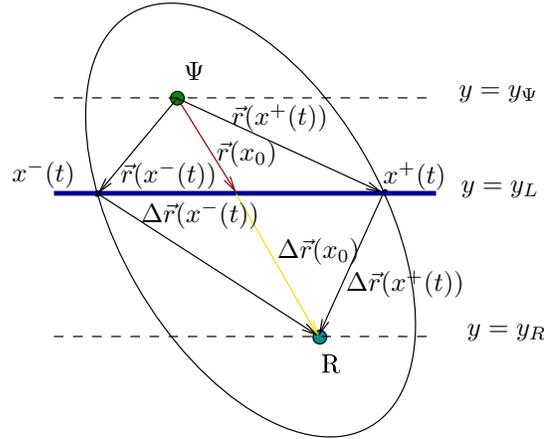


FIG. 2.4 – Contributions des demi-distributions $L_{\Psi_m, R}^-$ et $L_{\Psi_m, R}^+$ à un instant t

Il est difficile de donner une expression générale pour x^- et x^+ qui soit facilement interprétable. Toutefois, dans le cas où la distribution de retards à appliquer à la ligne est uniformément nulle (cas de "l'onde plane" se propageant perpendiculairement à la ligne L), Bruneau (Bruneau, 1998) donne une expression exacte

$$h_{PW0}(r_R, t) = \frac{1}{2\pi\sqrt{(c\zeta)^2 + 2|y_R|\tau}} Y(\tau), \quad (2.37)$$

où, $\zeta = t - t_0$. Le champ synthétisé présente alors une dépendance par rapport à la position d'écoute R ; pas seulement comme facteur d'atténuation global mais aussi vis-à-vis de la forme temporelle synthétisée.

Afin d'évaluer l'évolution lors de la propagation de la forme temporelle du front d'onde synthétisé, nous proposons de considérer l'inverse normalisée en temps et en niveau des réponses impulsionnelles $h_{\Psi_m, R}^-(t)$ et $h_{\Psi_m, R}^+(t)$.

$$\begin{aligned} \widetilde{h_{\Psi_m}^{inv,+}}(r_R, \zeta) &= Y(\zeta) Att_{\Psi_m}(r_R) \frac{4\pi\Delta r(x^+(\zeta + t_0))}{G_{\Psi_m}(x^+(\zeta + t_0))} \frac{dt(x^+(\zeta + t_0))}{dx} \\ \widetilde{h_{\Psi_m}^{inv,-}}(r_R, \zeta) &= -Y(\zeta) Att_{\Psi_m}(r_R) \frac{4\pi\Delta r(x^-(\zeta + t_0))}{G_{\Psi_m}(x^-(\zeta + t_0))} \frac{dt(x^-(\zeta + t_0))}{dx}, \end{aligned} \quad (2.38)$$

où, $Att_{\Psi_m}(r_R)$ correspond à la loi d'atténuation de la source Ψ_m défini par les équations 2.23 et 2.24.

On utilise les inverses normalisées car, par définition, $\frac{dt(x_0)}{dx} = 0$ ce qui donne $\frac{dx^+(t_0)}{dt} = \frac{dx^-(t_0)}{dt} = +\infty$. En effet, la fonction $t(x)$ est proportionnelle à la fonction de phase χ et x_0 est défini tel que $\chi'(x_0) = 0$.

Afin de quantifier l'évolution lors de la propagation de la forme des fronts d'onde synthétisés et mettre en évidence le phénomène de dispersion, on simule la valeur de $\widetilde{h_{\Psi_m}^{inv,+}}(r_R, \zeta)$ et $\widetilde{h_{\Psi_m}^{inv,-}}(r_R, \zeta)$ pour les points d'écoute de la figure 2.5. Ces simulations sont réalisées pour des sources virtuelles tests (source omnidirectionnelle focalisée en $(0, 1)$, source omnidirectionnelle en $(0, 3)$, source omnidirectionnelle en $(3, 5)$, onde plane à 0 degré.

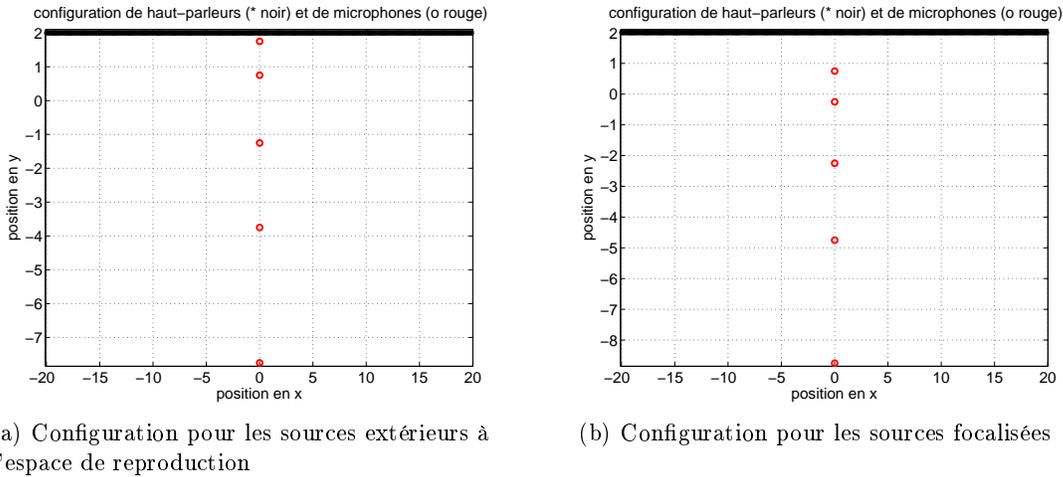


FIG. 2.5 – Configurations de tests pour l'estimation des erreurs introduites par l'approximation de la phase stationnaire.

L'ensemble des figures 2.6 représente les 10 premières millisecondes des inverses normalisées des réponses impulsionnelles pour les 4 sources virtuelles mentionnées. On utilise un banc linéaire "continu" (espacement des sources secondaires de 1 cm) de 40 m de long. Pour chacune de ces figures, le paramètre de variation correspond à la coordonnée y de la position d'écoute. On remarque que pour l'ensemble des figures, l'inverse normalisée tend vers 0 lorsque ζ tend vers 0 avec une pente relativement importante. Ceci signifie que la réponse impulsionnelle réelle est condensée autour de $\zeta = 0$. La plupart de l'énergie est ainsi émise sur ce premier pic. Les variations des contributions tardives sont plus lentes. Ce comportement traduit une émission plus tardive des basses fréquences.

Le lecteur notera que l'on s'attache principalement à décrire l'évolution de la courbe de l'inverse normalisée en fonction de la position d'écoute. Cette évolution traduit ainsi la modification de l'allure du front d'onde synthétisé entre les différentes positions d'écoute pour une source virtuelle donnée.

Pour l'ensemble des sources, l'inverse normalisée admet des valeurs plus faibles lorsque la distance du point d'écoute au banc de haut-parleurs augmente. Ceci correspond à un *étalement* plus prononcé de la réponse impulsionnelle. D'après ce que l'on vient de dire, cela signifie que l'émission

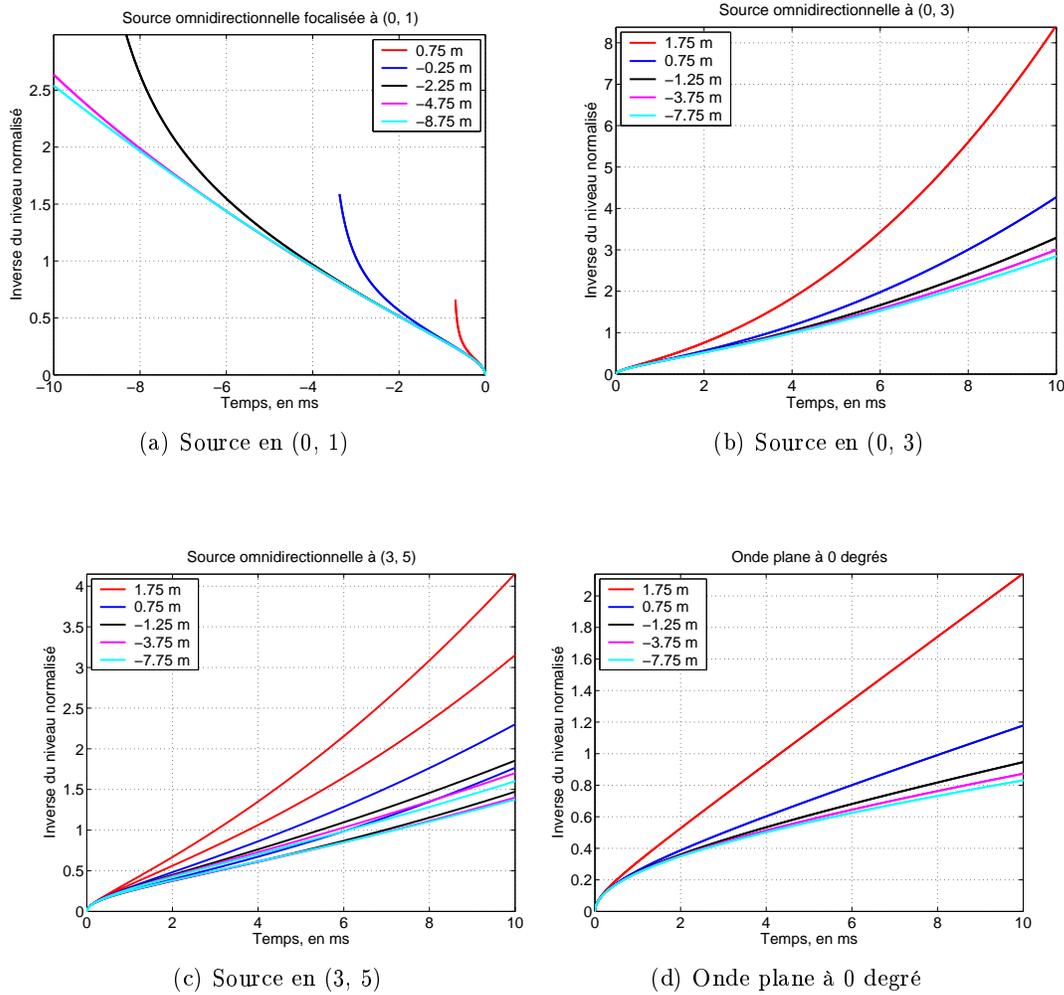


FIG. 2.6 – Évaluation de l'évolution de la forme temporelle du front d'onde synthétisé par l'inverse normalisée des réponses impulsionnelles $h_{\Psi_m, R}^-(t)$ et $h_{\Psi_m, R}^+(t)$ lors de la propagation. Paramètre de variation : coordonnée y de la position d'écoute.

de la partie basses fréquences est plus tardive dans le champ lointain que dans le champ proche du banc de haut-parleurs

La pente de l'inverse normalisée dépend de la source. Ceci s'apparente à un phénomène de dispersion lors de la propagation, l'ensemble des fréquences ne se propageant pas à la même vitesse et ne s'atténuant pas de la même façon.

Ainsi, le terme correctif appliqué en $\sqrt{\pm jk}$ ne peut corriger idéalement les caractéristiques du champ synthétisé par la distribution linéaire de haut-parleurs sur toute la zone d'écoute. Notons toutefois que les calculs présentés sont effectués dans le domaine analogique et consignent toutes les fréquences. Reste à savoir si les défauts qui subsistent sont audibles dans la bande de fréquence qui nous intéresse (20-20000 Hz). Ce point nécessite de réaliser des simulations dans le domaine fréquentiel pour un banc de haut-parleurs de taille "infinie" que nous effectuons dans la partie

suivante.

Pour la source focalisée, l'ensemble des contributions sont condensées autour de 0 pour la position d'écoute proche du point de focalisation ce qui correspond à une densité élevée de contributions (cf. figure 2.6(a)). Le niveau synthétisé en ce point est donc important. Ceci signifie que pratiquement *tous* les haut-parleurs apparaissent en phase à cette position. Le signal synthétisé est ainsi proche d'une distribution de Dirac et ne nécessite pas de correction supplémentaire.

Toutefois, pour les positions plus lointaines, l'inverse normalisée prend un comportement symétrique en comparaison avec les sources non focalisées. Ses valeurs sont de plus en plus faibles lorsque la distance au banc de haut-parleurs augmente. Ceci traduit ainsi un étalement du signal synthétisé qui s'accroît. Ce comportement est partiellement corrigé par le filtrage en $\sqrt{-jk}$, qui par contre, introduit des incohérences au niveau du point de focalisation.

Dans ce cas, l'effet de dispersion du champ synthétisé est maximal. Il ne requiert en effet *aucune* correction au point de focalisation. Lors de la propagation, le front d'onde synthétisé s'étale et doit être corrigé, par exemple, par le filtrage en $\sqrt{-jk}$.

2.2.3 Simulations : évolution spatiale du champ acoustique synthétisé

Dans cette partie, on s'attache à mettre en évidence la "qualité" du champ acoustique synthétisé par un banc de haut-parleurs de taille "infinie" dans l'ensemble de la zone d'écoute. Cette qualité est définie par rapport à une source WFS *idéale* caractérisée par

- une réponse fréquentielle "plate" pour n'importe quel point de l'espace de restitution dont le niveau est donné par l'équation 2.23 ou 2.24 en fonction de son emplacement ;
- une réponse en phase linéaire correspondant à un retard pur dont la valeur est donnée par la distance entre le point de réception R et la position de la source.

La réponse fréquentielle d'une telle source en R est alors donnée par,

$$H_{\Psi_m}^{id}(r(R), k) = \Phi_m(\phi) G_{\Psi}^{im}(r(nmic)) e^{-jk(\Delta r_0 \pm r_0)} \quad (2.39)$$

Dans cette partie, on choisit de ne considérer que des sources virtuelles omnidirectionnelles, ainsi on a $\Phi_{m=0}(\phi) = 1$ et on écrit uniquement Ψ afin de simplifier les écritures.

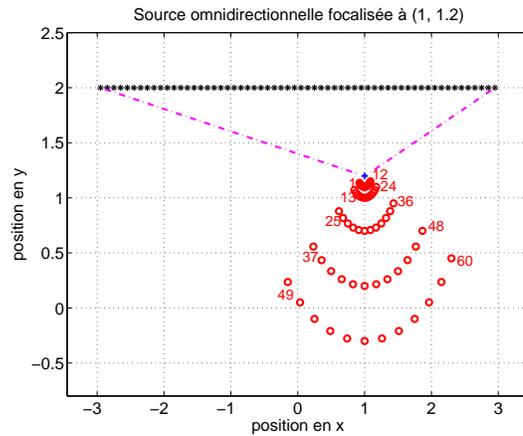
La fonction de qualité acoustique définie dans la partie 2.1 s'écrit pour la WFS en un point R de l'espace de restitution

$$Q_{\Psi}^{wfs}(r(R), k) = \frac{H_{\Psi}(r(R), k)}{H_{\Psi}^{id}(r(R), k)}. \quad (2.40)$$

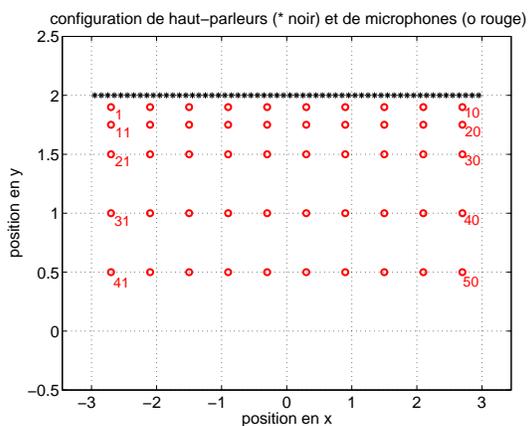
On examine alors le module de $Q_{\Psi}^{wfs}(r(R), k)$ et le temps de groupe relatif $TG_{\Psi}^{wfs}(r(R), k)$:

$$TG_{\Psi}^{wfs}(r(R), k) = -\frac{\partial \arg Q_{\Psi}^{wfs}(r(R), k)}{c \partial k}. \quad (2.41)$$

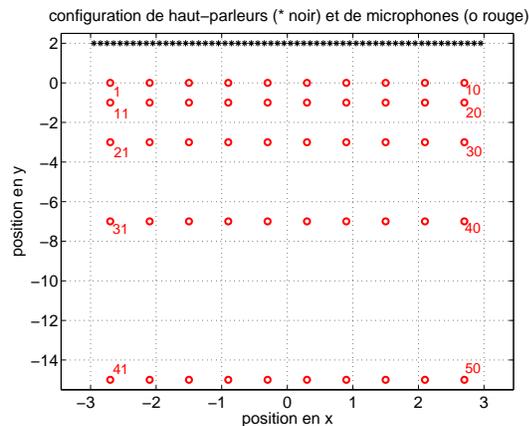
Celui-ci traduit le temps relatif d'arrivée de l'énergie aux différentes fréquences.



(a) configuration de proximité de la source focalisée



(b) configuration microphones proches



(c) configuration microphones lointains

FIG. 2.7 – Configurations de test avec 600 haut-parleurs (60 m, 10 cm d'espacement), seulement les 60 du centre sont représentés (* noirs); différentes positions de microphones (o rouges), proximité de la source focalisée, proches, lointaines.

Afin de tester la transparence du champ acoustique synthétisé sur une grande zone d'écoute pour la reproduction de sources virtuelles par un banc de haut-parleurs linéaire continu de taille infinie, on évalue la réponse fréquentielle d'un banc linéaire de 600 haut-parleurs (60 m, 10 cm d'espacement, $y_L = 2$ m), sous la fréquence d'aliasing spatial, lors de la reproduction d'un ensemble de sources virtuelles test sur un ensemble de microphones situés dans l'espace de restitution.

Les figures 2.7 représentent les différentes configurations de test pour lesquelles on ne représente que les 60 haut-parleurs du centre du dispositif :

- une configuration composée de 5 bancs de 10 microphones omnidirectionnels espacés de 60 cm situés en $y = 1.9; 1.75; 1.5; 1; 0.5$ appelée configuration de microphones *proches* (cf. figure 2.7(b)) ;
- une configuration composée de 5 bancs de 10 microphones omnidirectionnels espacés de 60

cm situées en $y = 0; -1; -3; -7; -15$ appelée configuration de microphones *lointains* (cf. figure 2.7(c));

- une configuration spécifique concernant la proximité de la source virtuelle focalisée à (1, 1.2) de 5 bancs de 12 microphones omnidirectionnels situés sur une portion de cercle centré sur la source de rayon $d = 0.1$ m; 0.25 m; 0.5 m; 1 m; 1.5 m appelée configuration de *proximité* de la source focalisée (cf. figure 2.7(a)).

On considère trois sources virtuelles de test :

- une source virtuelle omnidirectionnelle focalisée située en (1, 1.2);
- une source virtuelle omnidirectionnelle située en (3, 3), en face de la limite droite des bancs linéaires de microphones;
- une source virtuelle omnidirectionnelle en (0, 7), distante du banc de haut-parleurs.

Il ne s'agit certes pas d'un banc de haut-parleurs de taille infinie mais suffisamment grand pour pouvoir négliger l'effet de la diffraction sur les sources considérées. Les fonctions d'alimentation ont un niveau très faible et les distances de propagation sont très grandes ce qui rend le niveau des contributions des haut-parleurs manquants négligeable. On ne peut toutefois garantir la validité des résultats observés en basses fréquences (en dessous de $80 - 100$ Hz) pour les distances importantes au banc de haut-parleur.

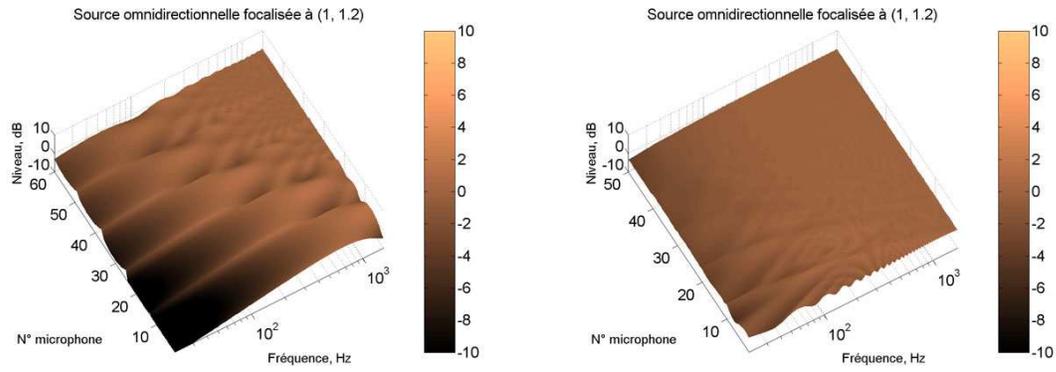
L'ensemble des figures 2.8 représente le module de $Q_{\Psi}^{wfs}(r(R), k)$. L'ensemble des figures 2.9 représente le temps de groupe relatif.

Ces simulations permettent, tout d'abord, de valider la loi d'atténuation des sources virtuelles, synthétisées par un banc de haut-parleurs linéaire de taille infinie (cf. équations 2.23 et 2.24). De manière plus précise, on voit qu'en dehors de la proximité immédiate du banc de haut-parleurs ou des sources, le module de la fonction de qualité reste en dessous des seuils d'audition (± 2 dB autour de la moyenne) sur toute la bande de fréquence considérée (20-20000 Hz). Pour les deux sources situées derrière le banc de haut-parleurs, on remarque que le niveau des basses fréquences a tendance à diminuer ($\simeq -5$ dB à 20 Hz à 10 cm du banc). On note aussi un fléchissement du temps de groupe traduisant une avance des basses fréquences comparativement aux hautes fréquences. Néanmoins ces variations restent largement en dessous des seuils de perception. Cet effet disparaît assez rapidement avec la distance au banc, ou devient non significatif, et reste limité aux très basses fréquences.

En ce qui concerne la source focalisée, on se rend compte que la réponse fréquentielle en grande proximité prend une pente importante avec une chute de plus de 10 dB à 20 Hz. En effet, au point de focalisation (position de la source), les contributions de l'ensemble des haut-parleurs sont cohérents et on retrouve ainsi la courbe du filtre en \sqrt{k} , soit +3 dB par octave. On a remarqué que pour cette source, les conditions pour l'approximation de la phase stationnaire ne sont pas respectées (équation 2.14). Le temps de groupe, étant proportionnel à la dérivée de la phase en fonction de la fréquence, tend vers une valeur nulle. Pour cette position, la phase est en effet constante conformément au facteur $e^{j\frac{\pi}{4}}$ appliqué à chacun des haut-parleurs.

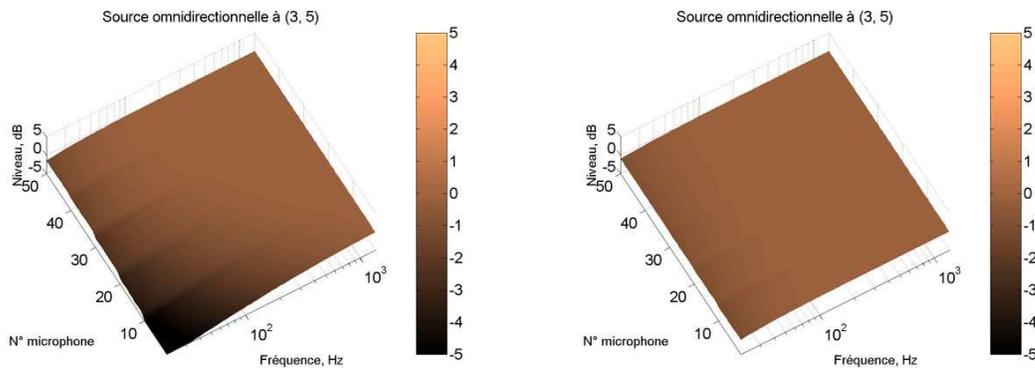
Lorsque le point d'écoute s'éloigne, on remarque que le module de la réponse fréquentielle devient progressivement plat en restant légèrement plus faible ($\simeq -2$ dB à 20 Hz) aux basses fréquences.

Le temps de groupe, quant à lui, présente un profil plus agité avec un léger retard relatif aux basses fréquences ($\simeq 2$ ms à 20 Hz) qui disparaît lorsque la fréquence et la distance à la source augmentent. Ce retard reste ainsi en dessous des seuils d'audition.



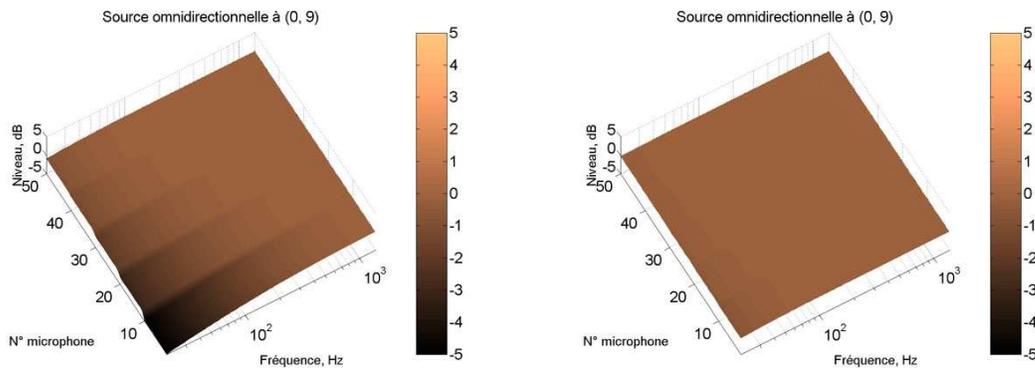
(a) Source omnidirectionnelle focalisée à (1, 1.2), échelle de niveau entre -10 et +10 dB, configuration de *proximité* de la source focalisée

(b) Source omnidirectionnelle focalisée à (1, 1.2), échelle de niveau entre -10 et +10 dB, microphones *lointains*



(c) Source omnidirectionnelle à (3, 5), microphones *proches*

(d) Source omnidirectionnelle à (3, 5), microphones *lointains*



(e) Source omnidirectionnelle à (0, 9), microphones *proches*

(f) Source omnidirectionnelle à (0, 9), microphones *lointains*

FIG. 2.8 – Module de $Q_{\Psi}^{wfs}(r(R), k)$ pour un banc de 600 haut-parleurs (60 m, 10 cm d'espacement) sous la fréquence d'aliasing spatial (20-2000 Hz) synthétisant différentes sources virtuelles omnidirectionnelles avec la WFS.

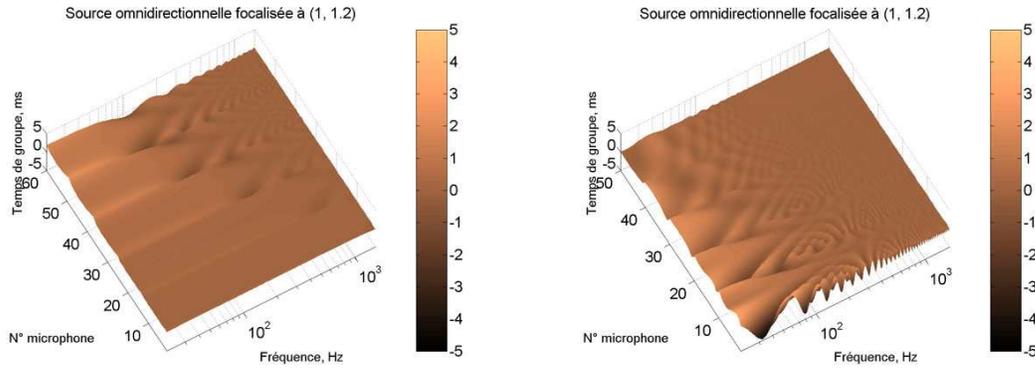
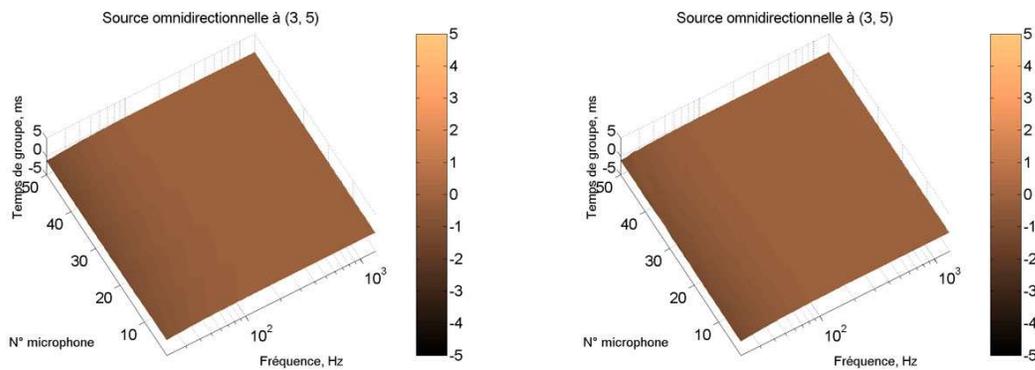
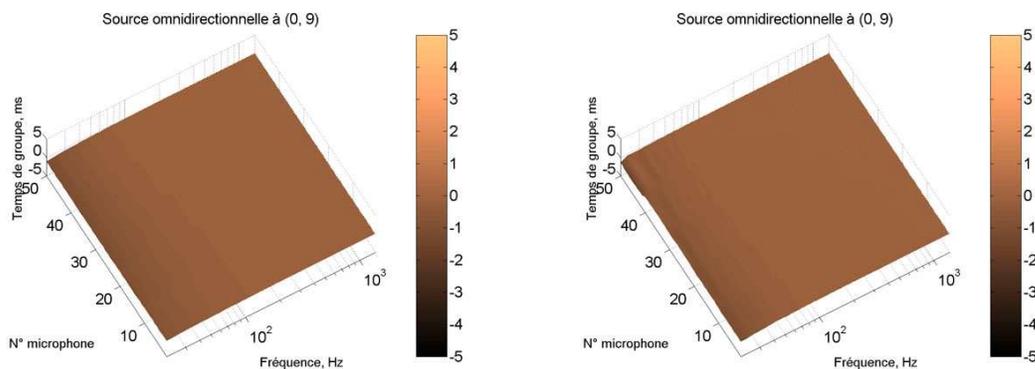
(a) Source omnidirectionnelle focalisée à $(1, 1.2)$, microphones *proches*(b) Source omnidirectionnelle focalisée à $(1, 1.2)$, microphones *lointains*(c) Source omnidirectionnelle à $(3, 5)$, microphones *proches*(d) Source omnidirectionnelle à $(3, 5)$, microphones *lointains*(e) Source omnidirectionnelle à $(0, 9)$, microphones *proches*(f) Source omnidirectionnelle à $(0, 9)$, microphones *proches*

FIG. 2.9 – Temps de groupe relatif associé à la réponse impulsionnelle d'un banc de 600 haut-parleurs (60 m, 10 cm d'espacement) sous la fréquence d'aliasing spatial (20-2000 Hz) synthétisant différentes sources virtuelles omnidirectionnelles avec la WFS.

2.3 Rayonnement d'un banc linéaire continu de taille finie : diffraction

La description du champ diffracté dû à la taille finie de la ligne de sources secondaires est décrit en détail par Nicol (Nicol, 1999) et Start (Start, 1997). Cette description se base sur une estimation dans le domaine fréquentiel du champ sonore émis par les parties manquantes de la ligne. Ils montrent alors que le champ diffracté peut être considéré comme émis par deux sources situées aux extrémités du segment. La directivité de ces sources est complexe et dépend de la source virtuelle synthétisée.

On propose ici deux descriptions complémentaires du phénomène de diffraction permettant d'apprécier son évolution au cours de la propagation du champ sonore synthétisé. On considère ici le phénomène de diffraction en tant qu'interruption de la construction du front d'onde. Pour cela, on reprend la description de la synthèse du champ sonore dans le domaine temporel effectuée dans la partie 2.2.2. Dans un deuxième temps, on utilise une description dans le domaine fréquentiel à partir des contributions relatives des différents haut-parleurs. Enfin, un ensemble de simulations permet d'évaluer l'influence de la diffraction sur la fonction de qualité $Q_{\Psi}^{wfs}(r(R), k)$ dans la zone d'écoute.

2.3.1 Domaine temporel : fenêtrage

Le champ diffracté $Diff_{\Psi_m}(r_R, t)$ dû à une distribution de sources secondaires de taille finie ($x \in [x_A, x_B]$, segment $[AB]$) en un point R est obtenu par un calcul dans le domaine temporel tel que présenté dans la partie 2.2.2. Afin de simplifier les expressions, dans la suite, on introduit,

$$\begin{cases} t_A = \Delta t_{\Psi_m}(r_R, x_A) \\ t_B = \Delta t_{\Psi_m}(r_R, x_B) \end{cases}. \quad (2.42)$$

Ainsi, si $x_0 \in [x_A, x_B]$,

$$Diff_{\Psi_m}(r_R, t) = -Y(t - t_A)h_{\Psi_m, R}^-(t) - Y(t - t_B)h_{\Psi_m, R}^+(t). \quad (2.43)$$

$x_0 \in [x_A, x_B]$ équivaut à ce que l'intersection de la ligne passant par Ψ_m et R avec la ligne L appartiennent au segment $[AB]$. L'ensemble des points R remplissant cette condition n'est autre que l'ensemble des points de la zone de visibilité de Ψ à travers le banc de haut-parleurs.

La fonction de transfert électroacoustique du segment de sources secondaires $h_{\Psi_m}^{AB}$ pour un point R de l'espace de restitution est alors donnée par,

$$h_{\Psi_m}^{AB}(r_R, t) = \Pi_{t_0}^{t_A}(t)h_{\Psi_m, R}^-(t) + \Pi_{t_0}^{t_B}(t)h_{\Psi_m, R}^+(t), \quad (2.44)$$

où $\Pi_{t_1}^{t_2}(t)$ est la fonction porte de largeur $\Delta T = t_2 - t_1$ décalée de $\tau = t_1 + \frac{T}{2}$ définie par,

$$\begin{cases} \Pi_{t_1}^{t_2}(t) = 1 \text{ si } t \in [t_1, t_2] \\ \Pi_{t_1}^{t_2}(t) = 0 \text{ sinon.} \end{cases} \quad (2.45)$$

L'effet de diffraction correspond à un fenêtrage de la forme temporelle synthétisée par chacune des deux sous-distributions de sources secondaires ($L_{\Psi_m, R}^-$ et $L_{\Psi_m, R}^+$).

La taille des fenêtres est donnée par

$$\begin{cases} T_{\Psi_m}^-(r_R) = t_A - t_0 \\ T_{\Psi_m}^+(r_R) = t_B - t_0, \end{cases} \quad (2.46)$$

où, $t_A - t_0$ et $t_B - t_0$ dépendent de la position de la source virtuelle (plus elle est proche de L , plus ces valeurs sont grandes), de la position d'écoute R (plus R est proche de L et elle est loin des limites de la zone de visibilité, plus ces valeurs sont grandes) et de la longueur du segment AB (plus il est grand, plus ces valeurs sont grandes).

L'effet de diffraction sur le champ synthétisé dépend ainsi de la portion d'énergie émise par $L_{\Psi_m, R}^-$ comprise entre t_0 et t_A et par $L_{\Psi_m, R}^+$ entre t_0 et t_B .

On a constaté dans la partie 2.2.2 que l'on observe un étalement de la réponse temporelle en rapport avec l'augmentation de la distance de R à L . On considère que cet étalement traduit une émission plus tardive des basses fréquences. La taille finie du banc de haut-parleurs réduit ainsi le niveau des basses fréquences. Le temps de groupe associé est aussi diminué.

D'autre part, un fenêtrage de type "porte" produit des oscillations dans la réponse fréquentielle obtenue. Ces oscillations peuvent être réduites en atténuant le niveau des haut-parleurs situés aux extrémités, solution retenue pour la WFS, afin d'"adoucir" la transition.

Dans le cas où $x_0 < x_A$,

$$h_{\Psi_m}^{AB}(r_R, t) = \Pi_{t_A}^{t_B}(t) h_{\Psi_m, R}^+(t). \quad (2.47)$$

Si $x_0 > x_B$,

$$h_{\Psi_m}^{AB}(r_R, t) = \Pi_{t_A}^{t_B}(t) h_{\Psi_m, R}^-(t). \quad (2.48)$$

En dehors de la zone de visibilité, la taille finie du banc de haut-parleurs effectue une sélection de la partie tardive de la réponse impulsionnelle. La réponse impulsionnelle synthétisée contient ainsi principalement des composantes en basses fréquences. La réponse fréquentielle associée présente des oscillations dues au fenêtrage de type "porte". Celles-ci sont réduites, là aussi, par l'atténuation des haut-parleurs situés aux extrémités du banc de haut-parleurs.

2.3.2 Domaine fréquentiel : contributions des haut-parleurs à une position d'écoute

Dans cette partie, on propose d'évaluer les contributions individuelles de chaque haut-parleur pour la synthèse du front d'onde associé à une source virtuelle en une position d'écoute donnée.

On a décrit dans la partie 2.2.1 la fonction de transfert électroacoustique d'un banc linéaire de haut-parleurs pour la synthèse d'une source virtuelle Ψ . Celle-ci s'écrit sous la forme d'une intégrale sur l'axe du banc de haut-parleurs (x dans ce cas). Dans cette intégrale, on distingue une fonction d'enveloppe $A(x)$ et une fonction de phase $\chi(x)$. La contribution $C_{\Psi_m}(x, k, r(R))$ au point R d'un haut-parleur situé en x pour la synthèse d'une source Ψ_m pour un nombre d'onde k s'écrit

$$C_{\Psi_m}(x, k, r(R)) = A(x)e^{-jk\chi(x)}. \quad (2.49)$$

Cette contribution présente un terme oscillatoire $e^{-jk\chi(x)}$ dont la valeur se stabilise autour du point de phase stationnaire $x = x_0$. En ce point, la fonction de phase χ présente un extremum global. La fonction χ étant strictement monotone de chaque côté de x_0 , des oscillations apparaissent en dehors de ce point. Le nombre d'onde k apparaît ainsi comme un facteur de compression/dilatation du motif d'oscillation.

La fonction de transfert électroacoustique au point R et à la fréquence k se construit en sommant les composantes provenant du lobe principal (autour de x_0), puis, par un jeu de sommations/annulations des contributions provenant des haut-parleurs situés de part et d'autre de x_0 . Pour une fréquence donnée, la troncature du banc de haut-parleurs interrompt abruptement la construction du front d'onde à un "emplacement" différent dans le motif d'oscillations de chaque côté du point de phase stationnaire (aux deux extrémités du banc). Le niveau synthétisé, pour une position et une fréquence données, varie en fonction du point d'arrêt dans le régime d'oscillation. Le module de la réponse fréquentielle synthétisée oscille alors autour d'une valeur moyenne. Afin de limiter les phénomènes de diffraction, Start (Start, 1997) propose d'atténuer les signaux d'alimentation des haut-parleurs situés aux extrémités du banc. Les gains d'atténuation sont définis par une fenêtre de type sinus/cosinus pour 10 % du nombre total de haut-parleurs à chaque extrémité. L'application de ces gains d'atténuation "adoucit" la rupture du motif d'oscillations et limite les fluctuations de la réponse fréquentielle.

Remarquons que cette description peut se rapprocher de ce que l'on appelle les zones de Fresnel. Les zones de Fresnel correspondent aux zones d'une surface qui apparaissent en phase ou en opposition de phase vis-à-vis d'un point de l'espace. Elles permettent de prédire le rayonnement de la surface. La description présentée ici donne une information supplémentaire sur le niveau et la phase exacte de contribution de chacun des haut-parleurs.

On applique cette analyse à un banc de haut-parleurs de 6 m de long orienté selon \vec{x} et centré par rapport à l'origine (cf. figure 2.10). On calcule la partie réelle des contributions de chacun des haut-parleurs, normalisée par sa valeur au point de phase stationnaire ($x = x_0$). Ce calcul est réalisé pour deux points d'écoute (un proche du banc de haut-parleurs et centré par rapport à celui-ci, un point plus lointain et excentré), pour 3 sources virtuelles différentes (une source omnidirectionnelle focalisée, une source omnidirectionnelle située derrière le banc et une onde plane), ceci, pour trois fréquences différentes (100, 300, 500 Hz).

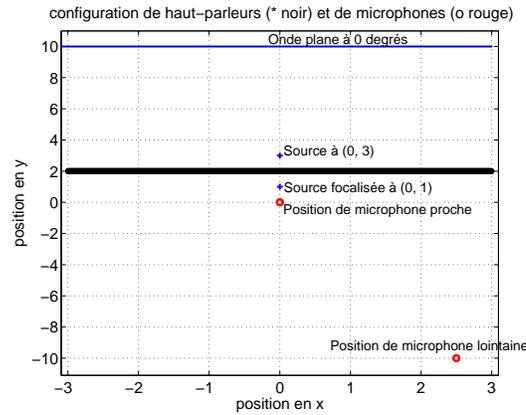


FIG. 2.10 – configuration de haut-parleurs et de microphone, sources virtuelles de test.

L'ensemble des figures 2.11 représente ces contributions normalisées ainsi que la fonction d'enveloppe normalisée (par sa valeur en x_0) à laquelle on applique l'atténuation des haut-parleurs situés aux extrémités.

La taille finie du banc et la fonction d'enveloppe réalisent une *sélection* d'une partie du motif d'oscillations qui varie en fonction de la fréquence et dépend de la position d'écoute.

A proximité de la source focalisée, à 100 Hz, la fonction d'enveloppe "sélectionne" le lobe principal (cf. figure 2.11(a)). A 300 Hz, l'atténuation porte principalement sur les parties négatives des lobes secondaires. Pour les fréquences plus élevées, elle concernera les lobes d'ordre supérieur. L'atténuation des haut-parleurs situés aux extrémités risque de n'avoir qu'un effet très marginal sur le champ synthétisé à cette position. Le niveau de la fonction d'enveloppe qui leur est associée est déjà très faible.

Pour la deuxième position d'écoute, le régime d'oscillation est plus marqué. La fonction d'enveloppe ne réalise une "sélection" faible que pour la position d'écoute proche. Elle est, d'autre part, dissymétrique par rapport au point de phase stationnaire. La rupture du motif d'oscillations est ainsi plus franche et l'effet de diffraction plus marqué. L'atténuation des haut-parleurs situés aux extrémités permet alors de lisser cet effet principalement pour les fréquences plus élevées (300, 500 Hz).

Pour la source omnidirectionnelle à $(0, 3)$, on peut faire le même type d'observations. Contrairement au cas de la source focalisée, le régime d'oscillations est plus rapide à proximité du banc et se ralentit pour une position plus éloignée.

L'onde plane présente un comportement spécifique avec une fonction d'enveloppe très peu "sélective" et, donc, une rupture abrupte des oscillations. Pour la position excentrée, une grande partie du lobe principal est tronquée aux basses fréquences ce qui provoque une diminution du niveau synthétisé. Dans le cas de l'onde plane, l'atténuation des haut-parleurs aux extrémités permet dans ce cas de limiter l'effet de la rupture, principalement pour les fréquences moyennes, mais va amplifier l'effet d'atténuation du niveau des basses fréquences.

Cette analyse met en évidence l'évolution spatiale de l'effet de diffraction dans la zone d'écoute. Il correspond à une rupture du motif d'oscillations dans la construction d'un front d'onde par un banc linéaire de haut-parleurs. On a ainsi vu que suivant la position d'écoute, la fonction d'enveloppe (niveaux relatifs des haut-parleurs au point d'écoute considéré) effectue déjà une sélection du motif d'oscillations. L'atténuation des haut-parleurs situés aux extrémités permet d'adoucir la rupture mais réduit le niveau synthétisé en basses fréquences et pour des positions d'écoute excentrées.

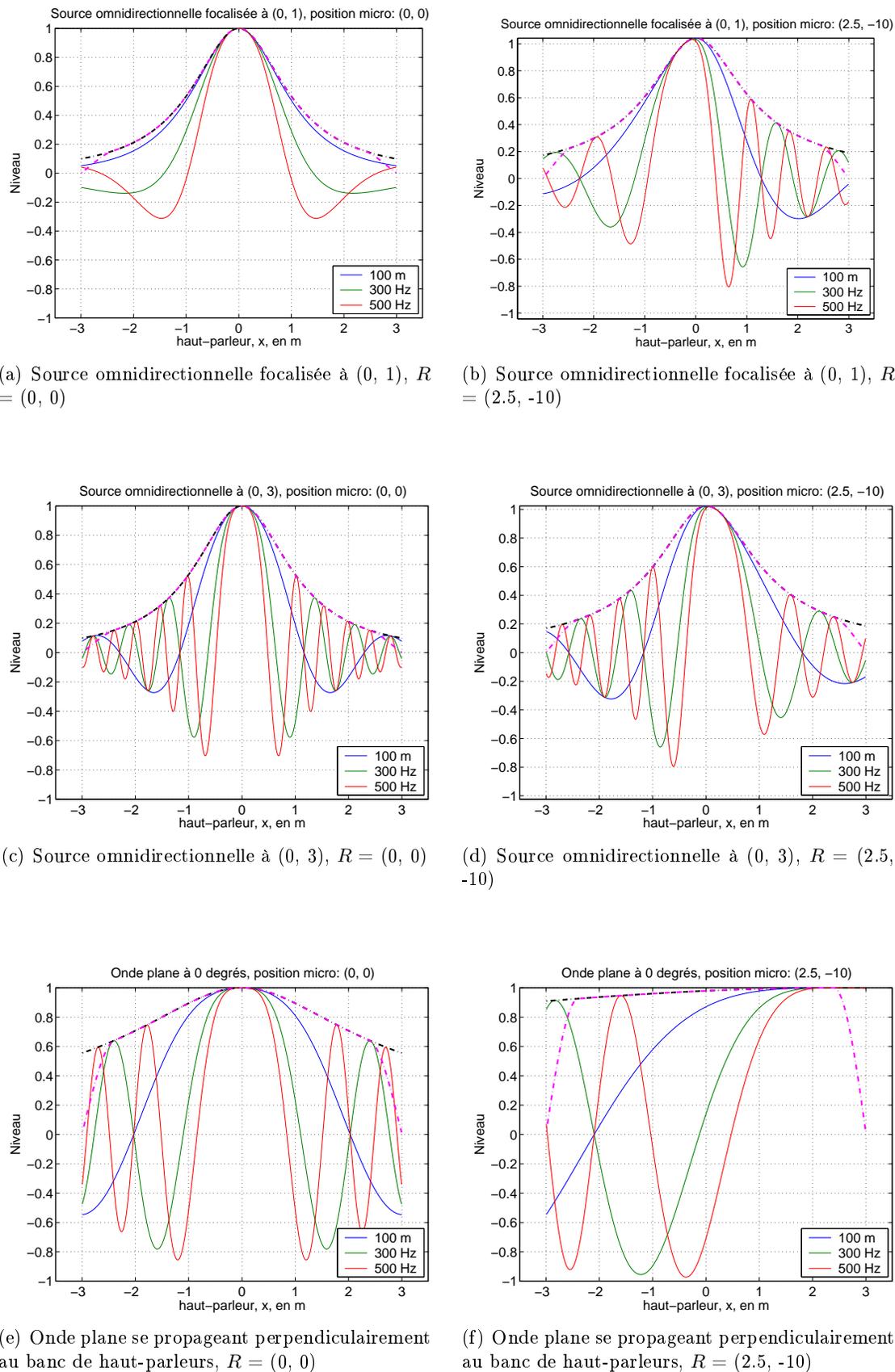


FIG. 2.11 – Partie réelle, en fonction de la fréquence, de la contribution des haut-parleurs d'un banc linéaire continu de 6 mètres de long, situé sur la ligne d'équation $y = 2$ et centré par rapport à l'origine pour un ensemble de sources virtuelles et de points de réception R . La ligne discontinue correspond à la fonction d'enveloppe à laquelle on applique l'atténuation des haut-parleurs situés aux extrémités.

2.3.3 Simulations : évolution spatiale du champ acoustique synthétisé

Dans cette partie on reprend le même type de simulations que celles effectuées pour un banc de haut-parleurs "infini" (60 m, 600 haut-parleurs) dans la partie 2.2.3. On utilise les mêmes configurations de microphones qui sont reproduites, ici, dans les figure 2.12. Le banc de haut-parleurs utilisés ici est composé de 60 haut-parleurs omnidirectionnels idéaux espacés de 10 cm, répartis sur une ligne suivant l'axe x et centré par rapport à $(0, 0)$.

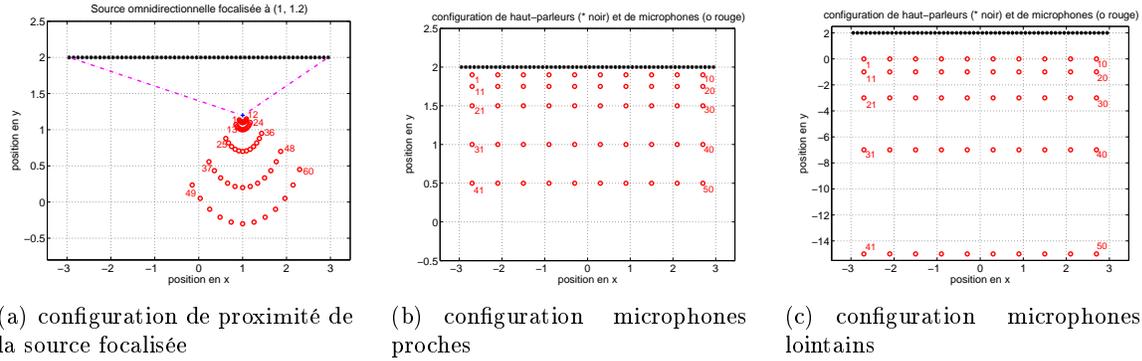


FIG. 2.12 – Configurations de test avec 60 haut-parleurs (6 m, 10 cm d'espacement, * noirs); différentes positions de microphones (o rouges)

Les ensembles de figures 2.13 et 2.14 représentent le module de la fonction de qualité $Q_{\Psi}^{wfs}(r(R), k)$ pour un banc de haut-parleurs de taille finie. Les simulations sont séparées en deux blocs de 6 figures qui rassemblent les situations où les microphones sont *proches* (cf. ensemble des figures 2.13) et les situations où les microphones sont *lointains* (cf. ensemble des figures 2.14). Dans chacun des blocs, sont mises face-à-face les réponses fréquentielles obtenues avec ou sans l'atténuation des haut-parleurs situés aux extrémités du banc.

Trois remarques d'ordre général :

1. Pour les microphones *proches*, la taille finie du banc de haut-parleurs n'a que peu ou pas d'influence sur le champ synthétisé (cf. figures 2.13). On note uniquement une diminution significative de niveau aux basses fréquences ($20 - 100Hz$).
2. L'atténuation des haut-parleurs situés aux extrémités permet de lisser les oscillations de la réponse fréquentielle, principalement pour les moyennes et hautes fréquences. On remarque que de légères oscillations subsistent. Elles dépendent assez fortement de la position d'écoute et restent en dessous des seuils d'audition.
3. On observe une modification de la loi d'atténuation aux basses fréquences, celles-ci s'atténuant plus vite que les hautes fréquences. Ce phénomène est d'autant plus marqué que la fréquence est basse.

Ce dernier point est un phénomène bien connu du rayonnement de sources linéaires de taille finie dans les applications de sonorisation. Urban (Urban *et al.*, 2003) définit une distance limite entre

ce qu'il appelle le champ proche et le champ lointain du banc de haut-parleurs. Dans le champ proche, l'atténuation est de -3 dB chaque fois que la distance à la ligne est doublée, et de -6 dB en champ lointain. La distance limite $d_B(f)$ (en m) dépend de la fréquence f (en kHz). Pour un banc de haut-parleurs de longueur H ,

$$d_B(f) = \frac{3}{2}fH^2 \sqrt{1 - \frac{1}{(3fH)^2}} \quad (2.50)$$

Cette formule donne une valeur limite de 4,5 m à 100 Hz et de 27 m à 500 Hz et donne un nombre complexe pour des fréquences inférieures à 60 Hz.

La source linéaire est équivalente, pour la WFS, au cas de la synthèse d'une onde plane se propageant perpendiculairement au banc. Ceci correspond aux figures 2.13(e), 2.13(f), 2.14(e), 2.14(f), pour lesquelles la modification de la loi d'atténuation est particulièrement visible. La fréquence limite dépend de la distance du point d'écoute au banc de haut-parleurs mais aussi de sa position par rapport au centre du banc. La fréquence limite est ainsi plus faible sur les côtés. L'atténuation des haut-parleurs situés aux extrémités du banc a tendance à amplifier le phénomène. Ce point rejoint les observations faites dans la partie 2.3.2.

La définition de la fréquence limite est plus compliquée que celle donnée par Urban. Son raisonnement se base sur les zones de Fresnel du banc en référence à une position d'écoute centrée. La formule est donnée par la distance à partir de laquelle l'ensemble des sources secondaires apparaît "en phase". Ce raisonnement pourrait se généraliser en considérant la sélection opérée par la fonction d'enveloppe et la taille finie du banc de haut-parleurs comme nous l'avons proposé dans la partie 2.3.2. On pourrait ainsi définir cette distance limite $d_B(f)$ comme celle à partir de laquelle le niveau de contribution des lobes secondaires devient négligeable par rapport à celui du lobe principal.

La modification de la loi d'atténuation pose un éventuel problème pour l'utilisation de la WFS aux basses fréquences (20-100 Hz) en champ libre. Toutefois, dans une salle d'écoute et à ces fréquences, il n'est pas possible de dissocier le direct et le champ réverbéré. Le problème se pose, en réalité, de façon plus complexe car, à ces fréquences, on ne peut distinguer le son direct des modes de la salle d'écoute. Nous abordons cette question dans la partie 8.1.2 mais cette question mériterait une étude spécifique.

La modification de la loi d'atténuation affecte de manière significative la réponse fréquentielle pour des positions d'écoute lointaines et en limite de la zone de visibilité des sources virtuelles. Le profil du module de $Q_{\Psi}^{wfs}(r(R), k)$ pour la source à (3, 5) est ainsi "vrillé" en limite de la zone de visibilité ; microphones 37-40, 46-50) avec des variations de niveau de l'ordre de 5 dB, le niveau synthétisé étant plus faible en basses fréquences (cf. figure 2.14(e)). Ces variations sont donc significatives d'un point de vue perceptif et réduisent la transparence de la synthèse de cette source.

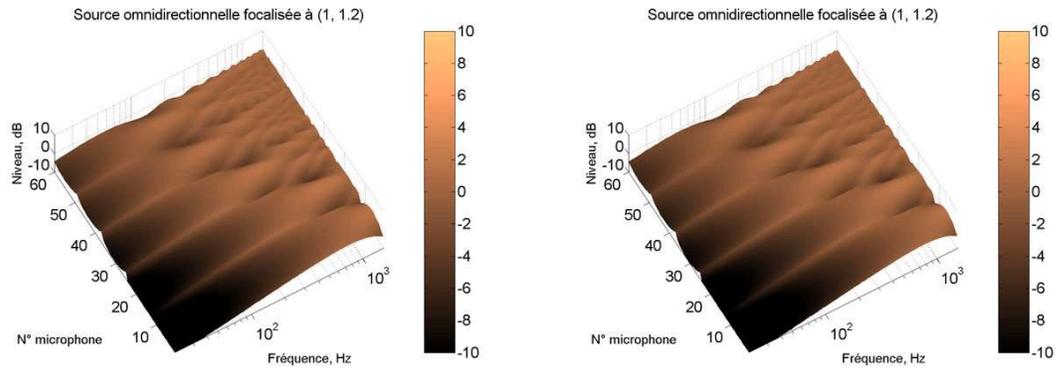
L'atténuation des haut-parleurs situés aux extrémités a tendance à amplifier ce phénomène. Ceci rejoint les remarques des sections 2.3.1 et 2.3.2 sur la sélection opérée sur le front d'onde

synthétisé par la troncature du banc en limite de la zone de visibilité. Ce phénomène s'observe aussi pour l'onde plane (cf. figures 2.14(e), 2.14(f)).

L'ensemble des figures 2.15 présente les variations de temps de groupe de la fonction de qualité $Q_{\psi}^{wfs}(r(R), k)$. Les sources virtuelles sont synthétisées par la WFS en atténuant les haut-parleurs situés aux extrémités du banc.

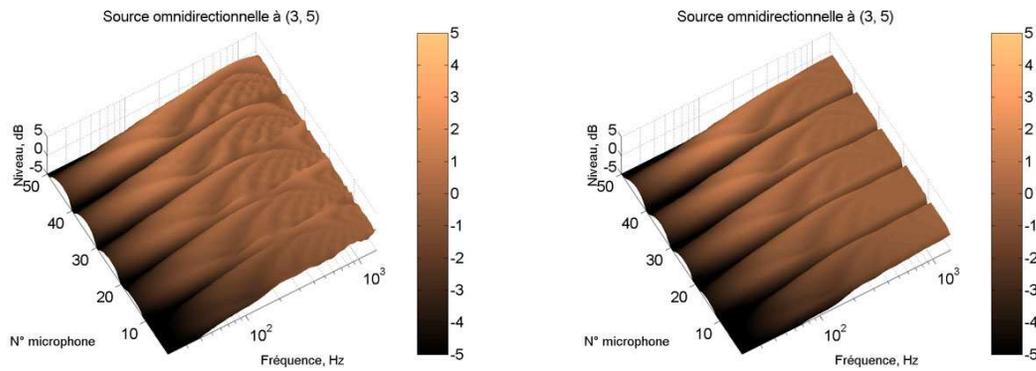
Indépendamment de la source virtuelle considérée, on observe des fluctuations relativement importantes du temps de groupe (2 à 3 ms) autour de la valeur de changement de mode d'atténuation. Toutefois, la bande fréquence dans laquelle se produit cette variation du temps de groupe est relativement faible. Il est donc probable que cette modification soit peu ou pas audible (cf. discussion sur l'audibilité des variations de temps de groupe, partie).

Le temps de groupe prend une valeur globalement plus faible aux basses fréquences (≈ -2 ms) pour l'onde plane (cf. figure 2.15(f)). Ces variations ont lieu sur une large bande de fréquence ce qui les rend potentiellement audibles. La largeur de cette bande de fréquence augmente avec la distance au banc de haut-parleurs ce qui accroît encore les risques de détection.



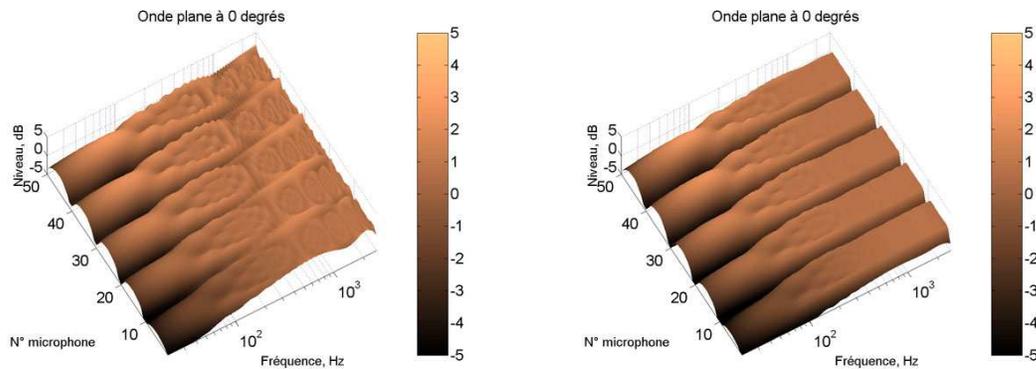
(a) Source omnidirectionnelle focalisée à (1, 1.2), sans atténuation haut-parleurs extrémités

(b) Source omnidirectionnelle focalisée à (1, 1.2), avec atténuation haut-parleurs extrémités



(c) Source omnidirectionnelle à (3, 5), sans atténuation haut-parleurs extrémités

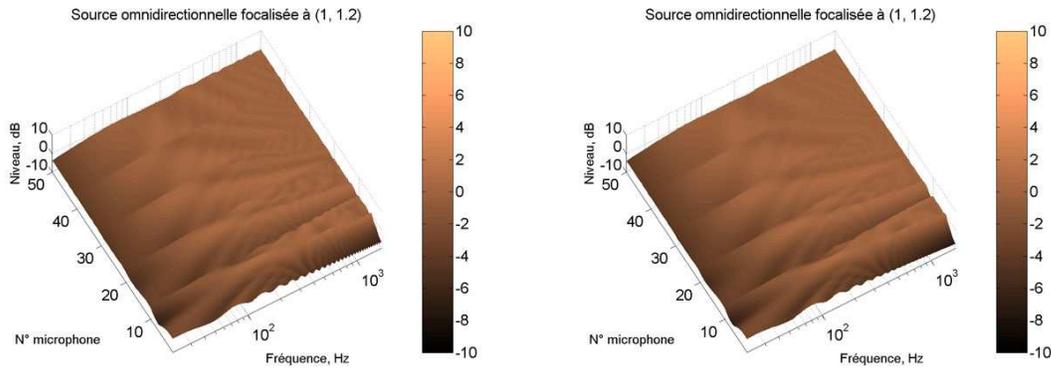
(d) Source omnidirectionnelle à (3, 5), avec atténuation haut-parleurs extrémités



(e) Onde plane se propageant perpendiculairement au banc de haut-parleurs, sans atténuation haut-parleurs extrémités

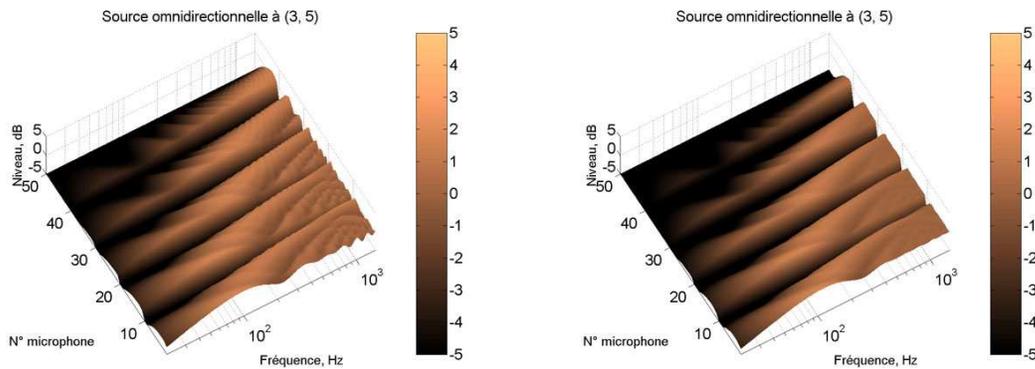
(f) Onde plane se propageant perpendiculairement au banc de haut-parleurs, avec atténuation haut-parleurs extrémités

FIG. 2.13 – Réponses fréquentielles avec compensation de niveau d'un banc de 60 haut-parleurs (6 m, 10 cm d'espacement) sous la fréquence d'aliasing spatial (20-2000 Hz) synthétisant différentes sources virtuelles omnidirectionnelles avec la WFS, microphones *proches*



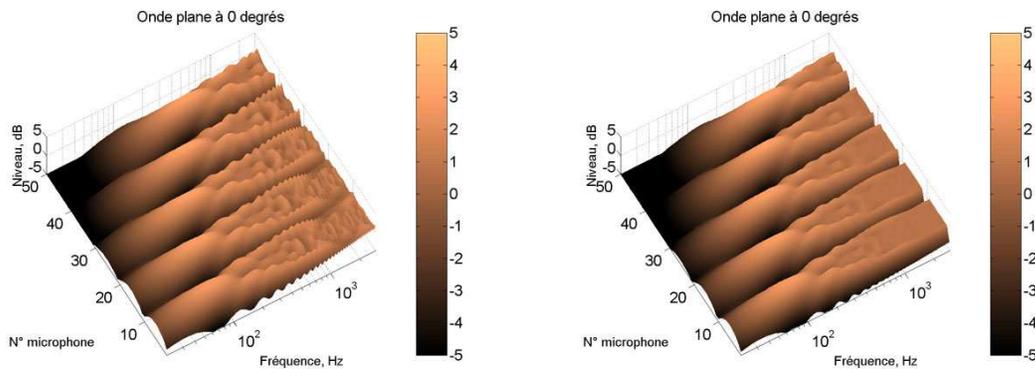
(a) Source omnidirectionnelle focalisée à (1, 1.2), sans atténuation haut-parleurs extrémités

(b) Source omnidirectionnelle focalisée à (1, 1.2), avec atténuation haut-parleurs extrémités



(c) Source omnidirectionnelle à (3, 5), sans atténuation haut-parleurs extrémités

(d) Source omnidirectionnelle à (3, 5), avec atténuation haut-parleurs extrémités



(e) Onde plane se propageant perpendiculairement au banc de haut-parleurs, sans atténuation haut-parleurs extrémités

(f) Onde plane se propageant perpendiculairement au banc de haut-parleurs, avec atténuation haut-parleurs extrémités

FIG. 2.14 – Réponses fréquentielles avec compensation de niveau d'un banc de 60 haut-parleurs (6 m, 10 cm d'espacement) sous la fréquence d'aliasing spatial (20-2000 Hz) synthétisant différentes sources virtuelles omnidirectionnelles avec la WFS, microphones *lointains*.

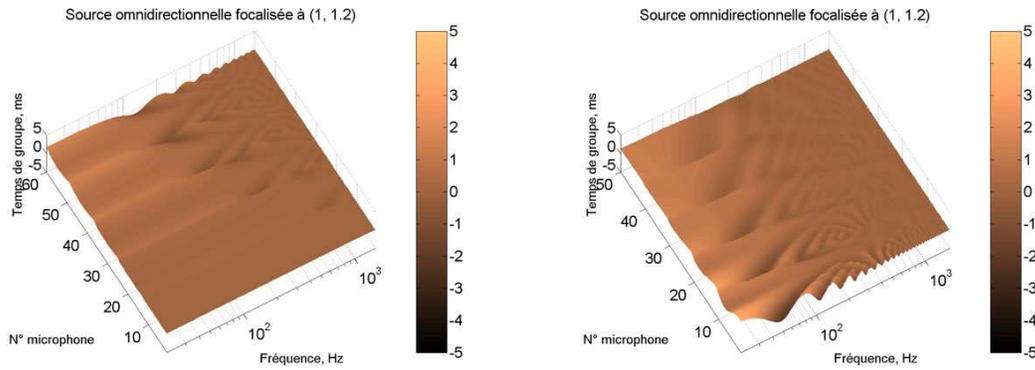
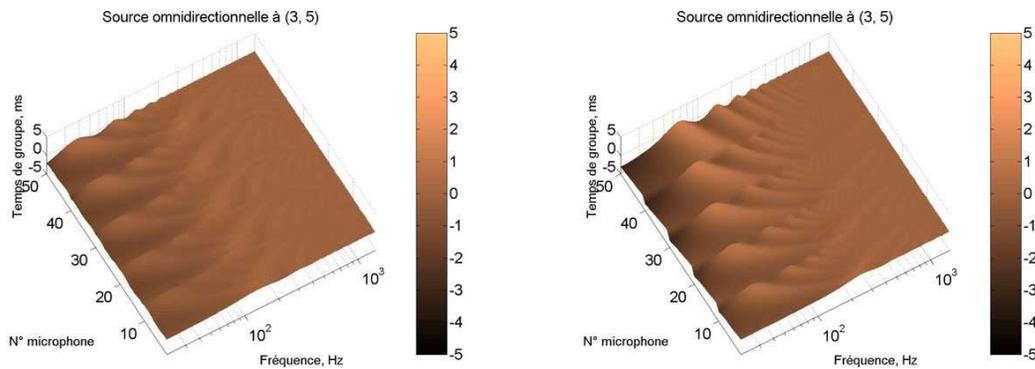
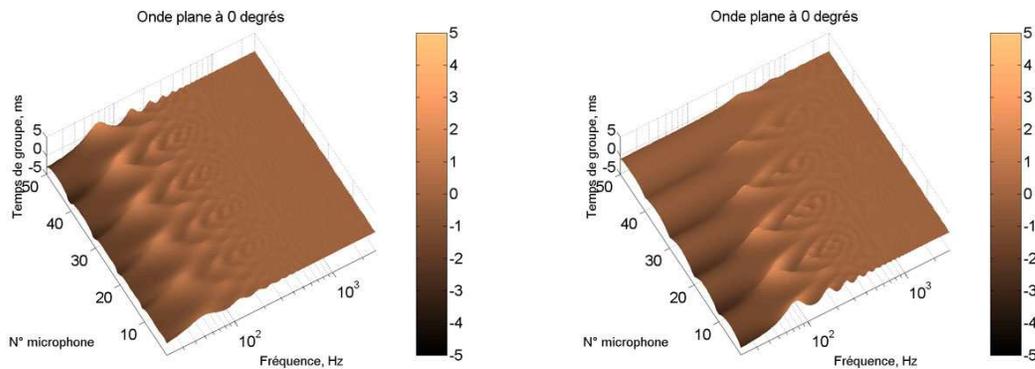
(a) Source omnidirectionnelle focalisée à $(1, 1.2)$, microphones *proches*(b) Source omnidirectionnelle focalisée à $(1, 1.2)$, microphones *lointains*(c) Source omnidirectionnelle à $(3, 5)$, microphones *proches*(d) Source omnidirectionnelle à $(3, 5)$, microphones *lointains*(e) Onde plane se propageant perpendiculairement au banc de haut-parleurs, microphones *proches*(f) Onde plane se propageant perpendiculairement au banc de haut-parleurs, microphones *lointains*

FIG. 2.15 – Temps de groupe relatif associé à la réponse impulsionnelle d'un banc de 60 haut-parleurs (60 m, 10 cm d'espacement) sous la fréquence d'aliasing spatial (20-2000 Hz) synthétisant différentes sources virtuelles omnidirectionnelles avec la WFS, avec atténuation haut-parleurs extrémités.

2.4 Rayonnement d'un banc linéaire échantillonné de taille finie : aliasing spatial

Le phénomène d'aliasing spatial est dû à l'échantillonnage de la distribution de sources secondaires. Dans la littérature, l'aliasing spatial est décrit dans la formulation de la synthèse d'une source virtuelle en WFS dans le domaine de Fourier spatial. On se place dans le cas d'une distribution linéaire de sources secondaires de taille infinie.

Ce phénomène apparaît comme un repliement des composantes spatiales du champ sonore cible (celui de la source primaire). La fréquence limite, à partir de laquelle ce phénomène apparaît, est appelée fréquence d'aliasing spatial. Dans le cas d'un banc de haut-parleurs de taille infinie, cette fréquence ne dépend pas de la position d'écoute R .

Les simulations présentées dans la section 1.4.1 ont montré que cette fréquence dépend aussi de la position d'écoute dans le cas de bancs de haut-parleurs finis.

Horbach, l'auteur et de Vries ont montré que la fréquence d'aliasing spatial a tendance à augmenter avec la distance dans le cas de la synthèse d'une onde plane se propageant perpendiculairement par rapport au banc de haut-parleurs (Horbach *et al.*, 2002). En effet, pour cette source particulière, le retard affecté à chacun des haut-parleurs est nul. Dans l'approximation de champ lointain, la différence de temps d'arrivée des différentes contributions de chacun des haut-parleurs ne dépend que de l'angle θ de la position d'écoute vis-à-vis de la normale au banc de haut-parleurs. Classiquement, le retard $\Delta t(n)$ de la contribution du haut-parleur n est donnée par

$$\Delta t(n) = n\Delta x \sin \theta. \quad (2.51)$$

On voit ainsi qu'en champ lointain, *tous* les haut-parleurs sont *en phase* et la fréquence d'aliasing associée est, a priori, *infinie*. La formule 1.16, pour une distribution de sources secondaires de taille infinie, donne une fréquence d'aliasing de $\frac{c}{\Delta x}$ Hz.

Le même phénomène s'observe pour les sources focalisées. En effet, par définition, tous les haut-parleurs sont en phase au niveau du point de focalisation (position de la source virtuelle). En ce point, la fréquence d'aliasing est aussi "infinie".

Dans cette partie, nous proposons un cadre mathématique qui permet de rendre compte de ces observations. Pour cela, nous utilisons la description temporelle du champ sonore rayonné proposée dans la section 2.2.2.

Considérons la fonction de transfert électroacoustique $h_{\Psi_m, \Delta x}^{diff}(r_R, t)$ en un point de l'espace R d'un banc de haut-parleurs pour la synthèse d'une source virtuelle Ψ_m . Le banc de haut-parleurs est composé de sources omnidirectionnelles idéales régulièrement espacées tous les Δx m. Ses extrémités sont situées en $x_A = n_A \Delta x$ et $x_B = n_B \Delta x$. Il s'agit ainsi d'une version échantillonnée

"spatialement" d'un banc de haut-parleurs continu compris entre x_A et x_B . Ceci donne

$$h_{\Psi_m, \Delta x}^{diff}(r_R, t) = \Delta x \sum_{n=n_A}^{n_B} G_{\Psi_m}(n\Delta x) \frac{\delta(t - \Delta t_{\Psi_m}(r_R, n\Delta x))}{4\pi\Delta r(n\Delta x)}. \quad (2.52)$$

Si la source virtuelle Ψ_m est située dans Ω_Ψ ("derrière" le banc de haut-parleurs), la fonction $x \rightarrow \Delta t_{\Psi_m}(r_R, x)$ est une fonction strictement décroissante $x^- \rightarrow t^-(x^-)$ pour $x^- = (x, x \in]-\infty, x_0])$, et, strictement croissante $x^+ \rightarrow t^+(x^+)$ pour $x^+ = (x, x \in [x_0, +\infty[)$. Dans le cas où Ψ_m est une source focalisée, les deux fonctions citées précédemment sont strictement monotones mais avec un sens de variation opposé. On définit $n_0 = \min_n(n\Delta x > x_0)$. L'équation 2.52 s'écrit sous la forme,

$$h_{\Psi_m, \Delta x}^{diff}(r_R, t) = \Delta x \sum_{n=n_A}^{n_0} G_{\Psi_m}(n\Delta x) \frac{\delta(t - t^-(n\Delta x))}{4\pi\Delta r(n\Delta x)} + \sum_{n=n_0}^{n_B} G_{\Psi_m}(n\Delta x) \frac{\delta(t - t^+(n\Delta x))}{4\pi\Delta r(n\Delta x)}. \quad (2.53)$$

On introduit les fonctions $h_{\Psi_m}^+(r_R, t)$ et $h_{\Psi_m}^-(r_R, t)$,

$$h_{\Psi_m, \Delta x}^{diff}(r_R, t) = \left(\Delta x \sum_{n=-\infty}^{n_0} \delta(t - t^-(n\Delta x)) \frac{dt^-(t^-(n\Delta x))}{dx^-} \right) \Pi_{t_A}^{t_0} h_{\Psi_m}^-(r_R, t) + \left(\Delta x \sum_{n=n_0+1}^{+\infty} \delta(t - t^+(n\Delta x)) \frac{dt^+(t^+(n\Delta x))}{dx^+} \right) \Pi_{t_B}^{t_0} h_{\Psi_m}^+(r_R, t). \quad (2.54)$$

L'échantillonnage spatial *régulier* du banc de haut-parleurs apparaît comme un échantillonnage *irrégulier* dans le domaine temporel des fonctions $h_{\Psi_m}^+(r_R, t)$ et $h_{\Psi_m}^-(r_R, t)$. La reconstruction du signal s'effectue en multipliant l'échantillon par l'inverse de la *densité locale d'échantillons* $\frac{dt^-(t^-(n\Delta x))}{dx^-}$ (resp. $\frac{dt^+(t^+(n\Delta x))}{dx^+}$). La fréquence de Nyquist associée correspond ainsi à la fréquence d'aliasing au point de réception R .

Pour un échantillonnage irrégulier, la fréquence de Nyquist est liée à la distribution temporelle des échantillons. Elle est néanmoins difficile à définir car elle dépend de la nature à court terme du signal échantillonné et de la méthode de reconstruction choisie. Pour la WFS, la reconstruction s'effectue en multipliant chaque "échantillon" par l'inverse de la densité locale d'échantillons et par sommation naturelle dans le domaine acoustique des contributions de tous les haut-parleurs.

Pour la synthèse du front d'onde par la WFS, on doit considérer deux distributions temporelles $t^-(n\Delta x)$ pour $n \leq n_0$ et $t^+(n\Delta x)$ pour $n > n_0$. Les écarts entre deux échantillons successifs $\Delta\tau^-(n)$ et $\Delta\tau^+(n)$ sont définis par

$$\begin{cases} \Delta\tau(n) = t^-((n-1)\Delta x) - t^-(n\Delta x) & \text{for } n_A < n \leq n_0 \\ \Delta\tau(n) = t^+((n+1)\Delta x) - t^+(n\Delta x) & \text{for } n_0 < n < n_B. \end{cases} \quad (2.55)$$

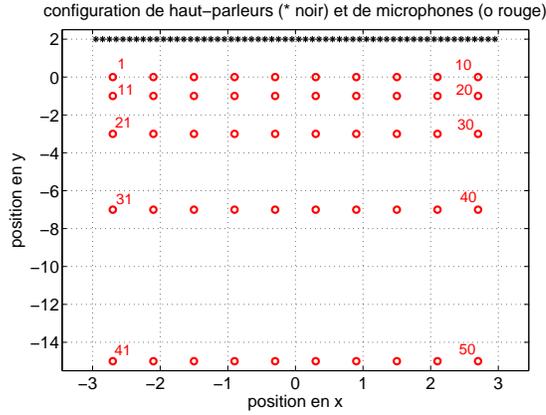


FIG. 2.16 – configuration de haut-parleurs et de microphones pour vérifier la validité du critère d'aliasing.

Pour un banc de haut-parleurs et une source virtuelle Ψ_m , nous proposons de définir,

$$f_{al}^{\Psi_m}(\Delta x, r_R) = \frac{1}{\max |\Delta\tau(n)|}. \quad (2.56)$$

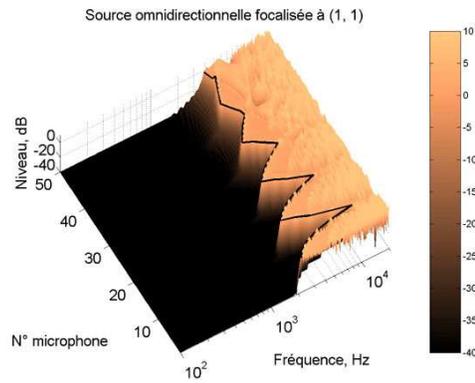
Afin de vérifier la validité de ce critère, on simule la réponse fréquentielle $H_{\Psi}^{\Delta x}(n_{mic}, f)$ avec des haut-parleurs espacés de Δx . On considère un ensemble de microphones réparti sur 5 lignes, comme indiqué sur la figure 2.4. On simule la réponse fréquentielle $H_{psi}^{\frac{\Delta x}{10}}(n_{mic}, f)$ d'un banc de haut-parleurs de même longueur mais avec 10 fois plus de sources, soit un espacement 10 fois inférieur.

La différence entre $H_{\Psi_m}^{\Delta x}(n_{mic}, f)$ et $H_{\Psi_m}^{\frac{\Delta x}{10}}(n_{mic}, f)$ correspond au "champ d'aliasing spatial" généré par le banc de haut-parleurs. L'ensemble des figures 2.17 présente ce "champ d'aliasing spatial" pour la reproduction de trois sources virtuelles différentes (une source focalisée, une source ponctuelle dans Ω_{Ψ} , et une onde plane se propageant perpendiculairement au banc de haut-parleurs). La ligne noire représentée sur les surfaces correspond au niveau fréquentiel du champ d'aliasing spatial à la fréquence d'aliasing estimée $f_{al}^{\Psi_m}(\Delta x, r_R)$ par notre critère.

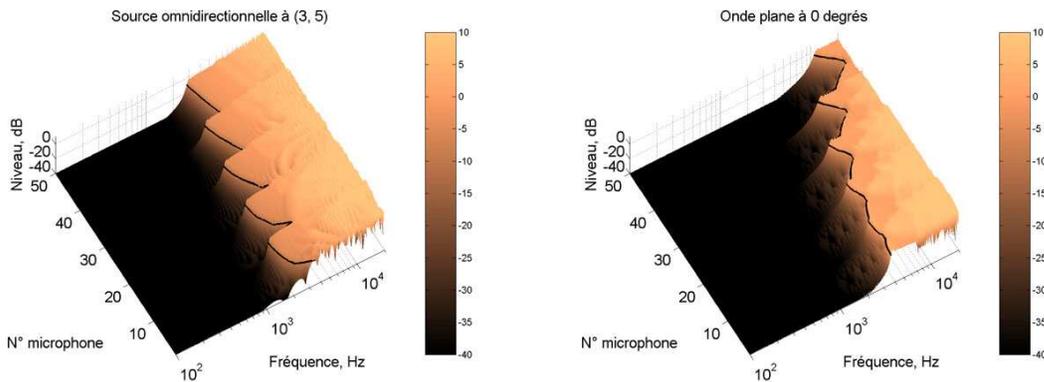
Ces simulations confirment la validité de ce critère en ce qui concerne l'estimation de la fréquence d'aliasing pour des bancs de haut-parleurs de taille finie. Nous avons effectué d'autres tests pour différentes configurations de haut-parleurs/microphones/sources virtuelles. Ils donnent des résultats similaires.

On remarque que cette formule tend vers la formule de la littérature (cf. équation 1.16) dans le cas d'un banc de haut-parleurs infini. La figure 2.18 présente la différence de temps d'arrivée des contributions reçues au niveau du point d'écoute R pour des haut-parleurs en champ lointain. La différence de temps est alors donnée par,

$$\Delta\tau = \frac{\Delta x(\sin \theta_{\Psi_m} + \sin \theta_R)}{c}. \quad (2.57)$$



(a) Source omnidirectionnelle focalisée à (1, 1), sans atténuation haut-parleurs extrémités



(b) Source omnidirectionnelle à (3, 3), sans atténuation haut-parleurs extrémités

(c) Onde plane se propageant perpendiculairement au banc de haut-parleurs, sans atténuation haut-parleurs extrémités

FIG. 2.17 – Validation de la définition de la fréquence d'aliasing spatial pour un banc de 60 haut-parleurs (6 m, 10 cm d'espacement) synthétisant différentes sources virtuelles omnidirectionnelles avec la WFS.

Si le banc de haut-parleurs est infini, quelque soit la position R considérée, si $x \rightarrow \pm\infty$ alors $\theta_R \rightarrow \pm\frac{\pi}{2}$ ce qui donne le résultat de l'équation 1.16.

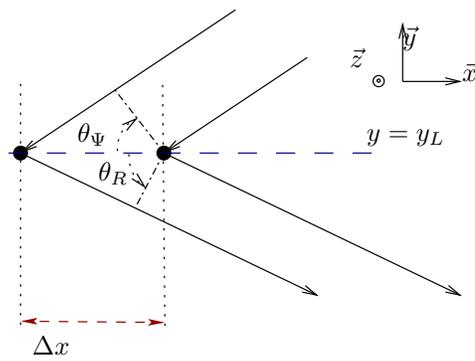


FIG. 2.18 – Calcul des différences de temps pour un banc de haut-parleurs infini

2.5 Discussion : qualité de la synthèse des sources virtuelles en WFS

Dans cette partie, nous avons mis en évidence et caractérisé les "défauts" de la synthèse du champ sonore en Wave Field Synthesis par un banc linéaire de haut-parleurs omnidirectionnels de taille finie.

Nous avons donné la loi d'atténuation des fronts d'onde synthétisés qui apparaît comme une combinaison de celle de la source virtuelle spécifiée et de celle du banc de haut-parleurs. Cette loi d'atténuation modifie ainsi le niveau *local* associé aux sources virtuelles. Cette modification n'affecte pas a priori la transparence de la reproduction. Elle risque cependant de provoquer des problèmes de cohérence lors de la présence de sources multiples et à proximité des bancs de haut-parleurs, où le niveau est a priori très élevé.

Nous avons d'autre part mis en évidence le caractère dispersif de la propagation des fronts d'onde synthétisés. Ce point est typique de l'utilisation d'une géométrie linéaire de sources acoustiques. On a vu cependant qu'elle n'affectait que très peu la transparence sauf à proximité immédiate du banc de haut-parleurs ou de la source dans le cas de sources focalisées. Remarquons que dans ces régions, l'approximation de la phase stationnaire, qui autorise le passage d'une géométrie planaire à une géométrie linéaire de sources secondaires n'est plus justifiée (cf. Annexe A.1).

Nous avons proposé une analyse des conséquences de la taille finie du banc. La diffraction qui en résulte a été décrite en examinant la rupture de la construction du front d'onde en un point de l'espace. On a alors mis en évidence deux conséquences :

1. des oscillations de la réponse fréquentielle en hautes fréquences ;
2. une baisse du niveau synthétisé en basses fréquences.

Nous avons alors montré que la solution retenue dans le cadre de la WFS qui consiste à atténuer le niveau alloué aux haut-parleurs situés aux extrémités permet d'adoucir la transition et de réduire les oscillations de la réponse fréquentielle. Le type de fenêtre proposée par Start (Start, 1997) permet de réduire les oscillations dans la réponse fréquentielle principalement au dessus de 300/400 Hz. La fonction de qualité n'est alors plus dans les limites fixées (± 1 dB) pour garantir la transparence de la reproduction à n'importe quelle position de la zone d'écoute.

Il conviendrait d'étudier d'autres types de fenêtrage (Hanning, Hamming, Blackmann, ...), comme suggéré par Nicol (Nicol, 1999), sur ce type de phénomène. Une dépendance fréquentielle du type de fenêtre utilisé pourrait aussi être étudiée afin d'améliorer la qualité de la restitution.

Nous avons enfin donné une définition originale de la fréquence d'aliasing qui permet de rendre compte de ses variations en fonction de la position d'écoute observées au chapitre 1. Remarquons que la formule donnée reste compatible avec la formule classique dans le cas d'un banc de haut-parleur de taille infinie.

Ce chapitre nous a ainsi permis de mettre en question l'acception selon laquelle la WFS permet de reconstruire un champ avec une qualité identique dans toute la zone d'écoute. Nous avons mis en évidence une *évolution* du champ sonore synthétisé. Parmi les phénomènes étudiés, certains ont un

impact perceptif non négligeable et sont susceptibles d'affecter la transparence de la reproduction. Pour d'autres, tels que la modification de la fréquence d'aliasing et de la loi d'atténuation, l'impact perceptif est difficilement quantifiable et nécessiterait la mise en place d'expériences perceptives.

Chapitre 3

Rayonnement en espace clos

La Wave Field Synthesis, en tant que technique de reproduction holophonique, est dérivée de l'intégrale de Kirchhoff-Helmholtz (cf. partie 1.2). Cette intégrale assure la synthèse exacte du champ sonore produit par un ensemble de sources du sous-espace des sources Ω_{Ψ} dans le sous-espace complémentaire Ω_R (sous-espace de reproduction). La distribution de sources secondaires qui effectue la synthèse du champ est située sur la frontière $\partial\Omega$ entre les deux sous-espaces complémentaires. Dans Ω_{Ψ} , le champ acoustique synthétisé par la distribution de sources secondaires est nul. Par conséquent, un tel système placé dans une salle d'écoute n'interagit aucunement avec celle-ci.

Cette propriété disparaît pour l'intégrale de Rayleigh 1. $\partial\Omega$ est alors un plan infini sur lequel ne sont disposées que des sources omnidirectionnelles. Cette intégrale admet, de fait, l'émission d'un champ *arrière*, symétrique par rapport au plan dans Ω_{Ψ} .

Dans les chapitres 1 et 2, ainsi que dans une grande partie de la littérature, l'analyse du champ synthétisé par le banc de haut-parleurs n'est effectuée qu'en champ libre. Les fronts d'onde synthétisés se propagent ainsi indéfiniment sans jamais se réfléchir sur un quelconque obstacle. Placé dans une salle de restitution (salle d'écoute), les fronts d'onde émis par le banc de haut-parleurs subissent de multiples réflexions sur les parois de la salle. Au bout d'un certain temps ("temps de mélange" (Polack, 1988) qui dépend du volume de la salle), le nombre de réflexions devient très important et on entre dans le régime de réverbération.

Suivant la nature des matériaux, les fronts d'onde se réfléchissent dans une direction symétrique à la direction d'incidence (réflexion spéculaire) et/ou sont répartis de manière uniforme dans l'ensemble des directions de l'espace (réflexion diffuse). On définit des coefficients d'absorption (resp. diffusion) pour déterminer quelle est la portion d'énergie réfléchie (resp. diffusée) par rapport à l'énergie incidente en fonction de la fréquence.

Dans une scène sonore classique, les sources virtuelles sont accompagnées d'un effet de salle synthétique additionnel. Lors de la reproduction, l'acoustique de la salle d'écoute va s'"ajouter", dans un processus de convolution énergétique, à l'ensemble des informations contenues dans la scène sonore. Le résultat perçu par l'auditeur ne correspond alors pas à l'objectif fixé par le

créateur de contenu.

Remarquons toutefois que le créateur de contenu peut l'anticiper lors d'une étape de monitoring dans une salle d'écoute de référence (sur ce point, voir les chapitres 1, 2, et 3 de (Pellegrini, 2001)). Cependant, l'acoustique de ce type de pièce est extrêmement bien contrôlée avec un niveau faible de réflexions précoces et un temps de réverbération relativement court, de l'ordre de 0.2 à 0.5 s indépendamment de la fréquence.

Ce type de conditions acoustiques ne se rencontre pas dans la plupart des pièces d'écoute réelles. Les parois sont généralement relativement réfléchissantes et la géométrie simple ce qui donne un niveau relativement élevé de réflexions précoces et crée des résonances qui réduisent la transparence de la reproduction.

Dans ce chapitre et plus généralement dans ce document, on se limite à la partie précoce du champ réfléchi ("premières réflexions").

La perception par le système auditif des réflexions est un phénomène complexe qui reste mal compris. Il est en effet difficile de présenter des situations réalistes desquelles on puisse tirer une conclusion générale et c'est sur ce point que butent la plupart des études de la littérature. Elles considèrent soit une situation simpliste en présentant une unique réflexion en plus du son direct (sur ce point voir par exemple (Burgtof, 1961) (Olive & Toole, 1989) (Blauert, 1999)), soit une situation complexe mais particulière et donc difficilement généralisable (Bech, 1995) (Bech, 1996) (Bech, 1998).

Toutefois, on peut identifier un certain nombre d'effets indésirables qui peuvent être réduits si le niveau des réflexions suffisamment bas. Pour des retards très faibles (< 1 ms, par rapport au temps d'arrivée du son direct), on assiste à une délocalisation de la source sonore. A partir de 1 ms, il existe un seuil en dessous duquel la réflexion est inaudible. Au-dessus de ce seuil, des effets de coloration apparaissent puis pour des niveaux supérieures un effet d'élargissement de la source sonore. Pour des niveaux encore supérieurs, la réflexion devient un événement auditif distinct que l'on nomme écho (Blauert, 1999).

Ces trois effets sont considérés comme des défauts pour la reproduction sonore qui réduisent la transparence. Le niveau de ces réflexions doit ainsi être maintenu faible, en particulier dans les 20/30 premières ms après l'arrivée du son direct afin d'éviter d'introduire de la coloration (Rubak, 2004) et/ou une modification de l'organisation spatiale de la scène sonore.

Dans ce chapitre, on se place dans le cas où le banc de haut-parleur est intégré dans les parois de la salle d'écoute. Cette situation spécifique est en fait la plus pratique étant donné le nombre important de haut-parleurs mis en œuvre.

Dans un premier temps, nous décrirons les modalités d'interaction d'un banc de haut-parleur linéaire avec une salle d'écoute de forme rectangulaire. Cette étude concerne principalement les réflexions précoces dans le plan horizontal. On montre ainsi que les limitations de la WFS (ondes de type cylindrique, fenêtrage, aliasing) engendrent une interaction avec la salle d'écoute qui dépend de la source virtuelle considérée.

Nous proposerons ensuite une méthode de *compensation* des réflexions du plan horizontal. Cette

méthode se présente dans une certaine mesure comme une extension de la Wave Field Synthesis. Elle requiert la présence de bancs de haut-parleurs intégrés dans chaque paroi verticale de la salle de restitution. Cette méthode fonctionne avec un certain nombre de limitations que nous exposerons et commenterons.

3.1 Interaction du système WFS avec la salle d'écoute

Cette partie propose de décrire l'interaction du système WFS avec la salle d'écoute par l'utilisation de la méthode des sources images. Cette méthode permet de décrire simplement les réflexions de premier ordre dans l'ensemble de la salle d'écoute soit dans l'ensemble de la zone de reproduction associée au système WFS.

On rappelle dans un premier temps les caractéristiques du rayonnement d'un banc de haut-parleur linéaire synthétisant une source virtuelle Ψ . La réflexion sur une paroi de la salle est ensuite décrite par une méthode dérivée de la méthode des sources images. Nous présentons enfin quelques simulations pour donner les caractéristiques principales de l'interaction du système WFS avec la salle d'écoute.

3.1.1 Rayonnement en 3 dimensions

Le rayonnement effectif en 3 dimensions d'un banc linéaire de haut-parleurs omnidirectionnels présente une symétrie de révolution autour de l'axe du banc de haut-parleurs (cf. partie 1.2.1). La synthèse d'une source virtuelle Ψ par un tel banc de haut-parleur provoque l'émission d'ondes à "symétrie cylindrique". Nous proposons alors d'exploiter cette caractéristique afin de déterminer la position de la source virtuelle Ψ_δ perçue par un auditeur situé en dehors de la zone d'écoute privilégiée (intersection du sous-espace de reproduction avec le plan horizontal). δ correspond au site de la position d'écoute par rapport au banc de haut-parleur. La zone d'écoute privilégiée est donnée par $\delta = 0$. Dans la suite, le site est donné en radian. La figure 3.1 représente les fronts

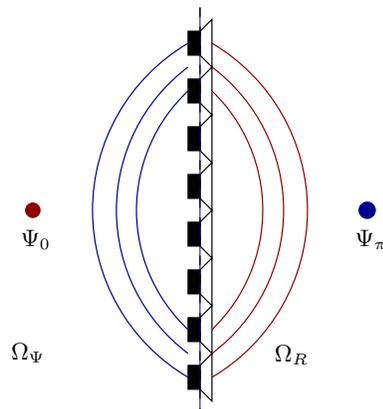


FIG. 3.1 – Champ synthétisé dans l'ensemble du plan horizontal

d'onde synthétisés par un banc de haut-parleurs omnidirectionnels dans le plan horizontal. Les

fronts d'onde synthétisés par le banc de haut-parleur dans Ω_R (zone d'écoute), correspondent à la position de source cible Ψ_0 . En revanche, de l'autre côté du banc de haut-parleurs, les fronts d'onde synthétisés semblent provenir d'une position symétrique à Ψ_0 , Ψ_π . Cette observation peut

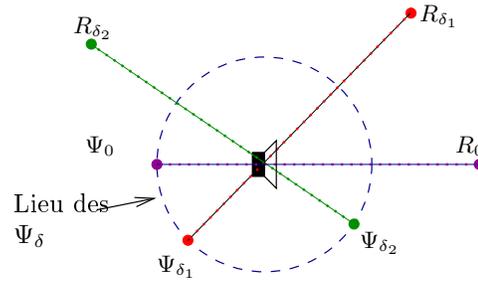


FIG. 3.2 – Localisation des sources virtuelles dans le plan vertical

se généraliser pour n'importe quel site d'écoute en exploitant la symétrie de révolution autour de l'axe du banc de haut-parleurs. Le site δ_Ψ de la source virtuelle perçue est ainsi donné pour un site d'écoute δ_R par $\delta_\Psi = \delta_R + \pi$. La figure 3.2 donne ainsi l'emplacement de la source virtuelle perçue Ψ_{δ_i} à une position d'écoute R_{δ_i} ainsi que le lieu des sources virtuelles Ψ_δ en fonction du site δ de la position d'écoute.

La synthèse correcte du champ sonore est limitée dans le plan horizontal à une zone de visibilité de la source virtuelle au travers du banc de haut-parleur si celui-ci a une taille finie. Par symétrie de révolution autour de l'axe du banc de haut-parleur, on peut définir un sous-espace de visibilité de la source virtuelle Ψ à travers le banc de haut-parleur de taille finie. En dehors de cette zone de visibilité, des ondes diffractées sont émises. Celles-ci semblent provenir des extrémités du banc de haut-parleur. Leurs caractéristiques (contenu fréquentiel, étalement temporel) dépend de la source virtuelle et de la position d'écoute (cf. partie 2.3).

3.1.2 Réflexion des fronts d'onde synthétisés

La réflexion d'un champ incident sur une paroi supposée infinie se décrit par la méthode des sources images (Cerveau, 1999). Dans la méthode des sources images, les fronts d'onde réfléchis sur une paroi sont émis par une source "image". La position de la source image est symétrique (par rapport à la paroi) à celle de la source primaire qui est l'origine des fronts d'onde. Cette source image d'"ordre" 1 va à son tour se réfléchir sur les parois de la pièce de restitution, créant des sources images d'ordre supérieur.

Le champ synthétisé dans la salle d'écoute est ainsi donné par la superposition des champs synthétisés par la source primaire, et par toutes les sources images à travers les parois de la salle d'écoute.

La figure 3.3 rend compte de la réflexion sur une paroi du champ sonore synthétisé par un banc de haut-parleur LS_{dir} synthétisant une source Ψ . Le champ réfléchi correspond à celui d'une source virtuelle image Ψ_{dir}^{im} synthétisée par un banc de haut-parleurs image LS_{dir}^{im} .

La visibilité de Ψ_{dir}^{im} est limitée, à la diffraction près, d'une part par la taille de LS_{dir}^{im} (uniquement

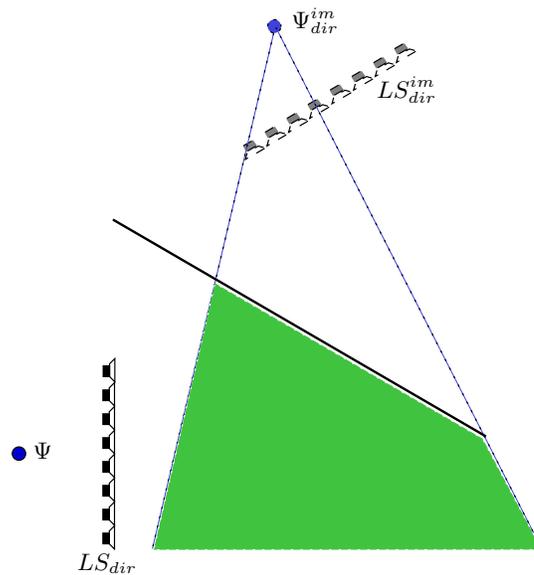


FIG. 3.3 – Réflexion du champ acoustique synthétisé sur une paroi verticale de taille finie

jusqu'à la fréquence d'aliasing spatial) et d'autre part par l'étendue spatiale de la paroi (pour n'importe quelle fréquence). La réflexion sur une paroi horizontale (cf. figure 3.4) est décrite de

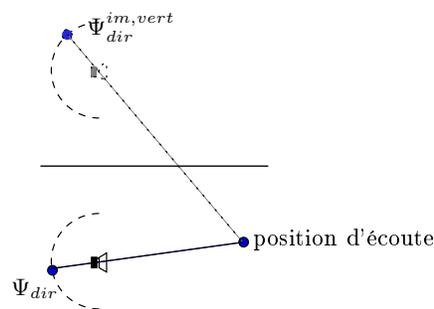


FIG. 3.4 –

manière similaire. L'exploitation de la symétrie de révolution permet alors de donner la position de la source virtuelle image perçue dans le plan vertical.

On peut alors décrire la réflexion des fronts d'onde émis par un banc de haut-parleurs, pour la synthèse d'une source virtuelle Ψ , comme émis par un banc de haut-parleur image pour la synthèse d'une source Ψ^{im} , image de Ψ par rapport à la paroi considérée.

3.1.3 Simulations : champ réfléchi sur les parois d'une salle rectangulaire

Considérons un banc linéaire de 60 haut-parleurs omnidirectionnels intégrés dans l'une des parois d'une salle rectangulaire de 6 mètres de large et de 4 mètres de profondeur. Les murs sont constitués de parois parfaitement réfléchissantes, le sol et le plafond étant parfaitement absorbants.

On caractérise la fonction de transfert électroacoustique du banc de haut-parleurs pour la synthèse d'une source Ψ dans la salle de restitution au niveau d'un ensemble de microphones. Les

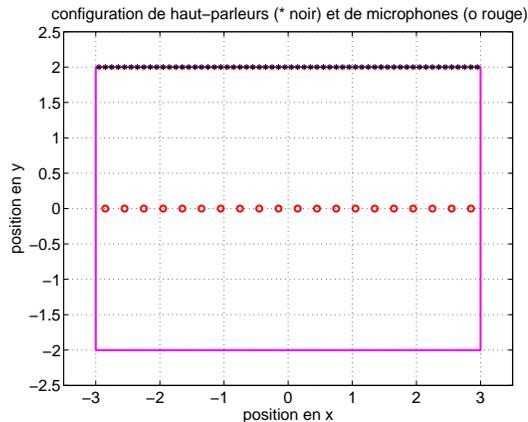


FIG. 3.5 – Position des haut-parleurs et des microphones dans la salle d'écoute

microphones sont situés sur une ligne parallèle au banc de haut-parleurs. Le champ réfléchi est obtenu en utilisant la méthode des sources images pour chaque haut-parleur du banc. La réponse impulsionnelle de chaque haut-parleur au niveau de chaque microphone est ainsi obtenue par propagation du champ émis par l'ensemble des sources images jusqu'à la position de microphone considérée en appliquant des critères de visibilité strictes. On applique à chacune des réponses impulsionnelles obtenues le filtre de correction en $\sqrt{j\bar{k}}$, le retard ainsi que l'atténuation pour la synthèse de la source Ψ par la WFS. La fonction de transfert électroacoustique est donnée en sommant l'ensemble des contributions de chacun des haut-parleurs au niveau du microphone considéré.

Les figures 3.6 représentent les réponses impulsionnelles filtrées passe-bas (sous la fréquence d'aliasing) obtenues en appliquant la méthode des sources images uniquement jusqu'à l'ordre 1. On ne considère que les sources images du plan horizontal dans un but de clarté de description. Les réflexions synthétisées correspondent effectivement à des sources virtuelles images de la source virtuelle cible. La visibilité de ces sources images à travers la paroi dépend de la position de la source virtuelle Ψ .

Pour la source virtuelle focalisée les fronts d'onde associés au son direct et aux composantes réfléchies sur les murs latéraux et le mur du fond sont bien visibles (cf. figure 3.6(a)). Néanmoins, en comparaison avec une source réelle située à cette position dans la pièce, la réflexion sur le mur frontal, dans lequel sont intégrés les haut-parleurs, est absente ainsi que toutes les réflexions d'ordre supérieur dont elle est à l'origine. La sensation que cette source est intérieure à la salle d'écoute risque ainsi d'être perturbée par cette partie absente du champ réfléchi. Cet effet peut toutefois être corrigé par la synthèse additionnelle d'une source virtuelle additionnelle symétrique par rapport au mur frontal (Caulkins *et al.*, 2003).

Le champ synthétisé pour la reproduction de la source virtuelle située au niveau de l'extrémité droite du banc de haut-parleur n'est pratiquement pas réfléchi sur le mur de droite mais se

réfléchi sur le mur de gauche et celui du fond (cf. figure 3.6(b)). On peut alors s'attendre à une délocalisation et/ou d'élargissement de la taille de la source sonore perçue ainsi qu'à une coloration du son direct en fonction de la position d'écoute.

Les figures 3.6(c) et 3.6(d) mettent en évidence un phénomène de concentration de l'énergie dans l'axe principal de la pièce pour des sources frontales lointaines qui n'éclairent pas ou peu les murs latéraux. Ainsi, l'énergie réfléchie précoce vient principalement des directions frontales et arrières. Pour des salles de grande taille et à cause de l'absence de réflexions précoces latérales, on risque d'"entendre" le mur du fond. La densité de réflexion précoce est alors sensiblement réduite. Ceci réduit les phénomènes de masquage entre les réflexions et provoque une augmentation des risques de perception de coloration ou d'écho.

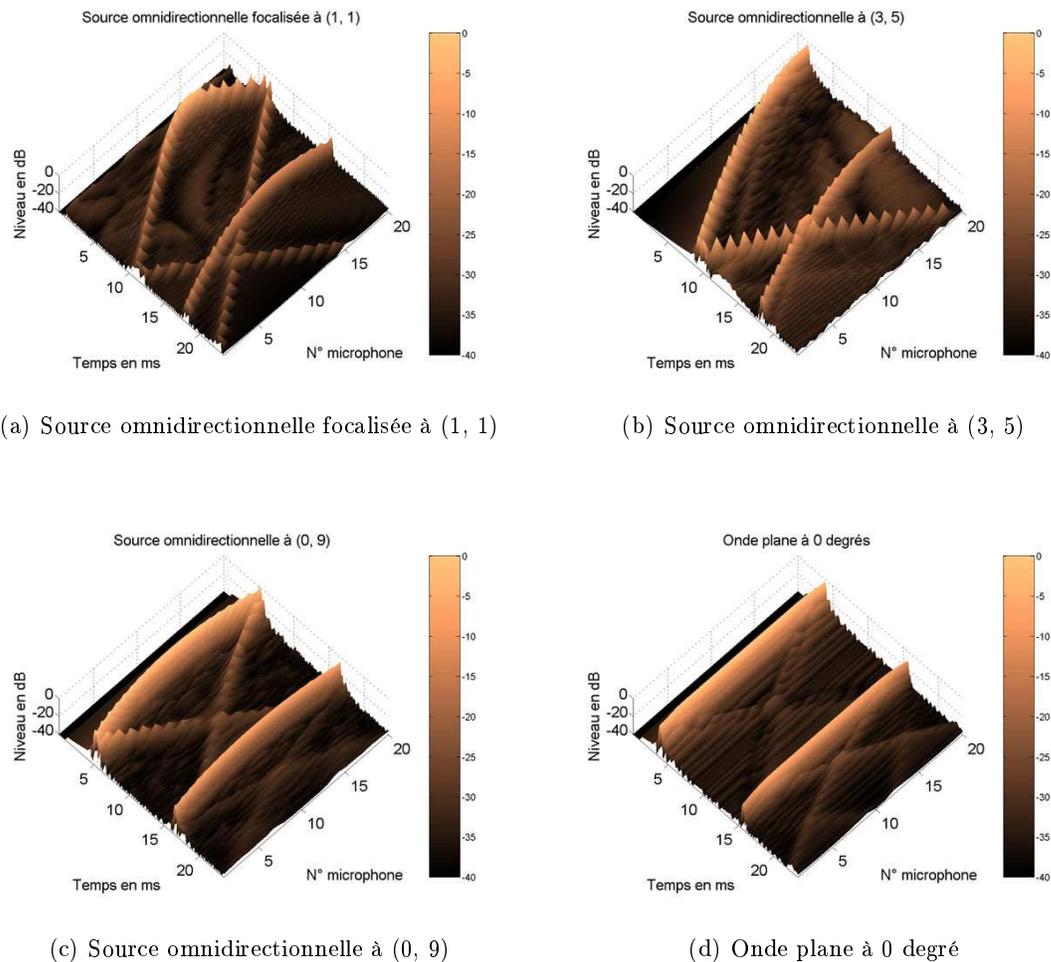


FIG. 3.6 – Réponses impulsionnelles filtrées passe-bas (sous la fréquence d'aliasing spatial) obtenues en utilisant la méthode des sources images jusqu'à l'ordre 1

Les figures 3.7 représentent les réponses impulsionnelles filtrées passe-haut (au-dessus de la fréquence d'aliasing) obtenue en appliquant la méthode des sources images jusqu'à l'ordre 4 uni-

quement dans le plan horizontal.

On remarque alors que les réponses impulsionnelles obtenues sont immédiatement denses et ne semblent pas privilégier de direction particulières. Ceci est dû au trainage introduit par le phénomène d'aliasing spatial et au caractère globalement plus "omnidirectionnel" du rayonnement synthétisé au-dessus de la fréquence d'aliasing spatial.

Par son caractère "diffus", le champ synthétisé au-dessus de la fréquence d'aliasing ne risque pas d'être perçu en tant qu'écho. L'aliasing spatial réalise un "étalement" temporel du signal musical diffusé qui limite les risques de modifications du timbre.

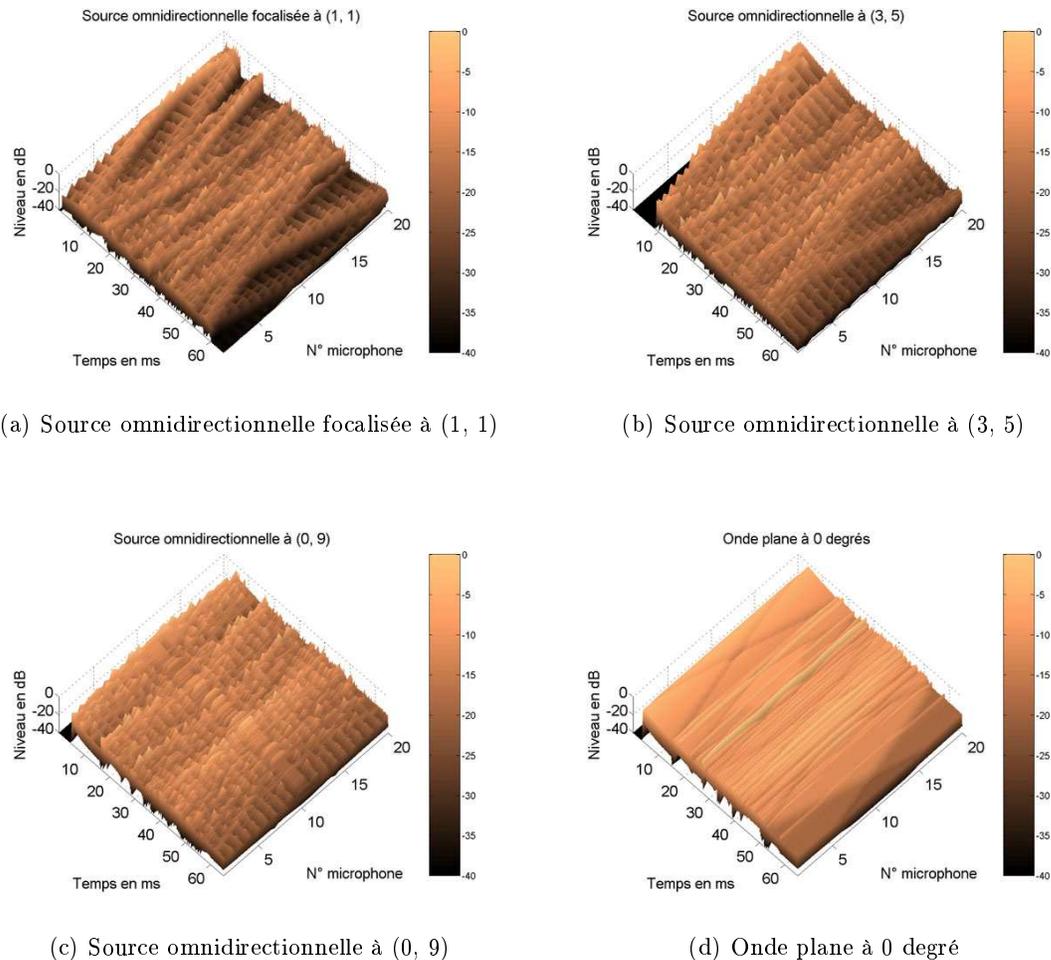


FIG. 3.7 – Réponses impulsionnelles filtrées passe-haut (au-dessus de la fréquence d'aliasing spatial) obtenues en utilisant la méthode des sources images jusqu'à l'ordre 1

3.1.4 Discussion : impression spatiale et source virtuelle synthétisée

Dans la partie précédent, nous avons mis en évidence la nature du champ réfléchi à l'intérieur de la salle pour la synthèse de différentes sources virtuelles par un banc linéaire de haut-parleurs omnidirectionnels. Nous avons alors discuté des risques de coloration du son direct par ces com-

posantes réfléchies. Nous avons aussi remarqué que la distribution *spatiale* du champ réfléchi *dépend* de la source virtuelle synthétisée. Reste alors à savoir quelle sont les conséquences en terme d'impression spatiale perçue par un auditeur.

La caractérisation objective de l'impression spatiale peut être donnée par un ensemble de critères qui sont calculés à partir de la distribution temps/fréquence/espace de l'énergie reçue au niveau d'un position d'écoute. Parmi ces critères, citons *l'efficacité latérale LE* (Barron & Marshall, 1981) proposée par Barron et Marshall. Ce critère porte uniquement sur l'énergie précoce d'une salle. Il est défini en *dB* par,

$$LE = 10 \times \log_{10} \left(\frac{\int_{t=5ms}^{80ms} p_{fig8}^2(t)}{\int_{t=0ms}^{80ms} p_{omni}^2(t)} \right), \quad (3.1)$$

où, $p_{omni}(t)$ est la réponse impulsionnelle mesurée par un microphone omnidirectionnel pour laquelle le son direct est donné à $t = 0ms$ et, $p_{fig8}(t)$ est la réponse impulsionnelle mesurée par un microphone figure 8 (à directivité dipolaire) dont le zéro est dans la direction de la source sonore. Cette mesure traduit ainsi le quotient de l'énergie précoce latérale par rapport à l'énergie précoce totale. La valeur limite de 80 ms est donnée par Barron et Marshall pour une salle de concert typique. Ils recommandent en réalité d'adapter cette limite temporelle afin de prendre en compte uniquement les réflexions précoces, c'est à dire, être donné par le temps de mélange de la salle.

La définition de cette mesure est dérivé des travaux consignés dans la thèse de Barron (Barron, 1974). Il y établit l'influence des réflexions latérales sur l'augmentation de l'impression d'espace (spaciousness en anglais). Barron et Marshall montrent que l'impression d'espace est une quantité qui augmente en relation avec *LE*. Au niveau fréquentiel, certains auteurs soutiennent l'importance particulière des réflexions latérales en dessous de 1,5 *kHz* pour l'impression d'espace (d'après (Blauert, 1999)).

Dans le cas de la WFS, on voit typiquement que *LE* subit des variations relativement importantes en-dessous de la fréquence d'aliasing. Ces variations dépendent de l'éloignement de la source virtuelle synthétisée. Plus celle-ci est proche, plus *LE* est élevé, plus la source est lointaine, plus *LE* est faible, particulièrement pour l'onde plane. Ces variations s'accompagne de modifications de la sensation d'espace associée à la salle d'écoute.

Encore une fois, une baisse du niveau des réflexions sur l'ensemble de la zone d'écoute permettrait d'atténuer ces variations.

3.2 Atténuation des réflexions par la WFS

Dans cette partie, nous proposons une méthode qui vise la compensation des réflexions sur les murs verticaux de la salle d'écoute. Cette méthode utilise la WFS pour synthétiser une source virtuelle de compensation (de la source virtuelle image) par un banc de haut-parleurs intégré dans la paroi sur laquelle se produit la réflexion.

Après avoir décrit la méthode proposée, nous l'illustrons par la compensation du champ réfléchi sur une paroi unique. On mettra ainsi en évidence un certain nombre de limitations qui réduisent son efficacité et sont liées à la WFS. En fin de partie, on donnera des indications pour la mise en œuvre de cette méthode en pratique et sa formulation en tant qu'extension de la Wave Field Synthesis.

3.2.1 Compensation du champ réfléchi par une paroi

Le champ synthétisé par un banc de haut-parleur LS_{dir} se réfléchit sur une paroi i de la salle d'écoute en formant un banc de haut-parleur image $LS_{dir}^{im,i}$. Ce banc synthétise une source virtuelle $\Psi_{dir}^{im,i}$, image de la source cible Ψ_{dir} . Nous proposons d'intégrer un banc de haut-parleurs $LS_{comp,i}$ dans la paroi i et de synthétiser une source virtuelle $\Psi_{comp,i}$ à la position de $\Psi_{dir}^{im,i}$ (cf. figure 3.8). $\Psi_{comp,i}$ est synthétisée en opposition de phase par rapport à Ψ_{dir} .

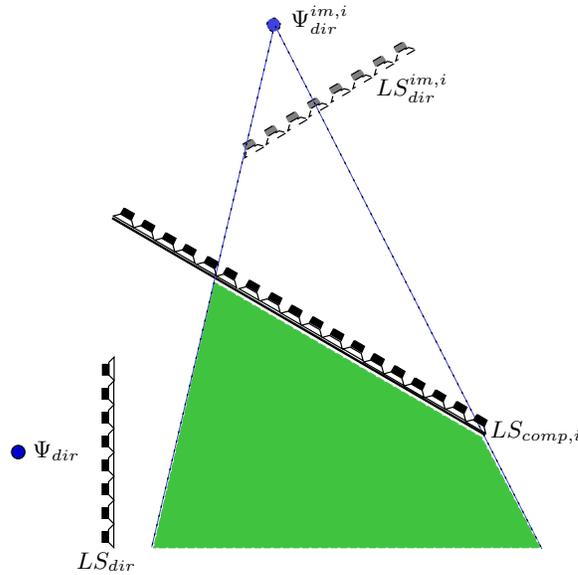


FIG. 3.8 – Compensation du champ réfléchi par l'utilisation d'un banc de haut-parleurs encastré dans la paroi

Dans le cas idéal où les deux sources sont parfaitement synthétisées (i.e. Ψ_{dir} , $\Psi_{comp,i}$ sont émises par une distribution planeaire continue infini de source secondaires omnidirectionnelles idéales), le champ acoustique associé à $\Psi_{comp,i}$ compense totalement le champ réfléchi (issu de $\Psi_{dir}^{im,i}$ dans l'ensemble de la salle d'écoute). Ainsi, les sources images d'ordre supérieur, issues de $\Psi_{dir}^{im,i}$, sont, à leur tour, *automatiquement* compensées par les sources images de la source de compensation $\Psi_{comp,i}$.

En réalité, l'utilisation de bancs de haut-parleurs linéaires échantillonnés de taille finie réduit les performances de cette méthode :

1. Tous les bancs de haut-parleurs utilisés sont linéaires, situés dans le plan horizontal. La compensation n'est ainsi effective que dans le plan horizontal pour des sources images

situées dans le plan horizontal.

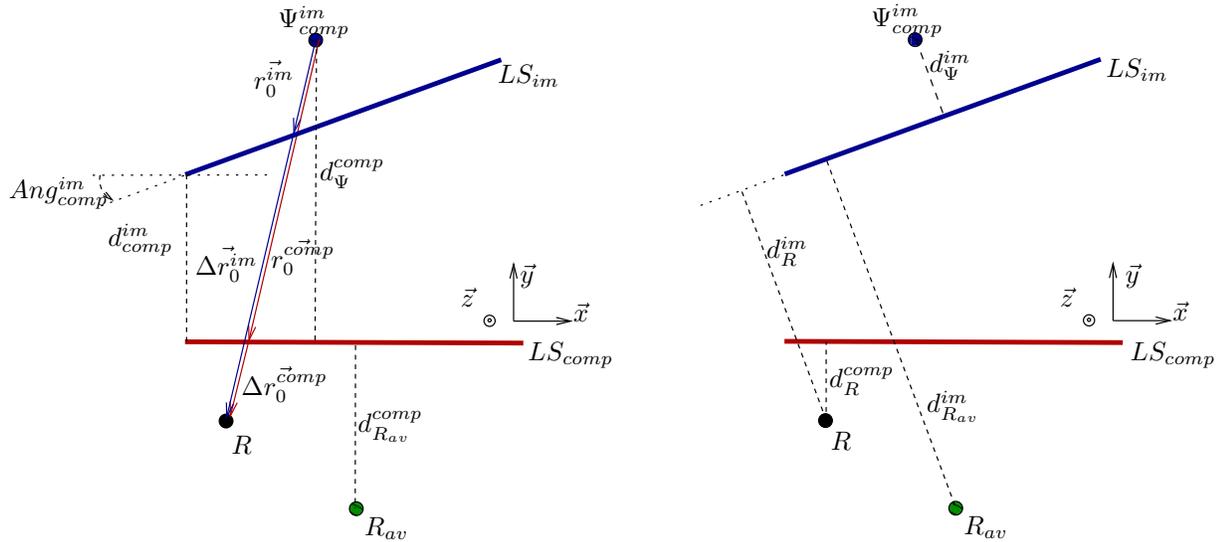
2. La loi d'atténuation que subissent les fronts d'onde associés aux sources virtuelles synthétisées dépend aussi de la position du banc de haut-parleurs utilisé (cf. partie 2.2.1). Ainsi le niveau des fronts d'onde synthétisés par le banc de haut-parleurs image et le banc de haut-parleurs de compensation ne subissent pas la même loi d'atténuation. Il n'y a donc pas compensation idéale possible dans l'ensemble de la zone d'écoute.
3. Il est impossible de tenter une quelconque annulation du champ réfléchi au-dessus de la fréquence d'aliasing spatial. Le champ synthétisé présente en effet une structure temporelle complexe qui évolue de façon "incohérente" dans l'ensemble de la zone d'écoute (cf. partie 2.4). La structure du champ d'aliasing spatial provoque cependant un étalement temporel des réflexions sur les parois de la salle d'écoute et réduit ainsi les risques d'introduction de coloration et de modification de l'impression spatiale (cf. partie 3.1.4).
4. L'utilisation de bancs de haut-parleurs de taille finie provoque une réduction de la visibilité de la source virtuelle à travers le banc. Un champ diffracté est émis dans et en dehors de la zone de visibilité qui dépend de la position de la source virtuelle, de la taille du banc de haut-parleur et de la position d'écoute (cf. partie 2.3). Il n'y a aucune raison que le banc de compensation soit de la même taille que le banc utilisé pour la synthèse du son direct. Sauf si les bancs sont coïncidents, ni la position de la source virtuelle, ni la position d'écoute relativement aux deux bancs de haut-parleurs ne sont les mêmes. D'autre part, on ne doit utiliser que les haut-parleurs de $LS_{comp,i}$ situés dans la zone de visibilité de $\Psi_{dir}^{im,i}$ à travers $LS_{dir}^{im,i}$ (cf. figure 3.8).

Afin d'évaluer l'influence de ces imperfections sur la méthode proposée, nous utilisons une configuration de test représentée dans la figure 3.9. Considérons une paire banc de haut-parleur image LS_{im} /banc de haut-parleurs de compensation LS_{comp} . La source image Ψ_{im} et la source de compensation Ψ_{comp} sont synthétisées par la WFS, en opposition de phase, par les bancs LS_{im} et LS_{comp} . La profondeur moyenne d'écoute, pour laquelle le niveau des fronts d'onde synthétisés est correct, est choisi de telle manière à ce que Ψ_{im} et Ψ_{comp} se compensent parfaitement en $(0, 0)$.

Dans la figure 3.9 ces deux sources coïncidentes sont nommées Ψ_{comp}^{im} . Les bancs de haut-parleurs présentent un angle relatif noté Ang_{comp}^{im} et sont espacés d'une distance d_{comp}^{im} pour $Ang_{comp}^{im} = 0$. Dans la suite, la paroi dans laquelle est intégré LS_{comp} est considérée comme parfaitement réfléchissante.

Loi d'atténuation respective des sources virtuelles

Considérons tout d'abord que LS_{im} et LS_{comp} sont des bancs de haut-parleurs continus infinis. La loi d'atténuation des fronts d'onde synthétisés pour Ψ_{comp} et Ψ_{im} est donnée par l'équation

FIG. 3.9 – Bancs de haut-parleurs *image* et *compensation*

2.23. On a alors

$$Att_{\Psi}^{im}(r_R) = \sqrt{\frac{d_{R_{av}}^{im}}{d_R^{im}}} \sqrt{\frac{d_R^{im} + d_{\Psi}^{im}}{d_{R_{av}}^{im} + d_{\Psi}^{im}}} \frac{1}{4\pi(\Delta r_0^{im} + r_0^{im})}, \quad (3.2)$$

et,

$$Att_{\Psi}^{comp}(r_R) = \sqrt{\frac{d_{R_{av}}^{comp}}{d_R^{comp}}} \sqrt{\frac{d_R^{comp} + d_{\Psi}^{comp}}{d_{R_{av}}^{comp} + d_{\Psi}^{comp}}} \frac{1}{4\pi(\Delta r_0^{comp} + r_0^{comp})}. \quad (3.3)$$

On définit alors la *réduction relative de niveau* de la réflexion

$$Red_{\Psi_{im}}(r_R) = \frac{Att_{\Psi}^{im}(r_R) - Att_{\Psi}^{comp}(r_R)}{Att_{\Psi}^{im}(r_R)}. \quad (3.4)$$

Il s'agit alors de ce l'on pourrait aussi nommer "l'atténuation" relative de la réflexion. Cependant pour éviter toute confusion entre "loi d'atténuation" et "atténuation relative des réflexions", on a choisit ce terme de réduction relative de niveau de la réflexion. Pour simplifier, dans la suite, on ne parlera que de réduction relative de niveau.

Remarquons que $\Delta r_0^{im} + r_0^{im} = \Delta r_0^{comp} + r_0^{comp}$ (cf. figure 3.9). On peut alors écrire

$$Red_{\Psi_{im}}(r_R) = 1 - \sqrt{\frac{d_{R_{av}}^{comp}}{d_R^{comp}}} \sqrt{\frac{d_R^{im}}{d_{R_{av}}^{im}}} \sqrt{\frac{d_R^{comp} + d_{\Psi}^{comp}}{d_{R_{av}}^{comp} + d_{\Psi}^{comp}}} \sqrt{\frac{d_R^{im} + d_{\Psi}^{im}}{d_{R_{av}}^{im} + d_{\Psi}^{im}}}. \quad (3.5)$$

La réduction relative de niveau est ainsi "complète" pour $r_R = r_{R_{av}}$. C'est en effet le point pour lequel le niveau des sources virtuelles Ψ_{im} et Ψ_{comp} est correctement reproduit. Il est donc égale pour les deux sources et celles-ci se compensent parfaitement. En dehors de ce point, la réduction relative de niveau dépend de,

- la position du point d'écoute de référence ($r_{R_{av}}$),
- la position relative des deux bancs de haut-parleurs (distance, angle relatif),

– la position de la source Ψ .

Dans la suite, le point d'écoute de référence $r_R = r_{R_{av}}$ sera situé à l'origine du repère cartésien.

Dans le cas où les deux bancs sont parallèles, remarquons que $d_{R_{av}}^{comp} + d_{\Psi}^{comp} = d_{R_{av}}^{im} + d_{\Psi}^{im} = |y_{\Psi}|$ et $d_R^{comp} + d_{\Psi}^{comp} = d_R^{im} + d_{\Psi}^{im} = |y_{\Psi} - y_R|$. L'équation 3.5 se simplifie largement et prend la forme

$$Red_{\Psi_{im}}(r_R) = 1 - \sqrt{\frac{|y_{LS_{comp}}|}{|y_{LS_{comp}} - y_R|}} \sqrt{\frac{|y_{LS_{im}} - y_R|}{|y_{LS_{im}}|}}. \quad (3.6)$$

La réduction relative de niveau dépend donc *uniquement* de la distance du point d'écoute aux deux bancs de haut-parleurs.

L'ensemble des figures 3.10 donne la réduction relative de niveau pour $Ang_{comp}^{im}=0$. Les 4 situations présentées correspondent à différentes positions du banc de compensation par rapport au centre du repère, point d'écoute de référence où la compensation est totale. Pour chaque figure, le paramètre de variation entre les courbes est la distance entre le banc image et le banc de compensation d_{comp}^{im} .

De manière générale, on note que la réduction relative de niveau est plus importante lorsque les deux bancs sont proches. Par ailleurs, le niveau de la source de compensation peut s'avérer supérieur à celui de la source image à proximité du banc de compensation. Ce phénomène de sur-compensation est limité lorsque le point d'écoute de référence est proche du banc de compensation. Dans ce cas, la réduction relative de niveau est moins importante pour des positions d'écoute plus éloignées du banc de compensation. Remarquons que ces positions d'écoute plus éloignées correspondent, à partir d'une certaine distance, aux sources images d'ordre supérieur. Lorsque le point de référence est *proche* du banc de compensation, on focalise la compensation sur les sources images d'ordre 1 au détriment de la compensation des composantes réfléchies d'ordre élevé.

Dans le cas où les deux bancs de haut-parleurs ne sont pas parallèles, la réduction relative de niveau dépend alors aussi de la position de la source synthétisée.

L'ensemble des figures 3.11 présente la réduction relative de niveau lorsque $Ang_{comp}^{im}=60$ degrés. Les 4 situations présentées correspondent à différentes valeurs de $y_{LS_{comp}}$ (distance du banc de compensation au point d'écoute de référence) et de d_{comp}^{im} (distance entre le banc image et le banc de compensation). Pour chaque figure, le paramètre de variation entre les courbes est la distance d_{Ψ}^{im} de la source au banc image.

Le comportement général de la méthode est similaire au cas où les bancs de haut-parleurs sont parallèles. Remarquons toutefois que la réduction relative de niveau est plus efficace pour des sources virtuelles proches du banc image. La différence est particulièrement marquée lorsque la distances entre les bancs de haut-parleurs est faible.

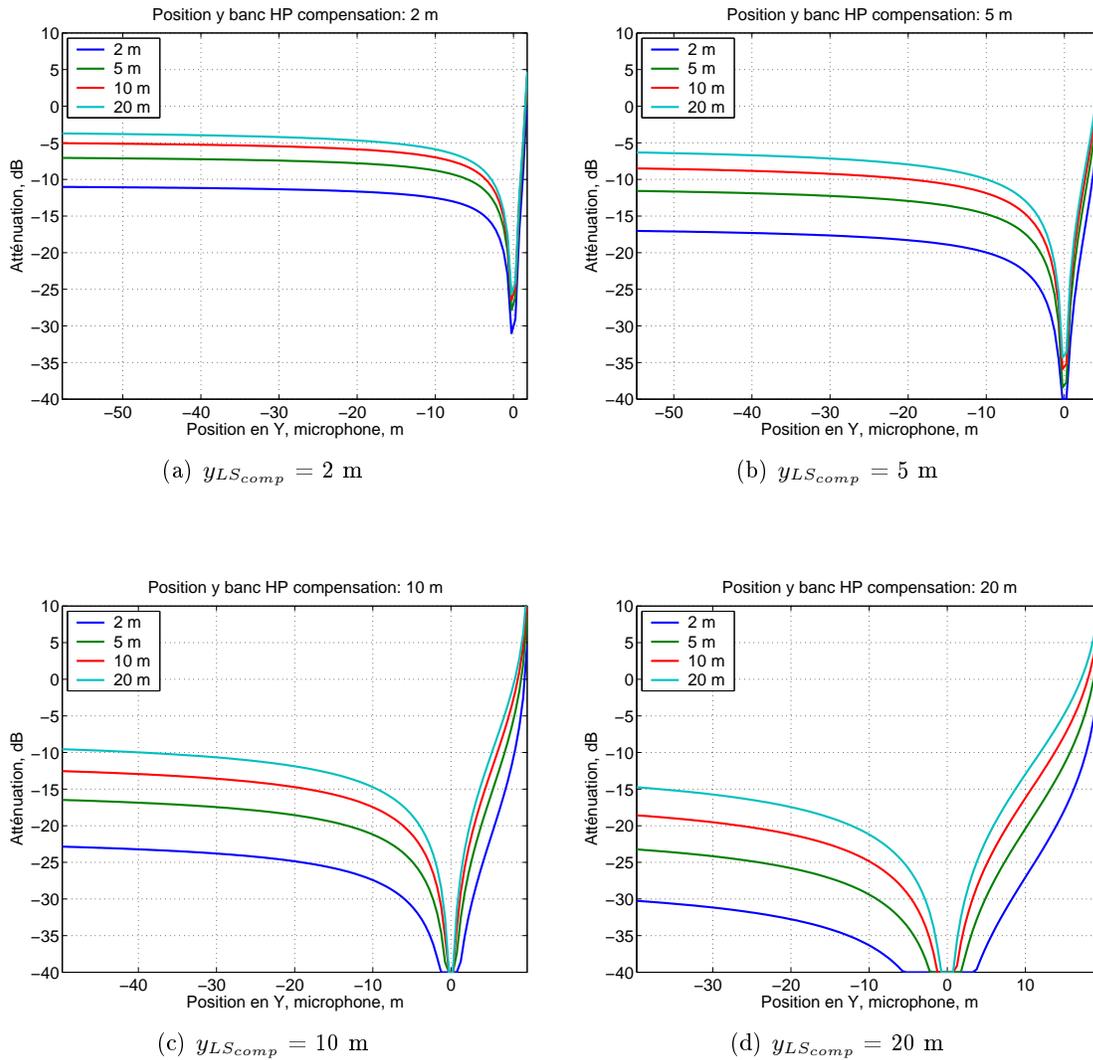


FIG. 3.10 – réduction relative de niveau de la réflexion dans le plan horizontal dans le cas où les deux bancs de haut-parleurs sont parallèles. Le paramètre de variation est la distance d_{comp}^{im} entre les deux bancs de haut-parleurs

La position du point de référence $r_{R_{av}}$ au banc de compensation apparaît comme donc une *variable de réglage* de la réduction relative de niveau en fonction de la position d'écoute.

- Lorsque le point $r_{R_{av}}$ est *proche* du banc de compensation, la réduction de niveau est maximale à proximité du banc de compensation et se dégrade au delà.
- Lorsque le point $r_{R_{av}}$ est plus distant du banc de compensation, il y a *sur-compensation* à proximité du banc de compensation. Cependant, la réduction relative de niveau est plus importante au delà.

Les positions d'écoute pour lesquelles la distance au banc de compensation est plus importante que la distance entre le banc image et le banc de compensation correspond en réalité à une position d'écoute *en dehors* de la "salle". Ainsi, la réduction relative de niveau observée correspond à celle d'une réflexion d'ordre supérieur. Suivant la position du point d'écoute de référence par rapport

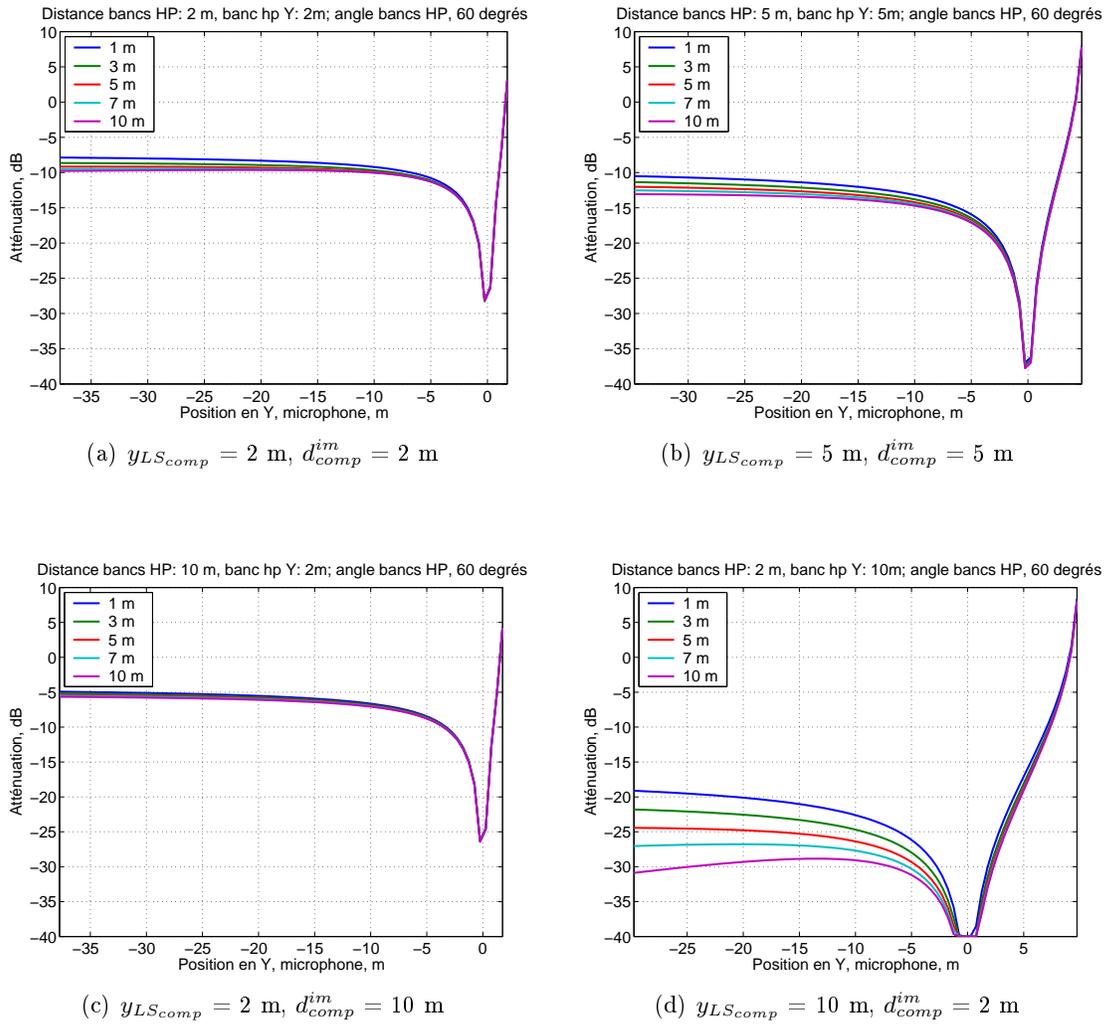


FIG. 3.11 – Réduction relative de niveau de la réflexion dans le plan horizontal dans le cas où $Ang_{comp}^{im} = 60$ degrés pour différentes valeurs de $y_{LS_{comp}}$ et d_{comp}^{im} . Le paramètre de variation est la distance d_{Ψ}^{im} de la source au banc image

au banc de compensation, on règle alors soit l'efficacité de la compensation des réflexions d'ordre 1 ($r_{R_{av}}$ à proximité du banc de compensation); soit l'efficacité de la compensation des réflexions d'ordre supérieur.

Nous nous sommes fixé comme objectif principal de réduire le niveau des réflexions précoces et ainsi les réflexions d'ordre 1. Nous recommandons ainsi de placer le point de référence à proximité du banc de compensation.

Performances en dehors du plan horizontal

Pour une position d'écoute située en dehors du plan horizontal, le champ synthétisé par les deux bancs de haut-parleurs étant "incorrect", il n'y a, a priori, pas de compensation possible. Néanmoins, rappelons que le champ rayonné par un banc linéaire de haut-parleurs omnidirectionnels présente une symétrie de révolution autour de son axe. Nous proposons alors d'évaluer le comportement des champs synthétisés en dehors du plan horizontal par le banc image et le banc de compensation.

Par symétrie de révolution autour de l'axe des bancs de haut-parleurs, nous définissons une situation *équivalente* dans le plan horizontal. La figure 3.12 exploite cette idée dans le cas où les deux bancs de haut-parleurs sont parallèles. Elle représente un plan de coupe pour une position d'écoute R à la verticale du point de référence R_{av} . Ψ_{comp}^R et Ψ_{im}^R représentent les sources virtuelles perçues au point R à travers, respectivement, les bancs LS_{comp} et LS_{im} . Par symétrie de révolution autour de l'axe des bancs LS_{comp} et LS_{im} , on définit des positions d'écoute équivalentes R_{comp} et R_{im} .

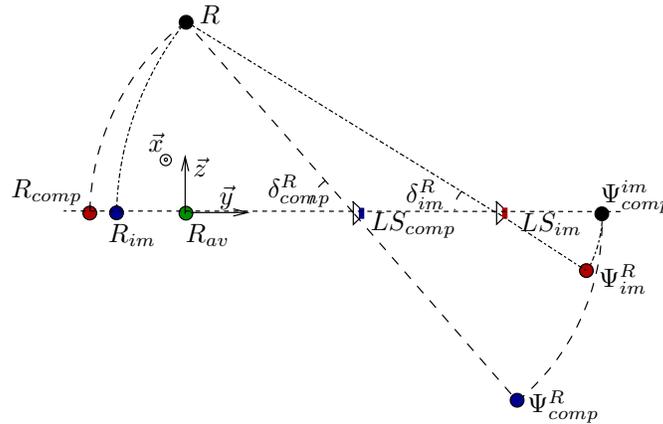


FIG. 3.12 – Positions équivalentes dans le plan horizontal pour une position d'écoute en dehors de ce plan

Les positions d'écoutes équivalentes se sont pas confondues. Les contributions des sources Ψ_{comp}^R et Ψ_{im}^R ne sont donc pas synchrones et n'ont pas un niveau équivalent. La différence de temps d'arrivée $\Delta\tau_{vert}(r_R)$ est donnée par,

$$\Delta\tau_{vert}(r_R) = \frac{d_{R_{comp}}^{\Psi} - d_{R_{im}}^{\Psi}}{c}, \quad (3.7)$$

où, $d_{R_{comp}}^{\Psi}$ (resp. $d_{R_{im}}^{\Psi}$) représente la distance entre R_{comp} (resp. R_{im}) et Ψ . La somme des contributions crée un effet de filtrage en peigne. On peut alors définir une fréquence limite $f_{comp}(r_R)$ à partir de laquelle les contributions des sources Ψ_{comp}^R et Ψ_{im}^R sont à nouveau en phase.

$$f_{comp}(R) = \frac{1}{2\Delta\tau_{vert}(R)}. \quad (3.8)$$

Au dessus de $f_{comp}(r_R)$ la réflexion n'est donc plus compensée mais amplifiée.

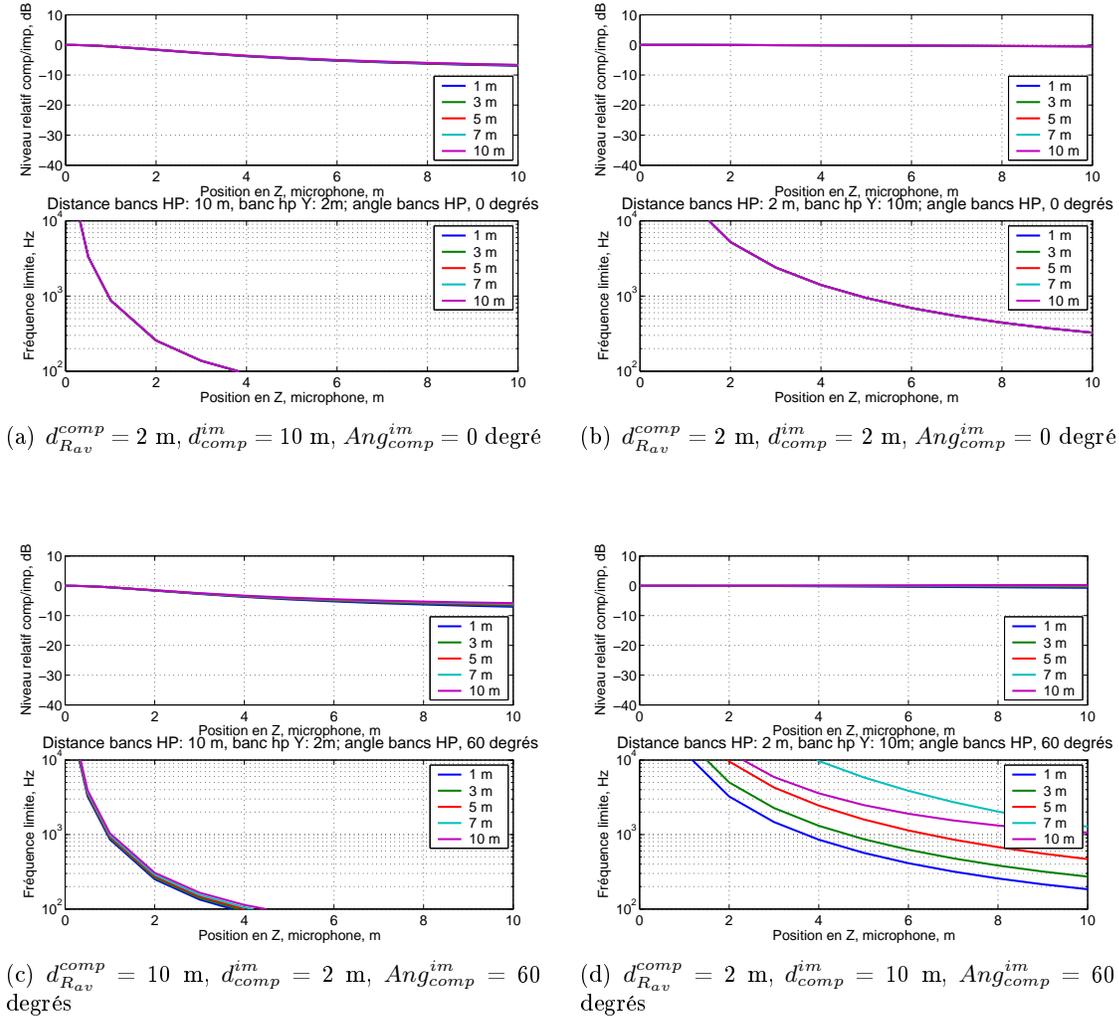


FIG. 3.13 – Efficacité de la compensation en dehors du plan horizontal. Niveau relatif des champs associés à Ψ_{comp} et Ψ_{im} , et fréquence limite de compensation en fonction de la hauteur de la position d'écoute. Le paramètre de variation pour les courbes d'une même figure est la distance d_{Ψ}^{im} de la source virtuelle Ψ au banc de haut-parleurs image LS_{im} .

L'ensemble des figures 3.13 présente la différence relative de niveau $\frac{Att_{\Psi}^{im}(r_R)}{Att_{comp}^{im}(r_R)}$ entre les fronts d'onde associés à Ψ_{comp}^R et à Ψ_{im}^R , et la fréquence limite $f_{comp}(r_R)$. Ces valeurs sont données en fonction de la hauteur (coordonnée z) de la position d'écoute, à la verticale du point de référence R_{av} (centre du repère cartésien). Sur chaque figure, les courbes correspondent à différentes distances d_{Ψ}^{im} de la source virtuelle Ψ au banc de haut-parleurs image LS_{im} . En dessous de chaque figure on indique les valeurs de $d_{R_{av}}^{comp}$, d_{comp}^{im} , et Ang_{comp}^{im} considérées. Dans les figures 3.13(a) et 3.13(b), LS_{comp} et LS_{imp} sont parallèles, dans les figures 3.13(c) et 3.13(d), ils présentent un angle relatif de 60 degrés.

De manière générale, la compensation en dehors du plan horizontal est plus efficace lorsque la distance du banc de compensation à la position de référence est importante. Dans ce cas, la différence entre le site de la position d'écoute vue du banc de haut-parleur image δ_{im}^R et vue du banc de haut-parleurs de compensation δ_{comp}^R est faible (cf. figures 3.13(b) et 3.13(d)). Les points R_{comp} et R_{im} sont ainsi pratiquement coïncidents. Les niveaux de Ψ_{comp}^R et Ψ_{im}^R sont ainsi équivalents et $f_{comp}(R)$ est élevée.

Pour des positions d'écoute plus proches, les niveaux relatifs deviennent sensiblement différents et $f_{comp}(R)$ décroît très rapidement (cf. figures 3.13(a) et 3.13(c)).

Dans le cas où les bancs de haut-parleurs sont parallèles, il n'y a pratiquement aucune différence entre les différentes valeurs de d_{Ψ}^{im} , que ce soit en terme de niveau relatif ou de fréquence limite. Pour un angle relatif de 60 degrés entre LS_{comp} et LS_{im} , la compensation semble plus efficace pour des sources éloignées du banc image.

Pour une hauteur d'écoute donnée z_R , la réduction relative de niveau est ainsi plus importante pour des positions d'écoute éloignées du banc de compensation. La compensation est alors plus efficace en dehors du plan horizontal pour les réflexions d'ordre supérieur. La méthode proposée est ainsi sensible à la hauteur de la position d'écoute par rapport au plan horizontal, en particulier lorsque les bancs de haut-parleurs image et de compensation sont proches.

Remarquons toutefois que dans toutes les situations présentées, pour des positions d'écoute comprises entre -50 et +50 cm, en-dessous et au-dessus du plan horizontal, la fréquence limite de compensation $f_{comp}(R)$ reste supérieure à la fréquence d'aliasing typique (≈ 2000 Hz). La compensation de la source image est ainsi simplement légèrement moins efficace en dehors du plan horizontal mais reste effective.

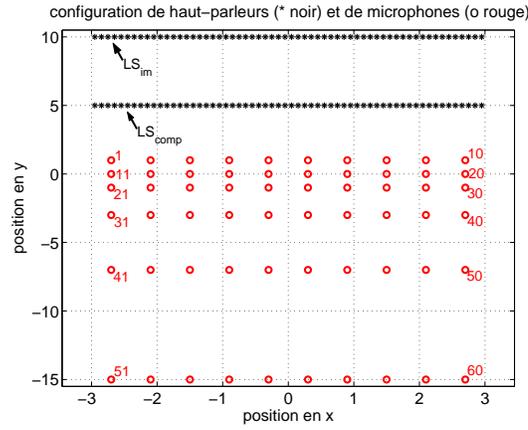


FIG. 3.14 – Configuration de test pour évaluer l'influence de la diffraction sur la méthode de compensation

Importance de la diffraction

Dans les deux parties précédentes, les fronts d'onde subissent une loi d'atténuation qui ne dépend pas de la fréquence. Ceci est vrai, en première approximation, pour un banc linéaire de haut-parleurs omnidirectionnels de taille infinie en dehors de la proximité immédiate des haut-parleurs (cf. partie 2.2).

Considérons maintenant des bancs de haut-parleurs de taille finie (6 m, 60 haut-parleurs omnidirectionnels idéaux, 10 cm d'espace entre les haut-parleurs) identiques pour LS_{comp} et LS_{im} . Dans cette partie, nous ne considérons que des bancs de haut-parleurs parallèles.

Les champs synthétisés par LS_{comp} et LS_{im} présentent des composantes diffractées (cf. partie 2.3) qui ne sont identiques que si LS_{comp} et LS_{im} sont coïncidents.

Simulons le champ synthétisé par LS_{comp} et LS_{im} au niveau de microphones situés dans la zone d'écoute (cf. figure 3.2.1). Dans la configuration étudiée, LS_{comp} est situé à 5 mètres du point de référence. La distance entre LS_{comp} et LS_{im} est de 5 mètres. Pour chacun des microphones, on évalue les réponses fréquentielles $H_{\Psi_{comp}}(r(nmic), k)$ et $H_{\Psi_{im}}(r(nmic), k)$ associées aux bancs LS_{comp} et LS_{im} .

La réduction relative de niveau peut s'exprimer en fonction de la fréquence,

$$Red_{\Psi_{im}}(r(nmic), k) = 1 - \frac{H_{\Psi_{comp}}(r(nmic), k)}{H_{\Psi_{im}}(r(nmic), k)}. \quad (3.9)$$

Afin d'évaluer uniquement l'influence de la diffraction, on définit

$$Red_{\Psi_{im}}^{diff}(r(nmic), k) = 1 - \frac{H_{\Psi_{comp}}(r(nmic), k)}{Att_{\Psi}^{comp}(r(nmic))} \frac{Att_{\Psi}^{im}(r(nmic))}{H_{\Psi_{im}}(r(nmic), k)}. \quad (3.10)$$

Cette quantité correspond à la réduction relative de niveau après compensation du niveau des sources virtuelles Ψ_{im} et Ψ_{comp} . Sans diffraction, $Red_{\Psi_{im}}^{diff}(r(nmic), k)$ serait nul pour l'ensemble des fréquences. Toutes les valeurs non nulles sont ainsi dues *uniquement* à la diffraction.

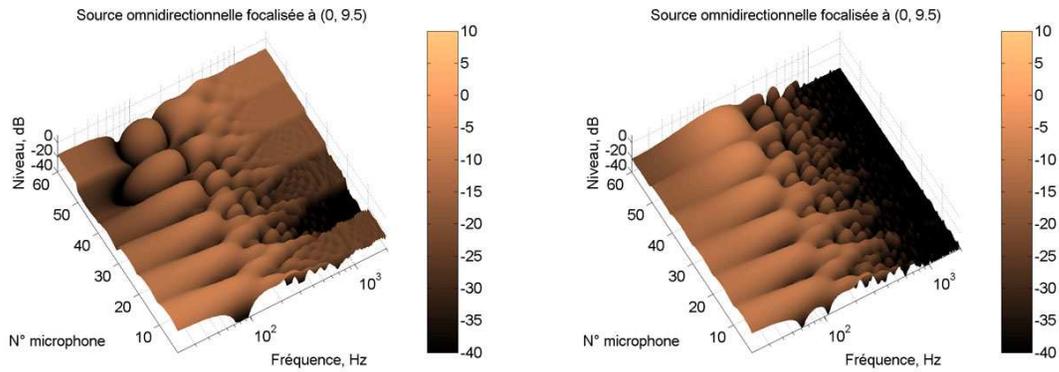
L'ensemble des figures 3.15 présente $Red_{\Psi_{im}}(r(nmic), k)$ et $Red_{\Psi_{im}}^{diff}(r(nmic), k)$ en dessous de la fréquence d'aliasing spatial pour trois sources virtuelles :

- une source virtuelle omnidirectionnelle située 50 cm devant le banc image (source focalisée),
- une source virtuelle omnidirectionnelle située 1 m derrière le banc image,
- une onde plane se propageant perpendiculairement au banc image.

Pour les deux sources omnidirectionnelles (cf. figures 3.15(a), 3.15(b), 3.15(c), 3.15(d)) les termes de diffraction n'ont pratiquement aucune influence au delà de 500-600 Hz. On a vu dans la partie 2.3 que la diffraction est très fortement réduite à ces fréquences par atténuation des haut-parleurs situés aux extrémités du banc.

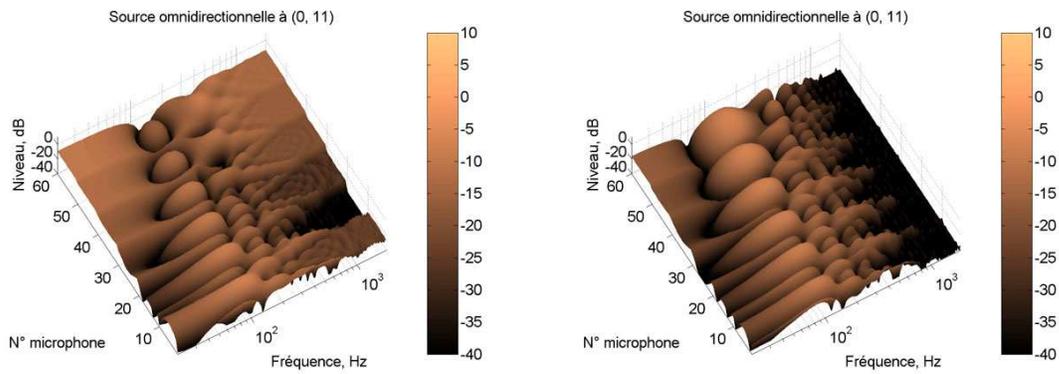
En dessous de cette fréquence, on remarque que l'effet de la diffraction est de l'ordre de la réduction de niveau obtenue sans compensation de niveau des sources virtuelles. En particulier, pour la ligne de référence ($y_R = 0$, microphones 11 à 20), la réduction relative de niveau n'est plus totale à cause de la diffraction.

Le même comportement se remarque pour l'onde plane. Ajoutons que l'effet de la diffraction s'étend ici sur toute la bande de fréquences considérée à un niveau élevé. Cet effet tend à disparaître pour des positions de microphone éloignées ($y_R = -15$, microphones 51 à 60), en particulier aux basses fréquences. On est toutefois face à un cas particulier car la source virtuelle synthétisée par LS_{im} et LS_{comp} est la même (onde plane se propageant perpendiculairement au banc de haut-parleurs), ce qui n'est pas le cas pour les autres sources considérées.



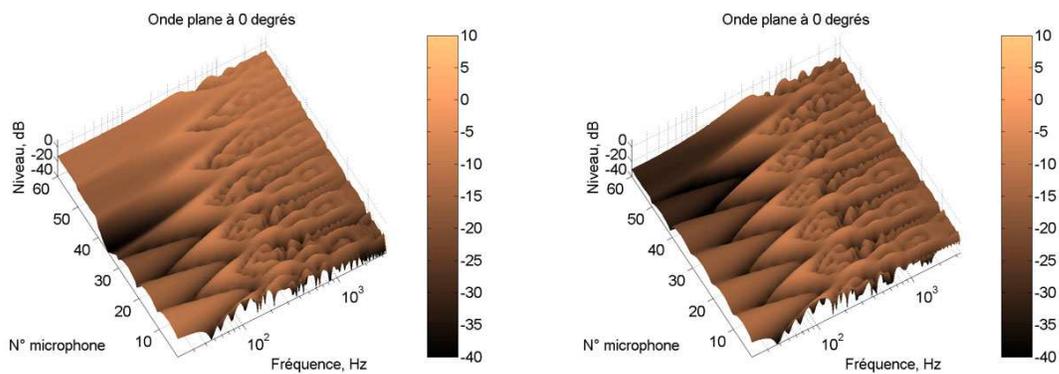
(a) Atténuation *sans* compensation du niveau, source omnidirectionnelle située 50 cm devant le banc image

(b) Atténuation *avec* compensation du niveau, source omnidirectionnelle située 50 cm devant le banc image



(c) Atténuation *sans* compensation du niveau, source omnidirectionnelle située 1 m derrière le banc image

(d) Atténuation *avec* compensation du niveau, source omnidirectionnelle située 1 m derrière le banc image



(e) Atténuation *sans* compensation du niveau, onde plane se propageant perpendiculairement au banc image

(f) Atténuation *avec* compensation du niveau, onde plane se propageant perpendiculairement au banc image

FIG. 3.15 – Réduction relative de niveau de la réflexion en fonction de la fréquence. Mise en évidence de l'erreur due à la diffraction.

3.2.2 Adaptation à des situations réelles

Dans la partie précédente, les bancs de haut-parleurs sont constitués de transducteurs idéaux. Dans ce cadre, nous avons montré qu'il est possible de compenser partiellement une réflexion de premier ordre issue d'une paroi parfaitement réfléchissante sur *toute l'étendue* de la zone d'écoute.

En pratique, on doit considérer des phénomènes supplémentaires :

1. les haut-parleurs utilisés n'ont pas des caractéristiques de rayonnement idéales,
2. la pièce est fermée et comporte un nombre supplémentaire de parois sur lesquelles se réfléchit le son direct et le champ rayonné par la paire banc de haut-parleurs image/banc de haut-parleurs de compensation,
3. les matériaux ont un coefficient de réflexion qui dépend de la fréquence, et la réflexion n'est pas totalement spéculaire,
4. les variations atmosphériques (température, humidité, ...) font varier la vitesse de propagation des ondes sonores,
5. la pièce comporte au moins un auditeur qui masque partiellement le son direct et modifie ainsi le champ réfléchi.

Les caractéristiques non idéales des transducteurs est un problème qui se rencontre aussi dans le cas de la synthèse du son direct. Cette approximation importante par rapport à la théorie requière la mise en place de techniques d'égalisation du champ sonore. Ces techniques doivent ainsi permettre de compenser les "défauts" des transducteurs pour la synthèse de fronts d'onde cohérents avec la cible. Dans les deux cas, la compensation doit être efficace dans *l'ensemble de la zone d'écoute*. La définition de telles techniques fait l'objet de la partie II de ce document.

Les variations atmosphériques influencent principalement la vitesse de propagation des ondes sonores. Il est alors possible en utilisant des capteurs d'humidité et de température de suivre les variations de la vitesse du son. Nous avons ainsi montré avec van Zon (van Zon *et al.*, 2004) que la prise en compte des modifications atmosphériques s'effectue simplement par la mise à jour des retards à affecter aux haut-parleurs.

La présence d'auditeurs dans la pièce est un phénomène bien plus complexe à évaluer dont ne sait estimer précisément les conséquences. Toutefois, les haut-parleurs sont a priori situés au niveau de la tête de l'auditeur qui apparaît alors comme un obstacle relativement "transparent" vis des ondes acoustiques en basses fréquences. Ce point mériterait cependant une étude plus approfondie et dépasse le cadre de ce document.

Le caractère absorbant des parois modifie les caractéristiques spectrales du "champ rayonné" par la source image. L'absorption peut se modéliser de façon simplifiée comme un filtrage linéaire. Ceci repose sur la mesure du coefficient d'absorption caractéristiques du matériau utilisé. Dans ce cadre, un simple filtrage du signal d'alimentation de la source de compensation permettrait de prendre en compte cet effet.

Toutefois, la mesure des coefficients d'absorption est classiquement réalisée par octave. La pré-

cision offerte est faible et ne permet pas d'accéder aux éventuels déphasages du signal réfléchi. De plus, la taille finie de la paroi introduit un effet de diffraction sur le champ réfléchi qui n'est pas modélisé par la méthode des sources images. La prise en compte de ces effets passe alors par une caractérisation in-situ du champ réfléchi.

Les réflexions diffuses sur la paroi ne sont pas non plus incluses dans notre modèle. D'un point de vue perceptif, ce phénomène n'est généralement pas considéré comme un "défaut" pour la reproduction sonore. La présence de diffusion dans le champ précoce a au contraire tendance à augmenter le seuil de détection des réflexions ainsi que la seuil de perception de l'écho. Nous conseillons ainsi de ne pas considérer la prise en compte de ce champ diffus précoce lors de la compensation.

Dans la suite de cette partie, nous aborderons les points 1 et 2 non encore commentés. Nous estimerons l'influence des parois supplémentaires sur la méthode de compensation proposée. Dans un premier temps, nous considérerons la réflexion de la paire banc image/banc de compensation sur les autres parois de la salle. Nous proposerons ensuite une généralisation de cette méthode à l'ensemble des parois de la pièce. Nous la présenterons alors comme une possible extension de la WFS.

3.2.3 Réflexion des haut-parleurs image et compensation sur les parois de la pièce

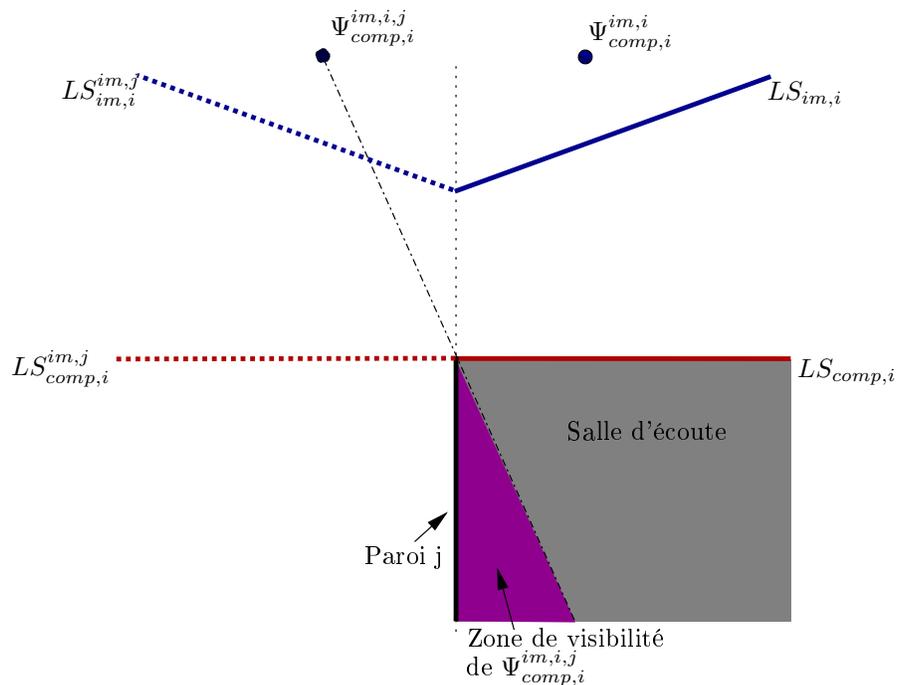


FIG. 3.16 – Réflexion de la paire banc image/banc de compensation sur une paroi horizontale adjacente

La réflexion de la paire banc image/banc de compensation sur une paroi adjacente verticale j est

présentée dans la figure 3.16. On s'aperçoit que la zone de visibilité de la source virtuelle $\Psi_{comp,i}^{im,j}$ (image des source virtuelles $\Psi_{comp,i}$ et $\Psi_{im,i}$ à travers la paroi j) est définie à partir l'extension spatiale de la paroi j . En dehors de cette zone de visibilité, les champs synthétisés sont,

- pour $LS_{comp,i}^{im,j}$, le champ de diffraction émis pour la synthèse WFS de la source $\Psi_{comp,i}^{im,j}$ en dehors de la zone de visibilité de $\Psi_{comp,i}^{im,j}$ à travers $LS_{comp,i}^{im,j}$;
- pour $LS_{im,i}^{im,j}$, le champ de diffraction de synthèse de la source $\Psi_{im,i}^{im,j}$ à travers la paroi j .

Les champs synthétisés par $LS_{comp,i}^{im,j}$ et $LS_{im,i}^{im,j}$ en dehors de la zone de visibilité sont par définition en opposition de phase mais n'ont aucune raison d'être identiques. Pour les murs adjacents, le phénomène de diffraction risque alors de réduire les performances de la méthode proposée pour les réflexions d'ordre supérieur. Cette diffraction résiduelle, étant issue de la "diffraction WFS", elle apparaît principalement en basses fréquences.

Nous avons suggéré dans la partie précédente, que dans le cas où de parois parallèles, la réduction relative de niveau des réflexions d'ordre supérieur reste efficace. Dans ce cas, la zone de visibilité de $\Psi_{comp,i}^{im,j}$ pour les deux bancs de haut-parleurs s'étend sur l'ensemble de la zone d'écoute.

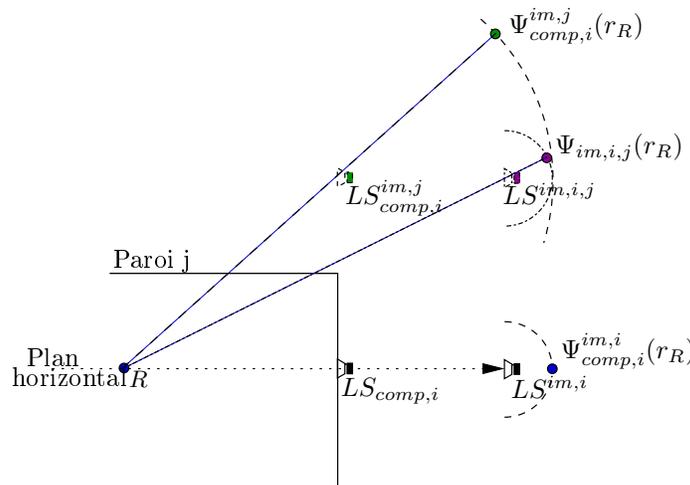


FIG. 3.17 – Réflexion de la paire banc image/banc de compensation sur une paroi verticale

Dans la figure 3.17, nous présentons la réflexion de la paire banc image/banc de compensation sur une paroi horizontale (réflexion dans la dimension verticale). Par symétrie de révolution autour de l'axe des bancs de haut-parleurs, on définit ainsi $\Psi_{im,i}^{im,j}(r_R)$ et $\Psi_{comp,i}^{im,j}(r_R)$ les sources virtuelles équivalentes synthétisée par réflexion des deux bancs de haut-parleurs sur la paroi j .

Cette situation est parfaitement équivalente à celle de la compensation des réflexions en dehors du plan horizontal. Sur ce point, on renvoie le lecteur à la partie 3.2.1.

Il s'agit d'une limitation importante de la méthode car cela un effet de filtrage en peigne sur les réflexions d'ordre supérieur. Ce filtrage en peigne dépend alors de la position relative de $\Psi_{im,i}^{im,j}(r_R)$ et $\Psi_{comp,i}^{im,j}(r_R)$, donc de la position d'écoute, de la position de la paroi j et surtout, dans le cas réel, de la directivité du haut-parleur considéré. Pour des haut-parleurs omnidirectionnels, le

niveau de la réflexion est atténué jusqu'à la fréquence limite $f_{comp}(R)$ (cf. partie 3.2.1) à partir de laquelle elle est amplifiée.

Ce type de problème risque d'introduire aussi de la coloration sur la réverbération tardive de la pièce d'écoute.

3.2.4 Généralisation à l'ensemble des parois verticales

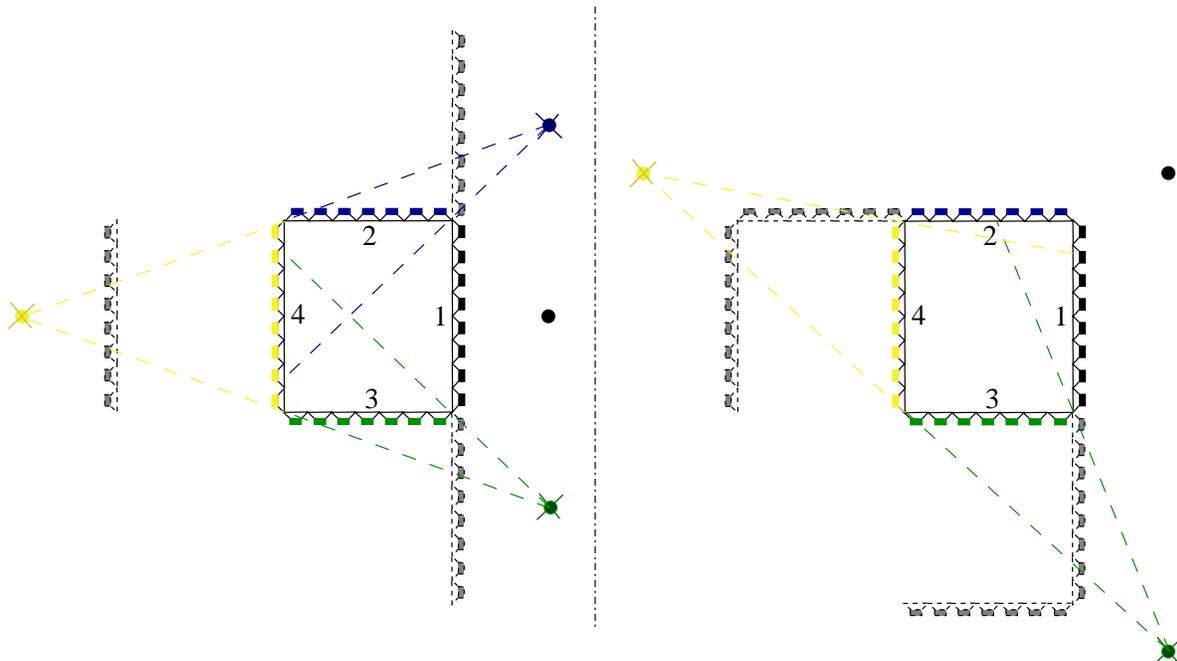


FIG. 3.18 – Compensation du champ réfléchi dans le cas d'une salle rectangulaire, visibilité des sources images

La méthode de compensation se généralise à l'ensemble des parois verticales de la pièce de reproduction. Il suffit pour cela de considérer pour chaque paroi de la salle d'écoute une paire banc de haut-parleur image/banc de haut-parleur de compensation.

Prenons l'exemple d'une pièce rectangulaire pour laquelle des bancs de haut-parleurs sont intégrés dans chacun des murs. En ne considérant que des sources virtuelles situées à l'extérieur de la pièce, on distingue deux situations (cf. partie 1.2.3) :

- la source virtuelle est synthétisée par l'un des quatre bancs de haut-parleur.
- le source virtuelle est synthétisée par deux bancs de haut-parleurs successifs.

La figure 3.18 représente l'application de la méthode proposée dans les deux cas cités. Dans les deux cas, l'ensemble des haut-parleurs sont utilisés. Ils ont alors "responsables" soit de la synthèse du son direct, soit de la compensation du champ réfléchi sur les parois verticales de la salle de restitution. Dans ce cas, tous les haut-parleurs de l'installation sont utilisés simultanément. Le cas des sources virtuelles focalisées est légèrement plus complexe mais on pourrait appliquer le même type de raisonnement. On montrerait ainsi que l'ensemble des haut-parleurs doivent

être utilisés dans les deux cas.

On sent bien que ceci pourrait se généraliser à des géométries plus complexes pour lesquelles les haut-parleurs seraient répartis continûment le long des parois. Pour l'instant, nous n'avons pas de démonstrations précises à apporter.

La méthode proposée peut se présenter comme une extension de la Wave Field Synthesis en exploitant, à nouveau, la symétrie du champ synthétisé.

Prenons l'exemple d'une source virtuelle extérieure à la pièce de restitution. On considère une paroi j dans laquelle est intégré le banc de haut-parleur $LS_{comp,j}$. La source virtuelle de compensation $\Psi_{comp,j}$ ainsi que la source image $\Psi^{im,j}$ se situent, par définition, à une position symétrique, par rapport à la paroi j , de la source virtuelle cible Ψ_{dir} . On rappelle alors que les ondes produites par un banc linéaire de haut-parleurs omnidirectionnels sont symétriques par rapport à l'axe dans le banc horizontal (cf. partie 3.1.1). Par symétrie, il est alors équivalent, vis-à-vis de la forme d'onde synthétisée, de spécifier la source Ψ_{dir} , afin de synthétiser la source de compensation $\Psi_{comp,j}$.

Si la source Ψ_{dir} est située dans le sous-espace de reproduction Ω_R^j associé au banc de haut-parleur intégré dans la paroi j , suivant les critères définis dans la partie 1.2.3, ce banc ne contribue pas à la synthèse du son direct et est donc un banc de haut-parleurs utilisé pour la compensation du champ réfléchi. Or, $\Psi_{dir} \in \Omega_R^j$ équivaut à ce que le facteur $\cos \theta_{0,j}$ soit de signe négatif. $\theta_{0,j}$ est en effet défini par l'angle entre le vecteur normal à la paroi j dirigé vers le sous-espace de reproduction associé Ω_R^j et le vecteur joignant une position de la paroi à la source Ψ (cf. figure 3.19, paroi j).

D'après la définition de la fonction d'alimentation des haut-parleurs (cf. équation 1.34), le front d'onde émis par le banc $LS_{comp,j}$ pour la synthèse de la source virtuelle Ψ_{dir} a une forme équivalente au front d'onde émis par $\Psi_{im,j}$. Ces deux fronts d'onde sont *naturellement* en opposition de phase.

En considérant que l'ensemble des haut-parleurs sont intégrés dans les parois de la salle d'écoute, on peut généraliser cette approche à l'ensemble des parois de la salle. En effet, d'après la remarque précédente, la WFS étendue à l'ensemble des haut-parleurs de la pièce synthétise à la fois les fronts d'onde associés au son direct et les fronts qui permettent la compensation des réflexions du plan horizontal.

Si le niveau des sources virtuelles était correctement synthétisé et si les parois étaient parfaitement réfléchissantes, la compensation du champ réfléchi serait automatiquement obtenue en spécifiant pour l'ensemble des haut-parleurs intégrés dans les parois de la salle d'écoute la synthèse de la source Ψ_{dir} . Toutefois, l'utilisation de bancs linéaires modifie la loi d'atténuation des fronts d'onde synthétisée (cf. partie 2.2.1). Le niveau synthétisé en WFS n'est conforme à celui de la source que sur une ligne parallèle au banc de haut-parleurs. Dans le cas de l'utilisation de plusieurs bancs de haut-parleurs non parallèles, il est nécessaire de définir une position d'écoute moyenne R_{av} , intersection des lignes sur lesquelles l'optimisation de niveau est réalisée.

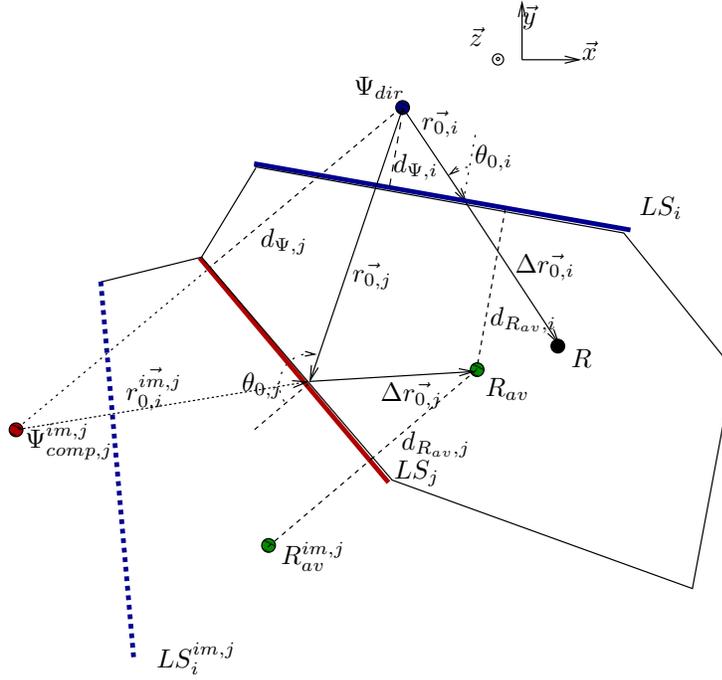


FIG. 3.19 – Extension du principe de la WFS pour la compensation des réflexions sur les parois de la salle d'écoute

Le signal d'alimentation d'un haut parleur situé à une position l_j appartenant à une paroi j , dans laquelle est intégré un banc de haut-parleurs LS_j , pour la synthèse de la source Ψ , est donné par (cf. équation 1.34),

$$U_{\Psi}^j(l_j, k) = S(\omega) \sqrt{\frac{k}{2\pi}} \text{Win}_{\Psi}^j(l_j) g_{\Psi}^i(d_{R_{av}}^j) \cos \theta_{0,j} \Phi_m(\phi) \frac{e^{-j(kr_{0,j} - \frac{\pi}{4})}}{\sqrt{r_{0,j}}}, \quad (3.11)$$

où $g_{\Psi,j}(d_{R_{av},j})$ correspond à,

$$g_{\Psi,j}(d_{R_{av},j}) = \sqrt{\frac{|d_{R_{av},j}|}{|d_{R_{av},j} + d_{\Psi,j}|}}. \quad (3.12)$$

$d_{R_{av},j}$ correspond à la distance de R_{av} à la paroi j , $d_{\Psi,j}$ étant la distance la source virtuelle à la paroi j .

Le terme $\text{Win}_{\Psi,j}(l_j)$ correspond à un terme de gain qui permet

- de sélectionner les haut-parleurs de la paroi j qui remplissent les critères de visibilité de $\Psi^{im,j}$ à travers $LS^{im,j}$;
- d'atténuer les haut-parleurs situés aux extrémités afin de réduire les effets de diffraction.

Toutefois, le champ synthétisé par la source $\Psi^{im,j}$ (image de la source Ψ à travers la paroi j) est correctement reproduit sur une ligne parallèle à $LS_i^{im,j}$ passant par $R_{av}^{im,j}$ (image du point d'écoute moyenne R_{av} à travers la paroi j , cf. figure 3.19). Le champ synthétisé par le banc j pour la synthèse de la source Ψ n'est pas normalisé vis-à-vis du même point de référence que

pour la source image $\Psi^{im,j}$ qu'il est censé compenser. . .

Le facteur de normalisation $g_{\Psi,j}(d_{R_{av},j})$ doit alors être modifié pour les bancs de haut-parleurs responsables de la compensation du champ réfléchi. Nous ne donnons pas ici la valeur modifiée du facteur de normalisation. La méthode proposée dans le chapitre 6 donnera un calcul automatique des fonctions d'alimentation pour une ligne de référence choisie.

Pour une source intérieure (source focalisée), le choix des haut-parleurs à utiliser pour la synthèse du son direct est arbitraire, car, par définition, celle-ci est située dans le sous-espace des sources associés à *tous* les bancs de haut-parleurs. Nous recommandons de sélectionner les haut-parleurs pour la synthèse du son direct par extension des critères donnés dans la partie 1.2.3 dans le cas d'une pièce rectangulaire. Il s'agit de minimiser la partie de la zone d'ombre (zone située entre les haut-parleurs et la source virtuelle) située dans la zone d'écoute. Dans la zone d'ombre les fronts d'onde ne sont en effet pas correctement synthétisés. Le raisonnement précédent reste alors valable pour le reste des haut-parleurs.

Un filtrage passe-bas supplémentaire doit être appliqué pour les haut-parleurs de compensation. Ceux-ci ne doivent, en effet, émettre de l'énergie qu'en dessous de la fréquence d'aliasing spatial. Nous proposons de fixer comme fréquence limite la fréquence d'aliasing minimale calculée sur la ligne de référence pour chaque paroi.

3.2.5 Discussion : validité de la méthode proposé

La méthode proposée dans ce chapitre permet une compensation des réflexions "du plan horizontal dans le plan horizontal". La méthode proposée se base sur la WFS et elle comporte ainsi les mêmes limitations (modification de la loi d'atténuation, diffraction). Elle se limite d'autre part, à la compensation des réflexions uniquement en dessous de la fréquence d'aliasing spatial. Nous avons fourni une analyse systématique de l'efficacité de la méthode sur les réflexions précoces d'ordre 1 et 2. Cette analyse a laissé apparaître plusieurs défauts.

1. Dans le plan horizontal, les sources de compensation synthétisées ne subissent pas la même loi d'atténuation que les sources images. La compensation n'est alors optimale que sur une zone de l'espace limitée. En dehors de cette zone, le champ réfléchi est compensé de manière imparfaite. On note même un phénomène de sur-compensation. C'est à dire que le niveau de la réflexion peut être parfois supérieur au niveau original, notamment, à proximité des parois. Cet effet peut être réduit en limitant le niveau émis par le banc de compensation. Ceci s'effectue en plaçant le point de référence, pour lequel la compensation est totale, à proximité du banc de haut-parleurs de compensation. La compensation des réflexions d'ordre supérieur est dégradée par cette opération ce qui est a priori moins "grave" vis-à-vis de la perception.
2. La qualité de la compensation se dégrade sensiblement en dehors du plan horizontal et n'est possible que jusqu'à une fréquence limite. Celle-ci dépend de la "hauteur" d'écoute, de la position relative des bancs de haut-parleurs image et de compensation, et de la source virtuelle.
3. Les réflexions d'ordre 2, ainsi que leur "descendance", issues des réflexions d'ordre 1 sur lesquelles on a appliqué la compensation, ne sont pas toujours proprement compensées à cause des effets de diffraction. Ceci est particulièrement remarquable lorsque la réflexion se produit sur une paroi adjacente perpendiculaire.

Nous avons proposé une formulation de cette méthode sous forme d'extension de la Wave Field Synthesis.

Une situation qui n'est pas abordée est celle des réflexions d'ordre 1 issues du plan vertical et de leur descendance. Celles-ci ne sont pas concernées par la méthode et sont ainsi laissées intactes. La méthode proposée dans ce chapitre nécessiterait donc une évaluation perceptive approfondie qui n'a pas été effectuée dans le cadre de la thèse. Il conviendrait aussi de déterminer l'influence de la méthode proposée sur la réverbération tardive et sur l'excitation des modes de la pièce, en particulier en basses fréquences.

Deuxième partie

Méthodes d'inversion multicanal pour
la Wave Field Synthesis

Dans la partie I, nous nous sommes attachés à décrire les caractéristiques du champ acoustique rayonné en champ libre et en espace clos lors de la synthèse de sources virtuelles par Wave Field Synthesis dans l'ensemble de la zone d'écoute. Nous avons considéré des situations idéales (haut-parleurs omnidirectionnels idéaux, parois parfaitement réfléchissantes) dans le but d'évaluer l'importance des inadéquations du champ reproduit par rapport à la cible. Nous avons pour cela utilisé des critères objectifs relatifs à la perception. Nous avons décrit des problèmes potentiels liés uniquement à la WFS, sans que l'on puisse mettre en cause les transducteurs utilisés. Nous avons mis en évidence une évolution de la qualité de la reproduction des fronts d'onde au cours de la propagation. Ceci se traduit, en champ libre, par une modification de la loi d'atténuation, une évolution de la fréquence d'aliasing spatial et des changements dans la réponse fréquentielle du système.

En espace clos, nous avons vu que l'interaction du système de reproduction avec la salle d'écoute dépend très largement de la source virtuelle synthétisée. Nous avons proposé une méthode de compensation des réflexions précoces qui vise à améliorer la transparence de la reproduction dans l'ensemble de la zone d'écoute. Cette méthode peut, dans une certaine mesure, se définir comme une extension de la Wave Field Synthesis et hérite des mêmes limitations.

1. La compensation n'est réellement effective que dans le plan horizontal.
2. La compensation ne peut être que partielle à cause de la géométrie linéaire des bancs de haut-parleurs employés. Il y a un risque de sur-compensation à proximité des bancs de haut-parleurs qui synthétisent les sources virtuelles de compensation.
3. La diffraction, due à la taille finie des bancs de haut-parleurs, dégrade les résultats.
4. Aucune compensation n'est possible au-dessus de la fréquence d'aliasing spatial.

Dans le cas réel, les haut-parleurs utilisés ont des caractéristiques de rayonnement qui s'écartent notablement du modèle idéal. Les parois ne sont pas parfaitement réfléchissantes ce qui va réduire l'efficacité de la méthode de compensation des réflexions. L'ensemble de ces caractéristiques, ainsi que les problèmes soulevés dans le cadre de la WFS, constituent autant de "défauts" qui réduisent la transparence de la reproduction.

Les *techniques d'égalisation* peuvent prendre en compte de ces "défauts", celles-ci consistent à "construire" un filtre qui est appliqué au signal d'entrée du ou des haut-parleurs du dispositif de restitution.

Les filtres d'égalisation sont *calculés* en référence à un *objectif* de reproduction à partir d'une *caractérisation* préalable du champ acoustique synthétisé .

La Wave Field Synthesis ayant l'ambition de reconstruire un champ sonore dans une zone étendue de l'espace, toute procédure d'égalisation dédiée à cette technique doit être efficace *dans l'ensemble de la zone d'écoute*.

Les études sur l'égalisation de bancs de haut-parleurs dans le cadre de la WFS sont peu nombreuses. Remarquons cependant la technique proposée par de Vries (de Vries, 1996) qui permet

la compensation des caractéristiques de rayonnement des haut-parleurs pour la WFS. Cette technique permet de garantir que le champ synthétisé est correct sur une ligne à l'intérieur de l'espace de restitution dans le cas où les haut-parleurs sont répartis sur une ligne. De Vries montre que cette technique est efficace dans le cas où les haut-parleurs sont identiques mais est mise en défaut si un seul des haut-parleurs du banc est différent.

Cette technique repose sur le calcul d'un ensemble de filtres en fonction de la source virtuelle considérée. Elle cherche ainsi à garantir le *résultat* en fonction d'une *cible*. Pour chaque source virtuelle synthétisée, on doit calculer un jeu de filtres à appliquer aux haut-parleurs du dispositif de reproduction. Cette technique sera expliquée plus en détail dans la partie 4.3.5.

La méthode proposée par de Vries se rapproche des techniques d'égalisation multicanal car elle vise à *assurer* que le champ acoustique cible est correctement reproduit. Il s'agit de contrôler le champ synthétisé par un ensemble de haut-parleurs au niveau d'un ensemble de points de contrôle réparties dans la zone d'écoute ou au niveau des oreilles de l'auditeur.

Ces techniques proposent un cadre général qui ne dépend pas a priori des caractéristiques de rayonnement ni de la disposition des transducteurs utilisés. L'enjeu consiste alors à *garantir* la validité de la solution obtenue dans *l'ensemble de la zone d'écoute*.

Nous verrons que ceci est possible à condition de placer convenablement les points de contrôle (microphones) dans la zone d'écoute, mais uniquement jusqu'à une fréquence limite dont la définition est donnée par des critères d'aliasing spatial. Cette fréquence dépend ainsi de la source virtuelle cible (objectif de reproduction) et de la disposition des haut-parleurs.

La question est de savoir comment améliorer la transparence de la reproduction dans l'ensemble de la zone d'écoute en prenant en compte et en compensant la directivité des sources secondaires (haut-parleurs) utilisés ainsi que l'acoustique de la pièce de restitution.

Cette deuxième partie du document vise à définir des techniques d'égalisation multicanal tirant partie et tenant compte des spécificités de la Wave Field Synthesis pour la prise en compte et la compensation des défauts de reproduction dans le cas réel.

Le chapitre 4 dresse un panorama non exhaustif des techniques d'égalisation de la littérature. Nous insisterons sur les spécificités de ces techniques en ce qui concerne la description du champ sonore, l'objectif de reproduction et la méthode de calcul des filtres employés.

Dans le chapitre 5, nous proposerons une méthode d'*égalisation multicanal* qui vise à compenser les défauts des haut-parleurs et de la WFS, notamment en terme de diffraction, dans une zone étendue de l'espace. Cette méthode sera validée dans le cadre de l'utilisation de haut-parleurs MAP (cf. partie 5.2.2) pour la WFS. Ces haut-parleurs ont des caractéristiques de développement qui s'éloignent de celles de sources omnidirectionnelles idéales.

Dans le chapitre 6, nous présenterons les différentes méthodes de compensation de l'effet de salle basées sur des techniques d'égalisation multicanal que nous avons proposées pendant la thèse ainsi que d'autres méthodes de la littérature. Nous proposerons une méthode originale qui vise le contrôle de la synthèse des sources virtuelles de compensation par la WFS dans la partie 3.2

afin de prendre en compte les caractéristiques de rayonnement des haut-parleurs et la nature des parois de la pièce de restitution.

Chapitre 4

Techniques d'égalisation pour la reproduction sonore

Les techniques d'égalisation visent à assurer la transparence de la reproduction sonore. Cette notion de transparence se définit d'un point de vue perceptif par l'absence de modification apportée par le système de transduction électroacoustique sur le signal original. Il s'agit en quelque sorte de faire "disparaître" le système de reproduction et de laisser place à une expérience perceptive spécifiée par le créateur de contenu.

Les techniques d'égalisation tentent de compenser les "imperfections" du système de reproduction. Celles-ci sont liées non seulement à la réponse des haut-parleurs mais aussi à l'acoustique de la pièce de reproduction. Dans les parties 2.1 et ans l'introduction 3 nous avons défini des critères objectifs qui permettent de juger de la transparence de la reproduction à partir de la caractérisation du système sous forme de fonction de transfert linéaire. Étant donné un *objectif* de reproduction défini au préalable, l'égalisation consiste à placer un *filtre* en tête de chaque haut-parleur afin de le réaliser et/ou de compenser les défauts du système.

Dans ce chapitre, nous présenterons deux classes de techniques d'égalisation : égalisation individuelle, égalisation multicanal. Ces méthodes visent l'égalisation d'un dispositif de reproduction multicanal pour lequel plusieurs haut-parleurs sont utilisés simultanément afin de réaliser un objectif de reproduction. Alors que l'égalisation individuelle considère *individuellement* les haut-parleurs du dispositif et vise à compenser leurs défauts *en moyenne*, l'égalisation multicanal tend à assurer la synthèse correcte de l'objectif de reproduction par l'ensemble des haut-parleurs du dispositif au moins en un nombre fini de points, appelés points de contrôle. Tandis que la première cherche à compenser les "défauts" des sources acoustiques, la seconde cherche à assurer la validité du *résultat final*.

Dans le cas de la WFS, nous préférons la seconde qui permet de garantir la validité de la synthèse au moins sur les points de contrôle. Un des enjeux consiste à étendre cette validité à l'ensemble de la zone d'écoute. Dans ce chapitre, nous tacherons de donner des indications sur les méthodes à mettre en œuvre pour assurer cette validité mais sans toutefois les expliciter. Ceci

fera l'objet du chapitre suivant.

Dans ce chapitre, nous mettrons l'accent sur l'égalisation multicanal. Nous présenterons les techniques d'inversion multicanal grâce auxquelles se calculent les filtres d'égalisation. Ce calcul est réalisé à partir d'une *caractérisation* du système et en fonction d'un *objectif* à réaliser. Nous présenterons une étude bibliographique sur les techniques d'égalisation multicanal appliquées à la captation et à la reproduction sonore. Ceci nous permettra d'engager une discussion sur la validité de l'égalisation multicanal en fonction du "contexte de reproduction" : support, description de la scène sonore, technique de reproduction, dispositions des haut-parleurs, environnement acoustique, zone d'écoute privilégiée.

4.1 Méthodes d'égalisation d'un système multicanal

Dans cette partie, nous présentons les deux classes de méthodes d'égalisation. Nous insistons sur la méthode d'égalisation individuelle. L'égalisation multicanal sera abordée plus longuement dans le reste du chapitre.

4.1.1 Égalisation individuelle

L'égalisation individuelle requiert une description du rayonnement de chaque haut-parleur (égalité en champ libre), et éventuellement de leur interaction avec la pièce de restitution (égalité en espace clos), par une unique fonction de transfert $c(z)$. Elle consiste à définir un filtre $h(z)$ qui permet de compenser tout ou partie des "défauts" de cette fonction de transfert.

Ce type d'égalisation ne permet qu'une prise en compte *locale* ou *en moyenne* des caractéristiques de rayonnement de chaque haut-parleur. Le filtre $h(z)$ est classiquement défini par :

$$h(z) = \frac{1}{c(z)}. \quad (4.1)$$

Cette formule donne, tout au moins, le *gabarit* du filtre à réaliser. Diverses méthodes de modélisation permettent de réaliser un tel filtre sous forme de filtre à réponse impulsionnelle finie (FIR) ou infinie (IIR).

Les filtres FIR permettent, a priori, une correction de la réponse en fréquence *et* en phase. Le filtre d'égalisation est obtenu en tronquant la réponse impulsionnelle associée à $h(z)$. Pour l'égalisation en champ libre, on utilise généralement des filtres à phase minimale ou à phase linéaire. En effet, on a vu dans la partie 2.3.3 que la compensation de la réponse en phase n'était pas forcément nécessaire. Ceci n'est pas le cas pour l'égalisation en espace clos pour laquelle la compensation des caractéristiques de phase de la fonction de transfert est profitable (Johansen & Rubak, 1996).

Les filtres IIR réalisent une compensation des résonances et des anti-résonances du filtre mais ne permettent pas, a priori, une compensation de la phase.

Nous ne cherchons pas à donner ici un panorama complet des méthodes de modélisation de filtres.

Sur ce sujet, on renvoie à (Karjalainen *et al.*, 1999), (Jot *et al.*, 1995) ou au chapitre 2 de la thèse de Véronique Larcher (Larcher, 2001).

La modélisation des filtres sous forme IIR permet une implémentation peu coûteuse dans le cadre d'applications temps-réel qui requièrent une faible latence. Cette implémentation a souvent été préférée pour de nombreuses applications. L'utilisation de ce type de filtre correspond à un compromis entre précision de traitement et coût de calcul temps-réel.

La technique de convolution partitionnée (Gardner, 1995) permet aujourd'hui l'utilisation de filtres FIR avec une latence acceptable et un coût de calcul temps-réel relativement faible. C'est ce que nous utiliserons ici.

Outre la qualité de la modélisation des filtres, la validité de l'égalisation individuelle est conditionnée par la caractérisation du rayonnement du haut-parleur et, éventuellement, par son interaction avec la pièce de restitution. Celle-ci est donnée par la fonction de transfert $c(z)$ et c'est à partir de $c(z)$ qu'est déterminé le gabarit du filtre d'égalisation.

La caractérisation du rayonnement des haut-parleurs est effectuée par des mesures de réponses impulsionnelles sur chaque haut-parleur en champ libre, ou en espace clos (dans la pièce d'écoute). Dans le second cas, la caractérisation effectuée prend également en compte l'acoustique de la salle d'écoute. Dans le cas idéal où ces mesures seraient indépendantes de la position de mesure (position du microphone utilisé), $c(z)$ pourrait être mesurée à n'importe quelle position de l'espace. Néanmoins, ce cas idéal ne se présente *jamais* en pratique. Les haut-parleurs classiques ont des caractéristiques de directivité complexes, en particulier en hautes fréquences. L'influence de la salle d'écoute dépend très largement de la position de mesures (temps d'arrivée des réflexions, modes). Il convient dans ce cas de choisir une position d'écoute préférentielle ou d'effectuer une caractérisation *en moyenne* de la réponse du haut-parleur dans l'espace considéré.

On préfère généralement la deuxième solution. On effectue alors une moyenne des réponses mesurées en plusieurs points. On peut aussi *simplifier* la réponse par lissage dans le domaine fréquentiel (voir par exemple (Hatziantoniou & Mourjopoulos, 2003)). D'autres techniques consistent à modéliser la réponse du haut-parleur et de la salle sous forme de filtre paramétrique (Mourjopoulos, 1994) afin de simplifier la description et d'obtenir immédiatement un filtre d'égalisation IIR.

L'égalisation individuelle ne permet donc d'obtenir qu'une égalisation en moyenne, ou bien à une position donnée, du champ synthétisé par chaque haut-parleur *indépendamment* des autres. L'égalisation individuelle ne permet pas de maîtriser l'erreur résiduelle dans l'ensemble de la zone d'écoute. Celle-ci est potentiellement importante si les haut-parleurs ont des caractéristiques de rayonnement complexes et suivant l'acoustique de la pièce de restitution.

L'égalisation individuelle individuelle ne permet pas de garantir que l'interaction entre les différents canaux du système synthétise le résultat voulu.

4.1.2 Égalisation multicanal

Contrairement à l'égalisation individuelle, l'égalisation multicanal a l'ambition de prendre en compte les caractéristiques du rayonnement de tous les haut-parleurs simultanément pour la

synthèse d'un champ acoustique cible.

Le champ acoustique synthétisé est mesuré par L microphones. Ces L points d'observation constituent L points de contrôle pour lesquels on assure, autant que possible, que le champ acoustique synthétisé correspond à la cible.

Le système constitué par les M haut-parleurs et les L microphones est équivalent, d'un point de vue traitement de signal, à un système linéaire MIMO ("multiple input/ multiple output system" en anglais). Un tel système est caractérisé par une matrice de $M \times L$ fonctions de transfert $C(z)$ qui contient les fonctions de transfert électroacoustique entre chacune des M entrées et des L sorties.

L'étape d'*inversion multicanal* consiste à calculer un ensemble de filtres à placer en tête de chacune des entrées du système afin de minimiser un critère d'erreur. Cette erreur correspond à la mesure de la différence entre les signaux de sortie réalisés et des signaux cibles (objectif à atteindre). D'un point de vue traitement de signal, cette étape revient à "inverser" la matrice de fonction de transfert $c(z)$; d'où le terme d'*inversion multicanal*.

Ce type de procédure prend naturellement en compte la dépendance de chaque signal de sortie par rapport à l'*ensemble* des entrées du système. Ceci permet de contrôler le champ sonore *réellement* synthétisé par la somme des contributions de tous les haut-parleurs en un nombre fini de points de l'espace.

Remarquons toutefois que cette procédure requiert la définition d'une cible sur l'ensemble des points de contrôle.

4.2 Inversion multicanal

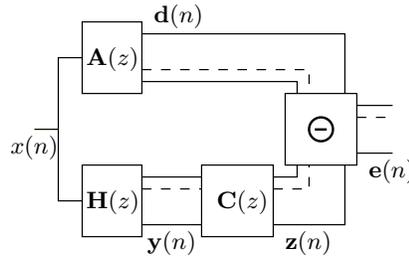


FIG. 4.1 – Schéma fonctionnel de l'inversion multicanal

La figure 4.1 présente la chaîne de traitement de signal associée au filtrage inverse multicanal. Les signaux d'entrées $y_m(n)$ du système sont obtenus par filtrage d'un signal unique $x(n)$ par $h_m(z)$; les M filtres associés à chacune des entrées étant regroupés dans le vecteur $H(z)$. Les signaux $y_m(n)$ sont transmis à travers le système (filtrés par $c_{ml}(z)$, l'ensemble des contributions filtrées des entrées étant sommées pour chaque sortie) pour donner les L signaux des sortie $s_l(n)$ groupés dans le vecteur $s(n)$. L'erreur associée au canal l est égale à la différence entre le signal cible $d_l(n)$ et le signal de sortie $s_l(n)$ du système MIMO. Les signaux cibles (formant le vecteur $d(n)$) sont obtenus par filtrage du signal $x(n)$ par un ensemble de fonctions de transfert $A_l(z)$

(formant le vecteur $A(z)$).

Les coefficients du vecteur $H(z)$ sont calculés afin de minimiser l'erreur $e_l(n)$ pour chacune des L sorties en rapport à une fonction de distance. La distance classiquement utilisée est la distance quadratique moyenne L_2 .

L'inversion multicanal vise ainsi à minimiser l'erreur entre les *sorties* du systèmes MIMO, dont la *description* est fournie par la matrice de fonctions de transfert $C(z)$ et une réponse *cible* $A(z)$ en agissant sur des *variables de contrôle* (coefficients des filtres $H(z)$).

Dans les sous-parties qui suivent, nous présentons les procédures d'inversion muticanal qui permettent le calcul des filtres $H(z)$, dans un premier temps dans le domaine fréquentiel puis dans le domaine temporel.

4.2.1 Inversion dans le domaine fréquentiel

Dans cette partie, l'ensemble des caractères gras correspondent à la transformée de Fourier des signaux temporels associés.

Dans le domaine de Fourier, le vecteur d'erreur entre les signaux cibles et les sorties du système MIMO à une impulsion de dirac ($x(n) = \delta(n)$, réponses impulsionnelles du système) s'écrit sous forme matricielle à chaque fréquence ω :

$$\mathbf{E}(\omega) = \mathbf{A}(\omega) - \mathbf{C}(\omega)\mathbf{H}(\omega). \quad (4.2)$$

Un calcul optimum des filtres pour chaque fréquence est donné par :

$$\mathbf{H}_0(\omega) = (\mathbf{C}^{*T}(\omega)\mathbf{C}(\omega))^{-1}\mathbf{C}^{*T}\mathbf{A}(\omega). \quad (4.3)$$

Si la matrice $\mathbf{C}^{*T}(\omega)\mathbf{C}(\omega)$ admet des singularités, ou plus généralement des valeurs propres faibles, le problème est mal conditionné et le filtre obtenu présente des résonances fortes. Ces résonances sont considérées comme des artefacts. Afin d'éviter ce type d'artefacts, on utilise une technique dite de régularisation (Kirkeby *et al.*, 1998). Celle-ci consiste à introduire une matrice de régularisation $\mathbf{B}(\omega)$ qui dépend de la fréquence ainsi qu'un gain de régularisation γ . Les filtres régularisés sont donnés par

$$\mathbf{H}_{0,reg}(\omega) = (\mathbf{C}^{*T}(\omega)\mathbf{C}(\omega) + \gamma\mathbf{B}^{*T}(\omega)\mathbf{B}(\omega))^{-1}\mathbf{C}^{*T}\mathbf{A}(\omega). \quad (4.4)$$

Le calcul des filtres dans le domaine fréquentiel est relativement simple et peu coûteux. Le nombre de fréquences N_{inv} sur laquelle le calcul est effectué est donné par le nombre de points utilisés par la transformée de fourrier discrète qui est au minimum la somme (moins un échantillon) de la longueur N_C des réponses impulsionnelles $C(z)$ qui décrivent le système, et de la taille voulue des filtres N_H .

Toutefois, les filtres obtenus par inversion dans le domaine fréquentiel font apparaître des contributions sur l'ensemble des N_{inv} points, au delà des N_H points prévus, après transformation dans

le domaine temporel par transformée de Fourier inverse. Ce phénomène est du à la nature circulaire de la transformée de Fourier qui se traduit dans le domaine temporel par du repliement. Les filtres ainsi calculés présentent des erreurs sous forme de précurseurs et de suiveurs par rapport au pic principal (Fielder, 2003) (Norcross *et al.*, 2004). En particuliers, les précurseurs risquent d'introduire des artefacts perceptifs auxquels l'oreille est particulièrement sensible. D'autre part, pour obtenir la taille de filtre, il est nécessaire de tronquer la réponse du filtre calculé de manière plus ou moins arbitraire et donc de dégrader, au moins d'un point de vue numérique, la qualité de l'inversion.

4.2.2 Inversion dans le domaine temporel

Dans le domaine temporel, une réécriture du problème permet de définir le critère d'erreur sous forme matricielle (Nelson *et al.*, 1996) :

$$e(n) = d(n) - R(n)w. \quad (4.5)$$

Le vecteur w contient les coefficients des M filtres de longueur I avec :

$$w^T = \left[h^T(0) \quad h^T(1) \quad \dots \quad h^T(I-1) \right].$$

La matrice R est donnée par :

$$R(n) = \begin{bmatrix} c_1^T(n) & c_1^T(n-1) & \dots & c_1^T(n-I+1) \\ c_2^T(n) & c_2^T(n-1) & \dots & c_2^T(n-I+1) \\ \vdots & \vdots & \vdots & \vdots \\ c_L^T(n) & c_L^T(n-1) & \dots & c_L^T(n-I+1) \end{bmatrix},$$

avec

$$c_l(n) = \left[c_{1l}(n) \quad c_{2l}(n) \quad \dots \quad c_{Ml}(n) \right]^T.$$

Le vecteur w_0 qui minimise l'erreur quadratique moyenne se calcule par une inversion au moindre carré sous la forme :

$$w_0 = (E[R(n)d(n)])(E[R^T(n)R(n)])^{-1}, \quad (4.6)$$

où, l'opérateur $E[.]$ désigne l'espérance mathématique.

L'inversion directe dans le domaine temporel s'avère ainsi très coûteuse, à la fois en terme de mémoire et de temps de calcul. En effet, dans le cas où le système comporte 32 entrées et 48 sorties, avec des réponses impulsionnelles de $C(z)$ de 1024 points et une longueur de filtres de 1024 points, l'inversion multicanal directe dans le domaine temporel équivaut à effectuer une inversion de matrice de taille 32768×32768 . La longueur des filtres sur lesquels le calcul est effectué peut être choisi arbitrairement contrairement à l'inversion dans le domaine temporel.

Le calcul itératif des filtres par un algorithme de filtrage adaptatif multicanal permet un calcul

plus simple. Ce type d'algorithme est classiquement employé en télécommunications afin de rendre le ou les canaux de transmission transparents vis-à-vis de l'information transmise. Il assure une *déconvolution* du système MIMO par lequel a transité le signal. Le filtre est calculé de façon itérative par descente de gradient sur la courbe d'erreur. L'aspect adaptatif permet à l'algorithme de s'"adapter" aux variations des caractéristiques du système MIMO au cours du temps.

Dans notre cas, on considère en première approximation que le système reste invariant au cours du temps. On effectue une caractérisation préalable du système par des mesures de réponses impulsionnelles. L'algorithme adaptatif multicanal corrige le vecteur w à chaque instant n , des réponses impulsionnelles en sortie du système par une descente de gradient, afin de minimiser l'erreur à chaque étape.

La littérature concernant les algorithmes de filtrage adaptatifs est très fournie. On peut se reporter par exemple à (Mitra & Kaiser, 1993) ou (Gansler & Benesty, 2000). On trouve un grand nombre d'algorithmes que l'on peut répartir en trois classes principales :

- l'algorithme LMS (Least Mean Square) et ses dérivés (NLMS, PNLMS, PNLMS++, ...)
- l'algorithme RLS (Recursive Least Square) et ses versions rapides (SQR-RLS, RLS-Laguerre, ...)
- l'algorithme APA (Affine Projection Algorithm) et ses versions rapides (FAP, MFAP, ...)

Ces algorithmes ont, pour la plupart, des versions en sous-bande (traitement par bande de fréquence) ou orthogonalisées (visant à décorrélérer les entrées du système, déjà effectué dans le cas de l'algorithme RLS) ce qui améliore la rapidité de convergence mais accroît la complexité d'implémentation.

La classe LMS est considérée comme une référence pour sa simplicité. Cet algorithme se dérive en effet de manière immédiate à partir du problème d'inversion multicanal (Nelson *et al.*, 1996). La vitesse de convergence de l'algorithme est faible et, n'ayant pas de mémoire des erreurs passées, l'algorithme peut converger vers un minimum local ce qui n'assure pas ainsi une solution optimale. L'algorithme RLS apporte une solution à ce problème. Il utilise les erreurs passées afin d'améliorer la décorrélation des signaux d'entrées. Cet algorithme converge vers des solutions optimales avec une rapidité bien supérieure à celle de la classe LMS. Par contre, le coût de calcul de cet algorithme est très important. Il requiert des inversions de matrices de grande taille. Ceci rend délicate son utilisation pour des applications temps réel. Des versions rapides de l'algorithme RLS ont été proposées mais sont complexes à mettre en œuvre et sont instables.

La classe des algorithmes APA est intéressante car elle profite des avantages des deux classes précédentes (vitesse de convergence proche du RLS, complexité de certaines versions proche du LMS). Dans ce travail, nous utiliserons une version multicanale de l'algorithme MFAP (Liu *et al.*, 1996), choisie pour sa simplicité d'implémentation et sa vitesse de convergence (cf. Annexe B.1).

4.3 Égalisation multicanal pour la captation et la reproduction

Dans cette partie, nous souhaitons dresser un panorama de l'application des techniques d'égalisation multicanal de le cadre de la captation et de la reproduction. Cette partie est une étude bibliographique qui nous permettra d'extraire des informations utiles pour la définition de notre méthode.

4.3.1 Application à la captation

L'utilisation de bancs de microphones trouve de nombreuses applications dans le domaine de la captation sonore. Les signaux de sorties sont obtenus par combinaison de versions filtrées des signaux captés par les microphones. Ceci permet soit de privilégier la captation d'une direction privilégiée de l'espace (formation de voie (Van Veen & Buckley, 1988), (Abhayapala, 1999)), afin d'isoler un signal "utile" de signaux "parasites"; ou bien de décrire l'organisation spatiale d'un champ sonore en le décomposant sur des fonctions de rayonnement de base (ondes planes, harmoniques sphériques ou cylindriques).

Les caractéristiques du *microphone équivalent* sont données par son diagramme de directivité. Celui-ci correspond à la réponse du banc de microphones à l'excitation d'une source, repérée par sa distance et son angle de provenance vu du centre du banc de microphones. On parle de diagramme de directivité en champ lointain lorsque la source se situe à une distance "infinie" par rapport au banc de microphone (i.e. la source est en champ lointain vis à vis de l'ensemble des cellules microphoniques) et de diagramme de directivité en champ proche dans le cas contraire. La frontière entre le champ proche et le champ lointain est donnée par $r = \frac{2L^2}{\lambda}$ (distance de Rayleigh, d'après (Abhayapala, 1999)) où L est la plus grande dimension du banc de microphones.

Les décompositions du champ sonore sur une base des fonctions de rayonnement s'effectuent en réalisant un "microphone équivalent" dont la figure de directivité en champ lointain est une des harmoniques cylindriques ou sphériques; ou alors en créant un "microphone équivalent" hyper-directif ("dirac directif" d'après (Laborie, 2000)) pour la décomposition en ondes planes. Les fonctions de directivité cibles sont définies indépendamment de la fréquence. Les filtres à appliquer à chacun des signaux captés par les microphones sont calculés en fonction de la configuration géométrique, de la directivité cible et de la fréquence.

Le formalisme Ambisonique ((Gerzon, 1973), (Gerzon, 1985) (Daniel, 2000)) se base sur le décomposition du champ sonore sur une base incomplète d'harmoniques cylindriques (Ambisonique 2D) ou sphériques (Ambisonique 3D). Cette technique permet une synthèse du champ sonore dans une zone limitée de l'espace qui, de manière générale, s'agrandit en fonction du nombre de fonctions de base considérées. Le dispositif de captation a longtemps été limité à une description du champ sonore au premier ordre par l'utilisation du microphone SoundField (Craven & Gerzon, 1977). Depuis quelques années, cependant, de nombreuses recherches ont abouti à des propositions de dispositifs microphoniques d'ordre supérieur basés soit sur des géométries circulaires (description 2D, harmoniques cylindriques, (Poletti, 2000)), soit sur des géométries sphériques

(description 3D, harmoniques sphériques, (Meyer & Elko, 2002), (Abhayapala & Ward, 2002), (Daniel & Moreau, 2004), (Cotterell, 2002), et (Daniel, 2000)), soit, enfin, sur des répartitions aléatoires à l'intérieur d'une sphère (description 3D, harmoniques sphériques, (Laborie *et al.*, 2003)).

La décomposition en ondes planes est singulière. Elle se rapproche de la formation de voies. Elle vise en effet à isoler un ensemble de directions de l'espace indépendamment de la fréquence. Le type de traitement effectué change selon la géométrie du banc de microphone utilisé. Pour des géométries planaires ou linéaires, on utilise, classiquement, une distribution de retard qui permet de focaliser le banc de microphones sur une direction particulière de l'espace (Flanagan *et al.*, 1985). Une généralisation de cette approche est donnée par la transformée de Radon ((Sonke, 2000), (Berkhout *et al.*, 1997)) qui effectue une décomposition en ondes planes d'un champ sonore capturé par une distribution linéaire (description 2D) ou planaire (description 3D) de microphones. Dans le cas de géométries circulaires ou sphériques, une première décomposition du champ sonore est effectuée dans le formalisme des harmoniques cylindriques ou sphériques qui sont ensuite combinées afin de former un microphone équivalent hyper-directif (cylindriques : (Hulsebos *et al.*, 2001) (Hulsebos & de Vries, 2002), (Hulsebos *et al.*, 2003); sphériques : (Meyer & Agnello, 2003), (Laborie *et al.*, 2004)).

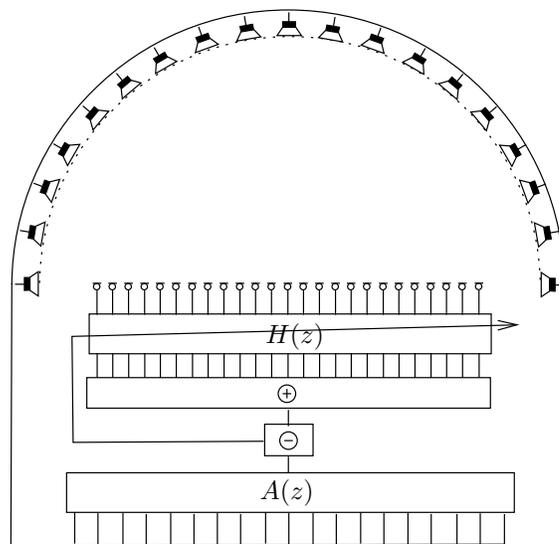


FIG. 4.2 – Inversion multicanal pour la captation sonore

Toutes ces techniques de traitement de signal donnent des filtres à appliquer à la sortie de chaque microphone afin de réaliser un diagramme de directivité donné. Elles se basent sur des capsules microphoniques idéales.

Le calcul des filtres peut se mettre sous une forme compatible avec l'inversion multicanal par un formalisme légèrement différent ((Abhayapala, 1999) reprenant (Ward & Williamson, 1999)) qui s'écrit simplement dans le domaine fréquentiel (cf. figure 4.2). On désigne par $b_r(\theta, \omega)$ le diagramme de directivité cible à une distance r pour un angle de provenance θ et une fréquence ω donnés. $d_r(\theta, \omega)$ représente le vecteur regroupant les réponses impulsionnelles mesurées au

niveau de chacun des microphones. On effectue un échantillonnage spatial des angles de provenance en P points sur lesquels on détermine $b_r(\theta_p, \omega)$ (regroupés dans le vecteur b_d) et $d_r(\theta_p, \omega)$ (regroupés dans la matrice D). En désignant par h l'ensemble des filtres à appliquer en sortie des microphones, le problème inverse multicanal revient alors à minimiser la norme du vecteur d'erreur E ,

$$E_r(\omega) = D_r h_r - b_d. \quad (4.7)$$

La minimisation de l'erreur est réalisée au sens des moindres carrés. On reconnaît une formulation de type inversion multicanal que l'on peut résoudre à l'aide des outils présentés dans la partie 4.2. Il est recommandé d'utiliser un nombre de points de description P supérieur au nombre de capsules microphoniques utilisées. Abhayapala (Abhayapala, 1999) note toutefois que les résultats sont dégradés au-dessus de la fréquence d'aliasing, laquelle est déterminée par la disposition spatiale des capsules. Il reproche au système son fonctionnement sous forme de boîte noire qui ne permet pas une compréhension intrinsèque du problème.

4.3.2 Application à la reproduction sonore

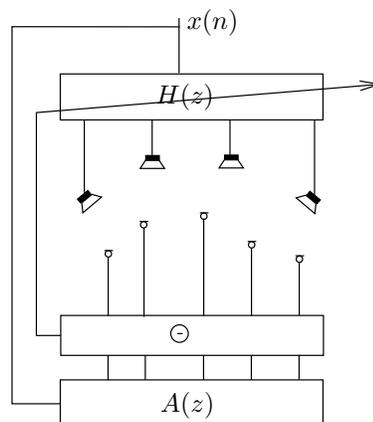


FIG. 4.3 – Inversion multicanal pour la reproduction sonore

Dans le cas de la reproduction sonore, le système MIMO est constitué d'un ensemble de M haut-parleurs dont le champ sonore est capturé au niveau de L microphones situés dans la zone d'écoute privilégiée (voir figure 4.3). L'objectif à réaliser correspond au champ acoustique cible capté par les points de contrôle. L'égalisation multicanal vise ici à minimiser l'erreur de reproduction, a priori uniquement au niveau des points de contrôle.

En reprenant la définition de Howe et Hawksford (Howe & Hawksford, 1991), on nomme "zone d'égalisation" (Region of Equalization) la portion de l'espace dans laquelle l'objectif de reproduction est convenablement synthétisé. Plus précisément, il s'agit de la zone dans laquelle, un auditeur ne peut faire la différence entre le champ acoustique cible et le champ acoustique synthétisé.

Pour une situation d'écoute particulière, la zone d'égalisation doit alors couvrir au minimum la

zone d'écoute privilégiée.

Dans cette partie, nous dressons un panorama des techniques d'égalisation multicanal de la littérature. Dans un premier temps, nous présentons un ensemble de techniques "historiques" qui se basent des techniques d'inversion multicanal pour un unique haut-parleurs mais plusieurs microphones de contrôle. Nous présentons ensuite, les techniques d'égalisation multicanal de la littérature qui utilisent tous les haut-parleurs afin d'"optimiser" la réalisation d'un objectif, par inversion multicanal, au niveau d'un certain nombre de points de contrôle et éventuellement au delà. Enfin, nous présentons deux techniques particulières qui ne basent pas sur l'inversion multicanal mais mettent en évidence l'importance de la distribution de haut-parleurs et de la technique de reproduction considérée pour la réalisation d'un objectif donné.

4.3.3 1 haut-parleur

Dans le cas le plus simple, on ne considère qu'un seul haut-parleur. Il s'agit alors d'un prolongement relativement direct de l'égalisation individuelle. Suivant la situation, on va chercher à compenser la réponse en un ou plusieurs points de l'espace ; en considérant, ou pas, l'acoustique de la pièce de restitution.

Le fait de ne pas considérer l'acoustique de la pièce de restitution revient à effectuer une égalisation de la réponse du haut-parleur dont le but est de corriger sa réponse en fréquence et en phase. Clarkson *et al.* (Clarkson *et al.*, 1985) ont été les premiers à proposer l'utilisation d'un algorithme d'inversion monocanal itératif pour le calcul du filtre d'égalisation. Cette technique n'utilise donc qu'un unique point de contrôle (mesure de la réponse impulsionnelle du haut-parleur dans l'axe) et sa "validité spatiale" dépend alors totalement de la position du point de contrôle et de la directivité du haut-parleur considéré qui n'est que rarement omnidirectionnelle. Wilson (Wilson, 1989) étend cette méthode à plusieurs points de contrôle et considère une moyenne pondérée de l'erreur pour le calcul itératif des filtres. Ceci permet de répartir l'erreur de façon plus ou moins uniforme sur l'ensemble des points de contrôle afin de prendre en compte la directivité du haut-parleur et éviter une détérioration des résultats en dehors de l'axe.

Lorsque l'égalisation concerne aussi l'acoustique de la pièce d'écoute, il s'avère utile, voir essentiel, d'utiliser plusieurs points de contrôle (Elliott & Nelson, 1989) (Elliott *et al.*, 1994). L'inversion multicanal réalise une minimisation uniforme de l'erreur à la fois en temps et en espace. Ceci permet de limiter les artefacts en dehors des points de contrôle.

Typiquement, si une seule mesure est utilisée, le microphone est susceptible de se trouver à l'emplacement d'un "zéro" d'un mode de la pièce. Le filtre d'égalisation obtenu a un niveau potentiellement très élevé à cette fréquence. En dehors de cette position précise, ce niveau élevé n'a alors plus de justification et introduit des artefacts.

L'utilisation d'un unique haut-parleur est en réalité la limitation majeure de ces procédures d'égalisation, qu'il s'agisse d'une égalisation en champ libre ou bien en espace clos. L'égalisation n'est rendue possible qu'en moyenne. Il s'avère notamment impossible d'annuler une réflexion sur une paroi de la salle de restitution dans une zone étendue de l'espace avec un unique haut-parleur.

4.3.4 Plusieurs haut-parleurs avec inversion multicanal

Dans le cas de l'utilisation de plusieurs haut-parleurs, Miyoshi et Kaneda (Miyoshi & Kaneda, 1988) ont montré qu'il existe une solution mathématique exacte au problème d'inversion multicanal à deux conditions :

- le nombre de haut-parleurs est strictement supérieur au nombre de points de contrôle,
- les fonctions de transfert entre les haut-parleurs et les microphones n'ont pas de zéro commun.

Ce résultat est connu sous le nom de théorème MINT (multi-input multi-output inverse theorem). Ils proposent une méthode associée qui permet de dériver les filtres qui réalisent l'inversion exacte. Leur longueur I est donnée par :

$$I = \frac{M(N_C - 1)}{M - L}, \quad (4.8)$$

où, N_C est la longueur des réponses impulsionnelles de $C(z)$. La synthèse de l'objectif est alors assurée de manière *exacte* au niveau des points de contrôle. Un retard additionnel permet d'assurer la causalité des filtres. Dans ce cas, rien n'indique que le résultat obtenu en dehors du point de contrôle ait une quelconque validité.

D'autres auteurs se sont intéressés aux conditions d'inversibilité de la fonction de transfert multicanal d'une pièce (voir par exemple (Putnam *et al.*, 1995) et (Flikkema, 2002)). Putnam *et al.* s'appuient sur les valeurs propres de la matrice de fonction de transfert pour chaque fréquence. Ils notent que plus celles-ci sont étalées, plus l'inversion de la matrice de fonction de transfert est complexe et plus le filtre nécessaire est long.

Dans le domaine de la reproduction sonore, l'égalisation multicanal est particulièrement employée pour la technique transaurale (Damaske, 1971) ou transaurale généralisée (Bauck & Cooper, 1996). Ces techniques visent la reproduction d'enregistrements binauraux à l'aide de haut-parleurs vers un (transaural) ou plusieurs auditeurs (transaural généralisé).

Contrairement à la diffusion au casque, chaque oreille d'un auditeur perçoit les signaux de l'*ensemble* des haut-parleurs par ce que l'on nomme les trajets croisés ainsi que par les multiples réflexions sur les parois de la pièce de reproduction. L'égalisation multicanal vise à annuler ces trajets, ainsi qu'éventuellement l'acoustique de la pièce de reproduction, au moins au niveau des oreilles des auditeurs. Ce problème est à l'origine des travaux de Miyoshi et Kaneda qui ont abouti au MINT.

Nelson *et al.* (Nelson *et al.*, 1992) sont les premiers à proposer d'utiliser les techniques itératives d'inversion multicanal grâce à l'algorithme LMS multicanal pour le calcul des filtres d'annulation des trajets croisés. Ils ne considèrent que le son direct. Ils se placent dans le cas général de la reproduction stéréophonique à deux canaux et utilisent deux microphones omnidirectionnels séparés de 18 cm en tant que point de contrôle.

Dans (Nelson *et al.*, 1996), Nelson *et al.* proposent une extension de ces travaux au transaural généralisé pour 2 auditeurs et 4 haut-parleurs. Leur méthode vise aussi la compensation de la position des haut-parleurs et/ou de l'auditeur. Ils suggèrent d'effectuer la caractérisation du

système en considérant les fonctions de transfert de la tête de l'auditeur ou bien en effectuant des mesures à l'aide d'une tête artificielle. Ceci permet de prendre en compte la présence de la tête et les phénomènes de diffraction et de masquage associés.

Dans (Nelson *et al.*, 1995) (ainsi qu'en moindre mesure dans (Nelson *et al.*, 1992) et (Nelson *et al.*, 1996)), Nelson *et al.* considèrent le problème de la zone d'égalisation en fonction de la fréquence qui semble, dans les cas étudiés (compensation de la position des haut-parleurs par rapport à une position d'écoute non centrale), de l'ordre de la longueur d'onde associée. Ils notent toutefois l'existence de configurations de source virtuelle cible/haut-parleurs/microphones qui étendent la zone d'égalisation (auditeurs peu excentrés par rapport aux haut-parleurs, sources virtuelles proches des haut-parleurs). Ces configurations correspondent à des situations privilégiées pour lequel le dispositif considéré est bien adapté. La reproduction est *physiquement* possible.

L'introduction des fonctions de transfert de la tête fait apparaître un problème de conditionnement de la matrice de fonction de transfert $C(z)$ (Kirkeby *et al.*, 1998). C'est particulièrement remarquable en basses fréquences dans le cas du "stereo dipole" (Kirkeby & Nelson, 1998), configuration de haut-parleurs écartés de 20 degrés au lieu des 60 degrés habituels qui permet d'élargir sensiblement la zone d'écoute privilégiée. Kirkeby *et al.* proposent d'introduire une matrice de régularisation pour une inversion multicanal dans un premier temps dans le domaine fréquentiel (Kirkeby *et al.*, 1998) puis dans le domaine temporel (Kirkeby & Nelson, 1999). La régularisation permet de limiter le niveau des filtres à certaines fréquences problématiques en augmentant artificiellement les valeurs singulières de la matrice de fonction de transfert jugées trop faibles.

Dans la plupart des techniques proposées précédemment, l'égalisation multicanal vise à assurer que les signaux sont correctement restitués au niveau des oreilles de l'auditeur. L'auditeur est sensé être totalement immobile. La validité des solutions proposées est uniquement locale et n'est assurée qu'à proximité des points de contrôle.

À notre connaissance, depuis les travaux sur le MINT, les techniques d'égalisation multicanal ont peu été employées dans les cas multi haut-parleurs où l'on considère aussi la compensation de l'acoustique de la salle d'écoute.

On peut toutefois citer le BSC (Boundary Sound Control) proposé par Ise (Ise, 1999). Cette technique s'appuie sur l'intégrale de Kirchhoff-Helmholtz afin d'assurer, en temps réel, la synthèse d'un champ sonore identique à l'original dans un volume de l'espace de reproduction. Ce principe se rapproche de celui du contrôle actif (Kuo & Morgan, 1999).

Cette technique consiste à capter sur une surface fermée les signaux de pression et de gradient de pression d'un champ acoustique cible. Idéalement, le dispositif de captation doit permettre de décrire parfaitement le champ de pression et de gradient de pression sur la surface. D'après l'intégrale de Kirchhoff-Helmholtz, la synthèse exacte de ces conditions aux limites sur une surface fermée de géométrie identique (surface de contrôle) grâce à un système électroacoustique permet de garantir que le champ restitué à l'intérieur de la surface est identique au champ original. Ceci n'est possible que dans la mesure où l'ensemble des sources acoustiques de l'espace de

reproduction sont situées à l'extérieur de la surface de contrôle. On reviendra sur ce point dans la partie 5.1.1 dans le cadre de la procédure d'égalisation multicanal proposée dans ces pages.

Dans l'idéal, cette technique permet une prise en compte automatique des caractéristiques de directivité des haut-parleurs utilisés ainsi que de l'acoustique de la pièce de restitution.

Ise propose d'effectuer un maillage de la surface. Dans chaque zone créée, on suppose que le champ de pression, capté par un microphone omnidirectionnel, et la composante normale du gradient de pression, captée par un microphone bidirectionnel, est constante sur toute la bande de fréquence audible.

Au niveau de la restitution, une répartition similaire de L microphones doit être disponible afin de capter, en temps réel, le signal reconstruit par un ensemble de M haut-parleurs situés en dehors de la zone d'écoute ainsi définie. Les signaux d'alimentation des haut-parleurs sont donnés par filtrage des signaux captés sur la distribution de microphones au niveau de la captation. Les $L \times M$ filtres sont actualisés en temps réel par un algorithme adaptatif multicanal afin de minimiser l'erreur de reconstruction du signal capté par les microphones situés dans l'espace de restitution.

Ce principe très général ne précise pas la position des haut-parleurs mais les simulations données dans (Ise, 1999) utilisent une répartition régulière de haut-parleurs situés, ainsi que les microphones, dans le plan horizontal. Il ne considère d'autre part que la synthèse du son direct.

En dehors de la puissance de calcul colossale nécessaire pour l'adaptation d'un nombre aussi important de filtres en temps réel, nous émettons de sérieux doutes sur la faisabilité de cette approche. D'un point de vue pratique, cela suppose de disposer des microphones tout autour de la zone de restitution qui doivent rester en place lors de l'écoute. Ces microphones doivent être répartis sur une surface fermée "suffisamment" discrétisée. L'hypothèse de champ constant sur chacun des éléments du maillage est en réalité peu crédible et le critère d'échantillonnage qui permet d'assurer la validité de la description n'est pas donné. Reste aussi à savoir comment l'auditeur peut pénétrer à l'intérieur d'une telle surface...

D'autre part, le fait de minimiser l'erreur au niveau des points de contrôle n'assure en rien que celle-ci est nulle et que le champ intérieur est convenablement restitué. L'erreur peut être minimale *mais* importante. Pour une prise en compte globale de l'acoustique de la pièce de restitution, il est nécessaire de disposer de haut-parleurs dans les trois dimensions de l'espace pour espérer annuler efficacement le champ réfléchi. Ce point n'est cependant pas mentionné dans l'article.

Malgré toutes ces critiques, remarquons tout de même que cette approche est la première à considérer une répartition de microphones avec pour objectif de garantir la validité de la solution proposée *en dehors* des points de contrôle.

4.3.5 Plusieurs haut-parleurs, méthodes alternatives

Dans cette partie, nous présentons deux méthodes d'égalisation multicanal singulières. Elles ne s'appuient pas sur une inversion multicanal mais sur une analyse des capacités du dispositif de reproduction employé.

Dans le cadre de la technique Ambisonique 3D limitée au premier ordre, Howe et Hawksford (Howe & Hawksford, 1991) proposent de s'appuyer sur la possibilité de synthèse physique du champ sonore afin de compenser les réflexions précoces de la salle d'écoute. C'est dans ce cadre qu'il définissent la zone d'égalisation ("region of equalization"). Celle-ci correspond dans le cas étudié à la zone d'écoute définie dans le cas de la reproduction Ambisonique premier ordre.

Par son lien avec l'analyse des capacités de la technique de reproduction employée, cette technique est à rapprocher de l'extension de la WFS à la compensation des réflexions précoces proposée dans la partie 3.2. Pour ces deux méthodes, c'est par une analyse des capacités du système de reproduction que la zone d'égalisation peut être définie proprement.

Une autre méthode doit retenir l'attention. Il s'agit de la méthode proposée par de Vries (de Vries, 1996) pour la prise en compte et la compensation des caractéristiques de rayonnement des haut-parleurs utilisés en pratique pour la WFS.

Cette méthode se base sur l'analyse du rayonnement d'un banc linéaire de haut-parleurs similaire à celle donnée dans la partie 2.2.1. Mais, à la différence de notre développement, les haut-parleurs sont à directivité arbitraire. On considère une ligne d'écoute préférentielle sur laquelle on détermine le champ synthétisé grâce à l'approximation de la phase stationnaire. On peut alors montrer que la contribution principale du front d'onde synthétisé est émise par le haut-parleur situé entre la source virtuelle Ψ et la position d'écoute (point de phase stationnaire). En modifiant la position d'écoute sur cette ligne, on peut alors déterminer la direction principale d'émission de chaque haut-parleur du banc (cf. figure 4.4) *pour la synthèse de la source Ψ* .

De Vries propose de calculer, pour chaque source synthétisée, un filtre d'égalisation qui vise à compenser la réponse du haut-parleur dans sa direction principale d'émission. Il montre que cette procédure permet de compenser les caractéristiques de directivité des haut-parleurs dans le cas où ils sont tous *identiques*. Cette technique est, par contre, mise en défaut lorsque les haut-parleurs n'ont pas tous la même directivité.

Cette technique assure une prise en compte et la compensation des caractéristiques de rayonnement des haut-parleurs utilisés par une analyse du champ acoustique synthétisé. Remarquons que cette technique se base sur l'approximation de la phase stationnaire et qu'elle est donc sujette aux erreurs dues à cette approximation. Elle permet d'autre part d'assurer le résultat sur une ligne de référence mais ne garantit pas la validité des résultats en dehors de cette ligne. Nous verrons cependant dans la partie 5.1.1 que le fait d'assurer la validité du champ synthétisé sur une ligne parallèle au banc de haut-parleurs permet de s'assurer, dans un certaine mesure, de la validité des résultats au delà de cette ligne.

Remarquons enfin que cette procédure d'égalisation dépend totalement de la source primaire considérée. Il s'agit alors d'une égalisation *en fonction du contenu*. Pour cette méthode, il est nécessaire de constituer une base de donnée de filtres qui seront utilisés en fonction de la source virtuelle synthétisée.

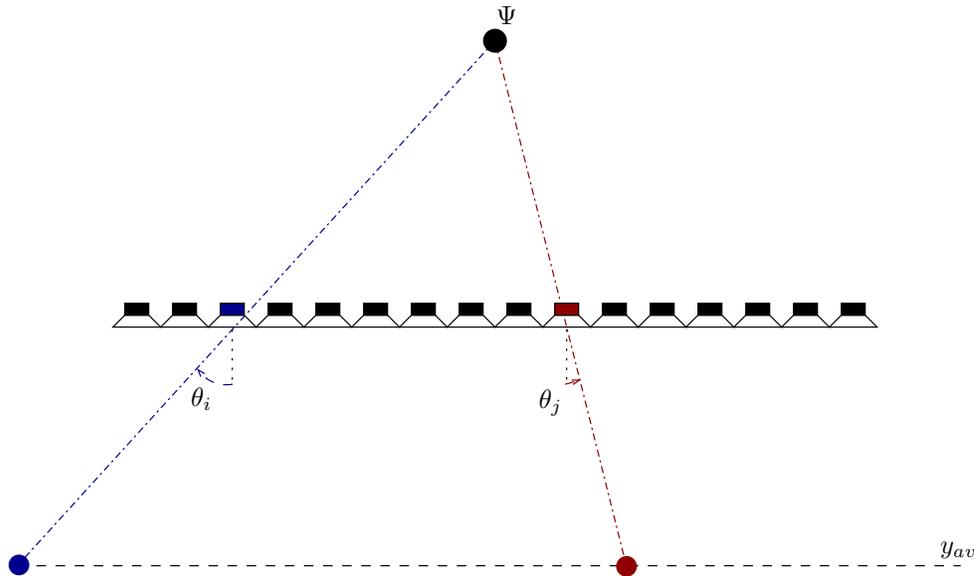


FIG. 4.4 – Égalisation multicanal pour la WFS, proposition de de Vries

4.4 Discussion : égalisation multicanal et contexte de reproduction

Dans le cadre de la reproduction sonore par un système multicanal, une méthode d'égalisation multicanal vise à assurer la synthèse d'un objectif de reproduction par un ensemble de M haut-parleurs. Elle est décrite par :

- la position et les caractéristiques de directivité des microphones employés ;
- la prise en compte ou non de l'acoustique de la pièce de restitution ;
- la définition d'un champ cible au niveau des microphones de contrôle.

La plupart des procédures d'égalisation que l'on vient de mettre en évidence se rapportent à un contexte de reproduction que l'on nommera "classique". Par contexte de reproduction, on entend support de diffusion (CD, DVD-A, DVD-V, SACD, fichiers informatiques, ...), description de la scène sonore, technique de reproduction, disposition des haut-parleurs, environnement acoustique, et zone d'écoute privilégiée.

Dans un contexte de reproduction dit "classique", le matériau sonore est délivré sous forme de canaux destinés à alimenter directement les haut-parleurs. La technique de reproduction et la description de la scène sonore sont ainsi figées et a priori inaccessibles. L'arrangement des haut-parleurs doit suivre une norme, par exemple, ± 30 degrés vis-à-vis de la position d'écoute privilégiée dans le cas de la stéréophonie à deux canaux.

Dans la plupart des applications classiques, les techniques de spatialisation utilisées sont basées sur les principes stéréophoniques (intensité, retards, combinaison) ou bien dans certains cas sur des principes transauraux ou binauraux. Ces techniques s'adressent à une zone d'écoute privilégiée réduite et à un nombre faible de haut-parleurs. Les techniques d'égalisation associées visent alors à assurer la transparence de chaque *canal* de reproduction indépendamment des autres.

Étant donné le faible nombre de haut-parleurs mis en jeu, la procédure d'égalisation mise en œuvre permet soit une égalisation non exacte dans une zone étendue, soit une égalisation plus efficace mais dans une zone réduite de l'espace. Le système de reproduction employé ne permet pas de réaliser une égalisation efficace dans une zone étendue.

La multiplication des systèmes de reproduction (Stéréophonie à 2 canaux, 5.1, 7.1, binaural, transaural, Ambisonique, Ambisonique ordres supérieurs (Near Field Compensated High Order Ambisonics, NFC-HOA) (Daniel, 2003), Ambiophonics (Glasgal, 2001), Wave Field Synthesis, ...) rend obsolète le paradigme de stockage classique (canaux destinés à alimenter directement les haut-parleurs). Non seulement, un mixage spécifique doit être réalisé pour chaque dispositif, mais aussi, pour les systèmes qui visent une synthèse physique du champ sonore (Ambisonique, NFC-HOA, Wave Field Synthesis), il n'existe pas d'arrangement de haut-parleurs standard. Le nombre de haut-parleurs et leur emplacement est spécifique à chaque installation. Il devient donc indispensable de s'appuyer sur un codage de plus haut-niveau de la scène sonore.

Pour les formats de type Ambisonique (Ambisonique, NFC-HOA), la solution choisie consiste à décomposer (encoder) le champ sonore sur un ensemble de fonctions de bases (harmoniques cylindriques ou sphériques) qui est synthétisé (décodé) sur le dispositif de reproduction spécifique à chaque installation.

Pour la Wave Field Synthesis, on dit parfois par abus de langage que la décomposition s'effectue dans le domaine des ondes planes. Toutefois, comme on a déjà pu le constater, les éléments de bases du système sont constitués de sources virtuelles spécifiées par leur emplacement et leurs caractéristiques de directivité. Les ondes planes correspondent à un cas limite de ces sources virtuelles. Ce "codage par le contenu" correspond à l'approche retenue pour le standard MPEG-4 (MPEG, 2000). L'intérêt de cette description par source sonore est d'être indépendante d'une technique ou d'un dispositif de reproduction particulier.

Dans ce cadre, une procédure d'égalisation multicanal doit assurer la transparence de chacune des fonctions de base de la technique de reproduction (harmoniques cylindriques ou sphériques pour Ambisonique, sources virtuelles pour la WFS). La synthèse de la scène sonore est ainsi réalisée,

- pour Ambisonique, en synthétisant les harmoniques cylindriques ou sphériques à partir des filtres obtenues par égalisation multicanal et en les alimentant avec les signaux affectés à chaque harmonique cylindrique ou sphérique ;
- pour la WFS, en synthétisant les sources virtuelles spécifiées dans la scène à partir des filtres obtenues par égalisation multicanal et en les alimentant avec les signaux associés à la description de la scène.

La procédure d'égalisation multicanal effectue un *contrôle du champ sonore* synthétisé par l'*ensemble* des canaux du dispositif de reproduction.

Pour ces techniques de reproduction, la zone d'égalisation associée doit être d'une taille au moins équivalente à celle de la zone d'écoute privilégiée. La disposition des points de contrôle doit donc

permettre une *description* du champ sonore et de l'objectif de reproduction valable dans l'ensemble de la zone d'écoute. On doit d'autre part s'assurer que le dispositif de reproduction est capable de synthétiser l'objectif de reproduction dans toute la zone d'écoute.

Dans la suite, nous proposerons une méthode d'égalisation multicanal qui tient compte et tire partie des spécificités de la Wave Field Synthesis. Dans un premier temps, cette technique sera définie puis appliquée au contrôle du rayonnement au champ libre d'un banc de haut-parleurs dont les transducteurs ont des caractéristiques de directivité arbitraires (chapitre 5). Nous en proposerons ensuite une évolution pour le contrôle de la compensation de l'acoustique de la pièce de reproduction (chapitre 6).

Chapitre 5

Contrôle du rayonnement en champ libre

Dans le cadre de la WFS, l'utilisation de haut-parleurs dont les caractéristiques de directivité s'éloignent de l'idéal (omnidirectionnel) est susceptible d'introduire de nombreuses erreurs. La méthode proposée par de Vries (de Vries, 1996) n'assure une compensation de la directivité des haut-parleurs que dans les cas où ils sont tous identiques.

Dans ce chapitre, nous nous plaçons dans un cas plus général et nous proposons une méthode *d'égalisation multicanal* spécifique à la Wave Field Synthesis. Le but de cette méthode est d'assurer ou tout au moins améliorer la transparence de la reproduction des sources virtuelles dans *l'ensemble* de la zone de reproduction.

Dans un premier temps, une description complète de la méthode est proposée. Nous mettons l'accent sur ses spécificités vis-à-vis des autres méthodes d'égalisation multicanal. Nous appliquons ensuite cette méthode à un banc linéaire de haut-parleurs omnidirectionnels idéaux. Ces simulations du cas idéal permettent de vérifier que le champ synthétisé est valide au delà des points de contrôle, idéalement dans *toute* la zone d'écoute.

Enfin, cette technique est appliquée à un banc de haut-parleurs MAP (Multi Actuator Panels), dont les caractéristiques de rayonnement complexes s'écartent des sources secondaires idéales considérées dans la théorie. Nous verrons que la méthode proposée permet effectivement d'améliorer la transparence acoustique du champ synthétisé par ces haut-parleurs dans le cadre de la WFS.

5.1 Égalisation multicanal pour la WFS

La Wave Field Synthesis repose sur le contrôle du champ sonore à la frontière de la zone d'écoute. Ce contrôle s'effectue à l'aide d'un ou de plusieurs bancs linéaires de haut-parleurs. La formulation, dérivée de l'intégrale de Rayleigh 1 (cf. partie 1.2), met en jeu des sources secondaires (haut-parleurs) omnidirectionnelles idéales, c'est-à-dire dont le rayonnement est donné par la

fonction de Green de l'espace à 3 dimensions.

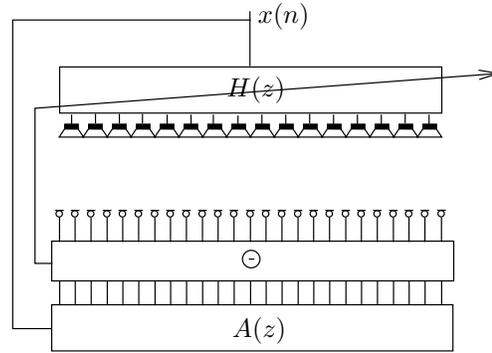


FIG. 5.1 – Inversion multicanal pour la Wave Field Synthesis

La méthode d'égalisation multicanal que nous proposons consiste à *reporter* le contrôle du champ sonore synthétisé par les haut-parleurs à une ligne de microphones située à distance du banc de haut-parleurs (voir figure 5.1). Le contrôle du champ sonore est ensuite réalisé par inversion multicanal dans le cadre de la synthèse d'une source virtuelle cible. Nous montrons qu'une disposition linéaire dans le plan horizontal des points de contrôle (microphones) permet, dans une certaine mesure, d'effectuer un contrôle efficace du champ sonore jusqu'à la fréquence d'aliasing spatial et ceci pour l'ensemble de la zone d'écoute. Au-dessus de cette fréquence limite, nous proposons une méthode alternative basée sur une procédure d'égalisation individuelle des haut-parleurs et un contrôle de type énergétique du champ sonore synthétisé.

L'égalisation multicanal comprend quatre étapes principales que nous détaillons dans cette partie :

1. identification du système MIMO ;
2. définition des réponses cibles en relation avec un ensemble de sources virtuelles ;
3. calcul de filtres par inversion multicanal sous la fréquence d'aliasing ;
4. égalisation du système au-dessus de la fréquence d'aliasing et composition des filtres.

5.1.1 Identification du système MIMO

L'identification du système MIMO s'effectue en mesurant les réponses impulsionnelles de chaque haut-parleur avec tous les microphones, en champ libre. Afin d'assurer un contrôle du champ sonore dans une large zone d'écoute, la mesure *locale* doit permettre une caractérisation *globale* du rayonnement de l'ensemble des éléments du système (haut-parleurs).

Supposons, dans un premier temps, que les microphones sont disposés sur une surface fermée $\partial\Omega$, située dans l'espace de reproduction et englobant la zone d'écoute. Les haut-parleurs ainsi que les parois de l'espace de reproduction sont situées à l'extérieur de la zone de reproduction. L'ensemble des sources sonores (haut-parleurs et sources images) est donc situé à l'*extérieur* de $\partial\Omega$. Suivant la nature des microphones, la mesure des réponses impulsionnelles réalise une caractérisation du champ de pression (microphones omnidirectionnels) ou du gradient normal de

pression (microphones bidirectionnels, "microphones figure 8") pour une surface fermée échantillonnée.

Dans le cas où l'on dispose simultanément du champ de pression et du gradient normal de pression sur la *totalité* de la surface $\partial\Omega$ (distribution continue de microphones), la description fournie peut se formuler comme un problème acoustique aux limites. Ce problème possède une solution *unique* à l'intérieur de $\partial\Omega$ (Bruneau, 1998). Le champ intérieur à $\partial\Omega$ se déduit des conditions aux limites par l'intégrale de Kirchhoff-Helmholtz.

Dans le cas particulier où $\partial\Omega$ est un plan infini et où l'ensemble des sources sonores est situé dans un demi-espace Ω_Ψ , les champs synthétisés, à partir de la pression ou du gradient de pression captés sur $\partial\Omega$, sont *identiques* dans Ω_R (sous-espace de reproduction) et, *identiques et en opposition de phase* dans Ω_Ψ (Bruneau, 1998). Ainsi, si l'ensemble des sources primaires sont dans Ω_Ψ , la connaissance *exclusive* de la pression *ou* du gradient de pression sur $\partial\Omega$ décrit de façon *unique* le champ sonore rayonné dans Ω_R (intégrales de Rayleigh 1 et 2, cf. partie 1.1).

La méthode d'égalisation multicanal vise le contrôle du rayonnement en champ libre d'un banc linéaire de haut-parleurs dont la directivité n'est pas connue a priori. On peut alors définir un plan infini qui sépare l'espace à trois dimensions et place l'ensemble des sources primaires (haut-parleurs) dans un des deux sous-espaces complémentaires ainsi définis.

Dans ce cadre, la connaissance de la pression *ou* du gradient normal de pression sur l'ensemble du plan infini assure une description *exacte* du champ sonore synthétisé dans *l'ensemble* du sous-espace de reproduction.

Nous proposons de considérer un banc de microphones de pression (omnidirectionnels), linéaire de taille finie à espacement constant, situé parallèlement au banc de haut-parleurs. Cette simplification de la géométrie du banc de microphones se rapproche des simplifications utilisées pour dériver la Wave Field Synthesis à partir de l'équation de Rayleigh 1 (cf. partie 1) :

1. Les bancs de haut-parleurs et les auditeurs étant situés dans le plan horizontal, le champ sonore ne doit être décrit précisément que dans ce plan. L'intégrale de Rayleigh 2 donne le champ synthétisé dans le sous-espace de reproduction à partir du signal de pression capté sur le plan infini. Une analyse de type phase stationnaire (cf. annexe A.1) sur le champ synthétisé dans le plan horizontal par chacune des colonnes du plan infini, proche de celle réalisée pour la Wave Field Synthesis (cf. partie 1.2.1), montre que la contribution principale du champ synthétisé est donnée par les microphones situés dans le *plan horizontal*.
2. Lors de la synthèse d'une source virtuelle par la Wave Field Synthesis, le champ émis dans le plan horizontal n'est correct qu'à l'intérieur d'une zone de visibilité déterminée par l'extension du banc de haut-parleurs et la position de la source virtuelle. En dehors de cette zone de visibilité, le champ est "nul" (à la diffraction près). On peut ainsi réduire la taille du banc de microphones et se concentrer sur la partie de la ligne de microphone située dans la zone de visibilité.

3. On verra dans la suite qu'un contrôle efficace du champ sonore n'est possible que jusqu'à la fréquence d'aliasing. Le banc de microphones ne doit donc décrire proprement le rayonnement de chacun des haut-parleurs que jusqu'à la fréquence d'aliasing maximale (pour l'ensemble des sources virtuelles cible et l'ensemble des positions de microphone).

Toutes ces simplifications ne permettent pas, a priori, de donner une description exacte du rayonnement des haut-parleurs en dehors des points de contrôles. Toutefois, plus le nombre de points de contrôles est important, plus le coût de calcul associé à l'inversion multicanal est élevé. On cherche donc plutôt à utiliser un nombre de microphones *suffisant* pour décrire les *contributions principales* du champ dans le plan horizontal mais pas plus.

Le nombre fini de microphones introduit un effet d'aliasing spatial dans la description du champ sonore. Dans la littérature des bancs de microphones, on considère classiquement une fréquence d'aliasing donnée par :

$$f_{al}^{Mes} = \frac{c}{\Delta x(1 + \sin \theta_{\Psi, max})}, \quad (5.1)$$

où, $\theta_{\Psi, max}$ est l'angle d'incidence maximal du champ sonore rayonné par les haut-parleurs.

5.1.2 Champ sonore cible

La deuxième étape de la méthode consiste à définir un ensemble de cibles associées à des sources virtuelles. Cet ensemble doit permettre de reproduire dans une zone d'écoute donnée *n'importe quelle* source virtuelle spécifiée par l'utilisateur.

Une source virtuelle est définie par sa position par rapport à un point de référence (centre la zone d'écoute considérée) et sa figure de directivité. Nous avons comme objectif que l'utilisateur ne perçoive pas la différence entre la source virtuelle qu'il a spécifiée et celle qui est effectivement reproduite. Dans cette partie, nous proposons des critères perceptifs qui nous permettent de dériver un ensemble suffisant de cibles.

Position des sources sonores

A tout banc linéaire de haut-parleurs, on associe une zone d'écoute privilégiée. Toute source synthétisée doit ainsi être visible, au sens acoustique du terme, à travers ce banc dans l'*ensemble* de la zone d'écoute. Ceci définit une zone limitée de l'espace dans laquelle peuvent se situer les sources reproduites qui remplissent ces critères "inversés" de visibilité. La "zone de visibilité des sources" dans laquelle celles-ci sont situées est définie par des critères géométriques simples (cf. figure 5.2).

Un nombre fini de positions de sources cibles est défini par un maillage de cette zone. La validité de ce maillage est donnée en considérant les facultés perceptives de localisation des sources sonores par le système auditif humain.

Nous considérons comme critères perceptifs le biais et le flou de localisation. Le biais de localisation est la différence entre la direction de la source sonore et la direction moyenne de l'événement

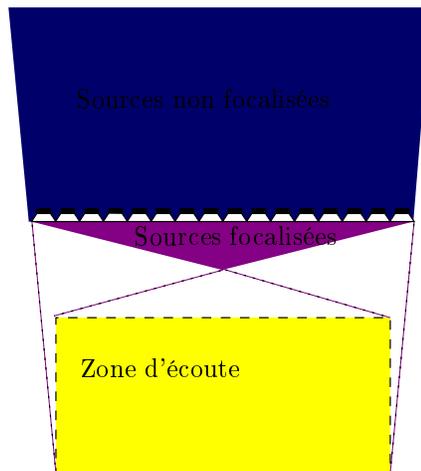


FIG. 5.2 – Zone de visibilité des sources virtuelles dans la zone d'écoute

auditif. Le flou de localisation (localisation blur) ou angle minimum audible (minimum audible angle) est défini comme le seuil différentiel de perception angulaire, c'est-à-dire la déviation angulaire pour laquelle 50% des sujets perçoivent un déplacement de l'événement auditif (repris de (Baskind, 2003), citant (Blauert, 1999)).

Le flou de localisation semble être le critère local le plus adapté pour juger de la validité du maillage. Blauert (Blauert, 1999) donne un flou de localisation pour des séquences de bruits blancs de 100 ms de ± 3.6 degrés en face et de ± 10 degrés de côté.

Cette distinction entre "en face" et "de côté" n'a cependant pas de sens pour la WFS. En effet, la WFS permet une navigation de l'auditeur à l'intérieur de la zone d'écoute et on ne connaît pas, a priori, l'orientation de la tête de l'auditeur. On considère la valeur la plus faible soit ± 3.6 . Pour une source Ψ_1 donnée, ce critère définit une zone autour de Ψ_1 dans laquelle toute source Ψ_2 est perçue à une position identique à Ψ_1 .

Nous proposons alors de définir un critère globale de validité du maillage de la zone de visibilité des sources : "pour toute source Ψ_i , il existe une source Ψ_j située à moins de 5 degrés de Ψ_i pour toute position R de l'auditeur dans la zone d'écoute". Ce critère correspond ainsi à un flou de localisation de ± 2.5 degrés, inférieur à la valeur donnée par Blauert, ± 3.6 degrés.

Directivité des sources sonores

On a vu dans la partie 1.3 que le rayonnement d'une source acoustique peut être décomposé sur un ensemble de fonctions de directivité élémentaires appelées harmoniques sphériques définies par un degré et un ordre. La synthèse du rayonnement d'une source est alors possible par la combinaison des harmoniques sphériques.

L'utilisation d'un banc linéaire de haut-parleurs pour la WFS ne permet la synthèse que d'un nombre réduit d'harmoniques sphériques (cf. partie 1.3) définies uniquement par leur degré m ($m \in \mathbb{Z}$). Les caractéristiques du rayonnement synthétisé s'écartent toutefois de la cible. Le champ synthétisé par un banc linéaire de haut-parleurs omnidirectionnels présente, en effet, une

symétrie de révolution autour de l'axe du banc quelque soit la directivité de la source virtuelle synthétisée. De manière générale, la directivité synthétisée en dehors du plan horizontal dépend principalement de la directivité des haut-parleurs.

La reproduction de la directivité par la Wave Field Synthesis est ainsi limitée. Le nombre d'harmoniques nécessaires pour chaque position de source virtuelle et la prise en compte des défauts de reproduction dépassent le cadre de cette étude.

Champ sonore cible

Le champ sonore cible est déterminé à partir de la position des sources et de leurs caractéristiques de rayonnement, ainsi que par la position des microphones et leurs caractéristiques de directivité. Ce champ est donc défini au niveau des points de contrôle et correspond aux contributions principales de la cible dans le reste du plan horizontal.

On a vu dans la partie 2.2.1 que la loi d'atténuation du champ sonore synthétisé par un banc linéaire de haut-parleurs présente une loi d'atténuation qui ne correspond pas à celle de la source virtuelle cible. Dans le cas de la WFS, le niveau n'est correct que sur une "ligne de référence" parallèle au banc de haut-parleurs et dont on peut choisir la position. Il est alors possible de choisir une distribution de niveaux sur la ligne de microphones qui garantit la synthèse d'un niveau correct à une distance quelconque du banc de haut-parleur. Nous recommandons de choisir pour ligne de référence une ligne parallèle au banc et située au milieu de la zone d'écoute. Nous proposons de normaliser en niveau (hors fonction de directivité) et en temps d'arrivée l'ensemble des sources virtuelles spécifiées au centre de la zone d'écoute. Dans la suite, on prendra ce point comme origine O du repère cartésien.

Le champ spécifié au niveau d'un microphone M situé au point (x_M, y_M) pour un banc de haut-parleurs L et une source virtuelle Ψ_m située en (x_ψ, y_ψ) est ainsi donné par :

$$A(x_M, y_M, t) = \sqrt{\frac{d_O^L}{d_M^L}} \sqrt{\frac{d_M^L + d_\Psi^L}{d_O^L + d_\Psi^L}} \frac{d_\Psi^O}{d_\Psi^M} \delta\left(t - \frac{d_\Psi^M - d_\Psi^O}{c} + \tau_d\right), \quad (5.2)$$

où d_A^B est la distance entre le point A et le point B et d_A^L la distance du point A à la ligne L et τ_d est un retard supplémentaire afin de garantir la causalité des filtres calculés.

Dans le cas d'une distribution de haut-parleurs plus complexe, on ne sait pas, pour le moment, caractériser complètement la loi d'atténuation. On pourra soit :

- choisir la ligne de mesure comme ligne de référence ;
- mener une optimisation de niveau en tenant compte de l'atténuation effective de l'ensemble, à condition de savoir la définir.

5.1.3 Inversion multicanal

Une fois le système identifié et les champs cibles définis, les filtres sont calculés par inversion multicanal. Celle-ci est réalisée dans le domaine temporel par l'algorithme adaptatif multicanal

MFAP (cf. annexe B.1).

L'inversion multicanal permet de minimiser l'erreur de reproduction, au sens des moindres carrés. Ceci permet de garantir que la contribution principale du champ synthétisé dans le plan horizontal, en aval de la ligne de contrôle, est "optimisée" au sens des moindres carrés. Afin de garantir la propagation du champ sonore "égalisé", l'erreur, au niveau des points de contrôle doit être aussi faible que possible.

Avant cela, nous proposons quelques étapes supplémentaires qui permettent d'améliorer les performances et le comportement de l'algorithme.

Choix des haut parleurs et des microphones

Une source virtuelle n'est visible à travers le banc de haut-parleurs que dans une portion limitée de l'espace. En dehors de cette zone de visibilité de la source à travers le banc, le champ synthétisé est "nul" (à la diffraction près). Comme nous ne cherchons pas à contrôler l'effet de diffraction, nous proposons de sélectionner les microphones situés à l'intérieur de la zone de visibilité de la source virtuelle.

Il est possible de déterminer une zone de visibilité à travers la distribution de microphones sélectionnés (cf. figure 5.3). Nous remarquons ainsi que certains haut-parleurs ne contribuent pas à la synthèse du champ associé à la source Ψ dans la zone déterminée par la distribution de microphones choisie. Nous proposons alors de ne pas les inclure dans le calcul des filtres par inversion multicanal. Ce choix s'avère judicieux dans les fréquences moyennes (500-2000 Hz) pour lesquelles l'effet de visibilité est très marqué (peu de diffraction). Ces haut-parleurs ne permettent en effet que de minimiser l'erreur au niveau des points de contrôles mais on remarque, en pratique que ceci provoque une sur-amplification inutile du niveau alloué aux fréquences moyennes. Ceci a tendance à dégrader le résultat de l'inversion multicanal en dehors de la fenêtre de visibilité, ce qui risque de provoquer des erreurs lors de la propagation naturelle au delà de la ligne de contrôle. Aux basses fréquences, ce choix est toutefois discutable car il revient à diminuer la taille effective du banc de haut-parleurs. Un traitement par bandes pourrait alors s'avérer plus efficace mais n'a pas encore été mis en œuvre.

En pratique, nous recommandons de sélectionner les microphones situés strictement à l'intérieur de la zone de visibilité et les haut-parleurs situés éventuellement légèrement en dehors de celle-ci.

Mise en forme de la matrice de fonction de transfert

La WFS fournit une description détaillée des fonctions d'alimentation à appliquer aux haut-parleurs d'un dispositif de reproduction donné. Ces signaux d'alimentation correspondent à des versions filtrées, atténuées et retardées du signal affecté à la source virtuelle. Les retards à appliquer sont donnés par le temps nécessaire pour la propagation des fronts d'onde depuis la position de la source virtuelle jusqu'à chaque haut-parleur. Ceci revient à allumer les haut-parleurs pro-

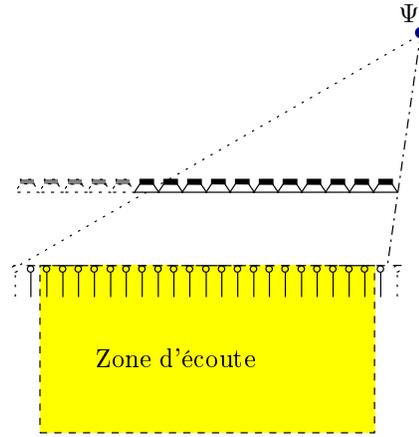


FIG. 5.3 – Choix des haut-parleurs et des microphones pour l'inversion multicanal

gressivement afin de synthétiser le front d'onde de proche en proche.

Lorsque l'on réalise l'inversion multicanal *directement* à partir des réponses impulsionnelles mesurées, on remarque que le pic principal d'énergie se situe précisément à la position de ces retards (Corteel, 2001). Ces retards sont en effet nécessaires pour la synthèse d'un tel front d'onde par le banc de haut-parleurs utilisé.

Dans le cas d'un banc de haut-parleurs linéaire d'une dizaine de mètre, la différence entre le retard le plus faible et le retard le plus important peut être de l'ordre de 1000 échantillons à 48 kHz. Ceci requiert des filtres de grande taille afin de permettre la formation du front d'onde et l'égalisation du système.

Afin de réduire la longueur des filtres nécessaires et d'"anticiper" la formation des fronts d'onde par le banc de haut-parleurs, nous proposons d'"appliquer" ces retards sur la matrice des fonctions de transfert $C(z)$. Cette opération de "mise en forme" de la matrice des fonctions de transfert donne une matrice modifiée $\widetilde{C}^\Psi(z)$ telle que :

$$\widetilde{C}_m^\Psi l(z) = z^{-\tau_m^\Psi} \times C_m l(z), \quad (5.3)$$

où, $C_m l(z)$ est la fonction de transfert du haut-parleur m mesurée par le microphone l , et τ_m^Ψ est le retard donné par l'équation 2.4 à affecter au haut-parleur l pour la synthèse de la source Ψ .

En plus de la réduction de la taille des filtres nécessaires, cette technique fournit des filtres qui sont alignés en temps quelque soit la source virtuelle synthétisée. Ceci est un avantage pour la reproduction en temps-réel de sources dynamiques car l'interpolation entre les différents filtres est plus facile à réaliser.

On pourrait être tenté d'appliquer aussi à $C(z)$ les atténuations données par la WFS. Ceci reviendrait à diminuer le poids de certains haut-parleurs dans la description du système. Ceci se traduirait par un étalement supplémentaire des valeurs propres de la matrice des fonctions de transfert. Nous ne recommandons pas d'effectuer cette opération car elle a tendance en réalité à dégrader le conditionnement du problème.

Fréquence limite de calcul

Dans la partie 2.4, on a vu que, pour un banc de haut-parleurs de taille finie, la fréquence d'aliasing spatial ne dépend pas uniquement de la position de la source sonore et l'espacement entre les haut-parleurs mais varie aussi suivant la position d'écoute. Une description dans le domaine temporel de la synthèse du champ synthétisé par la Wave Field Synthesis a permis d'en donner une formulation explicite. Cette formulation s'apparente à un échantillonnage irrégulier de la synthèse du front d'onde cible. La fréquence d'aliasing spatial est alors obtenue en examinant les temps d'arrivée successifs des contributions de chacun des haut-parleurs pour une position d'écoute donnée.

Pour l'inversion multicanal, on s'attend à ce que les filtres obtenus n'aient aucune validité, au-dessus de la fréquence d'aliasing, en dehors des points de contrôle. Le champ sonore synthétisé est, en effet, très perturbé au-dessus de cette fréquence limite et a un comportement uniquement *local*. Ce champ peut être vu, dans le domaine de Fourier spatial, comme un recouvrement de composantes spatiales, répétitions de celles de la source cible (Nicol, 1999) (Start, 1997). Il ne donne pas ainsi accès de manière *unique* au champ de la source cible. On cherche alors une fréquence limite à partir de laquelle les filtres calculés par l'inversion multicanal présentent un comportement local. Ils ne permettent une compensation du champ synthétisé qu'au niveau des points de contrôle.

Par analogie à la définition de la fréquence d'aliasing spatial pour une position d'écoute donnée (cf. partie 2.4), nous proposons de définir une fréquence d'aliasing spatial rapportée à chacun des haut-parleurs pour une position d'écoute R . Il suffit pour cela, pour un haut-parleur n et une position d'écoute R , d'examiner les temps d'arrivée des contribution qui proviennent du même demi-banc de haut-parleur (de chaque côté du point de phase stationnaire) juste avant et juste après celle du haut-parleur n . En général, il s'agit des haut-parleurs $n - 1$ et $n + 1$. En définissant $\Delta\tau(n)$ comme la différence de temps d'arrivée maximal entre les contributions de n et $n - 1$, et $n + 1$ et n . La fréquence d'aliasing $f_{al,ls}^{\Psi_m}(n, r_R)$ du haut-parleur n pour la synthèse d'une source Ψ_m au point R sera :

$$f_{al,ls}^{\Psi_m}(n, r_R) = \frac{1}{|\Delta\tau(n)|}. \quad (5.4)$$

On doit ainsi calculer la fréquence d'aliasing spatial "rapportée aux haut-parleurs" $f_{al,ls}^{\Psi_m}$ pour les M haut-parleurs du banc et les L microphones utilisés, pour chaque source virtuelle Ψ_m considérée. La fréquence limite de calcul des filtres pour chaque haut-parleur est donnée par la valeur la plus faible de :

- la fréquence d'aliasing de description (banc de microphones) ;
- la fréquence d'aliasing spatial pour le haut-parleur considéré vis-à-vis d'une source virtuelle donnée et de l'ensemble des points de contrôle.

D'un point de vue pratique, on recommande d'utiliser un espacement de microphone de l'ordre des deux tiers de celui des haut-parleurs pour que la fréquence d'aliasing de description soit toujours plus grande que celle rapportée aux haut-parleurs.

Le calcul des filtres par inversion multicanal dans le domaine temporel peut être accéléré par sous-échantillonnage des réponses impulsionnelles de $\widetilde{C}^\Psi(z)$. On choisit un facteur de sous-échantillonnage tel que la fréquence de Nyquist associée excède la fréquence limite de calcul la plus élevée.

Afin de vérifier la formule 5.4, on réalise des simulations avec un banc linéaire de 40 haut-parleurs omnidirectionnels idéaux espacés de 15 cm (6 m de longueur). L'identification du système MIMO (obtention de la matrice $C(z)$) est réalisée en simulant les fonctions de transfert de chaque haut-parleurs vers un ensemble des microphones. Les haut-parleurs et les microphones sont idéalement omnidirectionnels. La longueur totale du banc linéaire de microphones considéré est de 6 m. Celui-ci est centré, parallèle aux haut-parleurs à 2 m de distance. On considère deux espacements différents pour les microphones :

- 15 cm, $C_{40}(z)$, fréquence d'aliasing de description entre 1200 et 1400 Hz (cf. figure 5.6(a))
- 1,5 cm, $C_{400}(z)$, fréquence d'aliasing de description entre 11000 et 13000 Hz (cf. figure 5.7(a)).

Les filtres sont obtenus par inversion multicanal dans le domaine fréquentiel suivant la méthode donnée dans la partie 4.2.1. La matrice de régularisation est choisie telle que $\mathbf{B}^{*T}(\omega)\mathbf{B}(\omega) = I$, I étant la matrice identité ayant pour taille le nombre de haut-parleurs utilisés ; et $\gamma = 0.01$. On donne les résultats pour une onde plane se propageant perpendiculairement au banc de haut-parleurs (onde plane à 0 degré). Ceux-ci restent similaires sur d'autres types de sources et de configuration de haut-parleurs/microphones.

Après avoir calculé les filtres, on simule la réponse du système sur un banc de microphones

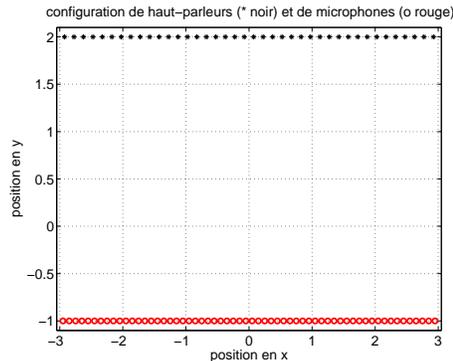


FIG. 5.4 – Configuration pour la vérification des résultats

distant (60 microphones espacés de 10 cm à 3 m du banc de haut-parleurs, cf. figure 5.4) dans le but de vérifier l'efficacité de l'égalisation multicanal en dehors des points de contrôle. La simulation est réalisée en déterminant les fonctions de transfert de chacun des haut-parleurs sur l'ensemble des microphones distant. La réponse du banc est alors obtenue par convolution des filtres calculés avec la réponse des haut-parleurs puis sommation de la contribution de l'ensemble des haut-parleurs sur chaque position de microphone.

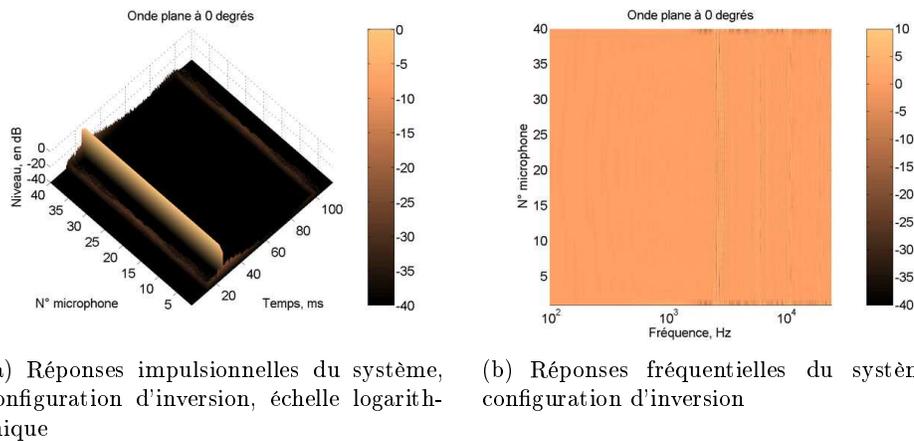


FIG. 5.5 – Inversion multicanal avec 40 microphones (espacement de 15 cm), réponse du système au niveau de la configuration d'inversion

On montre tout d'abord la réponse du système auquel on applique les 40 filtres calculés par inversion multicanal de $C_{40}(z)$ au niveau des microphones utilisés pour l'inversion (cf. figure 5.5). Les réponses impulsionnelles et fréquentielles sont presque parfaites. L'égalisation est ainsi "idéale" au niveau des points de contrôle. En effet, on est, à un microphone près, dans les conditions du théorème MINT qui garantit l'existence mathématique d'une solution exacte au niveau des points de contrôle.

Les ensembles de figures 5.6 et 5.7 représentent :

1. la configuration sur laquelle est effectuée l'inversion multicanal ;
2. les réponses impulsionnelles du système au niveau du banc de microphones distant, échelle logarithmique ;
3. ces mêmes réponses impulsionnelles auxquelles on applique un filtrage passe-bas avec une fréquence de coupure de 2000 Hz (fréquence inférieure à la fréquence d'aliasing spatial WFS pour toutes les positions de microphones), en échelle logarithmique ;
4. les réponses fréquentielles des filtres ainsi que la fréquence de coupure donnée par les multiples critères d'aliasing spatial (description du système, WFS rapportée aux haut-parleurs).

Pour l'inversion réalisée avec 40 microphones, au niveau du banc de microphones distant, les réponses impulsionnelles du système s'éloignent assez fortement du cas idéal (cf. figure 5.6(b)). Les réponses impulsionnelles s'étalent sur une centaine de millisecondes. Les réponses impulsionnelles restent toutefois correctes en dessous de la fréquence d'aliasing spatial WFS (cf. figure 5.6(c)), même si on note la présence de précurseurs de faible niveau. Ceci induit dans la réponse fréquentielle du filtre (cf. figure 5.6(d)) de légères variations juste au-dessus de la fréquence limite, en particulier pour les haut-parleurs situés aux extrémités. La fréquence limite est, dans ce cas, donnée par la fréquence d'aliasing spatial de description. Plus haut en fréquence, la réponse fréquentielle devient très agitée avec des pics de niveau élevé.

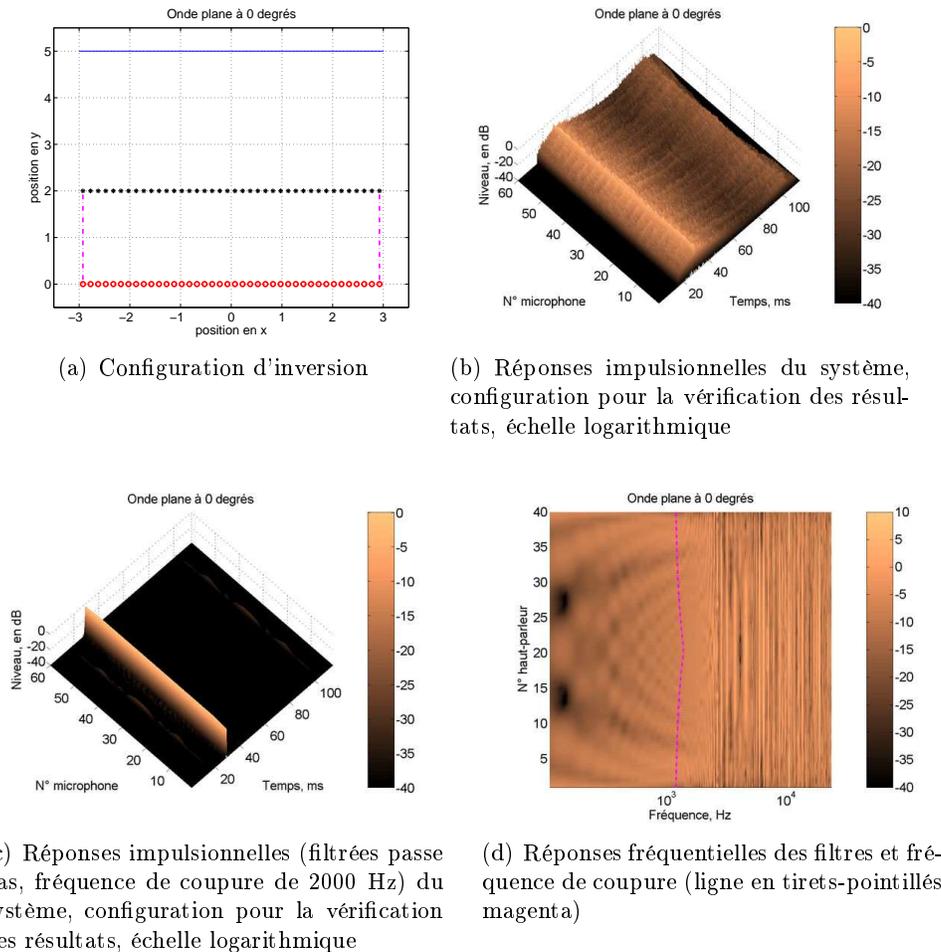
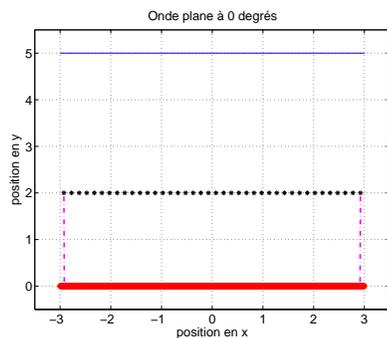


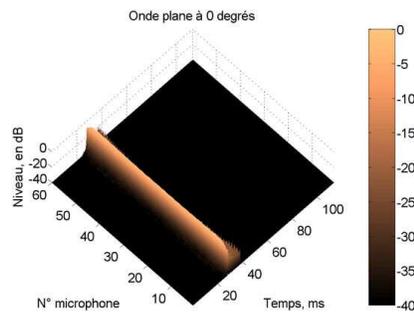
FIG. 5.6 – Inversion multicanal avec 40 microphones (espacement de 15 cm)

L'ensemble figures 5.7 présente le même type de résultats pour une inversion réalisée sur un banc de 400 microphones espacés de 1,5 cm. Les réponses impulsionnelles du système correspondent bien à celle de la source cible sur toute la bande des fréquences audibles (cf. figures 5.7(b) et 5.7(c)). En basses fréquences, les réponses fréquentielles des filtres associés sont semblables à celles des filtres calculés sur 40 positions de microphones (cf. figures 5.6(d) et 5.7(d)). On remarque une baisse sensible du niveau des filtres à partir de la fréquence d'aliasing spatial rapportée aux haut-parleurs (cf. équation 5.4). La réponse fréquentielle du système (non représentée ici) fait apparaître une baisse similaire du niveau synthétisé au-dessus de la fréquence d'aliasing spatial. L'optimisation des filtres réalisée par l'inversion multicanal consiste ainsi à diminuer le niveau d'émission des hautes fréquences. La réponse du système alors obtenue n'est plus parfaitement plate mais laisse apparaître une baisse sensible du niveau des hautes fréquences ce qui risque d'introduire une coloration notable du signal reproduit.

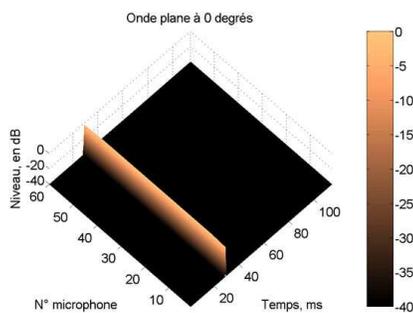
L'ensemble des figures 5.8 représente, en échelle logarithmique, les réponses impulsionnelles des filtres calculés par inversion multicanal pour les deux configurations proposées. On remarque que les filtres obtenus par inversion sur 40 microphones (cf. figure 5.8(a)) ont une réponse im-



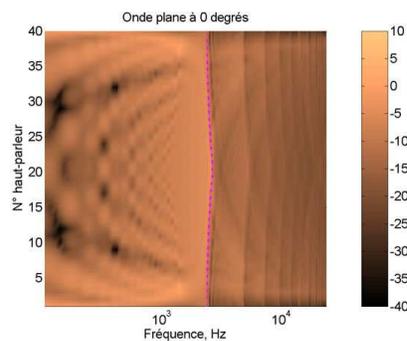
(a) Configuration d'inversion



(b) Réponses impulsionnelles du système, configuration pour la vérification des résultats, échelle logarithmique

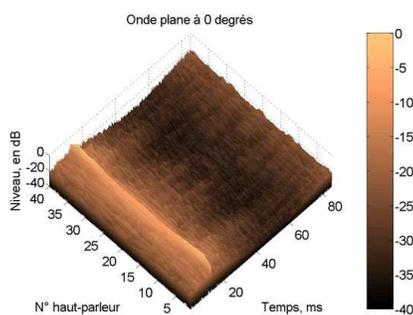


(c) Réponses impulsionnelles (filtrées passe bas, fréquence de coupure de 2000 Hz) du système, configuration pour la vérification des résultats, échelle logarithmique

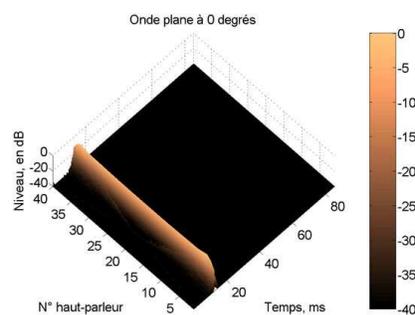


(d) Réponses fréquentielles des filtres et fréquence de coupure (ligne en tirets-pointillés magenta)

FIG. 5.7 – Inversion multicanal avec 400 microphones (espacement de 1,5 cm).



(a) Inversion multicanal réalisée avec 40 microphones



(b) Inversion multicanal réalisée avec 400 microphones

FIG. 5.8 – Réponses impulsionnelles des filtres obtenus par inversion multicanal, en échelle logarithmique.

pulsionnelle qui s'étend sur la totalité des 4096 points utilisés pour l'inversion. Ceci permet une synthèse quasi-parfaite du résultat voulu au niveau des points de contrôle mais produit des

artefacts en dehors (cf. figure 5.6(b)). En réalisant l'inversion sur 400 microphones, les filtres obtenus présentent une concentration d'énergie autour du pic principal. Ce pic principal se situe dans tous les cas et pour tous les filtres au niveau des retards autorisés pour l'inversion.

La solution donnée par l'inversion multicanal n'est valable que jusqu'à une fréquence limite différente pour chaque haut-parleur. Celle-ci dépend de la source virtuelle, de l'espacement entre les haut-parleurs, de l'espacement entre les microphones et de la position du banc de microphone. Au-dessus de cette fréquence limite, la solution obtenue n'a qu'une validité locale, au niveau des microphones de contrôle. Une procédure alternative doit donc être utilisée au-dessus de cette fréquence limite.

5.1.4 Au-dessus de la fréquence d'aliasing

Au-dessus de la fréquence d'aliasing, nous proposons d'utiliser des filtres d'égalisation individuelle, indépendamment de la source virtuelle considérée. Le filtre global sera composé à partir de la solution donnée par l'égalisation multicanal en basses fréquences et par l'égalisation individuelle. Nous proposons enfin d'effectuer un contrôle de type énergétique du champ sonore en hautes-fréquences pour assurer une égalisation moyenne sur les points de contrôle par bandes de fréquence.

Calcul des filtres d'égalisation individuelle

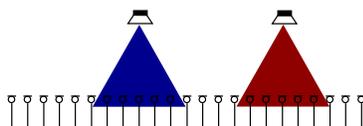


FIG. 5.9 – Sélection des mesures à utiliser pour l'égalisation individuelle des transducteurs

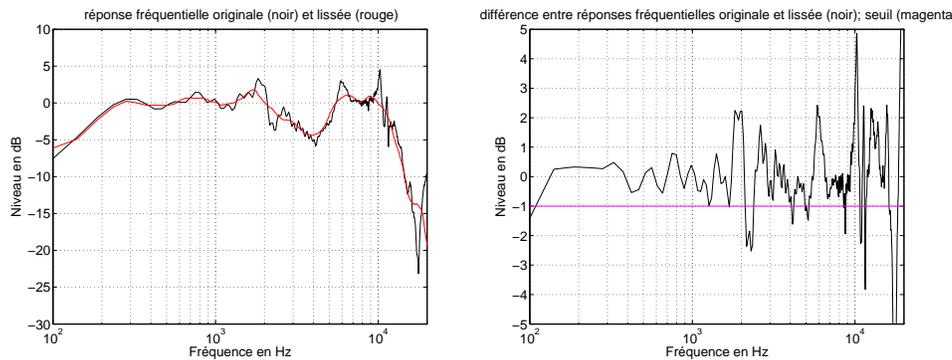
Pour le calcul des filtres d'égalisation individuelle, nous proposons une méthode qui vise à corriger en moyenne la réponse fréquentielle de chacun des haut-parleurs. Aucune correction de phase n'est apportée. L'égalisation de la phase d'un haut-parleur est un processus délicat et qui s'avère relativement inutile au-dessus de la fréquence d'aliasing spatial. En effet, le champ synthétisé par la WFS au-dessus de la fréquence d'aliasing spatial présente des caractéristiques de phase complexes qui dépendent très fortement de la position d'écoute.

La caractérisation du rayonnement des haut-parleurs s'effectue à partir des mesures réalisées pour l'identification du système MIMO (cf. partie 5.1.1). Pour chaque haut-parleur n , on sélectionne un ensemble de L_n mesures situées à l'intérieur d'un angle solide donné (par exemple 40 degrés) autour de l'axe principal du haut-parleur (voir figure 5.9).

Afin de compenser les différences de niveau entre les L_n mesures dues à la propagation, on multiplie celles-ci par la distance de propagation entre le haut-parleur n et le microphone l_n considéré.

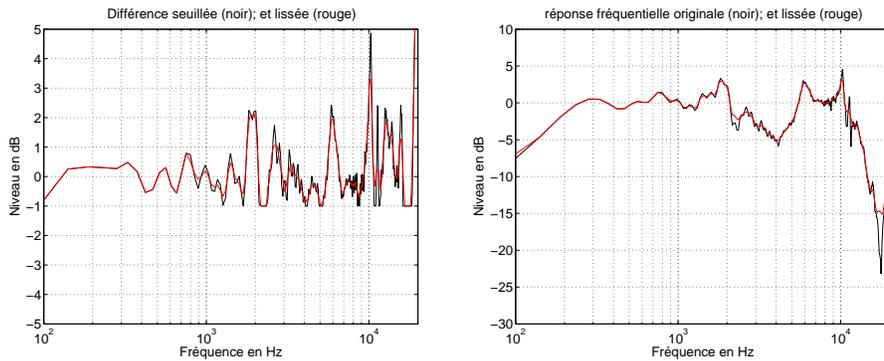
Pour chaque réponse impulsionnelle l_n , on calcule la réponse fréquentielle associée, à laquelle on applique une procédure non linéaire de lissage (cf. figure 5.10) qui consiste à :

1. effectuer un lissage en tiers d'octave de la réponse originale afin d'en extraire les variations globales (cf. figure 5.10(a)) ;
2. calculer la différence entre la réponse originale et la réponse lissée ;
3. sur cette différence, repérer toutes les valeurs situées en-dessous d'un seuil pour lequel nous proposons d'utiliser la valeur de -1 dB. Ces valeurs correspondent aux "zéros" de la réponse fréquentielle. On leur affecte alors la valeur du seuil (cf. figure 5.10(b)) ;
4. lisser la différence seuillée par demi-ton afin de réduire les résonances trop marquées (cf. figure 5.10(c)).



(a) Réponses fréquentielle originale (noir) et réponse fréquentielle lissée par tiers d'octave (rouge)

(b) Différence entre réponses fréquentielles originale et lissée (noir) ; seuil (magenta)



(c) Différence seuillée (noir) ; et différence seuillée lissée (rouge)

(d) Réponse fréquentielle originale (noir) ; et lissée finale(rouge)

FIG. 5.10 – Lissage des réponses impulsionnelles.

5. ajouter la réponse obtenue à la réponse fréquentielle originale lissée (cf. figure 5.10(d)).

La réponse fréquentielle résultante préserve l'allure générale de la réponse fréquentielle originale en réduisant légèrement les résonances trop marquées et en éliminant les zéros de la réponse fréquentielle. On a en effet vu que l'égalisation des zéros de la réponse fréquentielle est délicate et risque d'introduire des résonances dans les filtres d'égalisation qui dégradent le résultat de la procédure.

L'ensemble des N_i réponses lissées et compensées en niveau sont ensuite moyennées et inversées. Le résultat fournit le gabarit préliminaire du filtre d'égalisation.

Au-dessus de la fréquence de coupure haute et en dessous de la fréquence de coupure basse du haut-parleur, celui-ci n'est pas mécaniquement capable d'émettre de l'énergie. On spécifie alors dans ces régions fréquentielle, un niveau qui décroît rapidement dans le gabarit du filtre.

A partir de ce gabarit dans le domaine fréquentiel, on extrait un filtre à réponse impulsionnelle finie (RIF). Aucune correction de la réponse en phase n'étant proposée, on synthétise un filtre à phase linéaire ou à phase minimale.

Synthèse des filtres

L'étape suivante de la procédure consiste à fabriquer le filtre complet en utilisant, en basses fréquences, le filtre obtenu par inversion multicanal et, en hautes fréquences, le filtre obtenu par égalisation individuelle. La difficulté consiste alors à réaliser une transition douce entre les deux parties du filtre. Nous proposons de "corriger" le filtre d'égalisation individuelle pour chaque haut-parleur en estimant proprement, au niveau de la fréquence de coupure, le retard et le niveau du filtre obtenu par égalisation multicanal.

Dans notre cas, l'ensemble des filtres obtenus par inversion multicanal est aligné en temps grâce à l'opération de mise en forme de $C(z)$ (cf. partie 5.1.3), le retard est alors donné par le retard fixé pour l'égalisation du système.

Le niveau est plus difficile à estimer car il dépend du traitement effectué par l'inversion multicanal. Plusieurs solutions sont envisageables :

- utiliser les gains de la Wave Field Synthesis ;
- comparer le niveau des filtres obtenus par inversion multicanal avec les filtres d'égalisation individuelle sur une bande de fréquence réduite autour de la fréquence de coupure ;
- estimer le gain des filtres obtenus autour de la fréquence de coupure pour une configuration similaire de haut-parleurs/microphones idéaux par inversion multicanal dans le domaine fréquentiel.

En pratique, la dernière solution s'avère la plus précise. Toutefois, nous ne prétendons pas qu'il s'agisse de la solution optimale.

Compensation énergétique moyenne

La dernière opération consiste à estimer la réponse du système en convoluant les réponses impulsionnelles $C(z)$ mesurées avec les filtres calculés et en sommant les réponses obtenues au niveau de chaque microphone.

Au-dessus de la fréquence d'aliasing, on calcule, par bandes de fréquence l'énergie de la réponse du système, moyennée sur l'ensemble des microphones. La même procédure est réalisée sur le signal cible et les filtres sont corrigés globalement afin que l'*énergie moyenne* de la réponse du système corresponde à celle de la cible.

5.2 Résultats

Cette partie est consacrée à l'évaluation de la méthode d'égalisation multicanal pour la Wave Field Synthesis présentée dans ce chapitre. Nous nous attachons à évaluer la transparence des sources virtuelles synthétisées en comparaison avec des méthodes plus classiques.

Dans un premier temps, Nous considérons le cas de haut-parleurs omnidirectionnels idéaux. Ce cas d'étude permet de simuler la propagation du champ sonore synthétisé dans l'ensemble de la zone d'écoute et d'en évaluer la qualité. Nous comparons ainsi les résultats obtenus avec ceux

de la WFS "sans égalisation multicanal" présentés dans le chapitre 2. Dans la suite, on parle de "WFS" et de "WFS avec égalisation multicanal" pour faire la distinction entre la formulation de la WFS issue des travaux de l'université de Delft et de France Télécom R & D par simplification des intégrales de Kirchhoff-Helmholtz et Rayleigh (cf. partie 1.2) et la WFS avec la méthode d'égalisation multicanal que nous avons proposée.

Dans un deuxième temps, nous appliquons la procédure d'égalisation multicanal à un banc de haut-parleurs MAP (Multi-Actuator Panels) dont les caractéristiques de rayonnement s'écartent du cas idéal. On considère l'égalisation d'un ensemble de 4 haut-parleurs MAP (système de 32 voies) de 5,3 m de long. Ce système correspond à l'installation permanente réalisée dans le studio 4 de l'IRCAM. On évalue la qualité du champ acoustique synthétisé à l'intérieur de la zone d'écoute à partir de mesures réalisées en champ libre. Nous comparons les résultats obtenus par la "WFS avec égalisation multicanal" avec ceux de la "WFS avec égalisation individuelle", en appliquant le filtre d'égalisation individuelle aux haut-parleurs indépendamment de la source virtuelle synthétisée.

5.2.1 Résultats par simulation de haut-parleurs idéaux

Les résultats que nous présentons sont obtenus en considérant un banc linéaire de haut-parleurs omnidirectionnels idéaux similaire à celui employé dans la partie 2.3.3 (60 haut-parleurs espacés de 10 cm formant un banc linéaire de 6 m de long). Le système MIMO est identifié en simulant les réponses impulsionnelles des haut-parleurs sur un banc linéaire de 100 microphones omnidirectionnels espacés de 8 cm (cf. figure 5.11).

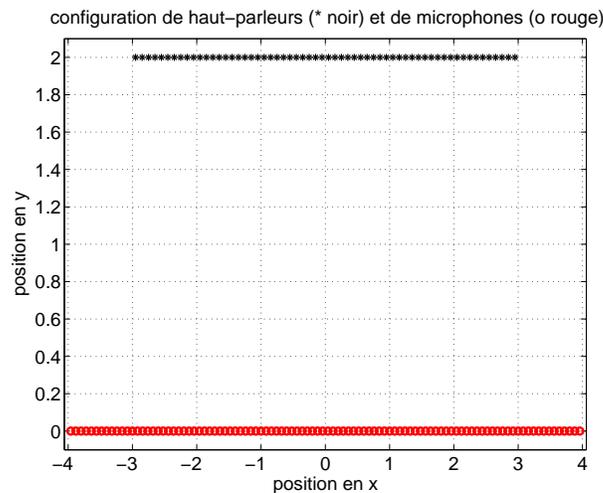


FIG. 5.11 – Configurations de haut-parleurs et de microphones utilisée pour l'identification du système

On reprend certaines sources utilisées pour les simulations avec la WFS (cf. partie 1.4), à savoir :

- source omnidirectionnelle focalisée située à $(1, 1.2)$, soit 80 cm *devant* le banc de haut-parleurs ;

- source omnidirectionnelle située (3, 5), soit un mètre *derrière* le banc de haut-parleur et en face de l'extrémité droite du banc ;
- une onde plane se propageant perpendiculairement au banc (onde plane à 0 degré).
- source d'ordre -1 (dipôle orienté parallèlement au banc de haut-parleurs) à (0, 1).
- source d'ordre 4 à (0, 1).

Dans un premier temps, nous donnons la réponse du système calculée sur une ligne de mesure à même distance du banc de haut-parleurs que les points de contrôle utilisés pour l'égalisation multicanal. Nous présentons ensuite l'allure des filtres synthétisés ce qui. Enfin, nous évaluons la qualité de la transparence de la reproduction dans le sous-espace de reproduction en définissant une fonction de qualité de manière similaire à ce qui a été fait pour la WFS (cf. partie 2.1).

Réponse du système

On estime la réponse du système sur une ligne située à 2 m du banc de haut-parleurs. Comme pour la WFS (cf. partie 5.11), la réponse du système aux filtres calculés par la procédure d'inversion multicanal est simulée, pour les sources virtuelles omnidirectionnelles, au niveau d'un ensemble de 60 microphones omnidirectionnels répartis sur une ligne située à 2 m en face du banc de haut-parleurs. Les microphones sont quant à eux régulièrement répartis tous les 30 cm (cf. figure 5.12(a)). Pour les sources d'ordre supérieur, la réponse est donnée pour un ensemble de 180 microphones espacés de 10 cm et situés à 2 m du banc de haut-parleurs (cf. figure 5.12(b)).

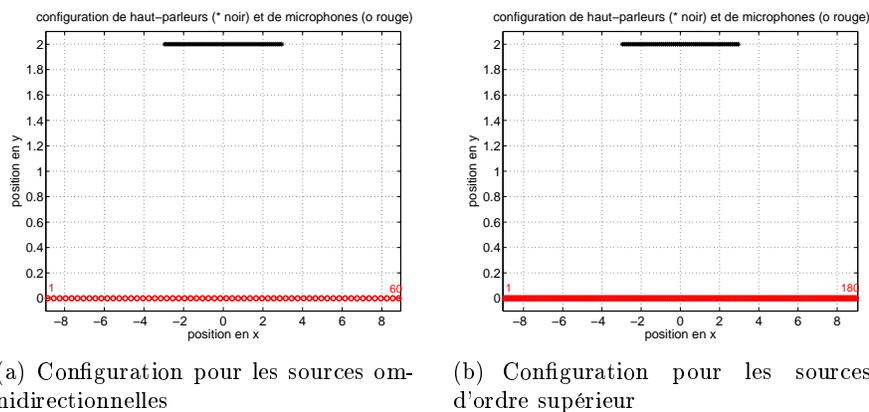


FIG. 5.12 – Configuration de haut-parleurs et de microphones permettant de vérifier la réponse du système

Pour l'onde plane (figure 5.13), le champ est idéalement synthétisé à l'intérieur de la fenêtre de visibilité sous la fréquence d'aliasing, en particulier dans le domaine fréquentiel (figure 5.13(a)). Dans le domaine temporel et sous la fréquence d'aliasing (figure 5.13(b)), on note toutefois la présence de deux fronts d'onde de très faible niveau (environ -40 dB par rapport au niveau maximal en dessous de 2000 Hz) en avance par rapport au front d'onde principal (moins de 5 ms). Ces fronts subissent ainsi le masquage rétro-actif ("pre-masking" en anglais) du front d'onde principal. Étant donné leur niveau et leur temps d'arrivée relativement au front d'onde principal,

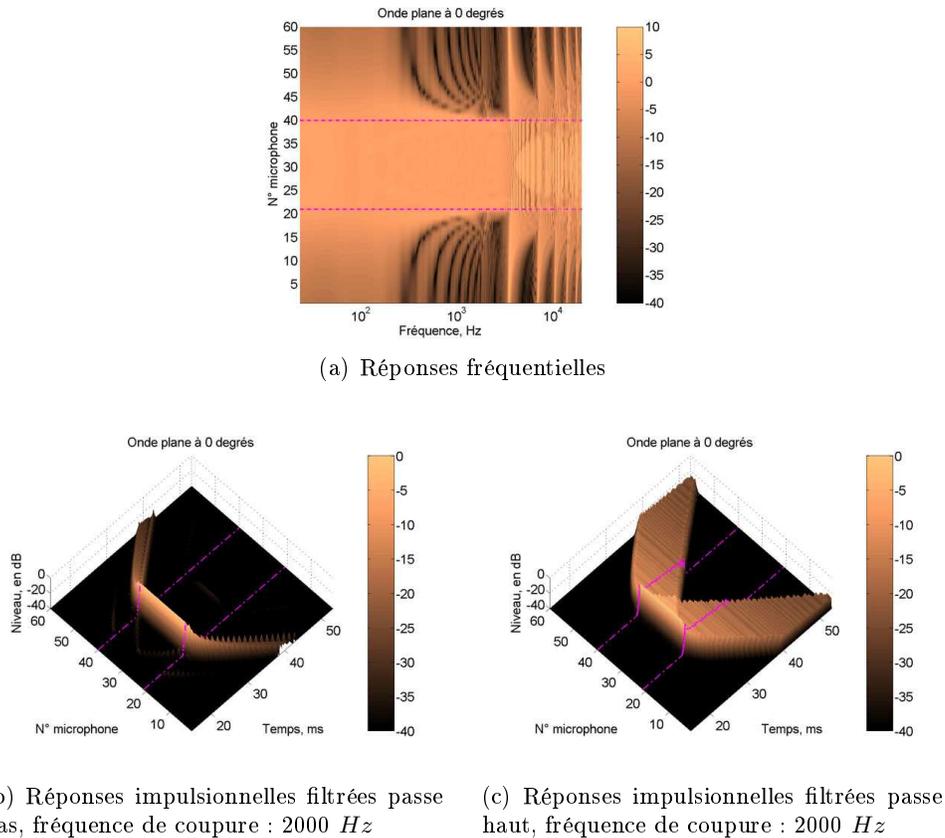


FIG. 5.13 – Champ synthétisé pour une onde plane se propageant perpendiculairement au banc de haut-parleurs.

ces fronts sont probablement inaudibles.

En dehors de la fenêtre de visibilité, dans le domaine fréquentiel (cf. figure 5.13(a)), le niveau du champ diffracté en basses fréquences (20-200 Hz) est plus élevé que pour la WFS (cf. figure 1.9(c)) et présente des oscillations aux fréquences moyennes (200-3000 Hz).

La réponse impulsionnelle au-dessus de la fréquence d'aliasing reste très similaire à celle de la WFS (cf. figure 1.9(c)).

Pour la source omnidirectionnelle à (3, 5) (cf. figures 5.14), les remarques faites pour l'onde plane restent justifiées. Le champ sonore est idéalement reproduit à l'intérieur de la fenêtre de visibilité aussi bien dans le domaine temporel que fréquentiel. On note de plus la présence de contributions de très faible niveau en avance par rapport au front d'onde principal, principalement en dehors de la fenêtre de visibilité (cf. figures 5.14(b)).

Pour la source omnidirectionnelle focalisée à (1, 1.2), la distinction entre les résultats à l'intérieur et à l'extérieur de la fenêtre de visibilité sont encore plus marqués. En effet, aussi bien les réponses impulsionnelles (figure 5.15(b)) que les réponses fréquentielles (figure 5.15(a)) sont idéalement synthétisées en dessous de la fréquence d'aliasing spatial à l'intérieur de la zone de visibilité avec des résultats sensiblement plus nets que pour la WFS (cf. figures 1.11(b) et 1.11(c)). Les résultats sont plutôt dégradés à l'extérieur de la zone de visibilité. Des fronts d'onde sont émis en avance

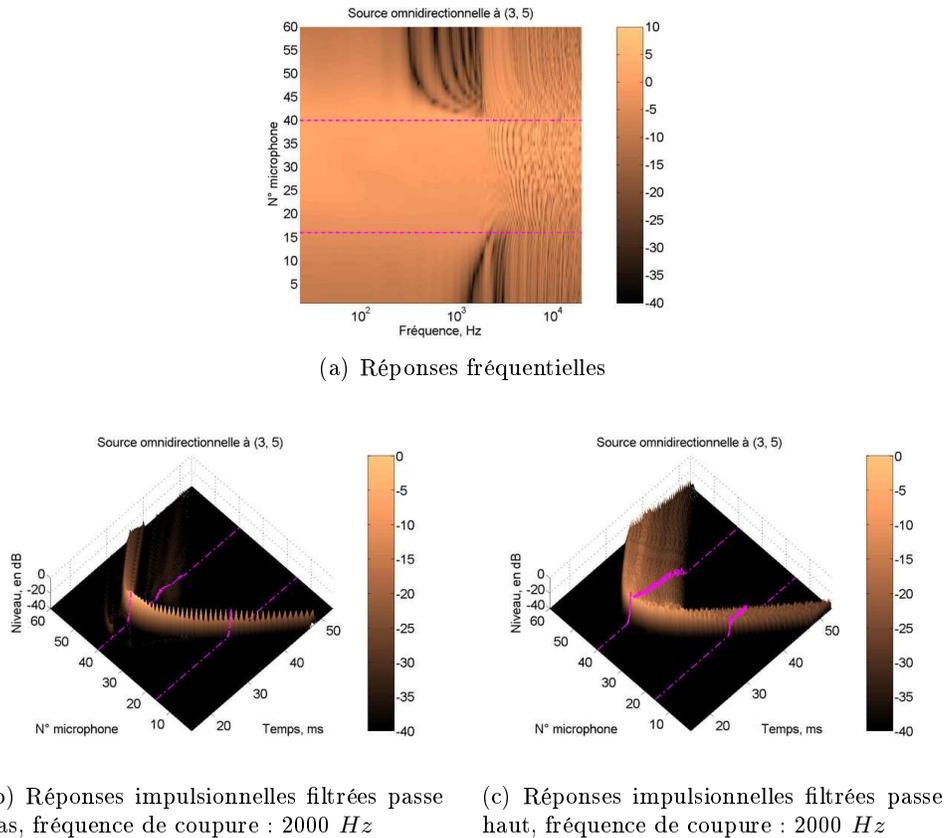


FIG. 5.14 – Champ synthétisé pour une source omnidirectionnelle situé 3 m derrière le banc de haut-parleurs et décentrée de 3 m sur la droite

(≈ 5 ms) par rapport au front d'onde principal avec un niveau, cependant, relativement faible (-20/-25 dB en dessous du niveau du front d'onde principal) (cf. figure 5.15(b)). Dans le domaine fréquentiel (cf. figure 5.15(a)), en dehors de la fenêtre de visibilité, on remarque des oscillations en dessous de la fréquence d'aliasing, en particulier pour les microphones 48 à 60.

Au-dessus de la fréquence d'aliasing, dans le domaine temporel (cf. figure 5.15(c)), on note une diminution de l'étalement temporel des contributions. L'énergie émise est condensée autour du front d'onde principal. Ceci permet, éventuellement, d'éviter des erreurs de localisation en hautes fréquences dues aux contributions précoces émises en avance par rapport au front d'onde principal pour les sources focalisées. Ces contributions proviennent du retournement temporel utilisé lors de la synthèse des sources focalisées en WFS (cf. partie 1.2.1).

Les résultats pour les sources de degré supérieur sont regroupés dans l'ensemble des figures 5.16. Comme pour la source omnidirectionnelle focalisée, les fronts d'onde sont idéalement reproduits à l'intérieur de la zone de visibilité aussi bien dans le domaine temporel (figures 5.16(b) et 5.16(d)) que fréquentiel (figures 5.16(a) et 5.16(c)). On remarque, en particulier, que les lobes sont fidèlement synthétisés dans le domaine temporel. Contrairement à la WFS (cf. figures 1.12(d) et 1.13(d)), il n'y a pratiquement pas de composante en opposition de phase. On note cependant une fusion des lobes en basses fréquences (en dessous de 100 Hz) pour la source virtuelle d'ordre

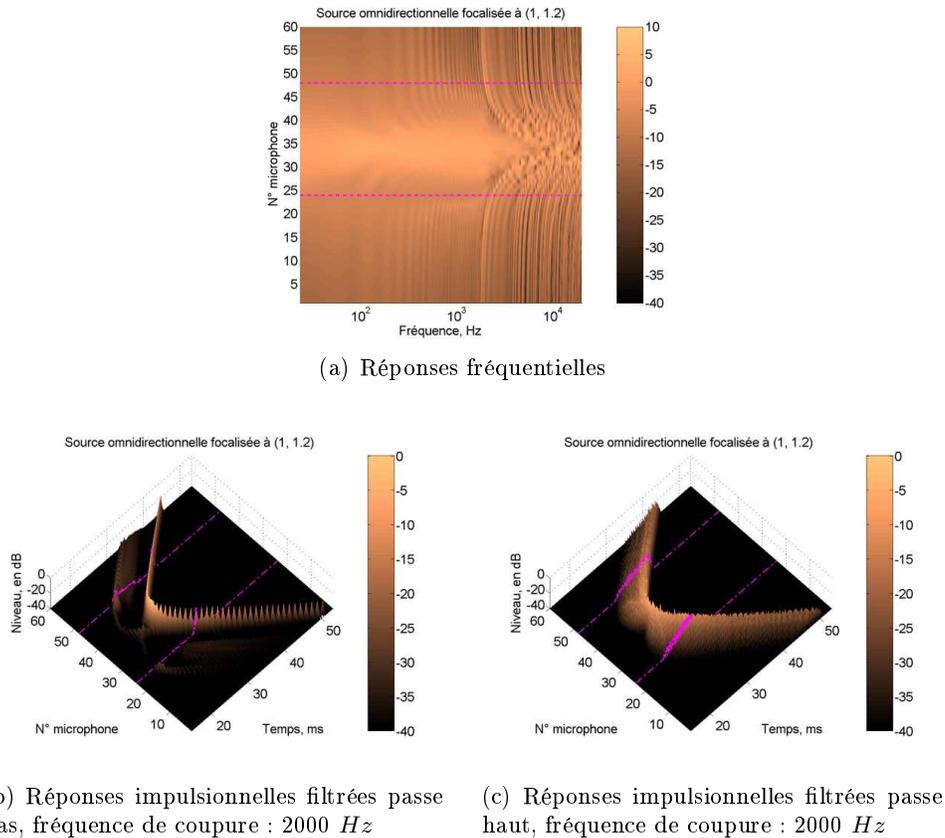
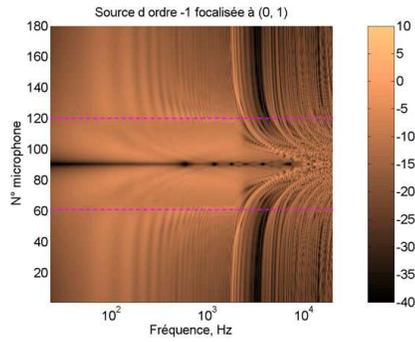


FIG. 5.15 – Champ synthétisé pour une source omnidirectionnelle focalisée situé 1 m devant le banc de haut-parleurs et décentrée de 1 m sur la droite

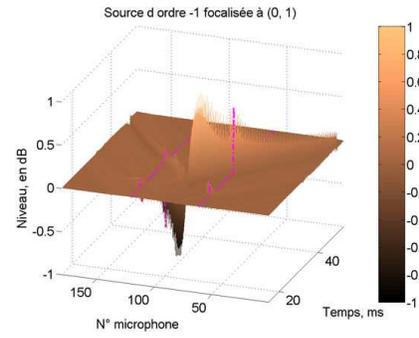
4. Le même phénomène existe pour la WFS (cf. figure 1.13(e)) à une fréquence, toutefois, plus élevée (400 Hz). Ce phénomène apparaîtrait ainsi comme une *impossibilité physique* pour un tel banc de haut-parleurs à reproduire une source virtuelle d'ordre 4 à cette position en dessous de 100 Hz.

En dehors de la fenêtre de visibilité, le champ synthétisé est, là aussi, relativement perturbé avec des oscillations faibles dans le domaine fréquentiel.

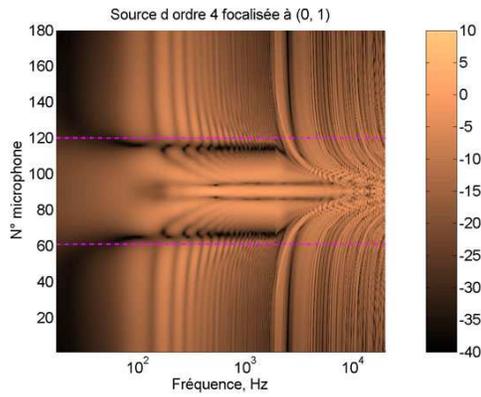
De manière générale, le champ synthétisé par la WFS avec égalisation multicanal est similaire à celui de la WFS. Les résultats sont toutefois plus propres, pour la WFS avec égalisation multicanal, à l'intérieur de la fenêtre de visibilité. Le champ diffracté en dehors de la zone de visibilité est plus important, notamment en basses fréquences. On remarque l'émission de fronts d'onde supplémentaires de faible niveau à l'intérieur de cette zone. Ceci semblent provenir de la correction spécifique de la diffraction par l'égalisation multicanal.



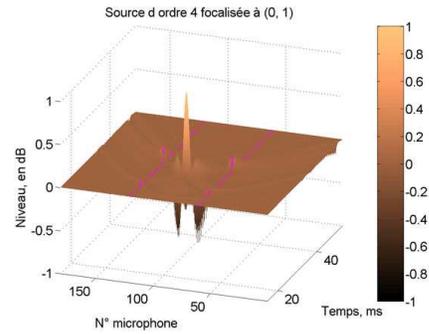
(a) Source d'ordre -1 située à $(0, 1)$, réponses fréquentielles



(b) Source d'ordre -1 située à $(0, 1)$, réponses impulsionnelles filtrées passe bas, fréquence de coupure : 2000 Hz



(c) Source d'ordre 4 située à $(0, 1)$, réponses fréquentielles



(d) Source d'ordre 4 située à $(0, 1)$, réponses impulsionnelles filtrées passe bas, fréquence de coupure : 2000 Hz

FIG. 5.16 – Champ synthétisé pour un ensemble de sources virtuelles de directivité dipolaire orientées parallèlement au banc de haut-parleurs

Allure des filtres

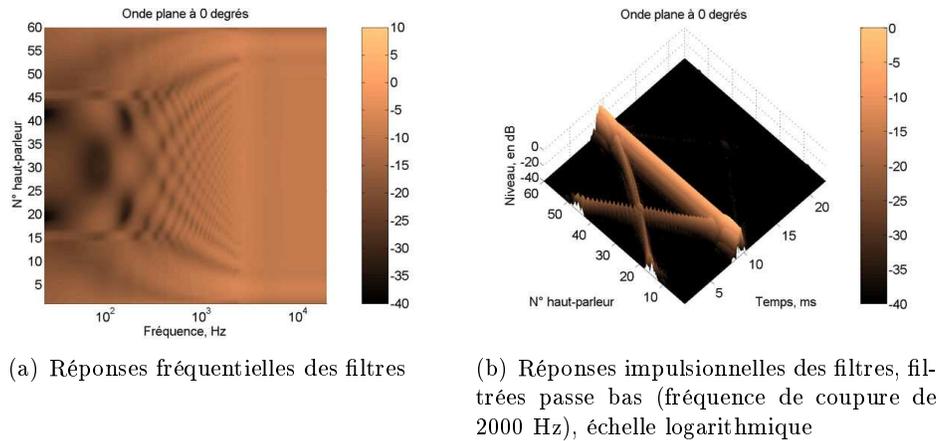


FIG. 5.17 – Filtres obtenus par la procédure d'égalisation multicanal pour une onde plane se propageant perpendiculairement au banc de haut-parleurs

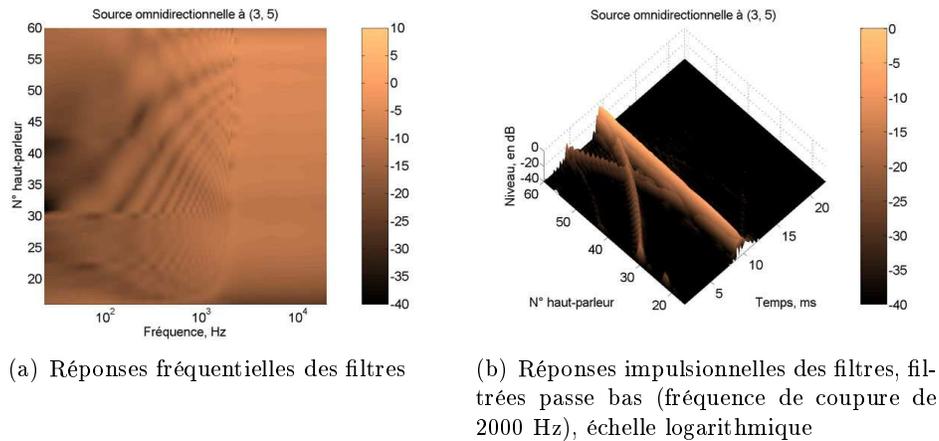


FIG. 5.18 – Filtres obtenus par la procédure d'égalisation multicanal pour une onde plane se propageant perpendiculairement au banc de haut-parleurs

Pour chaque source, on représente les réponses fréquentielles ainsi que les réponses impulsionnelles sous la fréquence d'aliasing (fréquence de coupure du filtre passe bas de 2000 Hz) pour les filtres calculés par l'égalisation multicanal.

De manière générale, les réponses fréquentielles en dessous de la fréquence d'aliasing sont plus complexes que celles de la Wave Field Synthesis (filtre en $\sqrt{\pm jk}$ soit +3 dB par octave, gain fonction de la position de la source virtuelle, de sa directivité et de la position du haut-parleur considéré).

Dans le domaine temporel, on remarque une composante principale dont le retard est indépendant de la source virtuelle et du haut-parleur considéré. Cette composante permet, comme pour la WFS, la synthèse du front d'onde principal associé à la source virtuelle considérée. On note deux composantes supplémentaires, émises en avance par rapport à la composante principale.

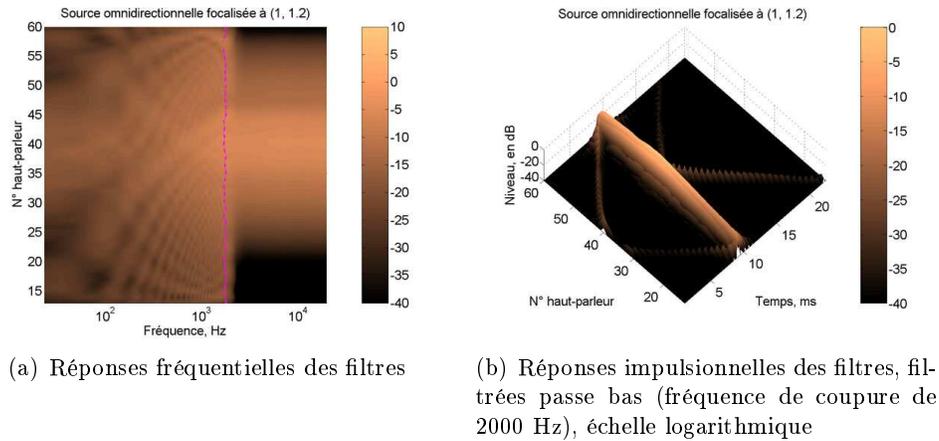


FIG. 5.19 – Filtres obtenus par la procédure d'égalisation multicanal pour une onde plane se propageant perpendiculairement au banc de haut-parleurs

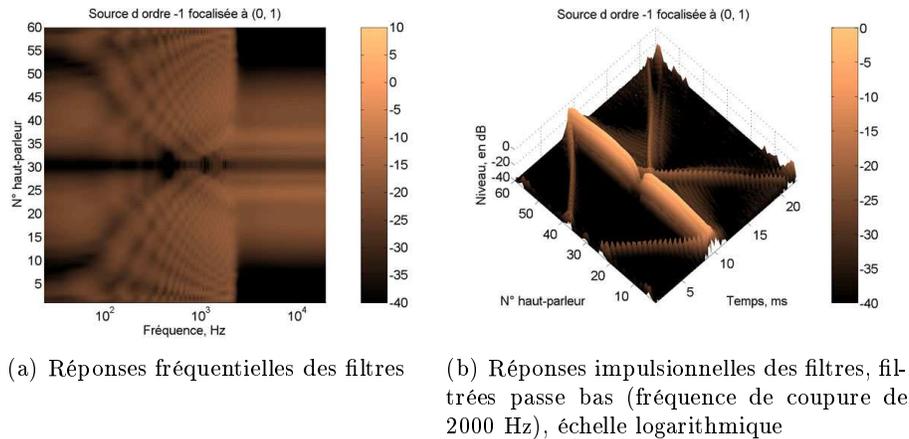
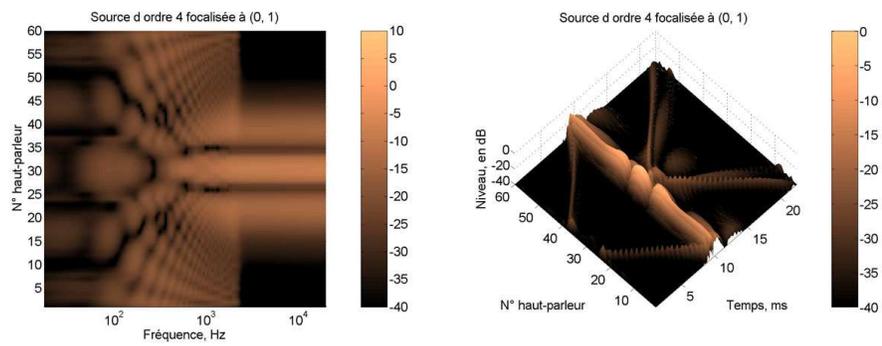


FIG. 5.20 – Filtres obtenus par la procédure d'égalisation multicanal pour une onde plane se propageant perpendiculairement au banc de haut-parleurs

La distribution des retards semble correspondre à des fronts d'onde issus de sources situées aux extrémités du banc de haut-parleurs. Cette observation doit être rapprochée de l'estimation du champ de diffraction à l'intérieur de la fenêtre de visibilité effectuée par Nicol (Nicol, 1999) et Start (Start, 1997). Ils montrent que celui-ci peut être approché par les contributions de deux sources situées aux extrémités du banc et ayant des caractéristiques de directivité très particulières. Les contributions supplémentaires que l'on observe dans les filtres calculés par l'égalisation multicanal peuvent ainsi être interprétés comme une synthèse, en opposition de phase, de celles des "sources de diffraction". Les fronts d'onde synthétisés semblent dépendre de la source virtuelle considérée.

Pour les sources focalisées, on note la synthèse de fronts d'onde situés derrière le front d'onde principal. Ces fronts d'onde semblent symétriques par rapport aux contributions émises devant le front d'onde principal. Il pourrait s'agir d'une contribution supplémentaire à la correction de



(a) Réponses fréquentielles des filtres

(b) Réponses impulsionnelles des filtres, filtrés passe bas (fréquence de coupure de 2000 Hz), échelle logarithmique

FIG. 5.21 – Filtres obtenus par la procédure d'égalisation multicanal pour une onde plane se propageant perpendiculairement au banc de haut-parleurs

la diffraction.

Qualité du champ synthétisé

Comme pour la Wave Field Synthesis (cf. partie 2.3.3), nous souhaitons vérifier la bonne propagation du champ acoustique "égalisé" au delà des points de contrôle. On représente alors l'évolution de la fonction de qualité $Q_{\Psi}^{Meq}(r(R), k)$ du champ sonore synthétisé en fonction de la position de l'auditeur dans le sous-espace de reproduction, de la fréquence et de la source virtuelle synthétisée. La fonction de qualité se définit en analogie avec l'équation 2.40 :

$$Q_{\Psi}^{Meq}(r(R), k) = \frac{H_{\Psi}^{Meq}(r(R), k)}{H_{\Psi}^{id}(r(R), k)}, \quad (5.5)$$

où, $H_{\Psi}^{Meq}(r(R), k)$ est la transformée de Fourier de la réponse impulsionnelle du système pour la synthèse de la source Ψ par l'égalisation multicanal. $H_{\Psi}^{id}(r(R), k)$ est la réponse "idéale" du système donnée par l'équation 2.39. Cette fonction de qualité traduit ainsi les variations de la réponse du système autour de la réponse idéale.

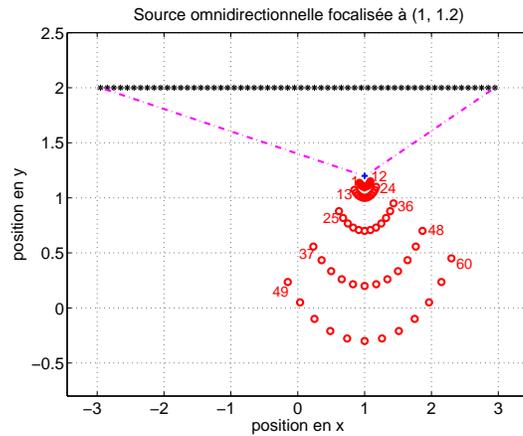
On simule la réponse du système $H_{\Psi}^{Meq}(r(R), k)$ sur un ensemble de positions de microphones pour deux configurations. Dans la première configuration les microphones sont situés à proximité du banc de haut-parleurs (cf. figure 5.22(b)). Dans la deuxième configuration, ils en sont éloignés (cf. figure 5.22(c)). Pour la source virtuelle focalisée en (1, 1.2), les microphones de proximité sont situés sur des cercles concentriques centrés sur la position de la source virtuelle (cf. figure 5.22(a)). Seuls les microphones situés dans la zone de visibilité de la source virtuelle sont utilisés. Remarquons que les configurations considérées sont ainsi les mêmes que pour la WFS (cf. figure 2.12).

L'égalisation multicanal est réalisée sur un banc de microphones situé à 2 m du banc de haut-parleurs. La distinction entre configuration à proximité et configuration de microphones lointains correspond donc au champ synthétisé, respectivement, en amont et en aval des points de contrôle.

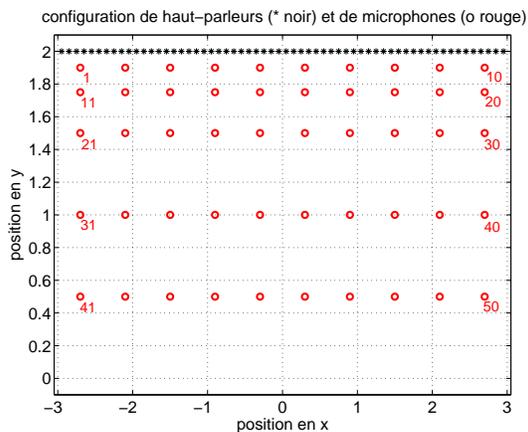
Les figures 5.23 représentent le module en dB de $Q_{\Psi}^{Meq}(r(R), k)$ pour les trois sources virtuelles et les deux types de configurations de microphones considérés. Le module de la fonction de qualité n'a, de manière générale, que très peu de variation sur la configuration des microphones lointains, et ceci indépendamment de la source virtuelle considérée (cf. figures 5.23(b), 5.23(d), et 5.23(f)). Celui-ci reste toujours à l'intérieur de ± 2 dB et pratiquement partout dans ± 1 dB.

Cette méthode réalise ainsi une correction des défauts de la WFS lors de la propagation du champ sonore au delà de la position des points de contrôle. Ceux-ci sont principalement dus à la diffraction (cf. partie 2.3.3).

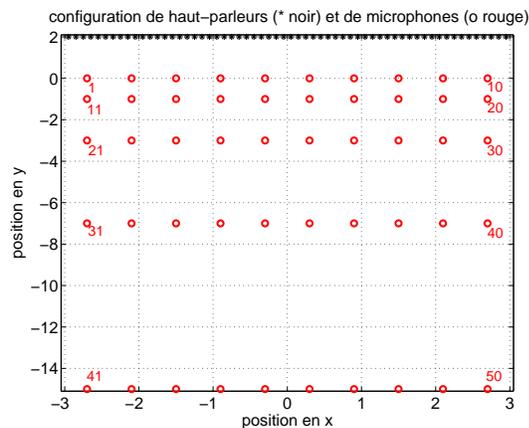
Pour la source focalisée, le module de $Q_{\Psi}^{Meq}(r(R), k)$ pour les microphones de proximité (cf. figure 5.23(a)) est similaire à $Q_{\Psi}^{Wfs}(r(R), k)$ pour la WFS (cf. figure 2.13(b)). Les résultats sont, au contraire, dégradés pour les configurations de microphones à proximité pour les autres sources virtuelles (cf. figures 5.23(d), et 5.23(f)). Ceux-ci restent toutefois acceptables pour les



(a) configuration de proximité de la source focalisée



(b) configuration microphones proches



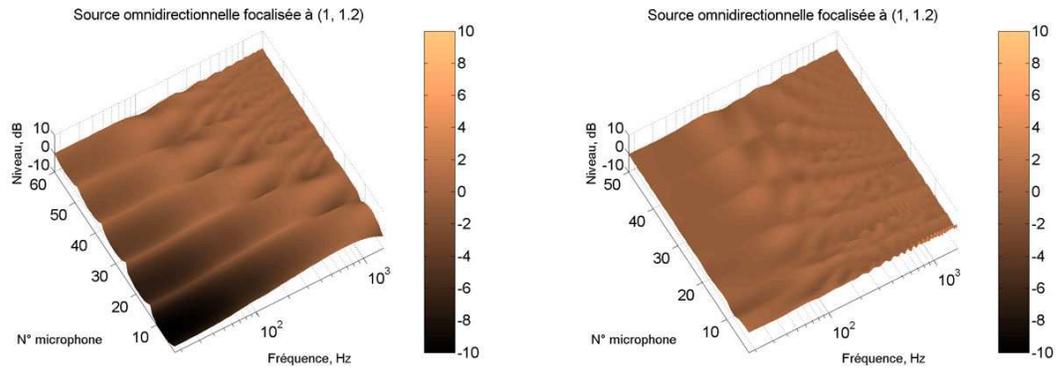
(c) configuration microphones lointains

FIG. 5.22 – Configurations de test avec 60 haut-parleurs (60 m, 10 cm d'espacement) ; différentes positions de microphones (o rouges), proximité de la source focalisée, proches, lointains

microphones 31 à 50 situés à 1 m et à 1,5 m du banc de haut-parleurs. Ils sont fortement dégradés pour les microphones 1 à 30 (10, 25, 50 cm du banc de haut-parleurs), positions auxquelles les auditeurs ne seront, toutefois, que rarement situés.

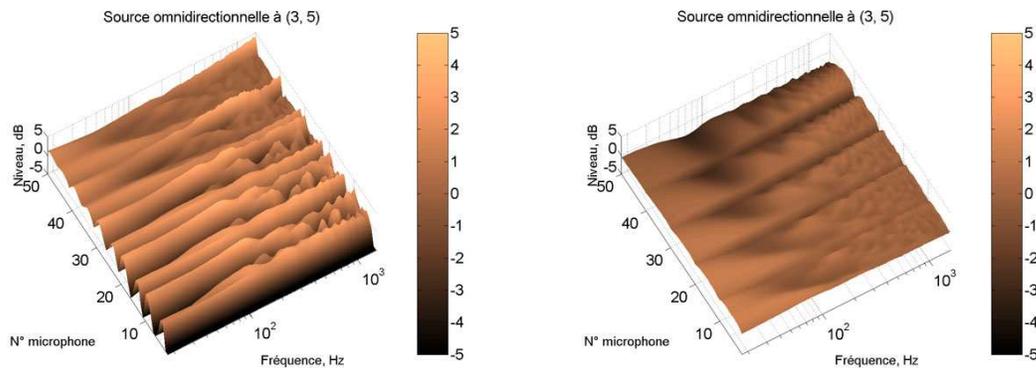
L'ensemble des figures 5.24 présente le temps de groupe associé à $Q_{\Psi}^{Meq}(r(R), k)$ pour les trois sources virtuelles et les deux types de configuration de microphones. Les résultats sur le temps de groupe confirment les remarques données pour le domaine fréquentiel. Les variations de temps de groupe observées autour de 0 sont extrêmement faibles pour la configuration de microphones lointains et sont plus importantes mais peu significatives à proximité du banc de haut-parleurs.

La procédure d'égalisation multicanal permet ainsi d'assurer la transparence de la reproduction du champ sonore en aval de la ligne sur laquelle sont situés les microphones de contrôle. Pour la



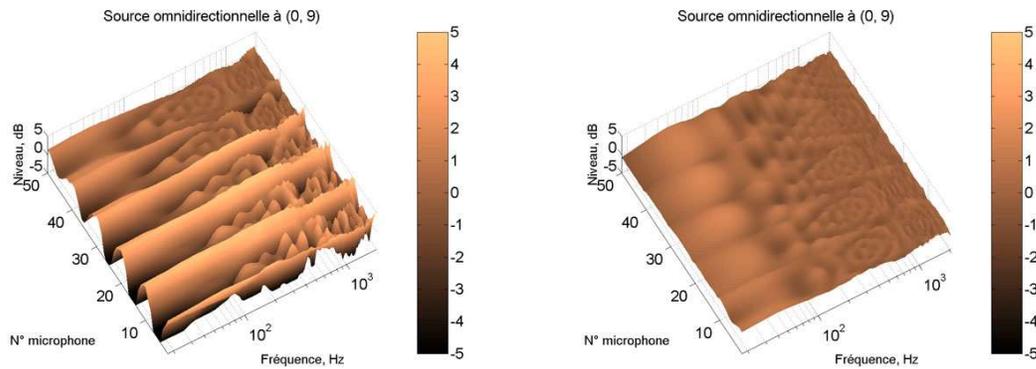
(a) Source omnidirectionnelle focalisée à $(1, 1.2)$, échelle de niveau entre -10 et $+10$ dB, configuration de *proximité* de la source focalisée

(b) Source omnidirectionnelle focalisée à $(1, 1.2)$, échelle de niveau entre -10 et $+10$ dB, microphones *lointains*



(c) Source omnidirectionnelle à $(3, 5)$, microphones *proches*

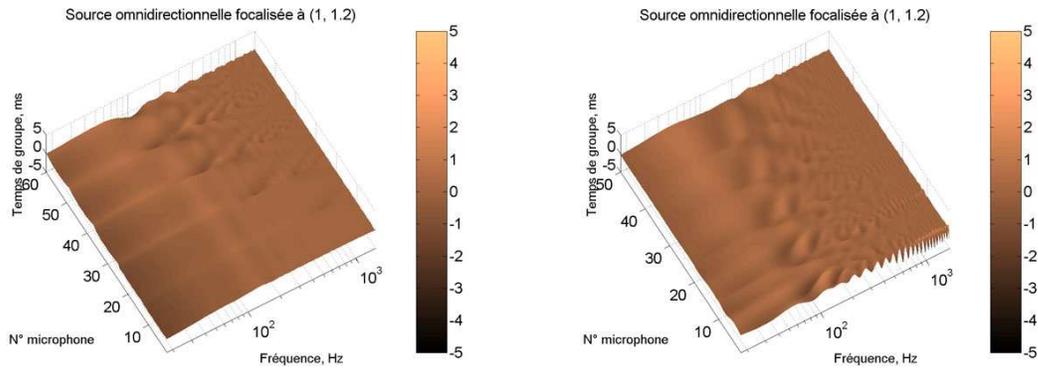
(d) Source omnidirectionnelle à $(3, 5)$, microphones *lointains*



(e) Source omnidirectionnelle à $(0, 9)$, microphones *proches*

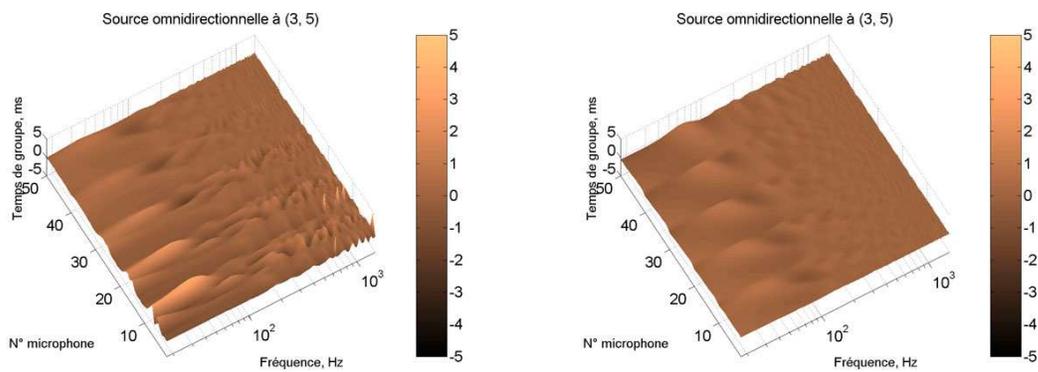
(f) Source omnidirectionnelle à $(0, 9)$, microphones *lointains*

FIG. 5.23 – Module de $Q_{\Psi}^{eq}(r(R), k)$ en dB pour un banc de 60 haut-parleurs (6 m, 10 cm d'espacement) sous la fréquence d'aliasing spatial (20-2000 Hz) synthétisant différentes sources virtuelles omnidirectionnelles avec égalisation multicanal



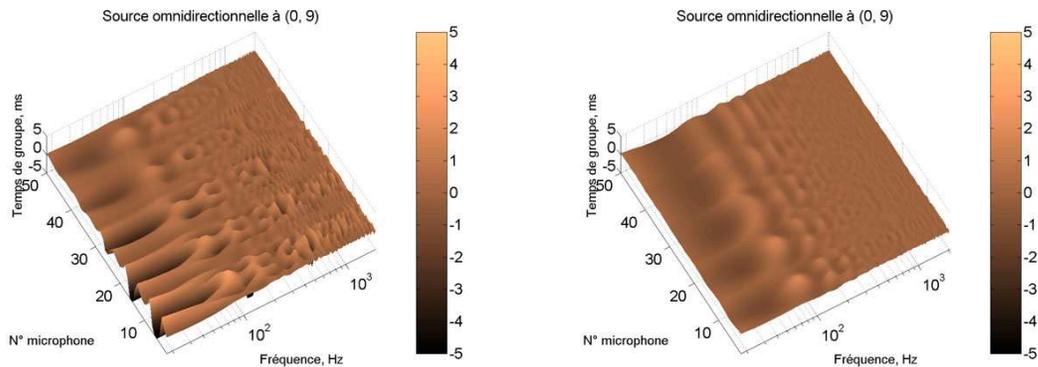
(a) Source omnidirectionnelle focalisée à (1, 1.2), échelle de niveau entre -10 et +10 dB, configuration de *proximité* de la source focalisée

(b) Source omnidirectionnelle focalisée à (1, 1.2), échelle de niveau entre -10 et +10 dB, microphones *lointains*



(c) Source omnidirectionnelle à (3, 5), microphones *proches*

(d) Source omnidirectionnelle à (3, 5), microphones *lointains*



(e) Source omnidirectionnelle à (0, 9), microphones *proches*

(f) Source omnidirectionnelle à (0, 9), microphones *lointains*

FIG. 5.24 – Temps de groupe associé à $Q_{\Psi}^{eq}(r(R), k)$ pour un banc de 60 haut-parleurs (6 m, 10 cm d'espacement) sous la fréquence d'aliasing spatial (20-2000 Hz) synthétisant différentes sources virtuelles omnidirectionnelles avec égalisation multicanal

configuration choisie, le champ synthétisé en amont reste correct environ 1 m devant cette ligne et se dégrade à proximité immédiate du banc de haut-parleurs.

5.2.2 Résultats avec des haut-parleurs MAP

Cette partie est dédiée plus spécifiquement à l'égalisation de la configuration de 4 haut-parleurs MAP installée de façon permanente dans le studio 4 de l'IRCAM. Après avoir présenté les caractéristiques des haut-parleurs MAP et l'installation elle-même, nous donnons les résultats de la WFS avec égalisation multicanal. Ceux-ci sont comparés à une technique plus simple qui utilise la WFS avec une égalisation individuelle des transducteurs indépendamment de la source virtuelle synthétisée.

Haut-parleurs MAP

Les haut-parleurs MAP sont dérivés de la technologie DML (Distributed Mode Loudspeaker) (Harris & Hawksford, 1997). La membrane de ce type de haut-parleur est une plaque constituée d'un matériau à la fois léger et rigide, mise en mouvement par un ensemble d'"excitateurs" (dispositifs de type électrodynamique dont la bobine est collée à la plaque).

A la différence des haut-parleurs électrodynamiques dont le fonctionnement est de type piston, les DML utilisent les modes de vibration d'ordre supérieur de la membrane. Pour obtenir un rayonnement efficace de nature "omnidirectionnelle" sur une large bande de fréquence (idéalement l'ensemble du spectre audible), la nature du matériau est critique. Ceci est possible à condition d'utiliser un matériau de type "sandwich" ayant des caractéristiques bien particulières (Angus, 2000).

On ne détaille pas le fonctionnement vibro-acoustique de ce type de haut-parleur qui est complexe et reste mal compris. Les matériaux utilisés sont, en effet, très inhomogènes et les conditions aux limites difficiles à modéliser. Elles ne sont ni fixes, ni libres.

Les MAPs utilisés à l'IRCAM correspondent à la deuxième génération de prototypes développés à Studer par Ulrich Horbach. La membrane est constituée d'un matériau de type "carton plume" de 135×70 cm. On colle huit excitateurs espacés régulièrement sur chaque plaque. Les plaques sont maintenues sur les côtés par un système de type pince/étai et sont bafflées. L'intérieur du baffle est recouvert de laine de roche comme pour des haut-parleurs électrodynamiques. Les travaux de Prokofieva montrent en effet que la présence d'un matériau poreux à l'arrière de la membrane permet de lisser sensiblement la réponse fréquentielle aux basses fréquences et d'améliorer la transduction électroacoustique (Prokofieva, 2003).

Chacun des excitateurs reçoit un signal indépendant ce qui permet de créer un système multi-voies avec une unique surface vibrante. On a pu montrer que le niveau de distorsion d'intermodulation est suffisamment faible ($\ll 1\%$) pour considérer le système comme linéaire et les excitateurs collés sur une même surface, indépendants entre eux (Corteel *et al.*, 2002).

L'avantage de ce type de haut-parleurs est de n'avoir qu'une très faible signature visuelle. Ils peuvent ainsi être intégrés dans un environnement existant de manière relativement aisée sans révéler la présence d'une ou plusieurs centaines de haut-parleurs, nécessaires pour la Wave Field Synthesis. D'autre part, la vibration de la surface est suffisamment faible pour ne pas gêner la



FIG. 5.25 – Dispositif de reproduction WFS réalisé au Forum Neues MusikTeater à Stuttgart en Mai 2004 dans le cadre d'un workshop sur l'utilisation de la WFS pour des applications musicales. La ligne de 6 MAPs situées dans le fond sont les MAPs appartenant à l'IRCAM. Les 8 autres MAPs sont les produits commercialisés par sonic emotion.

projection d'images. Ils peuvent ainsi être utilisés en tant qu'écrans de projection.

Toutefois, le comportement acoustique de ces haut-parleurs s'éloigne sensiblement de celui de sources ponctuelles omnidirectionnelles requises par la théorie de la WFS.

Configuration de tests et mesures

La configuration de tests utilisée dans cette partie correspond à l'installation réalisée dans le studio 4 de l'IRCAM. Cette installation est constituée de 4 haut-parleurs MAP alignés (32 voies ; 5,3 m de longueur). Le studio 4 est une pièce de 4 m de large et de 7,8 m de long. L'installation couvre une partie du mur droit, le fond de la salle étant occupé par d'autres systèmes de reproduction sonore. La zone d'écoute privilégiée choisie a une forme de trapèze, afin d'assurer la visibilité, à travers le banc de haut-parleurs, des sources situées sur les côtés. Elle démarre 1 m devant les haut-parleurs.

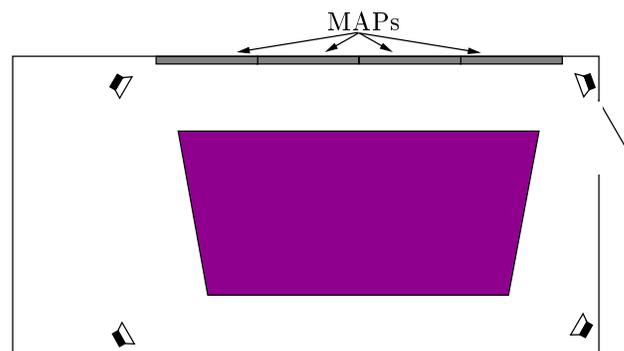


FIG. 5.26 – Installation réalisée dans le studio 4 de l'IRCAM.

La caractérisation du système est effectuée par des mesures de réponses impulsionnelle réalisées dans l'Espace de Projection (salle de concert à acoustique variable de l'IRCAM, en configuration

absorbante pour les mesures, $TR_{60} \approx 1 s$). Les panneaux sont suspendus par des élingues aux portiques mobiles et sont montés à trois mètres du sol. Ils sont éloignés des murs de côté afin de retarder l'arrivée de la première réflexion et d'extraire proprement le son direct (cf. figure 5.27).

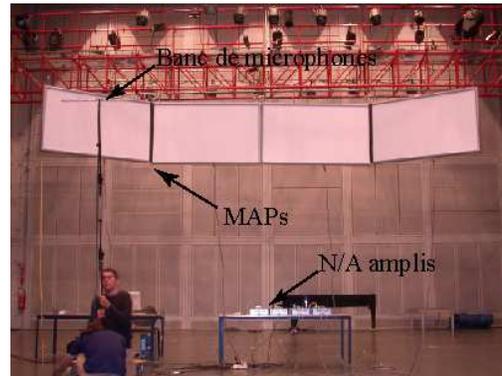


FIG. 5.27 – Mesures des MAPs dans l'espace de projection.

Les mesures sont effectuées à l'aide 3 bancs de microphones constitués, chacun, de 8 capsules omnidirectionnelles espacées de 10 cm. Les capsules microphoniques choisies sont les Monacor MCE 2000 dont la réponse fréquentielle peut être considérée comme plate dans la bande de fréquence qui nous intéresse (100-16000 Hz). Celles-ci sont montées dans un support en caoutchouc qui est attaché en avant d'une tige en aluminium recouverte de mousse de mélamine afin d'absorber la réflexion arrière (cf. figure 5.27). Ces cellules ont été calibrées en champ diffus.

La mesure simultanée de 24 réponses impulsionnelles est effectuée dans Max/MSP à partir de séquences de Golay. La chaîne de mesure complète est représentée dans la figure 5.28. Les séquences sont jouées successivement dans chaque haut-parleur en choisissant le canal de sortie de la carte son et en contrôlant un routeur ADAT par MIDI. La réponse au signal test est enregistrée sur les 24 microphones puis stockée. L'opération de déconvolution, qui permet de récupérer les réponses impulsionnelles à partir des signaux enregistrés, est effectuée en temps différé dans Matlab.

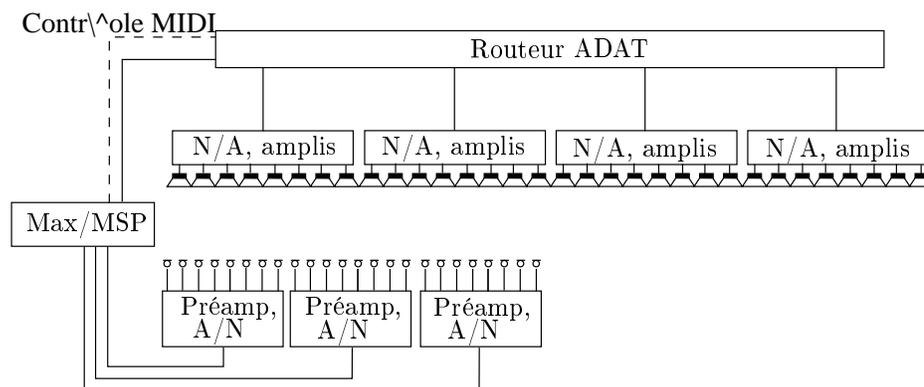


FIG. 5.28 – Dispositif de mesure des haut-parleurs MAP

Les bancs de microphones sont montés à 3 m de hauteur par des pieds perches. On réalise des

mesures pour deux configurations différentes :

- une configuration constituée de 96 positions de microphones espacées de 10 cm situés sur une ligne à 1,5m de distance des haut-parleurs (cf. figure 5.29(a)) ;
- une configuration constituée de trois lignes de 48 microphones situées respectivement à 1,5m ; 2m ; 2,5m des haut-parleurs (cf. figure 5.29(b)).

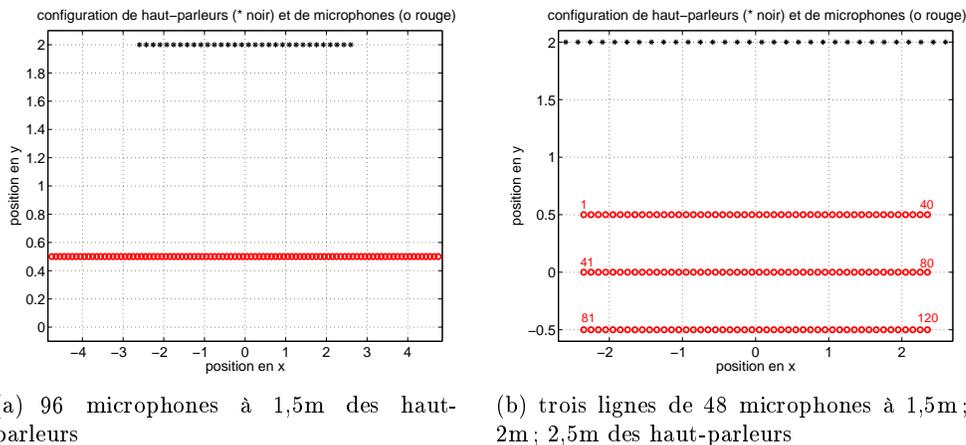


FIG. 5.29 – Configurations de mesure des haut-parleurs MAP

La première configuration est destinée à caractériser le système MIMO pour l'inversion multicanal. La deuxième configuration est utilisée, quant à elle, pour donner la fonction de qualité dans la zone d'écoute.

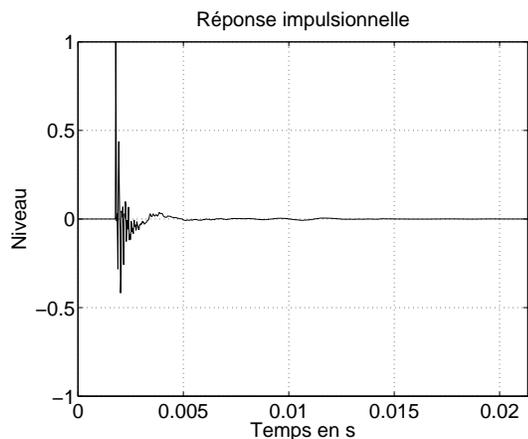
Caractéristiques de rayonnement des MAPs

La figure 5.30(a) représente une réponse impulsionnelle typique d'un canal d'un MAP mesuré dans l'axe, en l'occurrence, l'excitateur numéro 16 mesuré sur le microphone 48 de la configuration pour l'égalisation multicanal (cf. figure 5.29(a)). La contribution énergétique principale est située sur le premier pic. Cette contribution provient de la position de l'excitateur sur la plaque. C'est donc à partir de ce premier pic que peuvent se construire les fronts d'onde pour la Wave Field Synthesis (Boone & de Bruijn, 2000). Ce premier pic est suivi de près par des contributions de faible niveau et des oscillations lentes.

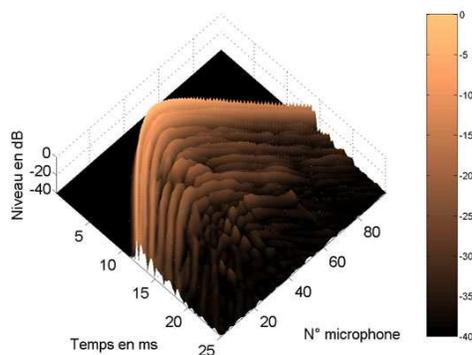
La figure 5.30(b) représente les 96 réponses impulsionnelles, en échelle logarithmique (dB), de ce même excitateur, mesurées sur la configuration de microphones de la figure 5.29(a). On applique un filtrage passe-bas (fréquence de coupure : 2000 Hz) aux réponses impulsionnelles. On remarque que le premier pic est suivi d'un régime d'oscillations dont les caractéristiques dépendent de la position de mesure. Ceci met en évidence les caractéristiques de directivité complexes de ce type de haut-parleur.

L'ensemble des figures 5.31 donne les réponses fréquentielles mesurées sur les 96 positions de microphones de la configuration de la figure 5.29(a) pour les excitateurs 15 et 16.

Les réponses fréquentielles des deux excitateurs ont un comportement globalement complexe



(a) réponse impulsionnelle de l'excitateur numéro 16 mesuré dans l'axe, microphone numéro 48

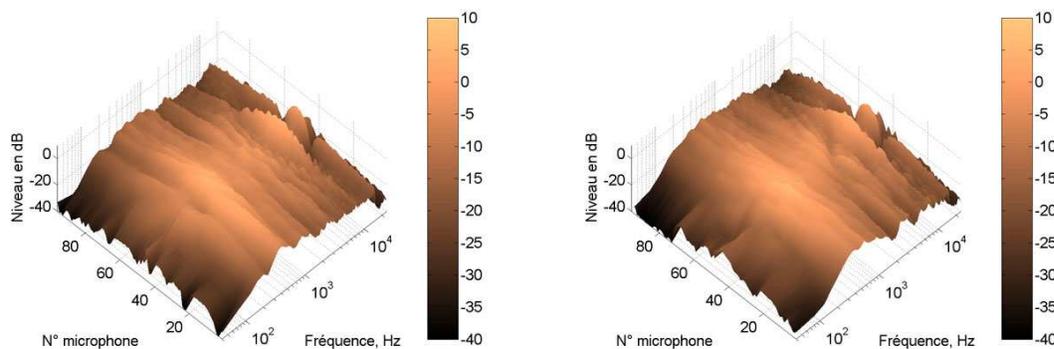


(b) réponses impulsionnelles de l'excitateur numéro 16 filtrées passe-bas, fréquence de coupure : 2000 Hz)

FIG. 5.30 – réponses impulsionnelles d'un excitateur à partir des mesures sur 96 microphones (cf. figure 5.29(a))

mais de même type. Ces haut-parleurs permettent de synthétiser de l'énergie acoustique dans la bande de fréquence 150-16000 Hz. A l'intérieur de cette bande de fréquence, il est intéressant de remarquer que le comportement des réponses fréquentielles est agité mais ne présente aucun zéro marqué qui serait susceptible de perturber l'inversion multicanal.

En comparant les deux réponses, on se rend compte qu'elles présentent des différences assez sensibles, particulièrement dans la bande de fréquences 200-400 Hz.



(a) réponses fréquentielles de l'excitateur numéro 15

(b) réponses fréquentielles de l'excitateur numéro 16

FIG. 5.31 – réponses fréquentielles de deux excitateurs pour les 96 microphones (cf. figure 5.29(a))

Réponse du système

Pour cette configuration, on calcule des filtres pour la synthèse des sources virtuelles suivantes par la procédure d'égalisation multicanal :

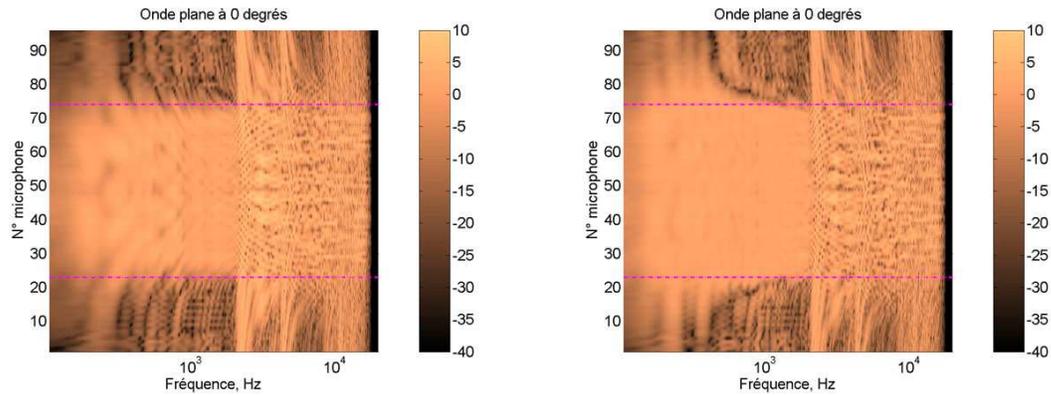
- source omnidirectionnelle focalisée située à $(0, 1.2)$, soit 80 cm *devant* le banc de haut-parleurs ;
- source omnidirectionnelle située $(3, 5)$, soit cinq mètres *derrière* le banc de haut-parleurs et au delà de l'extrémité droite du banc ;
- une onde plane se propageant perpendiculairement au banc (onde plane à 0 degré) ;
- source d'ordre -1 (dipôle orienté parallèlement au banc) à $(0, 1.2)$;
- source d'ordre 4 à $(0, 1.2)$.

On simule la réponse du système en "appliquant" les filtres calculés aux réponses mesurées sur la configuration d'égalisation multicanal. Il ne s'agit donc pas de mesures effectives du système égalisé. La linéarité du système et l'indépendance des voies entre elles nous permet cependant de garantir la validité des résultats présentés.

On compare la méthode d'égalisation multicanal avec une procédure plus classique, basée sur une égalisation individuelle des haut-parleurs indépendamment de la source virtuelle synthétisée. Les filtres d'égalisation individuelle sont obtenus par la méthode proposée dans la partie 5.1.4. Les sources virtuelles sont synthétisées par la Wave Field Synthesis en appliquant le filtrage en \sqrt{jk} jusqu'à la fréquence d'aliasing spatial, ainsi que les retards et les gains nécessaires (cf. partie 1.3).

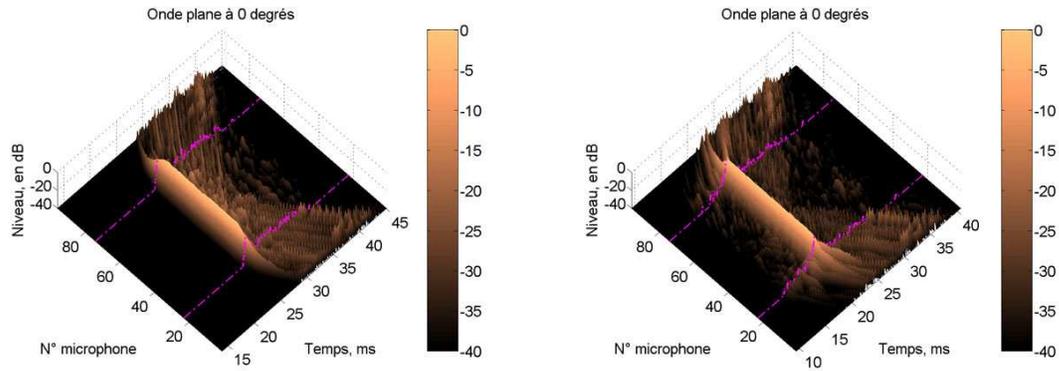
Quelque soit la source virtuelle considérée, la reproduction du champ sonore par la WFS avec égalisation multicanal est meilleure à l'intérieur de la zone de la visibilité que celle avec l'égalisation individuelle. Ceci est particulièrement visible dans le domaine fréquentiel mais reste aussi valable dans le domaine temporel. On note toutefois la présence de contributions de faible niveau devant le front d'onde principal. Cependant, le niveau de ces contributions reste très faible et subit très probablement le masquage rétro-actif du front d'onde principal.

En dehors de la fenêtre de visibilité, le champ acoustique synthétisé est globalement plus complexe pour l'égalisation multicanal que pour l'égalisation individuelle. Ceci est particulièrement marqué pour les sources focalisées (omnidirectionnelles et degré supérieur, cf. figures 5.34, 5.35, 5.36).



(a) Réponses fréquentielles, égalisation individuelle

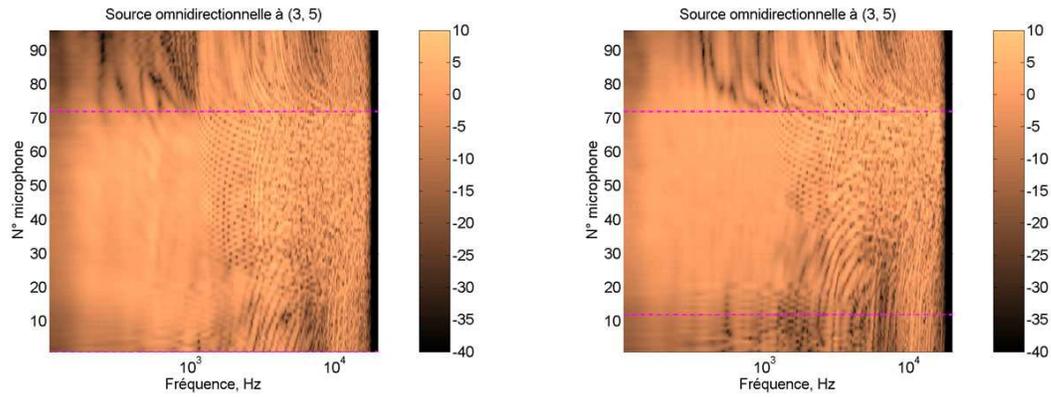
(b) Réponses fréquentielles, égalisation multicanal



(c) Réponses impulsionnelles filtrées passe bas (fréquence de coupure : 2000 Hz), égalisation individuelle

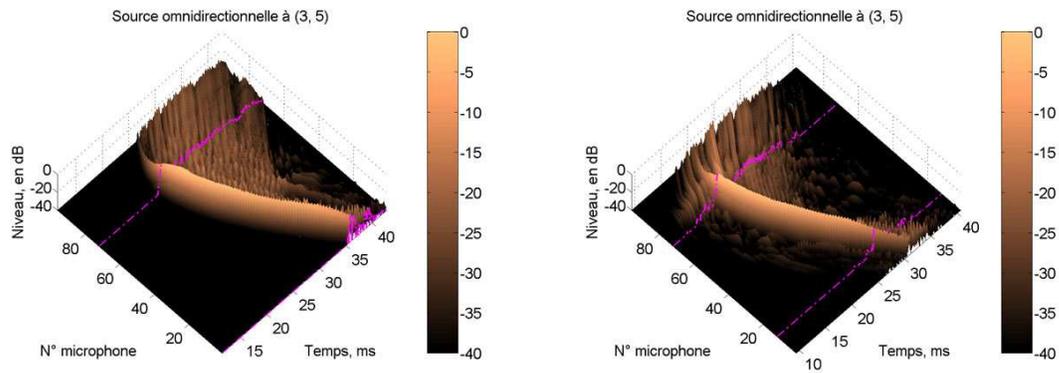
(d) Réponses impulsionnelles filtrées passe bas (fréquence de coupure : 2000 Hz), égalisation multicanal

FIG. 5.32 – Champ synthétisé pour une onde plane se propageant perpendiculairement au banc de haut-parleurs.



(a) Réponses fréquentielles, égalisation individuelle des haut-parleurs

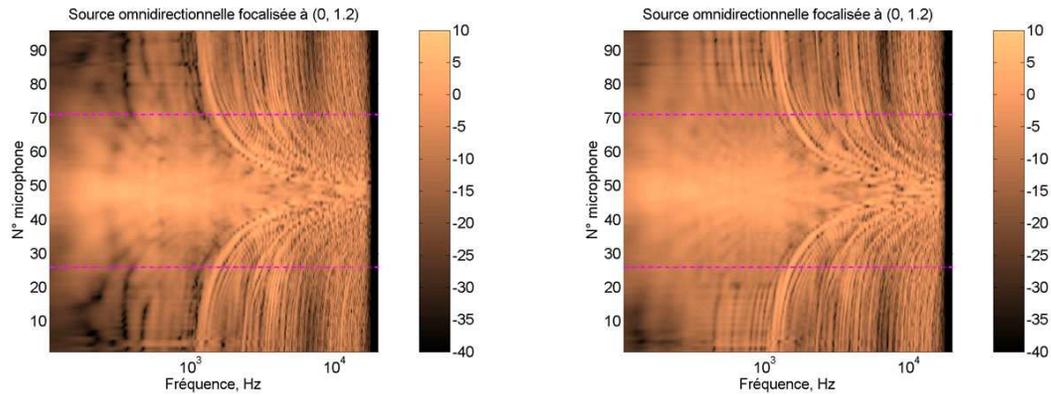
(b) Réponses fréquentielles, égalisation multicanal



(c) Réponses impulsionnelles filtrées passe bas (fréquence de coupure : 1500 Hz), égalisation individuelle

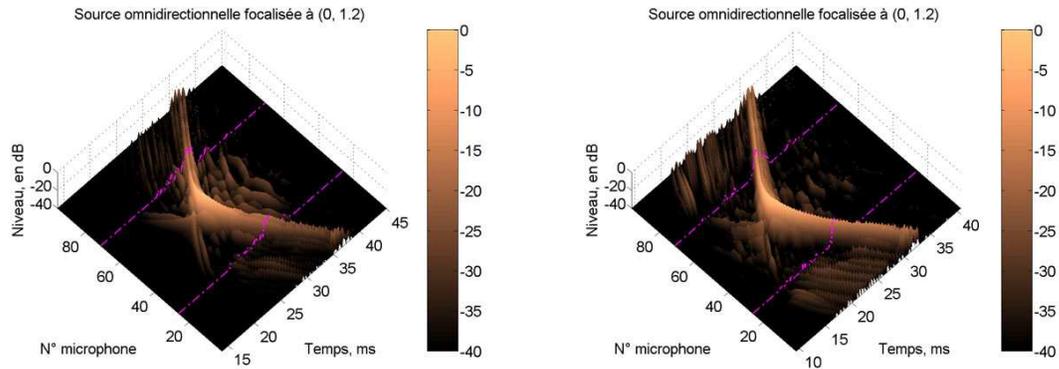
(d) Réponses impulsionnelles filtrées passe bas (fréquence de coupure : 1500 Hz), égalisation multicanal

FIG. 5.33 – Champ synthétisé pour une source omnidirectionnelle située 5m derrière le banc de haut-parleurs et à 3 m sur la droite.



(a) Réponses fréquentielles, égalisation individuelle des haut-parleurs

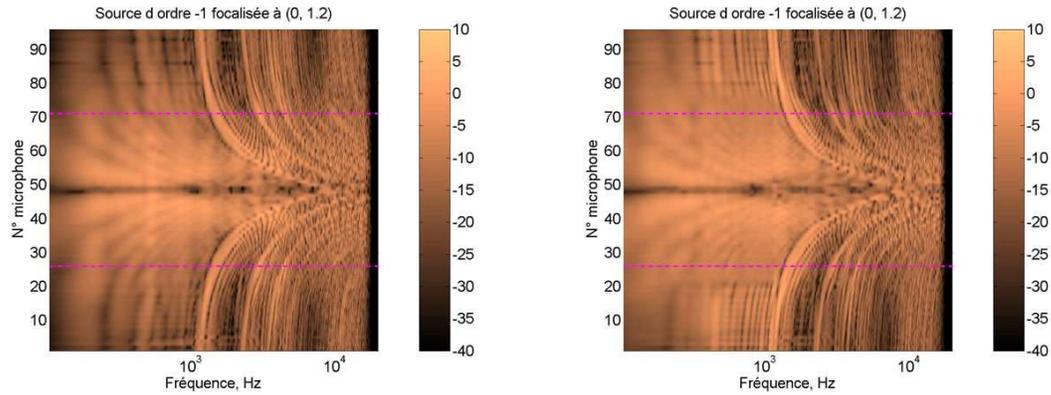
(b) Réponses fréquentielles, égalisation multicanal



(c) Réponses impulsionnelles filtrées passe bas, fréquence de coupure correspond à la fréquence d'aliasing, égalisation individuelle des haut-parleurs

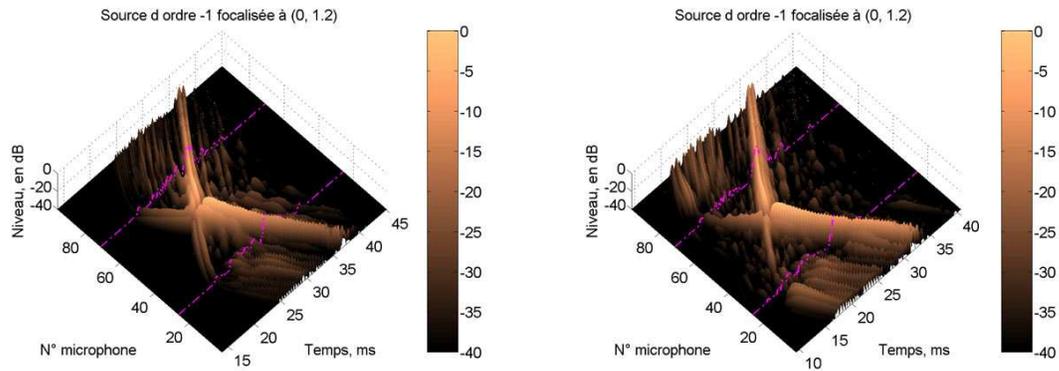
(d) Réponses impulsionnelles filtrées passe bas, fréquence de coupure correspond à la fréquence d'aliasing, égalisation multicanal

FIG. 5.34 – Champ synthétisé pour une source omnidirectionnelle focalisée situé 80 cm devant le banc de haut-parleurs.



(a) Réponses fréquentielles, égalisation individuelle des haut-parleurs

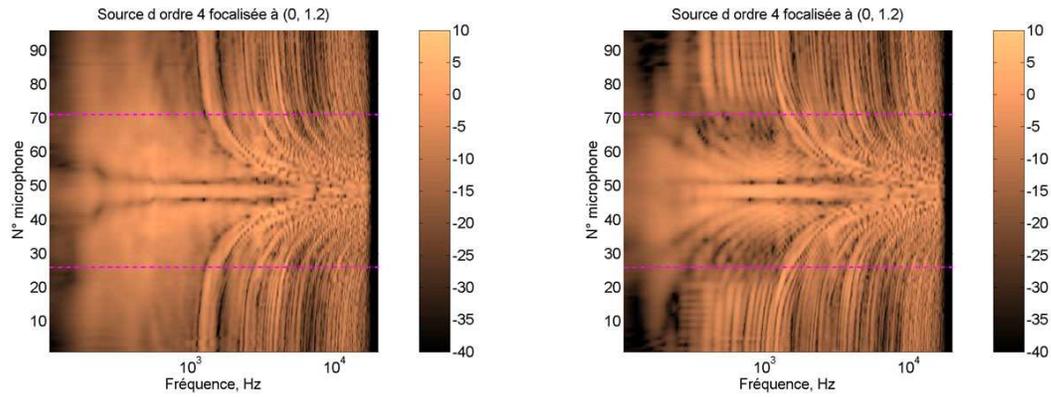
(b) Réponses fréquentielles, égalisation multicanal



(c) Réponses impulsionnelles filtrées passe bas, fréquence de coupure correspond à la fréquence d'aliasing, égalisation individuelle des haut-parleurs

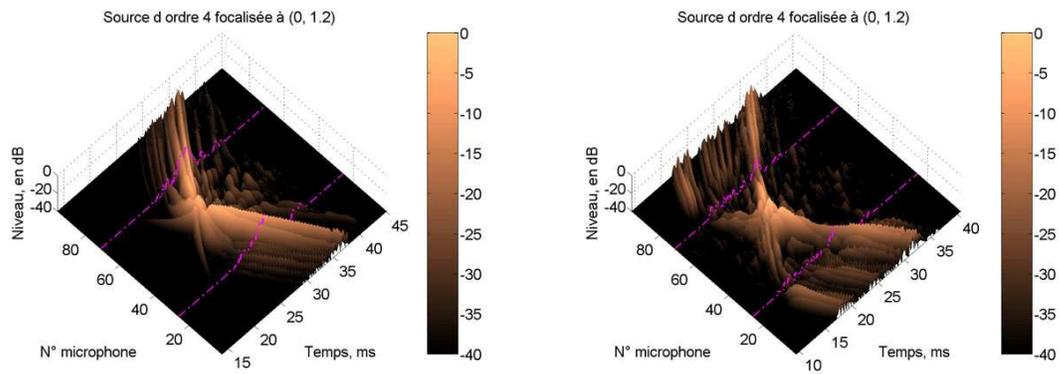
(d) Réponses impulsionnelles filtrées passe bas, fréquence de coupure correspond à la fréquence d'aliasing, égalisation multicanal

FIG. 5.35 – Champ synthétisé pour une source focalisée de directivité dipolaire orientée parallèlement au banc de haut-parleurs (directivité de degré -1) située en $(0, 1.2)$



(a) Réponses fréquentielles, égalisation individuelle des haut-parleurs

(b) Réponses fréquentielles, égalisation multicanal



(c) Réponses impulsionnelles filtrées passe bas, fréquence de coupure correspond à la fréquence d'aliasing, égalisation individuelle des haut-parleurs

(d) Réponses impulsionnelles filtrées passe bas, fréquence de coupure correspond à la fréquence d'aliasing, égalisation multicanal

FIG. 5.36 – Champ synthétisé pour une source focalisée de directivité de degré 4 située en $(0, 1.2)$.

Propagation du champ synthétisé

On se propose de comparer les fonction de qualité $Q_{\Psi}^{Meq}(r(R), k)$ ou $Q_{\Psi}^{IndEq}(r(R), k)$ qui se réfèrent à l'égalisation multicanal ou à l'égalisation individuelle. Ces fonctions de qualité sont évaluées à partir des mesures des 4 MAPs réalisées sur les trois lignes de microphones situées à l'intérieur de la zone d'écoute (cf. figure 5.29(b)).

On considère l'ensemble de sources suivant :

- source omnidirectionnelle focalisée située à $(0, 1.2)$, soit 80 cm *devant* le banc de haut-parleurs ;
- source omnidirectionnelle située $(3, 5)$, soit trois mètres *derrière* le banc de haut-parleurs et au delà de l'extrémité droite du banc de haut-parleurs ;
- source omnidirectionnelle à $(0, 9)$ soit sept mètres derrière le banc de haut-parleurs et centrée par rapport à la zone d'écoute.

L'ensemble des figures 5.37 présente le module de la fonction de qualité exprimé en dB. Les figures de gauche correspondent à la WFS avec égalisation individuelle ; les figures de droite, à la WFS avec égalisation multicanal. Rappelons que pour l'égalisation multicanal, les microphones 1 à 48 font partie des points sur lesquels a été réalisée l'égalisation multicanal. Ceci explique pourquoi le module de la fonction de qualité subit le moins de variations autour de la valeur cible (0 dB). De manière générale, on voit que l'égalisation multicanal permet de compenser les défauts principaux sur les trois bancs de microphones.

Cependant, la fonction de qualité obtenue n'est pas parfaite et fait apparaître des résonances et des anti-résonances dont le niveau excède le seuil perceptif de ± 1 dB. Ceci se remarque notamment pour l'onde plane autour de 700-800 Hz (cf. figure 5.37(f)) et la source focalisée entre 200 et 300 Hz (cf. figure 5.37(b)). La région fréquentielle 200-400 Hz semble problématique pour l'ensemble des sources considérées. C'est en effet la région fréquentielle dans laquelle le rayonnement des MAPs est le plus complexe et varie le plus entre les différents excitateurs. Ceci montre les limites de la méthode et pose la question de savoir quelles sont les caractéristiques de directivité que l'égalisation multicanal permet de compenser efficacement dans une zone étendue de l'espace. Le problème est extrêmement complexe et mériterait une étude spécifique.

L'ensemble des figures 5.38 présente le temps de groupe associé aux fonctions de qualité. Comme pour le module, les figures de gauche correspondent à la WFS avec égalisation individuelle ; les figures de droite à la WFS avec égalisation multicanal.

Les résultats confirment ceux sur le module de la fonction de qualité. La méthode d'égalisation multicanal permet d'améliorer le comportement en phase de la synthèse des fronts d'onde mais ne garantit pas une transparence parfaite dans l'ensemble de la zone de reproduction.

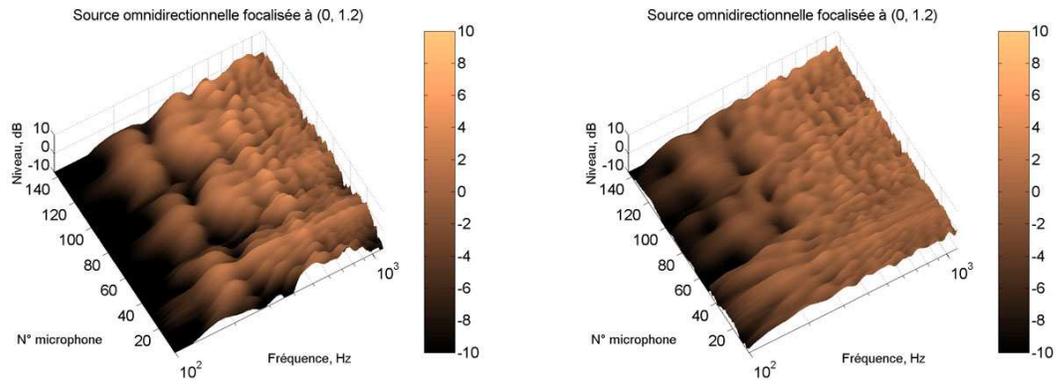
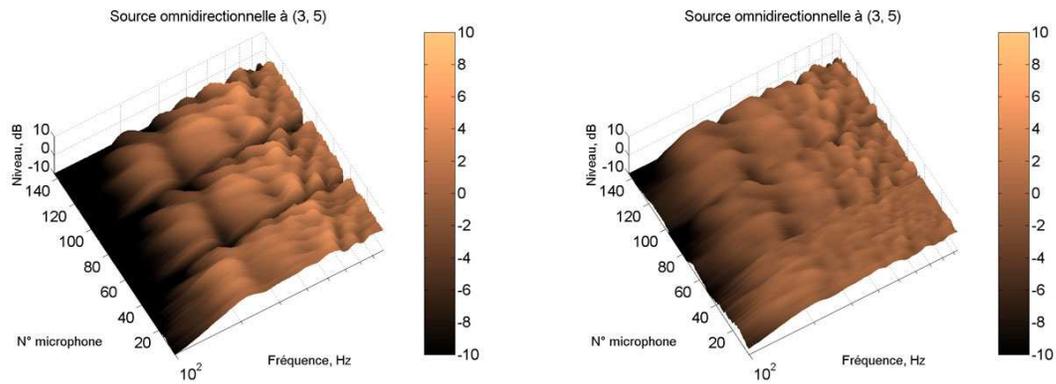
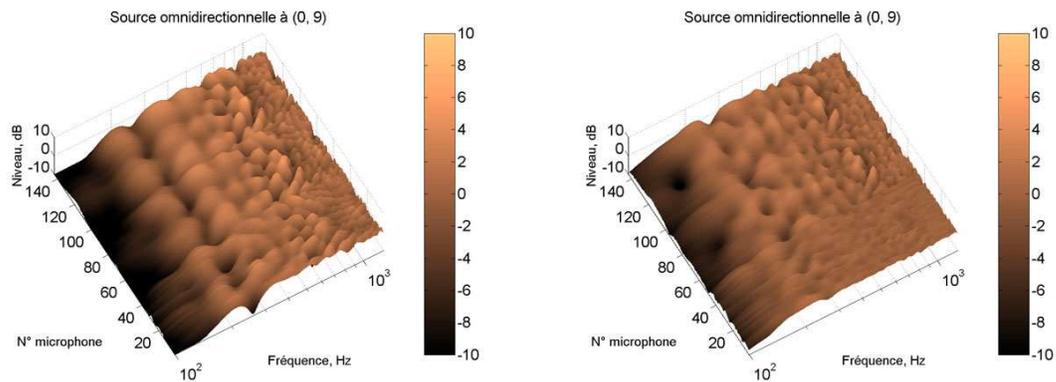
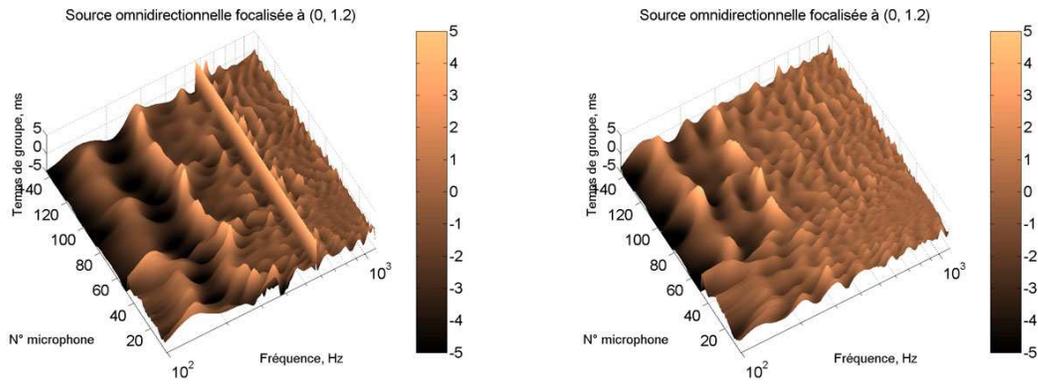
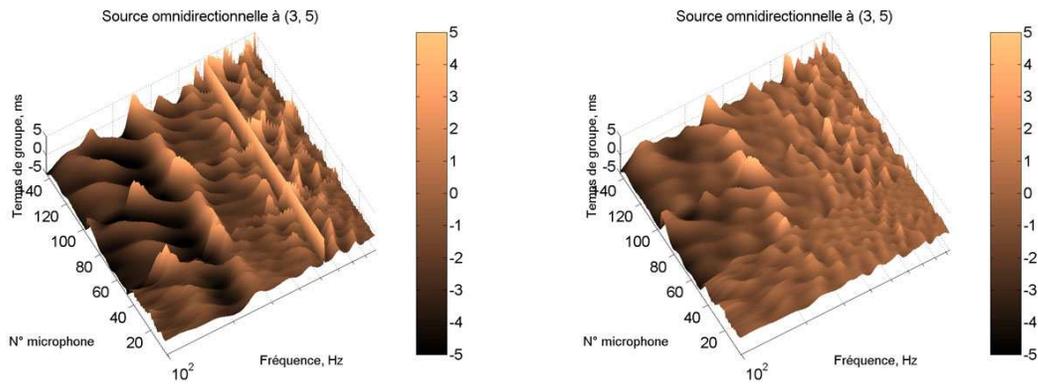
(a) Source omnidirectionnelle focalisée à $(0, 1.2)$, égalisation individuelle des haut-parleurs(b) Source omnidirectionnelle focalisée à $(0, 1.2)$, égalisation multicanal(c) Source omnidirectionnelle à $(3, 5)$, égalisation individuelle des haut-parleurs(d) Source omnidirectionnelle à $(3, 5)$, égalisation multicanal(e) Source omnidirectionnelle à $(0, 9)$, égalisation individuelle des haut-parleurs(f) Source omnidirectionnelle à $(0, 9)$, égalisation multicanal

FIG. 5.37 – Module de $Q_{\Psi}^{wfs}(r(R), k)$ pour la configuration de 4 haut-parleurs MAP sous la fréquence d'aliasing spatial (100-2000 Hz) synthétisant différentes sources virtuelles omnidirectionnelles avec égalisation individuelle ou égalisation multicanal



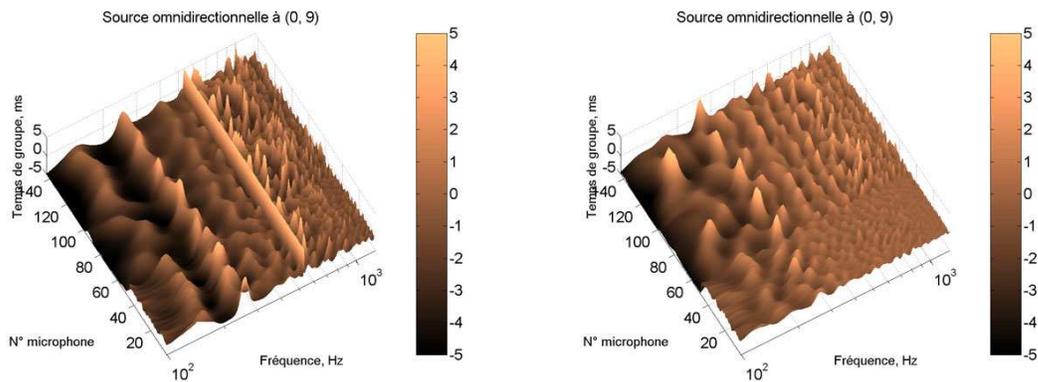
(a) Source omnidirectionnelle focalisée à (0, 1.2), égalisation individuelle des haut-parleurs

(b) Source omnidirectionnelle focalisée à (0, 1.2), égalisation multicanal



(c) Source omnidirectionnelle à (3, 5), égalisation individuelle des haut-parleurs

(d) Source omnidirectionnelle à (3, 5), égalisation multicanal



(e) Source omnidirectionnelle à (0, 9), égalisation individuelle des haut-parleurs

(f) Source omnidirectionnelle à (0, 9), égalisation multicanal

FIG. 5.38 – Temps de groupe associé à $Q_{\Psi}^{eq}(r(R), k)$ pour la configuration de 4 haut-parleurs MAP sous la fréquence d'aliasing spatial (100-2000 Hz) synthétisant différentes sources virtuelles omnidirectionnelles avec égalisation individuelle ou égalisation multicanal

5.3 Discussion : validité de la méthode d'égalisation multicanal

Nous avons présenté dans ce chapitre une méthode d'égalisation multicanal pour la Wave Field Synthesis. Cette méthode se base sur la caractérisation et le contrôle du champ acoustique émis par une distribution de haut-parleurs pour une ligne de microphones omnidirectionnels situés dans la zone d'écoute. Elle se rapproche des techniques de contrôle actif.

La particularité de cette technique est de réaliser une égalisation dépendante du contenu (de la source virtuelle reproduite) et d'assurer la transparence de reproduction de *chacune* des sources virtuelles reproduites sur un système WFS donné. Pour ce faire, il est nécessaire de constituer une base de données de filtres, chaque jeu de filtres étant associé à une source virtuelle, afin de permettre la reproduction d'une quelconque source virtuelle spécifiée par l'utilisateur.

Cette technique rend la reproduction sonore par Wave Field Synthesis en temps réel plus complexe. Elle requiert, en effet, un filtre RIF et un retard par haut-parleur et par source ; au lieu d'un filtre par haut-parleur pour toutes les sources, un retard et un gain par haut-parleur et par source pour la Wave Field Synthesis avec égalisation individuelle des transducteurs. La longueur des filtres requise reste cependant relativement faible (≈ 1000 points). La synthèse d'une source reste donc relativement peu coûteuse si on effectue la convolution dans le domaine fréquentiel ou si on utilise un algorithme de convolution partitionné.

Dans le cas de haut-parleurs idéaux, cette méthode permet de compenser les principaux défauts de la Wave Field Synthesis, en particulier la diffraction, dans une large zone d'écoute (cf. partie 5.2.1). Pour des haut-parleurs réels, par exemple des MAPs, cette méthode permet de compenser les principaux défauts en terme de directivité des haut-parleurs dans le plan horizontal. La transparence de la reproduction de la reproduction n'est pas parfaite mais bien meilleure que dans le cas d'une égalisation individuelle. Cette méthode ne permet cependant que le contrôle du champ sonore synthétisé à l'intérieur de la zone de visibilité et dans le plan horizontal. La qualité du champ sonore synthétisé en dehors du plan horizontal et dans le reste de la zone d'écoute dépend lui, totalement de la *directivité* des haut-parleurs que l'on utilise. Cette méthode ne permet en fait que le contrôle des *composantes principales* du champ synthétisé dans le plan horizontal.

Des études complémentaires seraient ainsi nécessaires afin d'estimer quelles sont les types de caractéristiques de directivité des haut-parleurs que cette méthode permet de contrôler et quelles sont celles sur lesquelles elle n'a pas d'action.

On peut aussi s'interroger quant à la validité du critère d'erreur (moindres carrés). Son lien avec la perception est en effet relativement limité puisqu'elle répartit l'erreur de manière uniforme sur toute l'étendue de la réponse impulsionnelle du système. Il est ainsi susceptible d'introduire des contributions erronées, en avance par rapport au premier front d'onde. Or, on sait que la perception est particulièrement sensible à ce type d'erreur. Une extension intéressante consisterait à définir de nouveaux critères d'erreur plus en lien avec la perception et de définir des algorithmes d'inversion multicanal répondant à ces critères.

Les premiers développements de cette méthode ont été réalisés lors du stage de DEA de l'auteur

(Corteel, 2001) à Studer Professional Audio en 2001 dans le cadre du projet européen Carrouso (Brix *et al.*, 2001). Ces développements préliminaires ont fait l'objet d'un dépôt de brevet aux USA par Harman International Industries (Horbach & Corteel, 2003) qui est actuellement dans sa phase d'examen, et de publications scientifiques (Corteel *et al.*, 2002) (Horbach *et al.*, 2002) (Horbach *et al.*, 2002).

Les développements réalisés au cours de la thèse concernent la mise en œuvre et le perfectionnement de la technique proposée, ainsi que le support de la synthèse de sources à directivité arbitraire. Une partie de ce travail a été menée en collaboration avec Terence Caulkins (Caulkins, 2003) (Caulkins *et al.*, 2003) (Warusfel *et al.*, 2004a). Les perfectionnements que nous avons apportés concernent la définition propre de la fréquence de coupure entre les deux régimes (multicanal/individuel) grâce aux travaux théoriques menés sur l'aliasing spatial, l'amélioration de la composition des filtres, l'introduction de la mise en forme de la matrice des fonctions de transfert, le choix des haut-parleurs et des microphones à utiliser et l'analyse de la propagation du champ synthétisé.

Cette technique est maintenant utilisée dans l'installation permanente du studio 4 de l'IRCAM. Elle a été présentée lors de nombreuses démonstrations publiques, depuis ses premiers développements, dans un nombre important de manifestations et de congrès scientifiques (19ème conférence internationale de l'Audio Engineering Society, Juin 2001, Allemagne ; International Funkausstellung 2001, Août 2001, Berlin Allemagne ; IST event 2001, Dusseldorf, Allemagne, Décembre 2001, Allemagne ; 112ème convention de l'AES, Munich, Mars 2002, Allemagne ; IRCAM Résonances 2002, Octobre 2002, France ; TonmeisterTagung 2002, Novembre 2002, Allemagne ; 114ème convention de l' AES, Mars 2003, Pays Bas ; IRCAM Résonances 2003, Octobre 2003, France ; workshop sur la Wave Field Synthesis, Forum Neues Musikteater, Stuttgart, Allemagne, Mai 2004 ; IRCAM Résonances 2004, Octobre 2004, France).

Chapitre 6

Prise en compte de la pièce de restitution

La prise en compte de la pièce de restitution par des méthodes de traitement de signal s'avère être un problème extrêmement complexe. Les caractéristiques du champ acoustique présent dans la salle dépendent énormément de la position considérée. Il s'avère véritablement impossible de modéliser l'interface haut-parleur/salle d'écoute par un unique canal acoustique que l'on chercherait ensuite à compenser. Ce type de procédure d'égalisation individuelle du haut-parleur *et* de la salle d'écoute est en effet voué à l'échec car il n'autorise qu'une compensation locale de l'acoustique de la pièce de restitution (Fielder, 2003).

Dans le chapitre 4, nous avons présenté des méthodes classiques d'égalisation multicanal pour la compensation de l'acoustique de la pièce d'écoute. Ces techniques s'appuient sur le théorème MINT qui autorise une compensation exacte de l'effet de salle en un nombre réduit de positions de la pièce de restitution. Toutefois, la compensation n'est effective qu'au niveau des points de contrôle et les résultats sont dégradés partout ailleurs. Elles s'appuient sur un nombre réduit de haut-parleurs qui servent *à la fois* à synthétiser le son direct et à compenser l'acoustique de la pièce de restitution.

Le problème inhérent à ces techniques est justement ce faible nombre de haut-parleurs (typiquement 2 à 5, stéréophonie). Ceci rend, par exemple, *impossible* l'annulation d'une réflexion précoce dans une large zone de l'espace. Il est de même inutile de vouloir tenter d'annuler ou même réduire la réverbération dans une zone étendue.

Les techniques les plus récentes utilisent des descriptions "simplifiées" de l'acoustique de la salle d'écoute (Hatziantoniou & Mourjopoulos, 2003), ou se concentrent sur la réduction globale du couplage des haut-parleurs aux modes de la salle d'écoute en dessous de la fréquence de Schroeder (Hershkowitz & Durlach, 1962) (Schroeder, 1996) par des techniques de filtrage simples (Pedersen, 2003).

Dans le cas où les haut-parleurs sont montés dans les parois de la salle d'écoute, on a vu que la Wave Field Synthesis permet de réduire le niveau des réflexions précoces *du* plan horizontal

dans le plan horizontal sur toute l'étendue de la zone d'écoute (cf. chapitre 3). Le nombre élevé de haut-parleurs qui interviennent ouvre en effet de nouvelles possibilités pour la compensation de l'acoustique de la salle d'écoute.

Ces dernières années, des méthodes basées sur l'inversion multicanal ont ainsi été proposées par l'auteur et Rozenn Nicol (Corteel & Nicol, 2003) et à l'université d'Erlangen par Spors *et al.* (Spors *et al.*, 2003) (Spors *et al.*, 2004). Ces méthodes sont issues du travail mené dans le cadre du projet Carrouso sur la compensation de l'acoustique de la pièce de restitution par inversion multicanal dans le cadre de la WFS. Dans ces méthodes, l'ensemble des haut-parleurs du dispositif de restitution est utilisé pour à la fois synthétiser le son direct *et* compenser l'acoustique de la pièce de restitution. Les filtres qui permettent de réaliser cet objectif sont calculés par inversion multicanal à partir de mesures réalisées sur une ceinture de microphones de contrôle qui entourent la zone d'écoute. Les méthodes diffèrent par l'identification du système MIMO et par le type d'algorithme d'inversion multicanal mis en jeu. Étant donné le nombre élevé de haut-parleurs et de microphones mis en jeu ainsi que la longueur des réponses impulsionnelles qui décrivent le système, la complexité de ces méthodes est relativement importante. Nous décrirons en détail de ce méthode au cours de ce chapitre.

Plus récemment van Zon et l'auteur (van Zon, 2003) (van Zon *et al.*, 2004), dans le cadre d'une collaboration entre l'université de Delft et l'IRCAM, ont proposé une méthode qui effectue une séparation entre la synthèse du son direct et la compensation de l'effet de salle. Dans le cas où les haut-parleurs sont intégrés dans les parois de la salle d'écoute, on peut en effet effectuer une distinction entre les haut-parleurs qui synthétisent le son direct et ceux qui sont utilisés pour la synthèse des sources virtuelles de compensation (cf. partie 3.2.1). La seconde particularité de cette méthode est de se concentrer sur la compensation des réflexions d'ordre 1. Cette méthode est présentée dans la suite et pour plus de clarté sera nommée méthode van Zon/Corteel.

Dans ce chapitre, nous proposons une méthode originale qui se base sur le contrôle de la synthèses des sources virtuelles de compensation des réflexions d'ordre 1 sur les murs verticaux de la salle d'écoute (cf. partie 3.2). Nous la nommons méthode de compensation individuelle des réflexions. Cette méthode peut être considéré comme une conséquence directe de la méthode van Zon/Corteel. Elle réalise une décomposition supplémentaire du problème en considérant séparément chacune des parois verticales et des réflexions d'ordre 1 correspondantes.

Au cours de ce chapitre, nous décrivons dans un premier temps les principales méthodes de compensation de la salle d'écoute par inversion multicanal, en suivant les mêmes étapes qu'au chapitre précédent (identification du système MIMO, champ sonore cible, inversion multicanal, au-dessus de la fréquence d'aliasing). Nous présentons ensuite des résultats qui permettent de juger de leur efficacité, en particulier pour la méthode de compensation individuelle des réflexions. Nous considérons des configurations composées de haut-parleurs et de parois idéales ; puis, une installation composée de haut-parleurs MAP afin d'éliminer les réflexions mutuelles sur les parois que forment les membranes (plaques rigides) de ces haut-parleurs.

6.1 Méthodes d'inversion multicanal pour la compensation de l'effet de salle

Dans la partie 3.1 du chapitre 3, nous avons présenté les caractéristiques du champ synthétisé en WFS par un banc de haut-parleurs intégré dans une des parois de la salle. On a vu que ces caractéristiques dépendaient largement de la source virtuelle synthétisée.

Nous avons ensuite proposé une méthode de compensation des réflexions précoces dans le cas où les parois étaient parfaitement réfléchissantes (cf. partie 3.2). Cette méthode se présente comme une extension de la Wave Field Synthesis et hérite des défauts inhérents à cette technique. Nous avons vu en particulier que la diffraction, due à la taille finie du banc de haut-parleurs, réduit ses performances, principalement aux basses fréquences (cf. partie 3.2.1).

En pratique, le problème présente des caractéristiques supplémentaires. Parmi celles-ci, la nature du matériau dont sont constituées les parois et la directivité des haut-parleurs utilisés pour la compensation sont aussi à prendre en compte.

La prise en compte de toutes ces caractéristiques par une caractérisation du champ acoustique présent dans la salle d'écoute qui doit être valable dans *toute la zone d'écoute* indépendamment de la source virtuelle synthétisée. Cette caractérisation doit décrire uniquement les réflexions sur les murs verticaux de la salle de restitution. C'est en effet les *seules réflexions* que le dispositif de reproduction employé est *capable* de compenser. L'objectif à réaliser est alors la compensation totale ou, de façon plus réaliste, partielle des réflexions sur les murs verticaux de la salle d'écoute dans le plan horizontal. Celui-ci peut être réalisé grâce à un ensemble de filtres calculés par inversion multicanal à placer en tête de chaque haut-parleur du dispositif de reproduction.

Dans cette partie, nous définissons les méthodes d'égalisation multicanal de la littérature ainsi que la méthode de compensation individuelle des réflexions que nous proposons dans ce document. Celles-ci sont décrites en reprenant les étapes de la partie 5.1 :

- identification du système MIMO ;
- champ sonore cible ;
- inversion multicanal ;
- que faire au-dessus de la fréquence d'aliasing ?

6.1.1 Identification du système MIMO

On a vu, dans la partie 5.1.1, que la connaissance de la pression et du gradient normal de pression rayonné sur une surface fermée $\partial\Omega$ décrit de manière exacte et *unique* un champ acoustique émis par des sources *primaires* situées à *l'extérieur* de $\partial\Omega$, dans le sous-espace Ω_R à *l'intérieur* de $\partial\Omega$. Dans le cas où $\partial\Omega$ est un plan infini, la connaissance de la pression sur $\partial\Omega$ est suffisante pour décrire le champ rayonné dans Ω_R . Ceci suppose que l'ensemble des sources, à l'origine du champ, soient situées dans un demi-espace limité par un plan infini.

Dans le cas où l'on veut décrire le champ associé au son direct *et* aux sources images correspon-

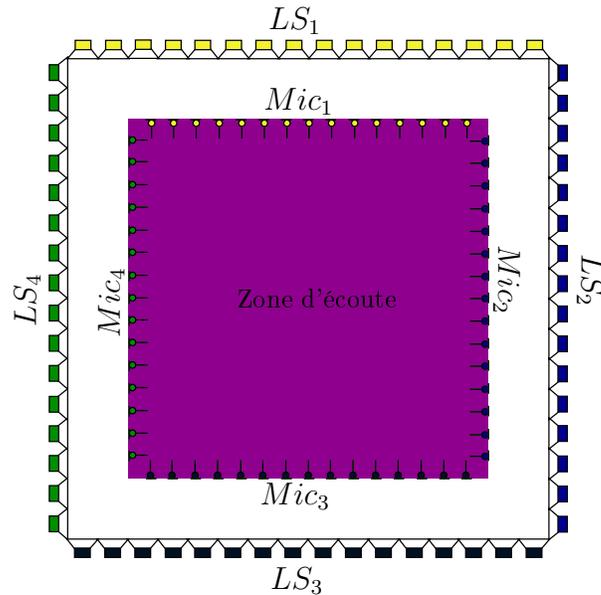


FIG. 6.1 – Haut-parleurs et microphones pour la compensation de l'effet de salle en WFS

dantes, il est impossible de définir un tel plan. On doit alors considérer une surface fermée sur laquelle il faut connaître la pression *et* le gradient normal de pression.

Rappelons de plus que la WFS ne permet "que" la compensation des sources images situées dans le plan horizontal et que celle-ci est principalement efficace dans le plan horizontal où sont, a priori, situés les auditeurs.

On peut alors entourer la zone d'écoute d'un nombre fini de microphones de pression et de gradient de pression (cf. figure 6.1) en reprenant les simplifications présentées dans la partie 5.1.1 . Il s'avère, en pratique, suffisant d'utiliser des microphones à directivité cardioïde orientés vers l'extérieur de la zone d'écoute. Le même type d'analyse que pour l'utilisation de plusieurs segments non colinéaires de haut-parleurs pour la Wave Field Synthesis montre que le "rôle" de la connaissance simultanée de la pression *et* du gradient de pression permet principalement de discriminer le champ entrant et le champ sortant mais n'apporte pas d'informations supplémentaires. Les microphones à directivité cardioïde effectuent cette distinction de façon plus simple en se concentrant sur les contributions entrantes. Ceci permet de diviser le nombre de microphones nécessaires par deux et réduire ainsi sensiblement le coût de calcul de l'inversion multicanal.

Comme pour l'égalisation multicanal, ce type de mesure n'assure pas une description *exacte* du champ sonore à l'intérieur de la zone d'écoute mais donne une description de ses *composantes principales*.

Ce type de mesure a été proposé par l'auteur et Nicol (Corteel & Nicol, 2003). Un point important consiste cependant à enlever des mesures toutes les réflexions ne provenant pas du plan horizontal. Ce point peut s'avérer relativement délicat en pratique et on ne propose pas de méthode pour le faire.

Spors *et al.* (Spors *et al.*, 2003) proposent de donner une description du système par une décom-

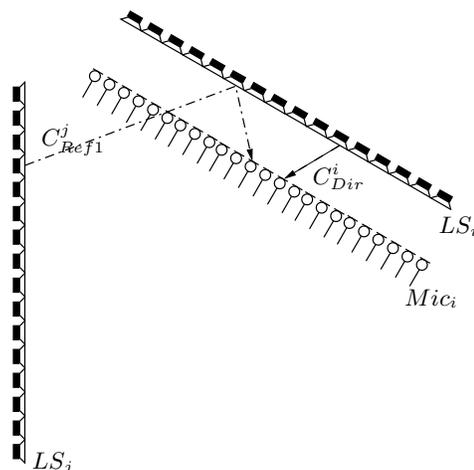


FIG. 6.2 – Caractérisation du champ direct et de la première réflexion pour la méthode proposée

position en ondes planes. Celle-ci est obtenue à partir de mesures effectuées sur un banc circulaire de microphones par la technique développée par Hulsebos (Hulsebos *et al.*, 2001) (Hulsebos & de Vries, 2002). Plus récemment (Spors *et al.*, 2004), ils proposent de réaliser l'identification dans le domaine des harmoniques cylindriques (Spors *et al.*, 2004) à partir du même type de mesure. D'après Spors *et al.*, la décomposition du champ sonore permet un meilleur conditionnement du problème en ajoutant des informations sur la direction de propagation des composantes. Remarquons tout de même que cette description n'est valable qu'à l'intérieur du cercle sur lequel ont été effectuées les mesures. Ceci ne permet pas de s'adapter à des géométries de zone d'écoute plus complexes. D'autre part, ce type de description n'évite pas le problème de l'élimination des composantes réfléchies qui ne proviennent pas du plan horizontal. En effet, si on ne les élimine pas, celles-ci sont distribuées sur *toutes* les composantes de base (ondes planes ou harmoniques sphériques).

Le but de ces méthodes est de réduire le niveau des réflexions précoces. Le fait de considérer des réponses impulsionnelles qui décrivent le champ réfléchi dans le plan horizontal et contenant des réflexions d'ordre élevé n'apporte pas, en réalité, d'informations supplémentaires pour la résolution du problème.

On a vu en effet que la réduction du niveau des réflexions d'ordre 1 entraîne une réduction de niveau *naturelle* pour les réflexions d'ordre supérieur. Nous proposons ainsi de "focaliser" la description sur chacune des réflexions d'ordre 1 sur les parois verticales de la salle d'écoute.

La méthode proposée ici vise à contrôler de manière efficace la synthèse des sources de compensation des réflexions de premier ordre par la WFS (cf. partie 3.2.1). Une réflexion de premier ordre est issue de la réflexion du son direct synthétisé par un ensemble de haut-parleurs sur une des I parois de la salle d'écoute. Nous proposons de considérer séparément chacune des parois de la salle de restitution et chacune des réflexions de premier ordre.

Supposons que l'on dispose d'un banc de haut-parleurs LS_i intégré à la paroi i considérée. Disposons un banc linéaire de microphones Mic_i parallèle à la paroi et mesurons (cf. figure 6.2) :

- le champ réfléchi $C_{Ref1}^i(z)$, au premier ordre, sur la paroi i par l'ensemble des haut-parleurs de l'installation, sauf, évidemment, ceux situés sur la paroi ;
- le son direct des haut-parleurs $C_{Dir}^i(z)$ situés sur la paroi, qui vont être utilisés pour compenser le champ réfléchi.

On remarque que *toutes* les sources qui produisent le champ sonore à caractériser sont situées dans un des deux demi-espaces séparés par le plan vertical passant par le banc de microphones Mic_i . On se trouve ainsi dans une situation de type "Rayleigh" et on peut se contenter d'utiliser des microphones de pression pour Mic_i . L'extraction des réflexions de premier ordre sur la paroi i est cependant facilitée par l'utilisation de microphones à directivité cardioïde. Ceux-ci permettent de "préfiltrer" naturellement le son direct provenant des haut-parleurs mesurés et de faciliter son extraction.

L'identification du système MIMO pour la méthode van Zon/Corteel se situe à l'intermédiaire entre les deux types de mesures présentées auparavant. Dans cette méthode, on effectue une distinction entre la synthèse du son direct et la compensation des réflexions. Cette méthode se base, elle aussi, sur le contrôle de la synthèse des sources de compensation des réflexions de premier ordre mais considère l'ensemble des parois en une seule étape. On utilise ainsi des bancs de microphones cardioïdes disposés autour de la zone d'écoute (cf. figure 6.1) avec lesquels on mesure :

- uniquement le champ réfléchi $C^{Room}(z)$ dans plan horizontal, sur toutes les parois *verticales* de la salle de restitution.
- le son direct *et* le champ réfléchi $C_{Dir}^{Ref}(z)$ dans plan horizontal, sur toutes les parois *verticales* de la salle de restitution, pour tous les haut-parleurs du dispositif de reproduction ;

6.1.2 Champ sonore cible

Pour les méthodes qui considèrent la synthèse du son direct et la compensation de l'effet de salle en une seule étape (Corteel & Nicol, 2003) (Spors *et al.*, 2003) (Spors *et al.*, 2004), le champ sonore cible correspond au rayonnement de la source virtuelle cible en champ libre capté par les microphones de contrôle. On renvoie ainsi le lecteur à la partie 5.1.2 pour plus de précisions.

Dans les méthodes proposées par Spors *et al.*, le signal spécifié correspond à la transformée des signaux simulés au niveau du banc circulaire, dans le domaine des ondes planes (Spors *et al.*, 2003) ou des harmoniques cylindriques (Spors *et al.*, 2004). Ces descriptions sont, a priori, avantageuses car elles s'effectuent sur des *bases* de fonctions de rayonnement acoustique de l'espace à trois dimensions. Ceci signifie que si l'on réalise l'optimisation sur chacune des fonctions élémentaires de la base (ondes planes, harmoniques cylindriques), on peut synthétiser, a priori, n'importe quelle source virtuelle par recombinaison des fonctions de base. Remarquons alors que l'optimisation sur la base des harmoniques cylindriques revient à réaliser un système Ambisonique aux ordres supérieurs et d'optimiser la synthèse des harmoniques cylindriques.

Le terme d'onde plane est utilisé de manière abusive dans le cadre de la Wave Field Synthesis. Il s'agit en réalité d'ondes à symétrie cylindrique autour de l'axe du banc linéaire de haut-parleurs considéré. Une recombinaison directe à partir de cette base risque ainsi d'être problématique. Pour les harmoniques cylindriques, ce problème est à rapprocher du travail de Daniel (Daniel, 2003) sur la compensation du champ proche pour les systèmes Ambisoniques horizontaux. Cette approche est ainsi séduisante mais n'est pas exempte de difficultés pratiques qui restent à explorer.

La méthode de compensation individuelle des réflexions ainsi que la méthode van Zon/Corteel supposent d'avoir calculé auparavant des filtres H_{Ψ_m} pour la synthèse du son direct associé à une source virtuelle Ψ_m . Ceci peut se faire, par exemple, par la méthode d'inversion multicanal proposée dans le chapitre 5. Ces filtres sont destinés aux haut-parleurs responsables de la synthèse de la source Ψ_m . Dans la méthode van Zon/Corteel, le champ cible $A_{\Psi_m}(z)$ est donné par :

$$A_{\Psi_m}(z) = -C^{Ref}(z) \times H_{\Psi_m}(z) \quad (6.1)$$

A_{Ψ_m} est un champ en opposition de phase avec le champ réfléchi sur les parois verticales de la salle de restitution.

Pour la méthode de compensation individuelle des réflexions, il s'agit de déterminer quelles sont les parois sur lesquelles se réfléchit le son direct au premier ordre. Dans notre cas, les haut-parleurs sont intégrés dans les parois de la salle de restitution. On a vu dans la partie 3.2.1 que l'on peut faire une distinction entre les bancs de haut-parleurs qui contribuent au son direct et ceux qui synthétisent les sources virtuelles de compensation. Ainsi, pour chaque paroi i qui sert à synthétiser une source virtuelle de compensation, le champ cible $A_{\Psi_m^i}(z)$ est donné par :

$$A_{\Psi_m^i}(z) = -C_{Ref1}^i(z) \times H_{\Psi_m} \quad (6.2)$$

$A_{\Psi_m^i}$ est un champ en opposition de phase avec le champ réfléchi sur la paroi i . Rappelons alors que la compensation du champ réfléchi n'est possible que sur une ligne à l'intérieur de la zone d'écoute car la loi d'atténuation de la réflexion et de la source virtuelle de compensation diffèrent (cf. partie 3.2.1). Pour la méthode de compensation individuelle des réflexions, la ligne de référence correspond à la position du banc de microphones Mic_i utilisé pour contrôler la synthèse des sources virtuelles de compensation de la réflexion de premier ordre sur la paroi i . Le choix de la position de ce banc de microphones conditionne ainsi l'efficacité de la compensation dans l'ensemble de la zone d'écoute.

6.1.3 Inversion multicanal

Dans cette partie, l'inversion multicanal est décrite en respectant les mêmes étapes que celles de l'égalisation multicanal (cf. partie 5.1.3) :

- choix des haut parleurs et des microphones à utiliser ;
- mise en forme de la matrice de fonctions de transfert ;
- fréquence limite de calcul.

Choix des haut parleurs et des microphones

Dans les méthodes proposées par Spors *et al.* (Spors *et al.*, 2003) (Spors *et al.*, 2004) et par l'auteur (Corteel & Nicol, 2003), on utilise tous les haut-parleurs et tous les microphones pour réaliser l'inversion multicanal.

Pour la méthode van Zon/Corteel, seuls les haut-parleurs qui ne contribuent pas à la synthèse du son direct sont utilisés pour l'inversion multicanal. Par contre, on réalise l'inversion multicanal sur tous les microphones (van Zon *et al.*, 2004).

Pour la méthode de compensation individuelle des réflexions, la sélection des haut-parleurs et des microphones est réalisée pour chaque paroi sur laquelle on réalise la compensation.

On utilise des critères de visibilité de la source virtuelle image à travers le banc de haut-parleurs image et la paroi sur laquelle se réfléchit le son direct (cf. figure 6.3). On renvoie le lecteur à la partie 3.2.1 pour le choix des haut-parleurs à utiliser lors de la compensation.

En ce qui concerne les microphones, on applique le même type de critère de visibilité. Pour la méthode d'égalisation multicanal du son direct, le choix de ces microphones est important afin de ne réaliser le contrôle du champ sonore qu'à l'intérieur de la zone de visibilité de la source virtuelle à travers le banc de haut-parleurs. Le contrôle du champ en dehors de cette zone de visibilité nécessiterait une modélisation aussi exacte que possible du champ diffracté. Il s'avère relativement délicat à mettre en place. Pour la méthode de compensation individuelle des réflexions, le signal spécifié en dehors de la zone de visibilité est le champ diffracté *réel* produit par réflexion du son direct sur la paroi considérée. Celui-ci peut être pris en compte en utilisant tous les microphones.

Mise en forme de la matrice de fonctions de transfert

Pour la synthèse des sources de compensation des réflexions d'ordre 1 par la WFS (cf. partie 3.2.4), les retards à appliquer aux haut-parleurs qui synthétisent les sources de compensation se dérivent de la même manière que pour le son direct. Ils traduisent le temps de propagation acoustique depuis la source virtuelle jusqu'à la position des haut-parleurs. Dans le cas de la compensation de l'effet de salle, ces retards dépendent de la position de la paroi et peuvent être relativement importants pour des pièces de grande taille.

Si on réalise l'inversion multicanal directement à partir de la matrice des fonctions de transfert qui décrit le système MIMO (comme dans les méthodes de Spors *et al.* et celle décrite par l'auteur dans (Corteel & Nicol, 2003), inversion multicanal directe), la longueur des filtres requis peut s'avérer très importante, tout simplement pour prendre en compte la propagation naturelle du champ sonore. Cette solution s'avère ainsi relativement inefficace pour réaliser ensuite la com-

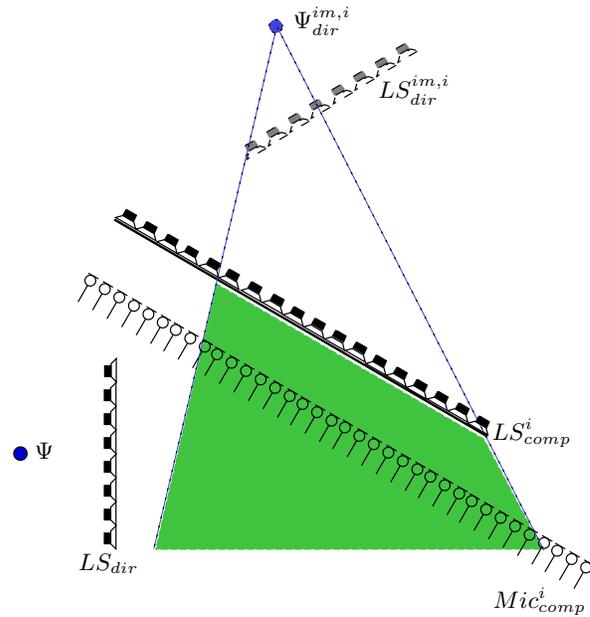


FIG. 6.3 – Choix des haut-parleurs et des microphones pour l'inversion multicanal

compensation de l'effet de salle en temps-réel.

La procédure d'inversion multicanal est, d'autre part, susceptible d'introduire des erreurs au début de la réponse impulsionnelle des filtres. Dans ce cas, pour certaines positions d'écoute, l'auditeur peut recevoir des contributions qui proviennent des haut-parleurs affectés à la compensation avant même l'arrivée du son direct.

Dans le cas où l'on réalise une mise en forme de la matrice de fonctions de transfert qui décrivent le système MIMO ($C_{Dir}^{Ref}(z)$ ou $C_{Dir}(z)$) en appliquant les retards naturels de propagation ($\tilde{C}_{Dir}^{Ref}(z)$ ou $\tilde{C}_{Dir}(z)$), le calcul des filtres est "focalisé" sur la compensation des réflexions d'ordre 1. La longueur des filtres requis n'est alors pas plus importante que dans le cas de l'égalisation multicanal (environ 1000 échantillons à 48 kHz). Ces filtres ne sont, en effet, utilisés que pour compenser les "défauts" de la Wave Field Synthesis et prendre en compte les caractéristiques acoustiques du matériau de la paroi et la directivité des haut-parleurs utilisés.

Ce type de mise en forme de la matrice des fonctions de transfert a été introduit par l'auteur dans la méthode van Zon/Corteel (van Zon *et al.*, 2004). Pour la méthode de compensation individuelle des réflexions, la mise en forme de la matrice de compensation est réalisée en appliquant les retards données par l'extension de la Wave Field Synthesis (cf. partie 3.2.4) pour la synthèse de la source Ψ_m à la matrice de fonctions de transfert décrivant le son direct $C_{Dir}^i(z)$.

Fréquence limite de calcul

Comme pour la procédure d'égalisation multicanal (cf. partie 5.1.3), la fréquence limite de calcul est donnée par les critères d'aliasing spatial WFS rapportés aux haut-parleurs. On a en effet remarqué dans la partie 3.2.4 que la synthèse de la source de compensation du champ réfléchi

est équivalente à la synthèse de la source "directe" dans le cas où les haut-parleurs sont intégrés dans les parois de la salle de restitution.

Rappelons toutefois que la compensation des réflexions n'est pas possible au-dessus de la fréquence d'aliasing spatial, calculée au niveau des microphones Mic_i pour le champ réfléchi sur la paroi i (cf. partie 3.2.1).

Inversion multicanal

Les filtres calculés n'étant valables que jusqu'à une certaine fréquence limite, nous recommandons d'effectuer un sous-échantillonnage des réponses impulsionnelles de $\tilde{C}_{Dir}^{Ref}(z)$ ou $\tilde{C}_{Dir}(z)$, afin de réduire le coût de calcul.

Spors *et al.* (Spors *et al.*, 2003) (Spors *et al.*, 2004) proposent de réaliser l'inversion dans le domaine fréquentiel. Ce choix peut s'avérer judicieux dans la mesure où la taille de la matrice des fonctions de transfert considérée peut être relativement importante. Pour les méthodes qui réalisent en une seule étape la synthèse du son direct et la compensation de l'effet de salle en utilisant l'ensemble des haut-parleurs (Spors *et al.*, 2003) (Spors *et al.*, 2004) (Corteel & Nicol, 2003), on utilise typiquement au moins une centaine de haut-parleurs et de microphones. Pour une salle de restitution de $4,5 \times 6$ m, avec un espacement entre les haut-parleurs de 15 cm, 140 haut-parleurs sont nécessaires pour couvrir toute la circonférence. Le nombre de microphones est du même ordre de grandeur. L'inversion dans le domaine temporel peut alors s'avérer relativement coûteuse et une inversion dans le domaine fréquentiel est plus efficace. Toutefois, les réponses impulsionnelles qui décrivent le système MIMO ($C(z)$) contiennent le champ réfléchi dans la salle de restitution et sont généralement relativement longues (plusieurs milliers de points à 48 kHz). Le calcul dans le domaine fréquentiel donne ainsi des filtres dont la longueur utile est du même ordre. On doit alors les couper de manière plus ou moins arbitraire pour une implémentation efficace en temps réel.

Le calcul d'inversion multicanal n'est pas soumis aux mêmes contraintes que la synthèse des sources en temps réel. On préfère réaliser l'inversion dans le domaine temporel afin de pouvoir choisir précisément la longueur utile des filtres calculés.

La méthode de compensation individuelle des réflexions d'ordre 1 présentée dans ce document réalise le calcul des filtres paroi par paroi. Ainsi, pour chaque inversion multicanal, le nombre de haut-parleurs et de microphones utilisés est bien plus faible (de l'ordre de quelques dizaines) et l'inversion multicanal dans le domaine temporel, par exemple avec la variante multicanal de l'algorithme MFAP présentée dans l'annexe B.1, reste un choix efficace.

6.1.4 Au-dessus de la fréquence d'aliasing

Comme précédemment, la compensation du champ réfléchi peut se mettre sous la forme du contrôle du champ sonore synthétisé par le banc de haut-parleurs encastrés dans la paroi. Or, on a vu dans la partie 5 que ceci n'est possible que jusqu'à la fréquence d'aliasing déterminée,

entre autres, par la géométrie du banc de haut-parleurs utilisés. Au dessus de cette fréquence d'aliasing, le champ n'est plus contrôlable dans une zone étendue de l'espace et la compensation du champ réfléchi n'est plus efficace.

Tous les filtres calculés doivent ainsi subir un filtrage passe-bas dont la fréquence de coupure est donnée par la plus faible des fréquences d'aliasing rapportées au niveau des microphones de contrôle, ou bien, au niveau des haut-parleurs.

6.2 Résultats

Dans un premier temps, nous considérons une situation idéale avec des haut-parleurs omnidirectionnels idéaux intégrés dans des parois parfaitement réfléchissantes. Nous présentons d'abord la compensation possible du champ synthétisé par un banc de haut-parleurs par un autre banc de haut-parleurs parallèle distant comme pour la WFS (cf. partie 3.2.1). Nous considérons ensuite deux parois verticales infinies dans lesquelles sont intégrés des bancs de haut-parleurs de taille finie. Enfin, nous présentons les résultats obtenus pour la compensation des réflexions de premier ordre sur les murs verticaux d'une pièce rectangulaire.

Dans un deuxième temps, nous montrons les résultats obtenus sur une configuration spécifique de haut-parleurs MAP. Le but est de compenser les réflexions mutuelles provoquées par la surface du haut-parleur lui même dans le cas où deux rangées de ces haut-parleurs se font face. Les résultats présentés ici sont issus des mesures effectuées par Rik van Zon dans la chambre anéchoïque de l'université de Delft en juillet 2003 auxquelles nous appliquons la méthode de compensation individuelle des réflexions.

6.2.1 Simulations

Paire de banc de haut-parleurs image/banc de haut-parleurs de compensation

Considérons une paire de banc de haut-parleurs image/banc de haut-parleurs de compensation, la même que celle utilisée dans la partie 3.2.1. Les bancs de haut-parleurs comportent chacun 40 monopoles espacés de 15 cm pour une longueur totale de 6 m. Ils sont situés sur deux lignes parallèles distantes de 5 m.

La source Ψ_{im} est synthétisée par le banc "image" LS_{im} en utilisant la méthode d'égalisation multicanal avec un banc de microphones similaire à Mic_{comp} , situé à 2 m de LS_{im} . La compensation de la source Ψ_{im} synthétisée par LS_{im} est réalisée par le banc "de compensation" LS_{comp} par la méthode de compensation individuelle des réflexions. Le banc de microphones de contrôle Mic_{comp} est situé à 2 m de LS_{comp} .

On utilise trois sources de test :

- une source virtuelle située 1 m devant LS_{im} et centrée,
- une source virtuelle située 3 m derrière LS_{im} et à 3 m sur la droite,
- une onde plane se propageant perpendiculairement à LS_{im} et LS_{comp} .

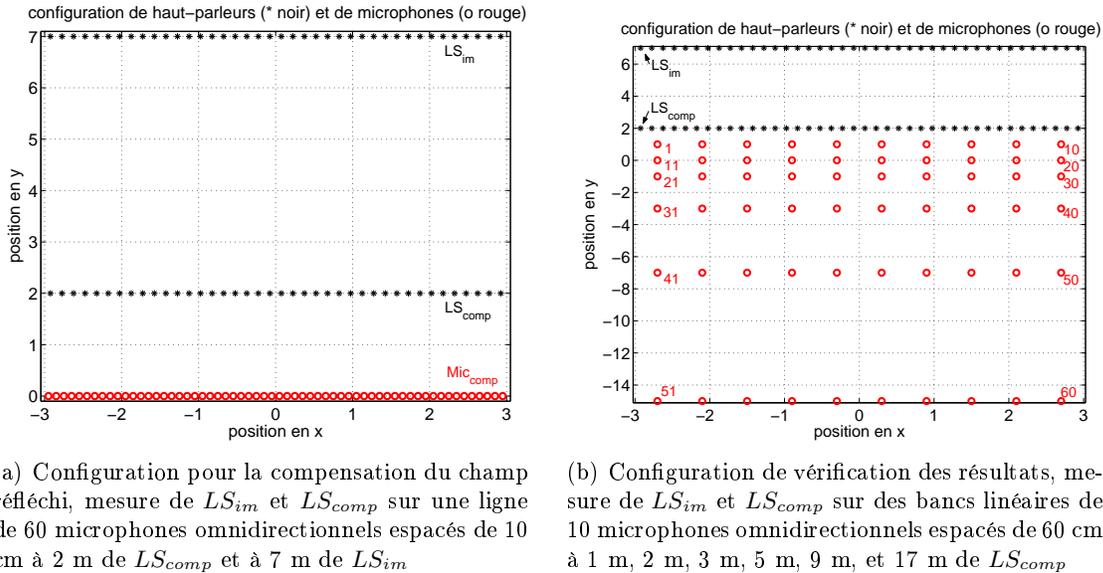
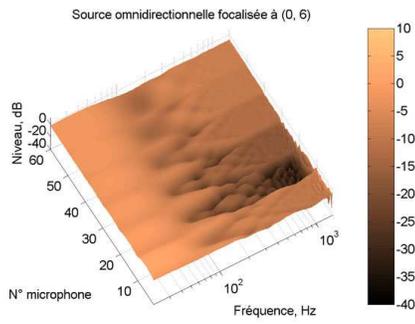


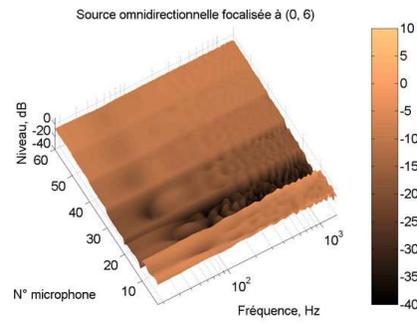
FIG. 6.4 – Configurations pour la réduction de niveau des réflexions

Les figures 6.5 représentent la réduction relative de niveau en fonction de la fréquence (cf. partie 3.2.1). Toutes les figures de gauche (cf. figures 6.5(a), 6.5(c), 6.5(e)) représentent la réduction relative de niveau obtenue pour des sources virtuelles Ψ_{im} et Ψ_{comp} synthétisées par la Wave Field Synthesis et normalisées en niveau en $(0, 0)$. Elles constituent une sorte de référence. Toutes les figures de droite (cf. figures 6.5(b), 6.5(d), 6.5(f)) représentent quant à elles la réduction relative de niveau obtenue en synthétisant Ψ_{im} par l'égalisation multicanal et Ψ_{comp} par la compensation individuelle des réflexions.

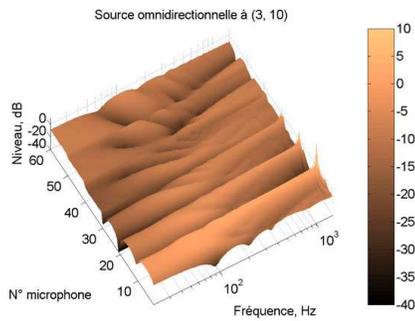
La méthode de compensation individuelle des réflexions permet de réduire le niveau des réflexions de manière uniforme en fonction de la fréquence. Cette méthode compense ainsi les erreurs dues à la diffraction de la Wave Field Synthesis. Nous avons en effet remarqué que le champ diffracté émis à l'intérieur de la zone de visibilité des sources virtuelles est différent pour Ψ_{im} et Ψ_{comp} , ce qui dégrade les performances de la réduction de niveau des réflexions, ce, principalement en basses fréquences (cf. partie 3.2.1). La réduction de niveau maximale obtenue se trouve bien pour la position du banc de microphones Mic_{comp} servant pour l'inversion multicanal (microphones 11 à 20 dans les figures 6.5).



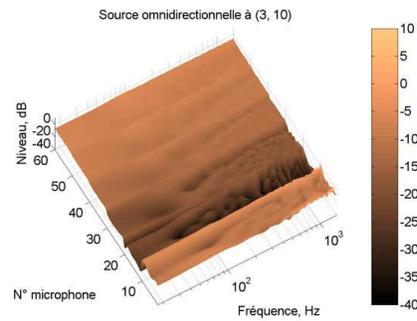
(a) Wave Field Synthesis, source omnidirectionnelle située 1 m devant LS_{im}



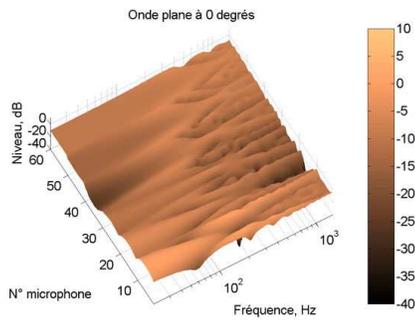
(b) Compensation individuelle de la réflexion par inversion multicanal, source omnidirectionnelle située 1 m devant LS_{im}



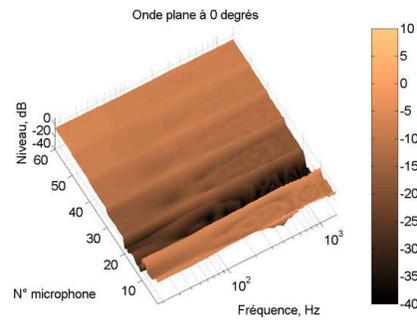
(c) Wave Field Synthesis, source omnidirectionnelle située 3 m derrière LS_{im} et décalée de 3 m sur la droite



(d) Compensation individuelle de la réflexion par inversion multicanal, source omnidirectionnelle située 3 m derrière LS_{im} et décalée de 3 m sur la droite



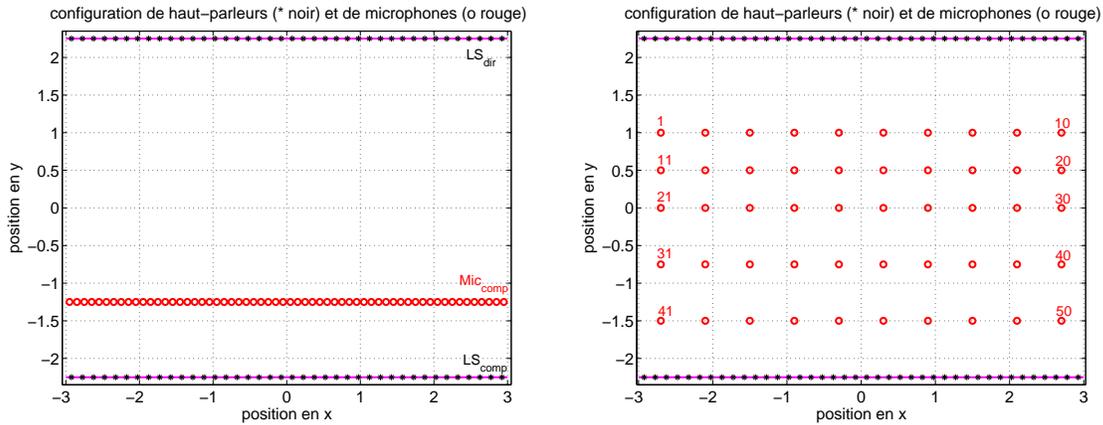
(e) Wave Field Synthesis, onde plane se propageant perpendiculairement à LS_{im} et LS_{comp}



(f) Compensation individuelle de la réflexion par inversion multicanal, onde plane se propageant perpendiculairement à LS_{im} et LS_{comp}

FIG. 6.5 – Réduction relative de niveau de la réflexion en fonction de la fréquence.

Deux parois parallèles



(a) Configuration pour la compensation du champ réfléchi, 1 ligne de microphones cardioïdes espacés de 10 cm situées à 1 m du banc de haut-parleurs arrière

(b) Configuration de vérification des résultats, simulation de la mesure de l'ensemble des haut-parleurs sur cinq lignes de 10 microphones omnidirectionnels espacés de 60 cm pour $y = 1$ m, 0.5 m, 0 m, -0.75 m, -1.5 m.

FIG. 6.6 – Configuration de haut-parleurs et de microphones pour la compensation du champ réfléchi sur la paroi arrière.

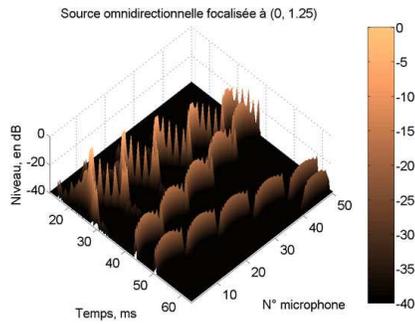
Considérons maintenant deux parois parallèles verticales infinies parfaitement réfléchissantes. Dans chaque paroi, est intégré un banc de 40 haut-parleurs omnidirectionnels idéaux (cf. figure 6.6(a)). Ces bancs de haut-parleurs sont ainsi les mêmes que ceux considérés dans la partie précédente. Les parois sont séparées de 4,5 m. Le banc de haut-parleurs LS_{dir} , intégré dans la paroi "avant", synthétise une source Ψ par la méthode d'égalisation multicanal. Le champ émis se réfléchit alors sur la paroi "arrière" en synthétisant une source virtuelle image Ψ_{im} qui à son tour se réfléchit sur la paroi "arrière" et ainsi de suite...

On considère les mêmes sources virtuelles que dans la partie précédente :

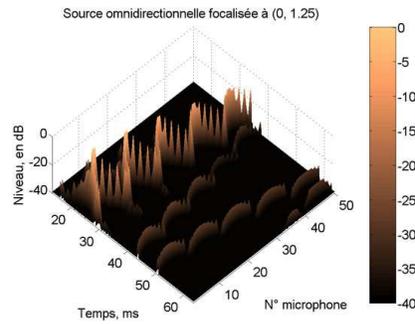
- une source virtuelle située 1 m devant LS_{dir} , centrée ;
- une source virtuelle située 3 m derrière LS_{dir} , à 3 m sur la droite ;
- une onde plane se propageant perpendiculairement à LS_{dir} et LS_{comp} .

L'ensemble des figures 6.7 présente les réponses impulsionnelles simulées au niveau de 5 lignes de microphones omnidirectionnels (cf. figure 6.6(b)) situées dans le même plan horizontal que LS_{dir} et LS_{comp} . Le champ réfléchi est simulé par la méthode des sources images. Les réponses impulsionnelles représentées ont subi un filtrage passe bas avec une fréquence de coupure inférieure à la fréquence d'aliasing pour toutes les positions de microphones. La fréquence de coupure est indiquée pour chaque figure.

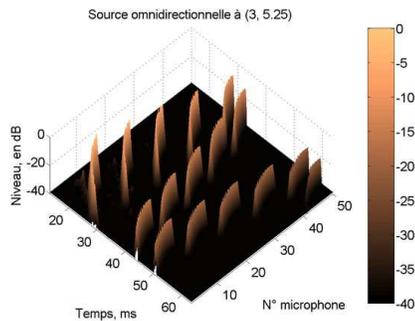
Les figures de gauche (cf. figures 6.7(a), 6.7(c), 6.7(e)) représentent les réponses impulsionnelles obtenues lorsque seul LS_{im} est actif. Il s'agit donc d'une situation *sans compensation*. Les figures de droite (cf. figures 6.7(b), 6.7(d), 6.7(f)) représentent au contraire les réponses impulsionnelles obtenues lorsque LS_{im} et LS_{comp} sont utilisés (situation *avec compensation*).



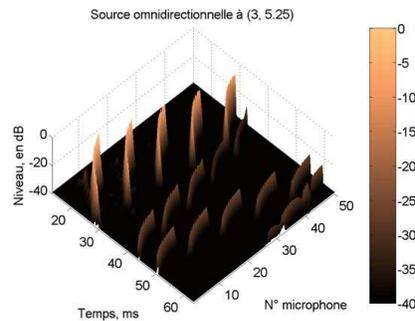
(a) *Sans* compensation, source omnidirectionnelle située 1 m devant le banc image , fréquence de coupure du filtre passe-bas : 1000 Hz



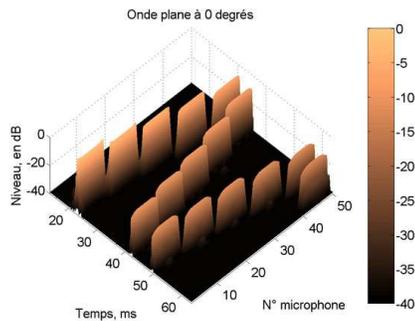
(b) *Avec* compensation, source omnidirectionnelle située 1 m devant le banc image , fréquence de coupure du filtre passe-bas : 1000 Hz



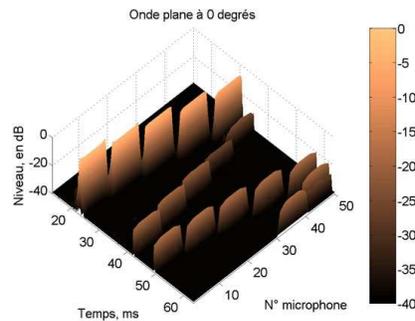
(c) *Sans* compensation, source omnidirectionnelle située 3 m derrière le banc image et décalée de 3 m sur la droite , fréquence de coupure du filtre passe-bas : 1000 Hz



(d) *Avec* compensation, source omnidirectionnelle située 3 m derrière le banc image et décalée de 3 m sur la droite , fréquence de coupure du filtre passe-bas : 1000 Hz



(e) *Sans* compensation, onde plane se propageant perpendiculairement au banc image , fréquence de coupure du filtre passe-bas : 1500 Hz

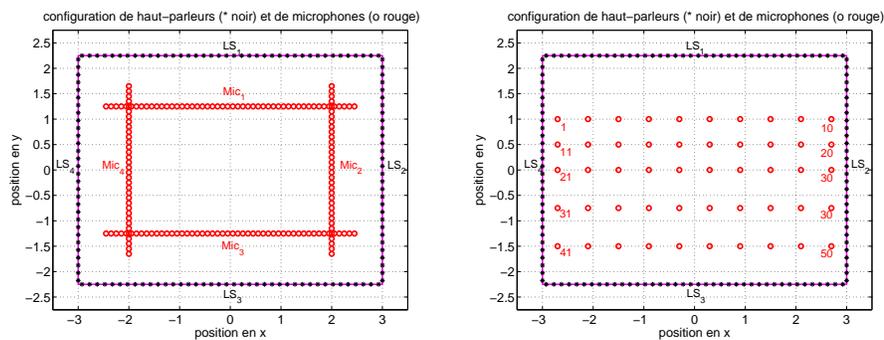


(f) *Avec* compensation, onde plane se propageant perpendiculairement au banc image , fréquence de coupure du filtre passe-bas : 1500 Hz

FIG. 6.7 – Réponses impulsionnelles, filtrées passe bas (fréquence de coupure indiquée sur chaque figure), du système de reproduction entre les deux parois (cf. figure 6.8(b)), avec ou sans compensation.

La réduction du niveau des réflexions est la plus efficace, indépendamment de la source virtuelle considérée, pour les positions d'écoute arrière (microphones 41 à 50). Il s'agit en effet des positions de microphone les plus proche du banc de microphone Mic_{comp} utilisé pour la compensation. Ceci se voit, en particulier pour la réflexion de premier ordre, où la réduction atteint 25 à 30 dB. Ce résultat est particulièrement intéressant car c'est à ces positions d'écoute que la première réflexion arrière est la plus importante avec un niveau similaire à celui du son direct et une différence de temps d'arrivée relativement faible ($\approx 2-3$ ms). Ce type de réflexion risque de provoquer une coloration excessive du son direct. Celle-ci peut ainsi être évitée en utilisant la méthode de compensation individuelle des réflexions. La compensation des réflexions d'ordre 2 (provenant de la paroi avant) est moins importante que celle des réflexions d'ordre 1 avec une réduction de niveau de l'ordre de 5 à 10 dB. Remarquons toutefois que le temps d'arrivée relatif de ces contributions par rapport au son direct est de l'ordre de 25 à 30 ms. Celui-ci ne dépend que faiblement de la position d'écoute considérée, contrairement aux réflexions de premier ordre. Ce temps d'arrivée relatif est du même ordre que le seuil perceptif à partir duquel les réflexions précoces ne contribuent plus directement à la coloration du son direct (Rubak, 2004). La méthode de compensation individuelle des réflexions permet ainsi de "nettoyer" la première partie de la réponse impulsionnelle pour la plupart des positions d'écoute entre les deux parois. Elle réduit, de manière uniforme, la coloration du son direct apportée par le régime de réflexions avant/arrière.

Compensation des réflexions d'une pièce complète



(a) Configuration pour la compensation du champ réfléchi, 4 lignes de microphones cardioïdes espacés de 10 cm situées à 1 m en face de chaque banc de haut-parleurs

(b) Configuration de vérification des résultats, simulation de la mesure de l'ensemble des haut-parleurs sur cinq lignes de 10 microphones omnidirectionnels espacés de 60 cm pour $y = 1$ m, 0.5 m, 0 m, -0.75 m, -1.5 m

FIG. 6.8 – Atténuation des réflexions pour une pièce complète

Ajoutons maintenant deux parois supplémentaires, parfaitement réfléchissantes, verticales, sur les côtés afin de "fermer" la salle (cf. figure 6.8(a)). Dans chacune de ces parois, on intègre un banc de 30 haut-parleurs omnidirectionnels idéaux espacés de 15 cm (LS_2 , LS_3 , et LS_4 sur la

figure 6.8(a)). En face de chacun des bancs de haut-parleurs, on place des bancs de microphone (Mic_2 , Mic_3 , et Mic_4 sur la figure 6.8(a)) à directivité cardioïde orientés vers la paroi à laquelle ils es rapportent.

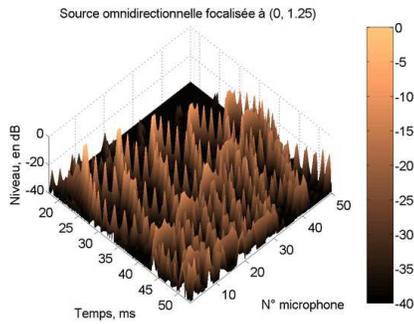
Les sources tests de la partie précédente sont synthétisées par le banc de haut-parleurs LS_1 à l'aide de filtres obtenues par la méthode d'inversion multicanal (cf. chapitre 5).

La compensation des réflexions de premier ordre est réalisée en utilisant la méthode de compensation individuelle des réflexions. On considère successivement les paires banc de haut-parleurs de compensation/banc de microphones de contrôle LS_2/Mic_2 , LS_3/Mic_3 , et LS_4/Mic_4 . Pour chaque paire, le champ cible pour le calcul des filtres correspond à l'opposé du champ réfléchi au premier ordre sur la paroi considérée.

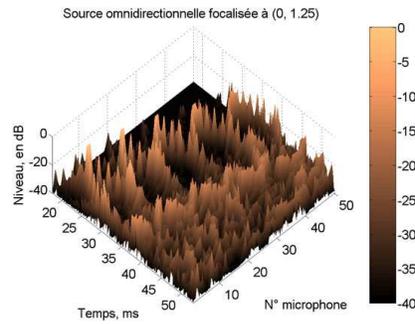
Comme pour les résultats de la partie précédente, l'ensemble des figures 6.9 présente les réponses impulsionnelles simulées au niveau de 5 lignes de microphones omnidirectionnels (cf. figure 6.8(b)) situées dans le même plan horizontal que les bancs de haut-parleurs et de microphones. Le champ réfléchi est simulé par la méthode des sources images. Les réponses impulsionnelles ont subi un filtrage passe bas avec une fréquence de coupure inférieure à la fréquence d'aliasing pour toutes les positions de microphone considérées. La fréquence de coupure utilisée est indiquée sur chaque figure.

Les figures de gauche (cf. figures 6.9(a), 6.9(c), 6.9(e)) représentent les réponses impulsionnelles obtenues lorsque seul LS_1 est utilisé (sans compensation). Les figures de droite (cf. figures 6.9(b), 6.9(d), 6.9(f)) représentent au contraire les réponses impulsionnelles obtenues *avec* compensation des réflexions précoces par LS_2 , LS_3 , et LS_4 .

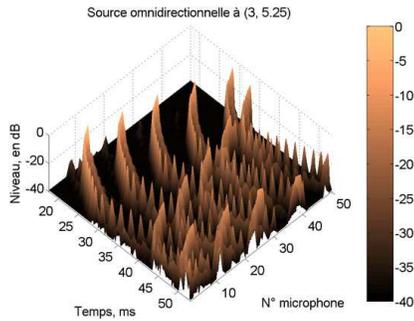
Comme pour la partie précédente, la méthode de compensation individuelle des réflexions permet de diminuer de façon significative le niveau des réflexions précoces dans la première partie de la réponse impulsionnelle (25/30 ms) et de réduire ainsi les risques de coloration du son direct dans l'*ensemble* de la zone d'écoute. Des "erreurs" subsistent pour les positions extrêmes de chaque banc de microphones (microphones 1/10/11/20/21/30/31/40/41/50). Ces positions ne sont en effet pas situées dans la zone d'écoute définie par les bancs de microphones utilisés pour la compensation des réflexions précoces. On ne peut alors garantir, pour ces positions, la validité des résultats.



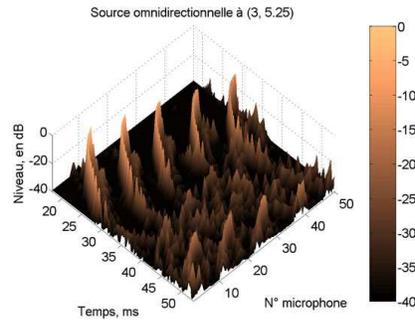
(a) *Sans* compensation, source omnidirectionnelle située 1 m devant le banc image , fréquence de coupure du filtre passe-bas : 1000 Hz



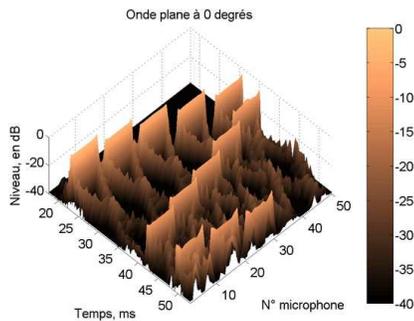
(b) *Avec* compensation, source omnidirectionnelle située 1 m devant le banc image , fréquence de coupure du filtre passe-bas : 1000 Hz



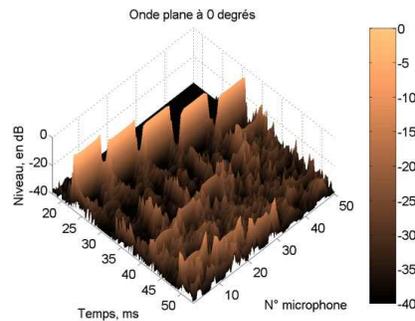
(c) *Sans* compensation, source omnidirectionnelle située 3 m derrière le banc image et décalée de 3 m sur la droite , fréquence de coupure du filtre passe-bas : 1000 Hz



(d) *Avec* compensation, source omnidirectionnelle située 3 m derrière le banc image et décalée de 3 m sur la droite , fréquence de coupure du filtre passe-bas : 1000 Hz



(e) *Sans* compensation, onde plane se propageant perpendiculairement au banc image , fréquence de coupure du filtre passe-bas : 1500 Hz



(f) *Avec* compensation, onde plane se propageant perpendiculairement au banc image , fréquence de coupure du filtre passe-bas : 1500 Hz

FIG. 6.9 – Réponses impulsionnelles, filtrées passe bas (fréquence de coupure indiquée sur chaque figure), du système de reproduction dans la pièce de restitution (cf. figure 6.8(b)), avec ou sans compensation.

L'ensemble des figures 6.10 précise le traitement effectué pour la source omnidirectionnelle située trois mètres derrière LS_1 et à trois mètres sur la droite. Cette source est ainsi à la limite du mur dans lequel est intégré LS_2 et un respect scrupuleux des critères de visibilité des sources virtuelles nous inciterait à ne pas utiliser LS_2 pour la compensation des réflexions précoces. Les figures 6.10(b), 6.10(c), 6.10(d) représentent les réponses impulsionnelles du système sur un banc de 20 microphones omnidirectionnels espacés de 30 cm situé au centre de la pièce, parallèle à LS_1 et LS_3 (à une position similaire à celle des microphones 21 à 30 dans la figure 6.8(b)). Le champ réfléchi est simulé par la méthode des sources images uniquement jusqu'à l'ordre 2. Les réponses impulsionnelles représentées ont subi un filtrage passe bas avec une fréquence de coupure inférieure de 1000 Hz.

La figure 6.10(b) représente les réponses impulsionnelles obtenues lorsque seul LS_1 est utilisé (sans compensation). On se rend compte que le champ diffracté qui se réfléchit sur le mur de droite est non négligeable et synthétise un second front d'onde qui suit immédiatement le son direct. De la même manière, la réflexion sur le mur arrière et le mur droit "ramène" le champ diffracté juste derrière la réflexion sur le mur du fond. Dans le cas où l'on n'utilise que LS_3 et LS_4 pour la compensation (cf. figure 6.10(c)), le niveau des réflexions sur le mur du fond et le mur de droite est réduit, mais on remarque que les réflexions du champ diffracté restent identiques. Celles-ci sont réduites, de manière significative, en appliquant la méthode de compensation individuelle des réflexions sur la paire LS_2/Mic_2 (cf. figure 6.10(d)).

Remarquons, que pour l'onde plane, la situation est similaire. On aurait dû en effet ne considérer que la paire LS_3/Mic_3 et ne pas utiliser les haut-parleurs sur les côtés. L'utilisation des paires LS_2/Mic_2 et LS_4/Mic_4 permet alors de diminuer significativement la diffraction, notamment juste derrière le front d'onde principal.

Les figures 6.11 représentent les réponses impulsionnelles du système, obtenues, pour différentes méthodes de compensation de l'effet de salle, au niveau des microphones de la figure 6.8(b). Le champ réfléchi est simulé par la méthode des sources images. Les réponses impulsionnelles représentées ont subi un filtrage passe bas avec une fréquence de coupure de 1000 Hz.

La figure 6.11(a) représente les réponses impulsionnelles obtenues sans compensation, lorsque seul LS_1 est utilisé. Les autres figures représentent les réponses impulsionnelles obtenues par :

- inversion multicanal sur l'ensemble des haut-parleurs et des microphones (synthèse du son direct et compensation du champ réfléchi en une seule étape) à partir de la matrice des fonctions de transfert mise en forme par les retards donnés par la WFS (évolution de la méthode proposée par l'auteur dans (Corteel & Nicol, 2003), figure 6.11(b)) ;
- la méthode van Zon/Corteel (figure 6.11(c)) ;
- la méthode de compensation individuelle des réflexions présentée dans ce document (figure 6.11(d)).

L'acoustique de la pièce de restitution est simulée par la méthode des sources images jusqu'à l'ordre 3. On voit ainsi que parmi les trois méthodes représentées, la méthode de compensation individuelle des réflexions est celle qui introduit le moins d'artefacts dans la première partie de la réponse impulsionnelle. Ceci permet d'assurer la transparence de la reproduction et d'éviter

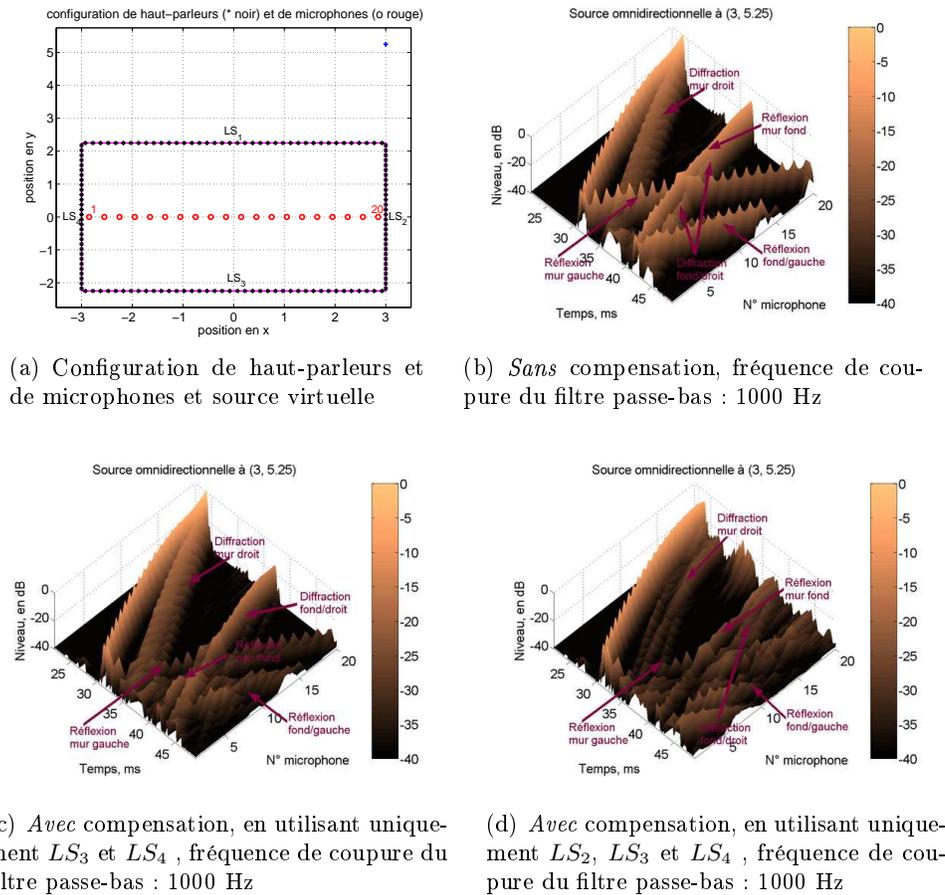


FIG. 6.10 – Réponses impulsionnelles, filtrées passe bas (fréquence de coupure indiquée sur chaque figure), du système de reproduction dans la pièce de reproduction sur le banc de microphone central (microphones 21 à 30, cf. figure 6.8(b)), avec ou sans compensation. Seules les sources images du plan horizontal jusqu'à l'ordre 2 sont considérées. Source virtuelle omnidirectionnelle située 3 m derrière LS_1 et décalée de 3 m sur la droite.

des colorations supplémentaires amenées par les autres méthodes.

La décomposition du problème permet ainsi, non seulement de réduire le coût de calcul de l'inversion multicanal, mais aussi de réaliser une compensation plus nette des réflexions précoces.

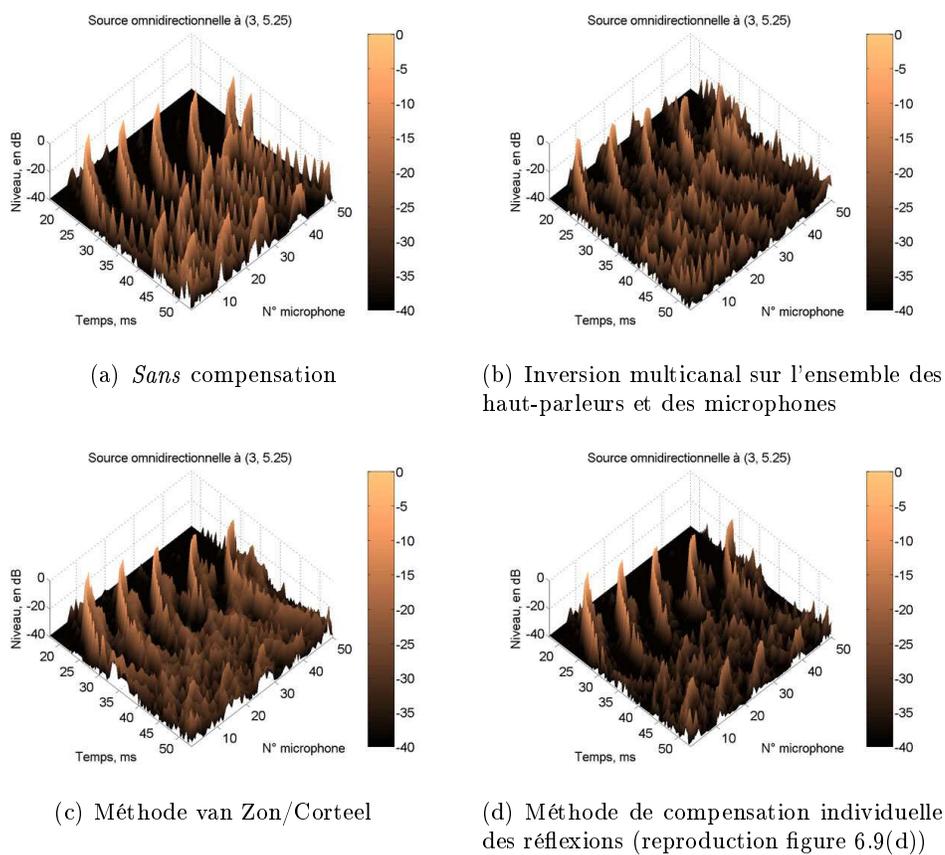


FIG. 6.11 – Réponses impulsionnelles, filtrées passe bas (fréquence de coupure : 1000 Hz), du système de reproduction dans la pièce de reproduction (cf. figure 6.8(b)), avec ou sans compensation.

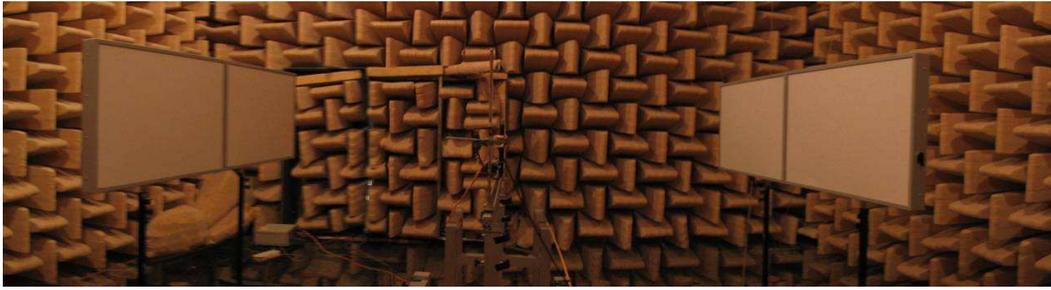


FIG. 6.12 – Configuration de mesure dans la chambre anéchoïque de l’université de Delft

6.2.2 Résultats avec des haut-parleurs MAP

Dans cette partie, on montre les résultats de la procédure d’inversion multicanal pour la compensation des réflexions mutuelles de deux rangées de 2 haut-parleurs MAP situées face à face. Ces haut-parleurs étant constitués de surfaces rigides de taille relativement importante (1,35 de largeur par 70 cm de hauteur), ils se comportent comme des réflecteurs acoustiques. Les haut-parleurs *forment* les parois sur lesquelles se réfléchit le son direct et sont par conséquent naturellement intégrés dans les parois réfléchissantes.

Le problème étudié, et qui a fait l’objet du travail de fin d’étude du Master of Science de Rik van Zon à l’université de Delft (van Zon, 2003), consiste à essayer de réduire les réflexions mutuelles de 2 rangées de 2 haut-parleurs MAP se faisant face. Les résultats présentés ici sont déduits des mesures effectuées pendant l’été 2003 dans la chambre anéchoïque de l’université de Delft en appliquant la méthode de compensation individuelle des réflexions.

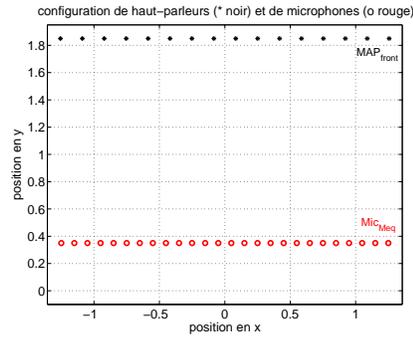
Mesures

La configuration de test est composée de 2 rangées (MAP_{front} et MAP_{back}) de 2 haut-parleurs MAP qui se font face et sont espacées de 3,7 m. La rangée MAP_{front} sert à synthétiser le son direct d’une source virtuelle qui se réfléchit sur la paroi que constitue MAP_{back} . On cherche à contrôler la synthèse de la source virtuelle de compensation Ψ_{comp} avec MAP_{back} .

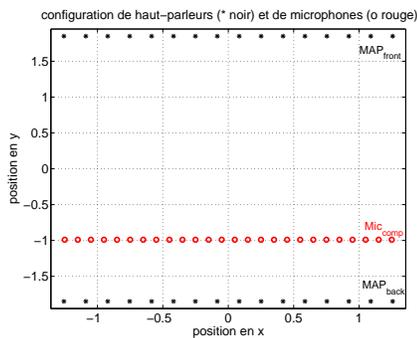
Cette configuration a été placée dans la chambre anéchoïque de l’université de Delft afin de pouvoir caractériser *uniquement* les réflexions mutuelles entre les deux rangées (cf. figure 6.12); et d’éviter d’incorporer dans les mesures des réflexions provenant des autres parois de la salle.

Trois séries de mesures ont été effectuées par Rik van Zon sur la configuration de MAPs pour caractériser :

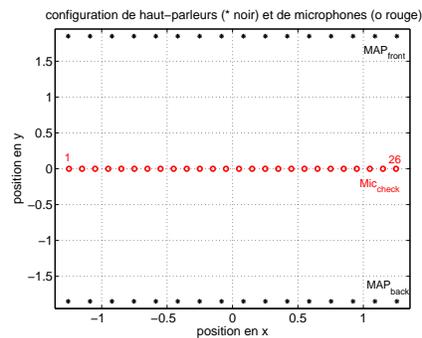
1. le son direct de la rangée MAP_{front} frontale de MAPs pour la synthèse du son direct par égalisation multicanal, mesure des 16 canaux MAP_{front} par 26 microphones omnidirectionnels espacés de 10 cm à 1,5 m de MAP_{front} (cf. figure 6.13(a));
2. les réflexions mutuelles entre les deux rangées de MAP pour l’inversion multicanal, mesure des 32 canaux sur une ligne Mi_{comp} de 26 microphones cardioïdes espacés de 10 cm à 86 cm de MAP_{back} et orientés vers MAP_{back} (cf. figure 6.13(b));



(a) Configuration pour l'égalisation multicanal de MAP_{front} , mesure de la rangée frontale de MAP par 26 microphones omnidirectionnels espacés de 10 cm à 1,5 m des MAP



(b) Configuration pour la compensation des réflexions mutuelles, mesure de MAP_{front} et MAP_{back} sur une ligne de 26 microphones cardioïdes Mic_{comp} espacés de 10 cm à 86 cm de MAP_{back} et orientés vers MAP_{back}



(c) Configuration centrale, mesure de MAP_{front} et MAP_{back} par 26 microphones omnidirectionnels espacés de 10 cm situés au centre de la zone d'écoute

FIG. 6.13 – Configurations de mesure des haut-parleurs MAP

3. le champ complet synthétisé (son direct+réflexions) entre les deux rangées de MAP pour la vérification des résultats, mesure des 32 canaux par Mic_{check} , une ligne de 26 microphones omnidirectionnels espacés de 10 cm situés au centre de la zone d'écoute (cf. figure 6.13(c)).

Les mesures ont été réalisées à l'aide d'un unique microphone monté sur un rail motorisé qui permet un déplacement automatique entre chaque série de mesures. Les microphones utilisés sont :

- soit un B& K 4006 omnidirectionnel (mesures 1 et 3),
- soit un B& K 4011 cardioïde (mesure 2).

Leurs caractéristiques de directivité peuvent être considérées comme idéales dans la bande de fréquence qui nous intéresse (100-16000 Hz).

A partir de la deuxième série de mesures (cf. figure 6.13(b)), on extrait les matrices des fonctions de transfert $C_{Dir}^{back}(z)$ et $C_{Ref1}^{front}(z)$ qui correspondent, respectivement, au son direct des canaux

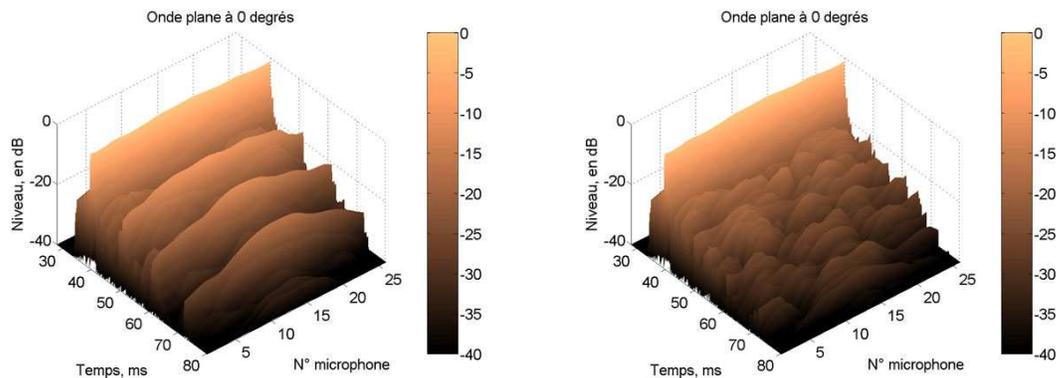
de MAP_{back} et du champ réfléchi (première réflexion uniquement) des canaux de MAP_{front} sur MAP_{back} . Cette opération s'effectue simplement en sélectionnant la partie utile des réponses impulsionnelles mesurées après estimation des temps de propagation de chacune des composantes. L'absence de réflexion sur les parois de la pièce et l'utilisation d'un microphone à directivité cardioïde permettent une extraction précise de ces composantes. On applique ensuite la méthode de compensation individuelle des réflexions pour la paire MAP_{back}/Mic_{comp} .

Résultats

Trois sources test sont utilisées :

- une onde plane se propageant perpendiculairement aux deux rangées de MAP, onde plane à 0 degré ;
- une source omnidirectionnelle à (1, 4.85), située 3 m derrière MAP_{front} et décalée de 1 m sur la droite ;
- une source focalisée à (0, 1.05) située 80 cm devant MAP_{front} .

Les figures 6.14, 6.15, et 6.16 représentent les résultats de la procédure de compensation. On simule la réponse du système à partir des mesures effectuées sur la configuration 3 (cf. figure 6.13(c)). Chaque groupe de figures représente les réponses impulsionnelles du système, simulées sur l'ensemble des positions des microphones de la configuration 3. On a appliqué aux réponses impulsionnelles un filtrage passe bas (fréquence de coupure inférieure à la fréquence d'aliasing). La figure de gauche présente les réponses impulsionnelles du système en utilisant que la rangée MAP_{front} , situation de référence *sans* compensation. Les figures de droite présentent les réponses impulsionnelles du système complet (MAP_{front} et MAP_{back}), situation *avec* compensation.

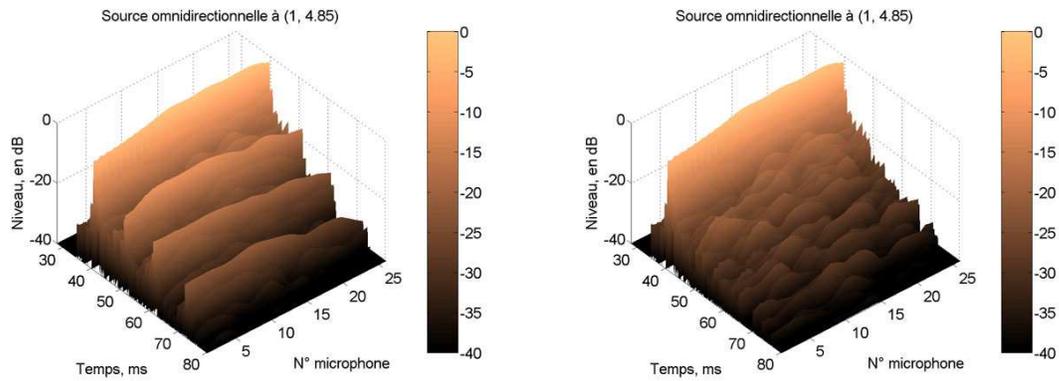


(a) Réponses impulsionnelles filtrées passe bas, fréquence de coupure correspond à la fréquence d'aliasing, sans compensation des réflexions mutuelles

(b) Réponses impulsionnelles filtrées passe bas, fréquence de coupure correspond à la fréquence d'aliasing, avec compensation des réflexions mutuelles

FIG. 6.14 – Résultats de la compensation sur la configuration centrale pour une onde plane se propageant perpendiculairement au bancs de haut-parleurs (onde plane à 0 degré)

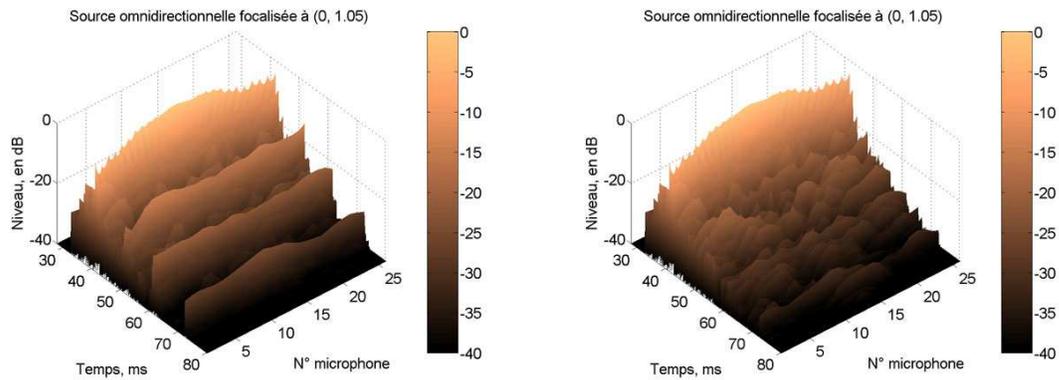
Pour les trois sources virtuelles considérées, on remarque une diminution très nette du niveau du champ réfléchi sur MAP_{back} , qui passe dans tous les cas sous le niveau de diffusion résiduelle



(a) Réponses impulsionnelles filtrées passe bas, fréquence de coupure correspond à la fréquence d'aliasing, sans compensation des réflexions mutuelles

(b) Réponses impulsionnelles filtrées passe bas, fréquence de coupure correspond à la fréquence d'aliasing, avec compensation des réflexions mutuelles

FIG. 6.15 – Résultats de la compensation sur la configuration centrale pour une source omnidirectionnelle située 3 m derrière la rangée frontale de MAP et décalée de 1 m sur la droite, source à (1, 4.85)



(a) Réponses impulsionnelles filtrées passe bas, fréquence de coupure correspond à la fréquence d'aliasing, sans compensation des réflexions mutuelles

(b) Réponses impulsionnelles filtrées passe bas, fréquence de coupure correspond à la fréquence d'aliasing, avec compensation des réflexions mutuelles

FIG. 6.16 – Résultats de la compensation sur la configuration centrale pour une source omnidirectionnelle focalisée située 80 cm devant la rangée frontale de MAP, source à (0, 1.05)

caractéristique des MAPs.

6.3 Discussion

Dans ce chapitre, nous avons présenté 4 méthodes de compensation des réflexions précoces de la salle de restitution pour la WFS par des méthodes d'égalisation multicanal. 3 de ces méthodes ont été proposées par l'auteur en collaboration avec Rozenn Nicol de France Télécom R & D (Corteel & Nicol, 2003) ou avec Rik van Zon de l'université de Delft (van Zon *et al.*, 2004). La quatrième correspond aux travaux de Spors *et al.* à l'université d'Erlangen en Allemagne.

Le but commun de ces méthodes est de contrôler le champ acoustique synthétisé dans le plan horizontal par une distribution de haut-parleurs situés dans le plan horizontal. Elles diffèrent par la nature de la description du MIMO, l'objectif de reproduction, le type d'algorithme d'inversion multicanal mis en œuvre et la longueur des filtres obtenus.

Parmi les méthodes que nous avons proposé, la méthode originale de compensation individuelle des réflexions précoces semble la plus adaptée à la compensation des réflexions spéculaires du son direct sur les parois *verticales* de la pièce de restitution dans le plan horizontal et sous la fréquence d'aliasing spatial. Elle permet de contrôler la synthèse des sources virtuelles de compensation par la Wave Field Synthesis (cf. partie 3.2.4).

Cette méthode est aussi une extension directe de l'égalisation multicanal proposée au chapitre précédent puisqu'elle réalise un contrôle du rayonnement "en champ libre" des bancs de haut-parleurs de compensation intégrés dans chaque paroi de la salle de restitution. Ceux-ci sont chargés de synthétiser une source de compensation dont les caractéristiques de rayonnement sont données par le champ acoustique réfléchi au premier ordre sur la paroi considérée.

Contrairement aux méthodes de compensation de l'effet de salle traditionnelles ou à celle proposée par Spors *et al.*, l'utilisation de la méthode proposée pour un rendu en temps-réel est relativement peu coûteuse. Les filtres calculés en temps différé sont en effet de longueur relativement courte (1000 points à 48 *kHz*) et associés à des retards qui tiennent compte du temps de propagation naturel des réflexions. Les filtres calculés sont de la même taille que ceux obtenus par la méthodes d'égalisation multicanal proposée au chapitre 5. Seul change le nombre de haut-parleurs mis en œuvre pour la synthèse d'une source virtuelle donnée. Dans le cas de la méthode de compensation individuelle des réflexions, tous les haut-parleurs du dispositif sont utilisés simultanément et pas uniquement ceux qui contribuent au son direct.

L'adaptation aux conditions atmosphériques (prise en compte de la modification de la vitesse du son) se fait simplement par ajustement des retards utilisés comme nous l'avons montré dans (van Zon *et al.*, 2004).

La méthode proposée héritant à la fois de la synthèse des sources de compensation par la WFS et de la méthode d'égalisation multicanal est aussi sujette à leurs défauts. En particulier, on ne maîtrise absolument pas ce qui se passe en dehors du plan horizontal. Comme pour la synthèse des sources de compensation par la WFS, les performances se dégradent en dehors du plan horizontal à cause de l'utilisation de bancs de haut-parleurs linéaires. Comme pour la méthode d'égalisation multicanal, ceci est de plus conditionné par les caractéristiques de directivité des haut-parleurs utilisés. Enfin, les conséquences de cette méthode sur la réverbération naturelle de

la salle de restitution restent à estimer.

Une évaluation perceptive de la méthode serait donc nécessaire pour déterminer dans quels cas elle s'avère bénéfique.

Au vu de toutes ces limitations, nous recommandons la plus grande prudence dans son utilisation. C'est pourquoi nous nous sommes attachés à réaliser la compensation des réflexions de la manière la plus propre et la simple possible en décomposant le problème au maximum. On peut en effet faire le reproche aux procédures d'inversion multicanal "aveugles" d'assurer un résultat le meilleur possible au niveau des microphones de contrôle sans s'assurer de la validité de la solution en dehors de ces points.

Troisième partie

Mise en œuvre et validation

Au cours des deux parties précédentes, nous sommes attachés à décrire les caractéristiques du champ sonore synthétisé en WFS par une distribution linéaire de haut-parleurs en champ libre et en espace clos. Nous avons aussi défini un ensemble de méthodes d'égalisation du champ sonore qui permettent de garantir, dans une certaine mesure, la transparence de la reproduction dans l'ensemble de la zone d'écoute.

Cette troisième et dernière partie est consacrée à l'utilisation pratique de la WFS dans un contexte de création et de restitution de scènes sonores spatialisées.

Nous proposons une mise en œuvre des techniques développées dans le reste du document au sein d'une chaîne de production sonore réalisée en collaboration avec *sonic emotion ag* à la suite du projet *Carrouso*.

Avant de rentrer dans les détails de la réalisation pratique, il est nécessaire de vérifier la possibilité de créer et manipuler une perspective sonore malgré les limitations inhérentes à la Wave Field Synthesis (loi d'atténuation non conforme à celle de la source, diffraction, aliasing spatial). Ce point fera l'objet du chapitre 7.

Nous divisons notre étude en deux parties. La première vise à objectiver les défauts potentiels dans la localisation des sources sonores synthétisées en WFS. Nous mettons en évidence des problèmes potentiels, principalement au-dessus de la fréquence d'aliasing. La deuxième partie de ce chapitre est consacrée à la présentation d'un test d'écoute interactif réalisé auprès d'ingénieurs du son. Les résultats de ce test nous permettront de valider la possibilité de créer et de manipuler une perspective sonore dans le cadre de la WFS malgré ses limitations.

Le chapitre 8 est consacré à la création, à la manipulation, et à la restitution de scènes sonores spatialisées pour la WFS. Nous donnons dans un premier temps les modalités de synthèse d'une scène sonore spatialisée en temps réel sur un dispositif WFS. Il s'agit notamment de décrire la synthèse d'un effet de salle additionnel en WFS. Notre analyse porte principalement sur la possibilité d'assurer un effet de salle cohérent dans l'ensemble de la zone d'écoute. Nous verrons qu'il existe cependant des limitations qui sont dues aux approximations réalisées dans la dérivation de la WFS.

Un autre point retiendra notre attention. Il s'agit de la persistance des effets en fonction de la taille et de la géométrie du dispositif de reproduction employé et de l'acoustique résiduelle de la salle de restitution. Ceci constitue une étape supplémentaire de prise en compte du dispositif de reproduction et de la pièce d'écoute. Nous mettrons principalement en évidence les problèmes potentiels et présenterons les solutions de la littérature applicables immédiatement ou bien qu'il serait nécessaire d'adapter.

Enfin, nous présentons la chaîne de production et de restitution proprement dite.

Chapitre 7

Validation perceptive de la perspective sonore en WFS

Dans l'introduction générale, nous avons abordé la notion de perspective sonore. La perspective sonore est caractéristique de la synthèse de scènes sonores en zone étendue. Elle traduit les variations naturelles de l'organisation spatiale de la scène sonore *perçue* en fonction de la position d'écoute.

Les limitations inhérents à la technique (loi d'atténuation des sources sonores non conforme, champ synthétisé incorrect au-dessus de la fréquence d'aliasing spatial) sont susceptibles de modifier le ressenti de cette perspective. L'objet de ce chapitre est donc de vérifier la possibilité de gérer cette perspective en WFS.

Ceci peut se faire en deux temps. Il s'agit, tout d'abord, d'estimer la validité des signaux binauraux transmis aux auditeurs et de voir dans quelle mesure ceux-ci donnent des informations en terme de localisation qui correspondent à celles de la source cible. Il faut, ensuite, déterminer quels sont les paramètres qui permettent de créer et de manipuler cette perspective sonore et dans quelle mesure ceux-ci sont utilisables par le créateur de contenu.

Dans un premier temps, nous nous plaçons du point de vue de l'auditeur. Nous analysons la validité des signaux binauraux restitués par la WFS. Notre analyse ne cherche pas à déterminer la localisation effective de l'événement auditif car les mécanismes de décisions mis en œuvre par le système auditif restent encore mal compris. Nous préférons estimer le degré de cohérence des signaux reçus au niveau des oreilles de l'auditeur avec "ce à quoi" on s'attendrait venant de la source cible. Cette estimation est réalisée par le calcul des *indices de localisation* à partir des signaux binauraux qui permettent de modéliser le traitement par le système auditif de la localisation d'une source.

Cette approche vise à "objectiver" la nature des erreurs de localisation potentielles. Nous verrons que les indices sont bien restitués en-dessous de la fréquence d'aliasing spatial. Au-dessus de cette fréquence, nous mettrons en évidence certaines incohérences entre les indices binauraux fournis par la WFS et ceux qu'aurait fournis la source cible.

Dans une deuxième partie, nous nous intéressons plus particulièrement à la création et à la manipulation de la perspective sonore. Nous cherchons à intégrer le maniement des paramètres qui permettent de créer l'impression de perspective aux tâches de mixage. Pour cela, nous donnons les résultats d'un test d'écoute interactif réalisé auprès d'ingénieurs du son auxquels est demandé de recréer une perspective sonore. Ce test permet effectivement de valider l'utilisation de ces paramètres et de vérifier a posteriori que, malgré ses limitations, la WFS permet bien de rendre la perspective. Par rapport aux situations traditionnelles, nous définirons un paramètre supplémentaire qui permet la manipulation de l'organisation spatiale des sources. Enfin, nous montrons la nécessité de développer des interfaces de contrôle qui permettent au créateur de contenu de se déplacer dans la zone d'écoute afin de vérifier la validité de la scène créée.

7.1 Validation objective de la localisation des sources en WFS

Dans la première partie de ce chapitre, nous proposons d'estimer les signaux binauraux restitués au niveau des oreilles des auditeurs et de voir dans quelle mesure ceux-ci sont susceptibles de donner une localisation correcte de l'événement auditif.

Après avoir rappelé les modalités de localisation des sources sonores par le système auditif par l'évaluation des indices binauraux, nous montrerons dans quelle mesure ceux-ci sont correctement perçus par l'auditeur pour la synthèse de sources virtuelles en WFS.

7.1.1 Localisation des sources sonores

La localisation angulaire des sources sonores par le système auditif humain se fait grâce à l'analyse des différences des signaux binauraux (signaux captés par les deux oreilles). Des expériences psychoacoustiques (voir par exemple la partie 2.1 de (Blauert, 1999)) réalisées dès le début du vingtième siècle ont permis de mettre en évidence l'exploitation de ces différences par le système auditif.

Dans cette partie nous mettons en évidence les différents mécanismes qui permettent d'estimer la localisation des sources sonores par comparaison entre les signaux binauraux. L'objectivation de ces différences est liée à la définition d'*indices binauraux de localisation* que nous présentons dans un premier temps. Nous donnerons ensuite les modalités d'extraction de ces indices à partir des signaux binauraux.

Indices de localisation des sources sonores

On distingue trois indices binauraux principaux qui permettent au système auditif d'estimer la localisation de l'événement auditif dans différentes régions fréquentielles :

1. la différence interaurale d'intensité (IID, Interaural Intensity Difference) qui n'apparaît qu'à partir de 500Hz (au-dessous, la tête est considérée comme un obstacle "transparent" en comparaison avec les longueurs d'ondes mises en jeu) et joue un rôle dominant dans la

localisation à partir de 1500 Hz ;

2. la différence interaurale de phase (IPD, Interaural Phase Difference) qui n'est perceptible que jusqu'à environ 1500 Hz (suivant la taille de la tête de l'individu), au delà, il y a ambiguïté entre plusieurs directions ;
3. la différence interaurale d'enveloppe (IED, Interaural Envelop Difference) qui permet de détecter au-dessus de 1500 Hz une différence de temps pour des signaux ayant une attaque brève ; cet indice est cependant considéré comme de relativement faible influence par rapport aux autres.

L'IPD et l'IED sont généralement regroupés sous le terme d'ITD (Interaural Time Difference), IPD en-dessous de 1500 Hz, IED au-dessus de 1500 Hz. Les données expérimentales de Kuhn (Kuhn, 1977) permettent d'estimer une forme algébrique pour ces indices :

$$IPD_{f \leq 1500 Hz} \simeq 3 \frac{a}{c} \sin \theta, \quad (7.1)$$

$$IED_{f \geq 1500 Hz} \simeq 2 \frac{a}{c} \sin \theta, \quad (7.2)$$

où, a est le rayon en m de la tête pour une modélisation sphérique de la tête et, c la vitesse du son dans l'air, et θ l'azimut. Cette approximation de l'ITD est justifiée dans le cas où la source sonore se situe dans le plan horizontal ou pour des valeurs faibles de site ; au delà, une modélisation en forme d'ovale de la tête donne une meilleure approximation (Larcher, 2001).

Cette définition montre que l'ITD permet une localisation de l'azimut des sources sonores tout en laissant une ambiguïté avant/arrière et sur le site, suivant une courbe appelée cône de confusion. Cette ambiguïté est partiellement levée par l'estimation de l'IID en fonction de la fréquence et/ou par des mouvements de la tête.

Les indices de localisation sont évalués après analyse tonotopique de la cochlée que l'on peut assimiler à un filtrage par bandes de fréquences. Plusieurs modélisations de ce filtrage ont été effectuées dont on retiendra le formalisme des bandes rectangulaires équivalentes (ERB, Equivalent Rectangular Bandwidth) et leur implémentation sous forme de cellules *ARMA* du quatrième ordre appelée filtres *gammatone* (Slaney, 1993) (Patterson, 1994) dont la largeur de bande augmente avec la fréquence.

Cette analyse, visant à effectuer un traitement séparé des fréquences, est effective à tous les niveaux de l'audition, des noyaux du tronc cérébral jusqu'au cortex auditif (d'après (Baskind, 2003)). Ceci indique bien que l'évaluation des indices binauraux est effectuée par bandes de fréquence.

La localisation de sons complexes se base ainsi sur des mécanismes de fusion des différents indices binauraux de localisation en fonction de la fréquence. Ceux-ci restent encore mal compris. Ce sujet est en effet connexe à la ségrégation de flux auditifs qui permet de séparer les "messages" associés à plusieurs sources dans une scène complexe. Il semble alors que le synchronisme ainsi que la durée de l'attaque des sons joue un grand rôle dans les mécanismes de fusion ((Baskind, 2003) reprenant (Perrot *et al.*, 1970)). L'analyse de ces phénomènes dépasse largement le cadre

de notre étude.

Une autre modalité importante de l'audition spatiale est l'effet de précédence (autrement appelé "effet Haas", "loi du premier front d'onde", ...). Ce mécanisme permet, notamment, à l'audition de se focaliser sur la localisation du son direct dans un environnement réverbérant afin d'estimer précisément la position de la source sonore. Il fonctionne suivant un mécanisme complexe d'inhibition qui n'est pas encore parfaitement compris d'un point de vue physiologique (Blauert, 1999).

Il est possible, pour des signaux familiers, d'estimer une localisation très approximative à l'aide d'une seule oreille (écoute monoraule) en analysant la coloration spécifique apportée à la source (diffraction de la tête, réflexions spécifique du pavillon, et résonances du conduit auditifs) en fonction de la position. L'influence de cet indice semble toutefois faible par rapport aux indices binauraux de localisation dans la localisation des sources sonores. Il est toutefois essentiel pour les personnes atteintes de surdit  sur une des deux oreilles. Il leur permet ainsi de d terminer une localisation approximative des sources sonores (Blauert, 1999). Celle-ci est facilit e par les mouvements de la t te qui permettent, comme dans le cas binaural, une lev e partielle de l'ind termination par la variation coh rente des indices sp cifique   une zone de provenance de l' v nement auditif.

Extraction des indices binauraux   partir de mesures de HRIR

La caract risation objective des signaux binauraux est r alis e par la mesure des HRIR (Head Related Impulse Responses), dont la transform e de Fourier est appel e HRTF (Head Related Transfer Functions). Les HRIR correspondent   la mesure en champ libre des r ponses impulsionnelles capt es aux niveaux des deux oreilles d'un individu pour une source sonore   une position donn e de l'espace. Celles-ci d pendent de la position de la source sonore et de la morphologie de la personne (forme et taille de la t te, torse, ...) ce qui les rend sp cifiques   chaque individu. L'extraction des indices binauraux de localisation (IID, IPD, IED) dans le syst me auditif est effectu e par bandes de fr quences apr s analyse tonotopique des signaux binauraux. On donne dans cette partie, les op rations math matiques qui permettent d'extraire les valeurs de ces indices binauraux   partir de la mesure des HRIR. Ces op rations sont issues de mod lisation du traitement effectu es dans le cortex auditif.

Dans la suite, les signaux $x_{L,k}$ et $x_{R,k}$ repr sentent les signaux capt s respectivement par l'oreille gauche et par l'oreille droite dans la bande de fr quence k .

Estimation de l'IPD L'extraction de l'IPD est d riv e   partir de la corr lation interaurale   court-terme par :

$$IPD(t) = \operatorname{argmax}_{\tau} \int_{-\infty}^{+\infty} x_{L,k}(t' - \frac{\tau}{2}) x_{R,k}(t' + \frac{\tau}{2}) w(t' - t) dt', \quad (7.3)$$

où, τ est généralement limité dans l'intervalle $-1 \text{ ms} < \tau < 1 \text{ ms}$ qui correspond aux valeurs maximales admises par l'IPD. $w(t)$ est une fenêtre qui permet de sélectionner une étendue temporelle finie sur laquelle on calcule l'autocorrélation et de prendre en compte la mémoire limitée du système auditif. Martin (Martin, 1995) propose de lui donner la forme paramétrique $Ate^{-\frac{t}{t_0}}$ dont la largeur effective est de l'ordre de 2 à 3 ms. Une valeur de t_0 qui remplit ces critères est donnée par $t_0=1.5 \text{ ms}$.

Une estimation de l'IPD est effectuée lors de la détection d'une attaque. Le maximum de la fenêtre est centré sur le pic d'attaque de chaque fréquence. L'effet de précédence peut se modéliser de manière simplifiée en empêchant tout calcul de l'IPD au niveau des attaques détectées entre 2 à 10 ms après un calcul valide. Un mécanisme simple de masquage rétroactif peut aussi être implémenté afin d'éviter de prendre en compte des maxima locaux immédiatement suivis par un pic de plus grande ampleur.

Pour les HRIR, une solution simple consiste à effectuer la détection sur le maximum de la réponse impulsionnelle (Martin, 1995).

Estimation de l'IED L'estimation de l'IED est effectuée de manière identique à celle de l'IPD en utilisant l'enveloppe du signal. Celle-ci est obtenue, pour chaque bande de fréquence, en portant le signal au carré et en effectuant un filtrage passe-bas avec une fréquence de coupure de 800 Hz. Une autre technique consiste à effectuer une rectification demi-onde et à appliquer le même filtrage passe-bas (Blauert & Cobben, 1978).

Estimation de l'IID L'estimation de l'IID est effectuée dans chaque bande de fréquence par la différence d'intensité entre l'oreille droite et l'oreille gauche. L'IID est donné (Martin, 1995) à un temps t et pour une bande de fréquence k par :

$$IID_k(t) = 20 \times \log_{10} \left(\frac{E_{R,k}(t)}{E_{L,k}(t)} \right), \quad (7.4)$$

où, $E_{R,k}(t)$ et $E_{L,k}(t)$ sont les signaux d'enveloppe respectivement associés aux oreilles droite et gauche pour la fréquence k .

Dans le cas d'un signal stationnaire ou d'une analyse par réponses impulsionnelles, l'IID est donné par :

$$IID_k = 20 \times \log_{10} \left(\frac{\langle x_{R,k}(t) \rangle}{\langle x_{L,k}(t) \rangle} \right), \quad (7.5)$$

où, $\langle x_{R,k}(t) \rangle$ et $\langle x_{L,k}(t) \rangle$ sont les valeurs quadratiques moyennes des signaux associés aux deux oreilles (sans extraction d'enveloppe) pour la bande de fréquence k .

7.1.2 Cohérence des indices binauraux de localisation pour la WFS

L'évaluation de la localisation de sources virtuelles synthétisées par un banc de haut-parleurs peut être effectuée par la mesure du champ synthétisé à une position donnée par une tête arti-

ficielle. Nous proposons une méthode dérivée de celle proposée par Pulkki (Pulkki *et al.*, 1999) qui permet de simuler les signaux binauraux captés au niveau des oreilles d'un auditeur pour la synthèse d'une source virtuelle par un ensemble de haut-parleurs.

Nous présentons ensuite un ensemble de simulations réalisées pour la synthèse de sources virtuelles en WFS avec ou sans égalisation multicanal pour un banc linéaire de haut-parleurs idéaux à directivité omnidirectionnelle de 6m de long. Plutôt que d'utiliser une théorie unifiée de la localisation dont les résultats sont sujets à caution, on compare les indices binauraux extraits des simulations par rapport à ceux donnés par les signaux binauraux pour la source cible.

Simulations binaurales d'un banc de haut-parleurs

On cherche ici à estimer les signaux binauraux captés au niveau des oreilles d'un auditeur à une position donnée de l'espace pour la synthèse d'une source virtuelle par un banc de haut-parleurs. Les signaux binauraux sont dérivés de la mesure des fonctions de transfert entre chacun des haut-parleurs et des microphones situés dans les conduits auditifs de l'auditeur. En considérant le système comme linéaire, on applique le filtrage nécessaire à la synthèse de la source virtuelle cible aux fonctions de transfert associées aux haut-parleurs et on somme l'ensemble des contributions au niveau de chacune des oreilles.

Dans le cas de haut-parleurs idéaux, Pulkki *et al.* proposent de dériver ces mesures à partir de HRTF mesurées en chambre anéchoïque. Il utilise les mesures effectuées sur le mannequin KEMAR par Gardner et Martin (Gardner & Martin, 1994). Pulkki *et al.* considèrent une position d'écoute centrale par rapport au dispositif, c'est à dire vis-à-vis de laquelle l'ensemble des haut-parleurs sont équidistants et à des positions angulaires simples figurant dans la base de données des mesures du KEMAR au MIT.

Dans notre cas, les haut-parleurs sont placés sur une ligne et aucune position d'écoute réaliste ne se trouve "au centre" du dispositif. Il s'agit alors de retarder et d'atténuer les HRTF pour tenir compte de l'atténuation et du temps de propagation naturels. La validité de cette approche semble acquise tant que les haut-parleurs sont omnidirectionnels et que la distance de l'auditeur à un haut-parleur est supérieure à 1 m. Au delà de cette distance, il est communément admis que les HRTF ne varient pas ou peu (Larcher, 2001) (Pellegrini, 2001).

D'autre part, les bases de données des HRTF disponibles, dont celles du KEMAR, offrent une résolution angulaire minimale de 5 degrés qui s'avère insuffisante dans notre cas. Il est nécessaire de procéder à une interpolation des HRTF afin de simuler les mesures aux positions intermédiaires requises.

Dans la suite, on considère que les haut-parleurs sont situés dans le plan horizontal et on utilise les données du KEMAR. L'interpolation est réalisée dans le domaine fréquentiel à l'aide des deux positions du plan horizontal les plus proches. Cette interpolation s'effectue sur les réponses fréquentielles correspondantes pour lesquelles on a extrait au préalable le retard interaural. Celui-ci est simplement estimé à partir de la formule de Kuhn (cf. équation 7.1) de l'ITD basse fréquence. Les pondérations affectées à chaque réponse fréquentielle s'obtiennent à partir de la différence

angulaire entre les deux HRTF de la base de données correspondant aux positions les plus proches et la position angulaire cible. Une fois les réponses fréquentielles estimées, on ajoute le retard interaural cible calculé la formule de Kuhn (cf. équation 7.1).

Conformément aux résultats de Pellegrini (Pellegrini, 2001), cette interpolation simple permet d’obtenir de très bons résultats lorsque la discrétisation de la base de données d’origine est suffisamment fine. Dans le cas où la base de données est plus lâche, il convient alors d’utiliser des méthodes d’interpolation plus complexes faisant intervenir des spline sphériques par exemple, voir Pellegrini (Pellegrini, 2001). Sur l’interpolation des HRTF, on renvoie aussi à l’analyse de Larcher (Larcher, 2001) qui s’applique aux différentes implémentations temps-réel de la synthèse binaurale.

Analyse de la validité des signaux binauraux captés par l’auditeur

Comme annoncé dans l’introduction de cette partie, le modèle utilisé ici n’a pas vocation à estimer la localisation réelle de la source virtuelle synthétisée pour une position de la zone d’écoute. Notre approche consiste à estimer la variation des indices par rapport aux indices issus de l’analyse des HRTF correspondant à la position cible.

On effectue alors un calcul de l’IPD et de l’IED pour chaque position d’écoute i selon la méthode proposée dans la partie précédente en utilisant, d’une part, les fonctions de transfert pour la simulation de la synthèse de la source virtuelle cible à l’aide du banc de haut-parleurs (IPD_{synth}^i et IED_{synth}^i), et d’autre part, les HRTF correspondant à la position de la source cible (IPD_{source}^i et IED_{source}^i). Le calcul est effectué dans 9 bandes de fréquences réparties sur l’ensemble du spectre audible, l’analyse temps/fréquence étant réalisée par un banc de filtres Gammatone (Slaney, 1993). La fréquence centrale la plus basse considérée est ici de 150 Hz. On effectue le calcul de l’IPD uniquement sur les quatre premières bandes dont la fréquence centrale est située sous la limite de validité perceptive de l’IPD (1500 Hz). L’IED est alors calculé sur les cinq bandes restantes.

La fonction d’intercorrélation calculée sur des signaux à bandes étroites est susceptible de posséder de nombreux maxima locaux dont la valeur est très proche, en particulier au dessus de 500 à 700Hz (sur ce point, voir partie 2.2 du chapitre 2 et 4.4 du chapitre 4 dans (Baskind, 2003)). On choisit d’exclure les bandes pour lesquelles l’erreur est supérieure à ± 30 degrés. On considère alors qu’il y a une indétermination que l’on assimile davantage à un flou de localisation.

Le calcul de l’IID est réalisé pour 42 bandes de fréquence ERB conformément à la définition de Patterson (Patterson, 1994). On propose alors simplement de donner la différence, en dB, entre l’IID calculé pour la synthèse par le banc de haut-parleurs et l’IID calculé pour la source cible.

Estimation objective de l’erreur de localisation pour la WFS

Dans cette partie, on considère un banc de 40 haut-parleurs idéaux omnidirectionnels espacés de 15 cm. On définit dans la zone d’écoute 50 positions disposées sur 5 lignes parallèles au banc

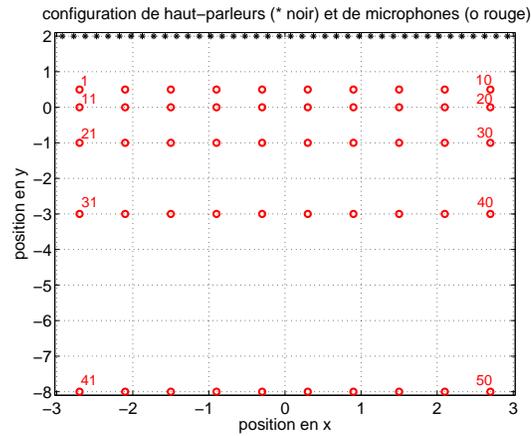


FIG. 7.1 – Positions des têtes artificielles sur lesquelles on simule les signaux binauraux pour la synthèse d’une source virtuelle par la WFS avec ou sans égalisation multicanal.

de haut-parleur pour différentes profondeurs d’écoute (1-10 : 1,5 m ; 11-20 : 2 m ; 21-30 : 3 m ; 31-40 : 5 m ; 41-50 : 10 m) sur lesquelles on estime les signaux binauraux et les indices binauraux de localisation (cf. figure 7.1). On considère que la tête est orientée perpendiculairement au banc de haut-parleur (tous les auditeurs le regardent). On considère deux sources test :

- une onde plane se propageant perpendiculairement au banc de haut-parleurs (onde plane à 0 degré) ;
- une source virtuelle focalisée située un mètre devant le banc de haut-parleurs et centrée par rapport à celui-ci.

Ces sources sont synthétisées par un jeu de filtres WFS ou obtenues par la WFS avec égalisation multicanal (cf. partie 5.1). L’égalisation multicanal est réalisée en prenant comme points de contrôle un banc de 60 microphones omnidirectionnels espacés de 10 cm et situés à 2 m du banc de haut-parleurs.

Pour chaque source et chaque méthode de synthèse, on estime les indices binauraux de localisation à partir des signaux binauraux d’après la méthode de la partie précédente.

L’ensemble des figures 7.2 présente l’erreur de localisation en degrés par bandes de fréquence. Les bandes 1 à 4 sont issues du calcul de l’IPD, et les bandes 5 à 9 du calcul de l’IED. Ces figures représentent l’erreur de localisation pour les deux sources virtuelles synthétisées par la WFS avec égalisation multicanal. Les valeurs d’erreur de localisation sont limitées à ± 90 degrés. On remarque des pics d’erreur correspondant à une erreur de localisation maximale que l’on attribue à l’ambiguïté d’estimation du maximum de la fonction d’intercorrélacion.

Les figures 7.2(c) et 7.2(d) montrent que pour les deux sources test et les deux méthodes de synthèse (avec ou sans égalisation multicanal) l’IPD (bandes 1 à 4) est globalement bien restitué pour les 4 bandes de fréquence et la localisation est ainsi correcte pour toutes les positions d’écoute. On remarque toutefois une petite erreur de localisation pour l’onde plane sur les côtés

de la zone d'écoute (microphones 1/10/11/20/21/30/31/40/41/50). Cette erreur peut être attribuée à la diffraction. Elle est accentuée à proximité du banc de haut-parleurs pour l'égalisation multicanal, c'est à dire en amont des points de contrôle ce qui est conforme avec les observations de la partie 5.2.1.

En considérant les bandes 5 à 9, on voit que l'erreur de localisation due à l'IED reste faible dans un nombre relativement important de bandes de fréquences. Ceci est d'autant plus vrai quand la distance de la position d'écoute au banc de haut-parleur augmente. Ce point est cohérent avec, d'une part, l'augmentation de la fréquence d'aliasing avec la distance au banc de haut-parleurs (cf. partie 2.4), et d'autre part, la cohérence du *premier* front d'onde reçu au dessus de la fréquence d'aliasing avec le front d'onde cible.

Enfin, on remarque que les résultats sont globalement équivalents que l'on utilise ou pas l'égalisation multicanal. Ce résultat est peu surprenant car l'apport de l'égalisation multicanal pour le cas de haut-parleurs idéaux reste limité à une correction de la diffraction par synthèse de sources virtuelles de compensation des ondes diffractées. Le niveau des ondes diffractées est toutefois trop faible, dans la configuration utilisée, pour créer une variation significative d'IPD.

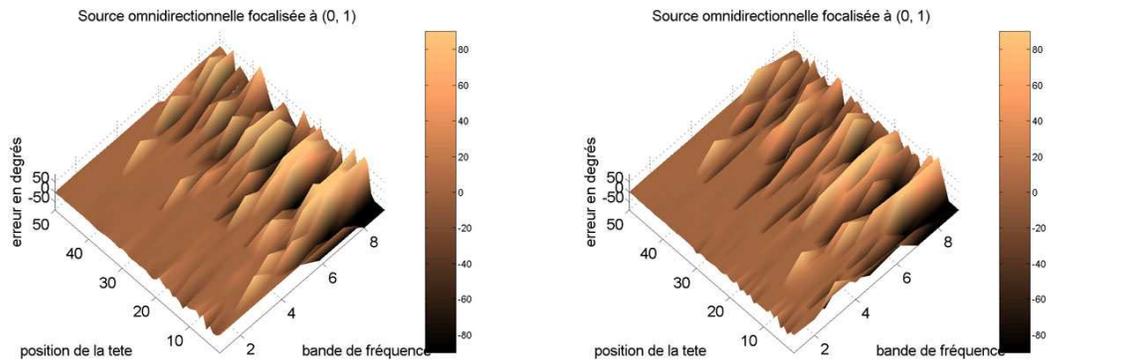
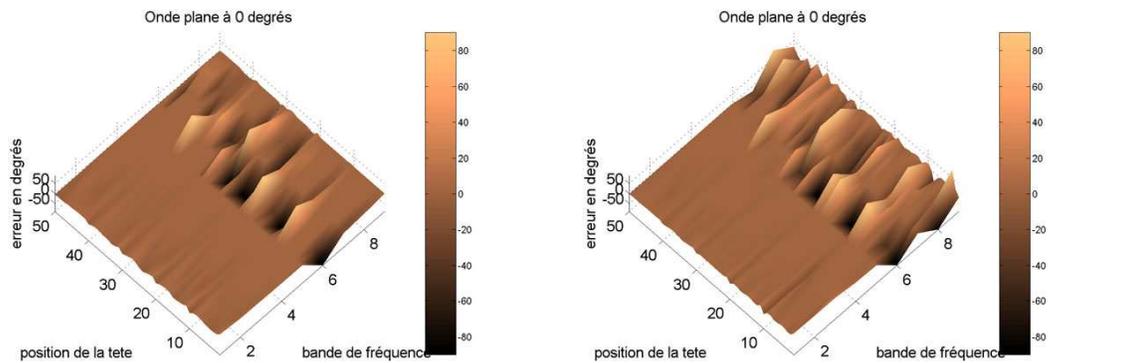
On observe le même type de comportement pour la source focalisée (cf. figures 7.2(a) et 7.2(b)). On remarque tout de même une restitution moins robuste de l'IED (bandes de fréquences 5 à 9), notamment, à proximité du banc de haut-parleurs (microphones 1 à 20). Ce comportement s'explique par le fait que le "champ d'aliasing spatial", pour ce type de source, est situé temporellement *devant* le front d'onde principal (cf. partie 1.4.1). Ce champ est constitué des fronts d'onde successifs, provenant des différents haut-parleurs, qui ne fusionnent pas au dessus de la fréquence d'aliasing en un front d'onde unique. Pour une position d'écoute donnée, seule la dernière contribution est cohérente avec le front d'onde cible. L'IED détecté a ainsi de grandes chances de ne pas correspondre à l'IED cible.

Toutefois, cette dernière contribution est aussi la contribution maximale et est donc détectée comme attaque par le modèle simple de détection (cf. partie 7.1.1. Ainsi, la fenêtre de calcul de l'autocorrélation à court terme est centrée sur ce maximum et le modèle se focalise alors sur cette direction. La cohérence de ce résultat avec la perception peut raisonnablement être mis en doute, et marque les limites du modèle simpl(iste) utilisé.

L'erreur de localisation en hautes fréquences pour les positions d'écoute proches du banc de haut-parleurs (microphones 1 à 20) est tout de même relativement constante sur l'ensemble des 5 bandes de fréquence d'IED (bandes 5 à 9). Il est alors possible de considérer cette erreur comme une délocalisation de la source à ces fréquences.

L'ensemble des figures 7.3 présent la différence en dB dans 42 bandes de fréquences, entre l'IID calculé pour une source virtuelle synthétisée par le banc de haut-parleurs et l'IID calculé pour la source cible.

Pour l'onde plane à 0 degré (cf. figures 7.3(c) et 7.3(d)), l'erreur sur l'IID est faible jusqu'à une fréquence limite correspondant à la fréquence d'aliasing. Celle-ci augmente ainsi avec la distance

(a) Source omnidirectionnelle focalisée à $(0, 1)$, WFS sans égalisation multicanal(b) Source omnidirectionnelle focalisée à $(0, 1)$, WFS avec égalisation multicanal

(c) Onde plane à 0 degré, WFS sans égalisation multicanal

(d) Onde plane à 0 degré, WFS avec égalisation multicanal

FIG. 7.2 – Erreur de localisation par bandes de fréquence par l'estimation de l'ITD (IPD en dessous de 1500 Hz, bandes 1 à 4 ; IED au dessus de 1500 Hz, bandes 5 à 9)

au banc de haut-parleurs. En dessous de cette fréquence limite, de légères fluctuations d'IID apparaissent sur la figure 7.3(c) (sans égalisation multicanal) qui restent toutefois inférieures au seuil perceptif de 1 dB (Pellegrini, 2001). Ces fluctuations disparaissent pour la WFS avec égalisation multicanal à partir de 2 m de distance, distance à laquelle sont situés les microphones de contrôle (cf. figure 7.3(d)).

Au-dessus de la fréquence d'aliasing, on remarque des variations plus importantes d'IID qui dépendent de la position par rapport au banc. L'erreur admet un niveau positif sur la gauche et négatif sur la droite. On rappelle que, d'après la définition de l'IID (cf. équation 7.5) et celle de l'erreur, un niveau positif de l'erreur équivaut à un niveau trop élevé perçu par l'oreille droite et a donc tendance, a priori, à décaler la localisation de l'événement auditif sur la droite, et inversement.

Cette erreur d'IID peut s'expliquer par l'asymétrie des contributions reçues au dessus de la fréquence d'aliasing. Celle-ci est due à la taille finie du banc de haut-parleurs. Le masquage physique de la tête à ces fréquences donne ainsi un niveau cumulé différent suivant la position

d'écoute considérée. En effet, pour les positions d'écoute extrêmes et une orientation de la tête de l'auditeur perpendiculaire à la direction du banc de haut-parleurs, l'ensemble des haut-parleurs est situé du même côté de la tête de l'auditeur. Le niveau cumulé des contributions des haut-parleurs au niveau de chaque oreille est alors plus élevé du côté où ils sont situés. Le résultat dépend alors de l'*orientation* de la tête de l'auditeur et mériterait une étude approfondie. Ce problème potentiel renvoie à l'analyse perceptive du champ d'aliasing spatial qui reste encore à réaliser.

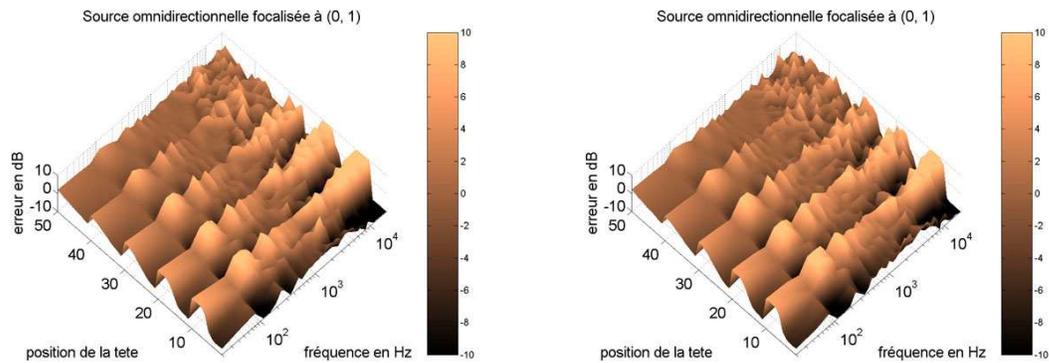
Pour la source focalisée (cf. figures 7.3(a) et 7.3(b)) l'erreur de synthèse de l'IID possède le même type de fluctuations en fonction de la position d'écoute que pour l'onde plane au dessus de la fréquence d'aliasing spatial. Le motif est, en revanche, *inversé* par rapport à l'onde plane (voir par exemple, les microphones 1 à 10 pour les figures 7.3(a) et 7.3(c)). L'erreur est négative sur la gauche du banc de haut-parleurs et positive sur la droite et ce, pratiquement *indépendamment de la fréquence*. L'erreur est toutefois plus forte au dessus de la fréquence d'aliasing. Le phénomène s'atténue avec la distance.

Cette erreur d'IID est en contradiction, d'un point de vue perceptif, avec les critères temporels (IPD et IED). En hautes fréquences, la source virtuelle risque alors d'être délocalisée vers les haut-parleurs les plus proches de la position d'écoute. Encore une fois, il conviendrait d'effectuer des études plus poussées pour estimer l'influence de cette restitution incorrecte de l'IID sur la localisation observée et notamment à estimer l'influence de l'orientation de la tête.

Ce résultat est à rapprocher de ceux de Wittek (Wittek, 2004) qui a effectué des mesures d'IID du champ synthétisé pour des sources focalisées en WFS. Il ne considère qu'une position de tête orientée *perpendiculairement* au banc de haut-parleurs et centrée par rapport à celui-ci. Les sources virtuelles synthétisées sont elles aussi centrées et situées sur la *droite* de l'auditeur. Il analyse alors principalement les différences d'atténuation de sources virtuelles en champ proche et n'observe pas, contrairement au cas de sources réelles, de variations significatives d'IID en basses fréquences. Il attribue, pour sa part, ce phénomène à la taille de la tâche focale synthétisée. A cause du retournement temporel, la taille de cette tâche, en fonction de la fréquence, est de l'ordre de $\frac{\lambda}{2}$ m, λ étant la longueur d'onde.

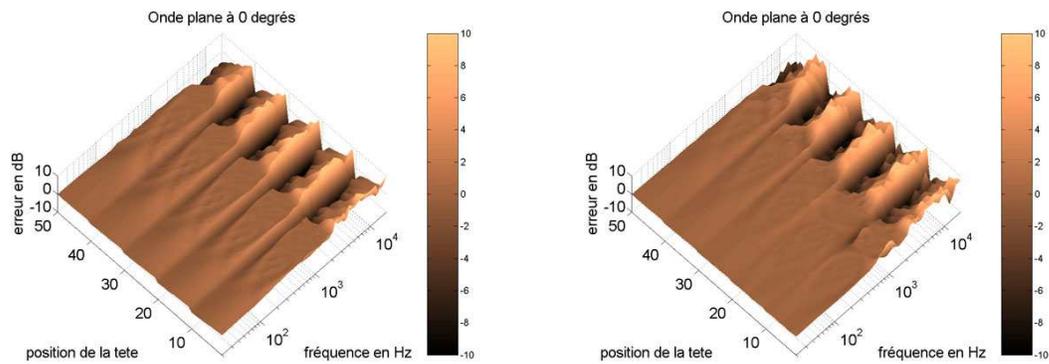
Malgré les quelques réserves que nous avons formulées pour la localisation en hautes fréquences, l'IPD est correctement reproduit dans l'*ensemble* de la zone d'écoute et des bandes de fréquences. Ceci valide d'un point de vue perceptif "objectif" la synthèse *physique* des fronts d'onde associés à une source virtuelle par la WFS. Cette étude fait écho aux expériences perceptives de localisation reportées dans les thèses de l'université de Delft (Vogel, 1993) (Start, 1997) (Verheijen, 1997) que nous avons repris dans (Corteel, 2000).

La localisation effective des sources virtuelles synthétisées en WFS au-dessus de la fréquence d'aliasing reste encore mal définie et mériterait des études complémentaires.



(a) Source omnidirectionnelle focalisée à (0, 1), WFS sans égalisation multicanal

(b) Source omnidirectionnelle focalisée à (0, 1), WFS avec égalisation multicanal



(c) Onde plane à 0 degré, WFS sans égalisation multicanal

(d) Onde plane à 0 degré, WFS avec égalisation multicanal

FIG. 7.3 – Estimation de l'erreur en dB dans 42 bandes de fréquences entre l'IID calculé pour la synthèse d'une source virtuelle synthétisée par la WFS avec ou sans égalisation multicanal et l'IID calculé pour la source cible pour les 50 positions d'écoute

7.2 Exploration spatiale des scènes sonores

La WFS permet la reproduction des propriétés acoustiques d'un champ sonore dans une zone étendue de l'espace. Elle peut, en particulier, reproduire la courbure des fronts d'onde en rapport avec la position de la source virtuelle.

Il est admis que la courbure des fronts d'onde ne fournit pas, par elle-même, d'indice perceptif efficace pour l'estimation de la distance (Blauert, 1999). Toutefois, la courbure *locale* du front d'onde permet l'estimation de la direction de la source virtuelle (cf. partie précédente) ainsi que la création d'une loi d'atténuation qui n'est cependant pas celle de la source virtuelle cible et dépend de sa position par rapport au banc de haut-parleurs (cf. partie 2.2.1).

Si l'installation sonore est suffisamment grande pour permettre une déambulation, les auditeurs font l'expérience de variations dynamiques de l'impression de distance des sources virtuelles, associées à la variation de leurs directions perçues et de leur niveau, en cohérence avec leur déplacement.

Par rapport aux techniques de reproduction stéréophoniques, on passe ainsi d'une situation égocentrée à une scène sonore "exocentrée" dont certains des paramètres ne s'apprécient, a priori, que par la navigation. On peut parler de création d'effets de parallaxe en relation avec une perspective sonore cohérente pour l'ensemble des "points de vue" de la scène (cf. introduction générale).

Pour la WFS, ces effets sont disponibles grâce à la manipulation de la position exacte de la source sonore et non plus uniquement de son azimut comme dans le cas des techniques traditionnelles. Apparaît ainsi un nouveau paramètre, correspondant à la distance de la source virtuelle par rapport au centre du dispositif que nous proposons de nommer "distance holophonique" (Noguès *et al.*, 2003).

La manipulation de ce paramètre n'est possible que pour les techniques de reproduction en zone étendue (WFS, Ambisonique d'ordre supérieur, et dans une certaine mesure binaural avec repérage de la position et l'orientation de la tête).

Pour la création d'une scène sonore, ceci pose la question de la manipulation et de l'utilisation de ce *nouveau* paramètre. D'autre part, la création d'une perspective sonore donne la possibilité à l'utilisateur d'adopter de nombreux points de vues de la scène. L'étape de mixage ne peut alors plus se faire d'une unique position, au centre du dispositif. Le risque est que cette position redevienne une position d'écoute préférentielle ce qui remet tout simplement en cause les avancées permises par la WFS. La tâche de mixage doit alors être considérée comme exocentrée de telle manière à ce que la scène créée reste cohérente quelque soit la position de l'auditeur dans l'installation sonore.

Ce "nouveau" type de mixage a fait l'objet, pendant le premier semestre 2003, du mémoire de fin d'étude de Mathieu Noguès (étudiant ingénieur du son, CNSMDP) et d'une publication pour le DAFX03 (Noguès *et al.*, 2003). Ces résultats ont été repris pour le DAGA/CFA04 (Warusfel *et al.*, 2004a).

Une première partie de cette étude concerne la gestion des effets de parallaxe par la manipulation

exclusive de la *distance holophonique* (sans effet de salle additionnel).

Une deuxième partie de cette étude s'intéresse aux relations entre la distance holophonique et les réglages traditionnels de l'effet de distance (impression d'éloignement de la source) par ajout d'effet de salle synthétique.

Ces deux points ont fait l'objet d'un test d'écoute interactif réalisé auprès de 14 ingénieurs du son.

Ce test a permis de valider l'utilisation de la distance holophonique pour la création d'une perspective sonore. Au contraire des situations naturelles, il a été montré que la distance holophonique peut être utilisée de manière relativement indépendante par rapport à l'impression d'éloignement de la source.

Dans cette partie, nous décrivons les deux expériences perceptives mises en œuvre. Nous en donnerons les résultats ainsi que les limites de l'approche considérée.

7.2.1 Utilisation de la distance holophonique pour la création d'une perspective sonore

Dans cette partie, nous reprenons tout d'abord l'exemple de création de perspective sonore donné dans l'introduction générale que nous replaçons dans le contexte du test d'écoute interactif. Nous présentons ensuite la première expérience perceptive ainsi que ses résultats.

Distance holophonique, effets de parallaxe

Prenons l'exemple d'un ensemble constitué de trois guitares et d'une voix. Ce type d'ensemble est classiquement situé sur deux plans. Les guitares sont placées en arrière plan de la voix. Nous proposons alors trois "situations" de la perspective sonore par la manipulation *exclusive* de la distance holophonique. Celles-ci sont représentées dans la figure 7.4.

La première situation consiste à placer les trois guitares et la voix à une distance holophonique faible (à proximité du banc de haut-parleurs). Un auditeur qui se déplace dans l'installation sonore peut alors "visiter" chacune des sources sonores en se rapprochant physiquement d'une des sources sonores de la scène.

La deuxième situation typique consiste à laisser la voix à une distance holophonique faible et à synthétiser les guitares sous forme d'ondes planes (distance holophonique "infinie"). On rappelle ici que les ondes planes ne sont en réalité, pour la WFS, que des ondes de type cylindrique émises par un banc de haut-parleurs linéaire. Les fronts d'onde reçus dans le plan horizontal sont ainsi plans mais l'onde subit l'atténuation du banc de haut-parleur linéaire.

Comme on l'a vu dans la partie 7.1.2, une telle source virtuelle permet une localisation *angulaire* de la source sonore quelque soit la position de l'auditeur. Les trois guitares ont alors une répartition angulaire identique dans l'ensemble de la zone d'écoute et forment un plan figé qui "glisse" sur les haut-parleurs et "suit" l'auditeur lors de ses déplacements. La voix, au contraire, reste à une position fixe et l'auditeur peut alors choisir le point de vue qu'il souhaite adopter vis-à-vis

d'elle.

La troisième situation consiste à reproduire toutes les sources sonores sous forme d'ondes planes. On offre ainsi une perspective identique pour n'importe quel point d'écoute. C'est l'ensemble de la scène qui suit l'auditeur lors de ces déplacements.

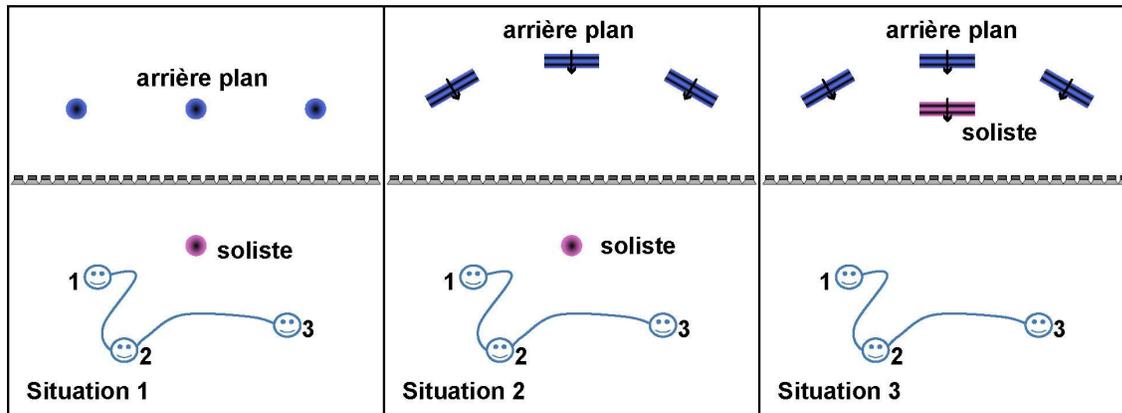


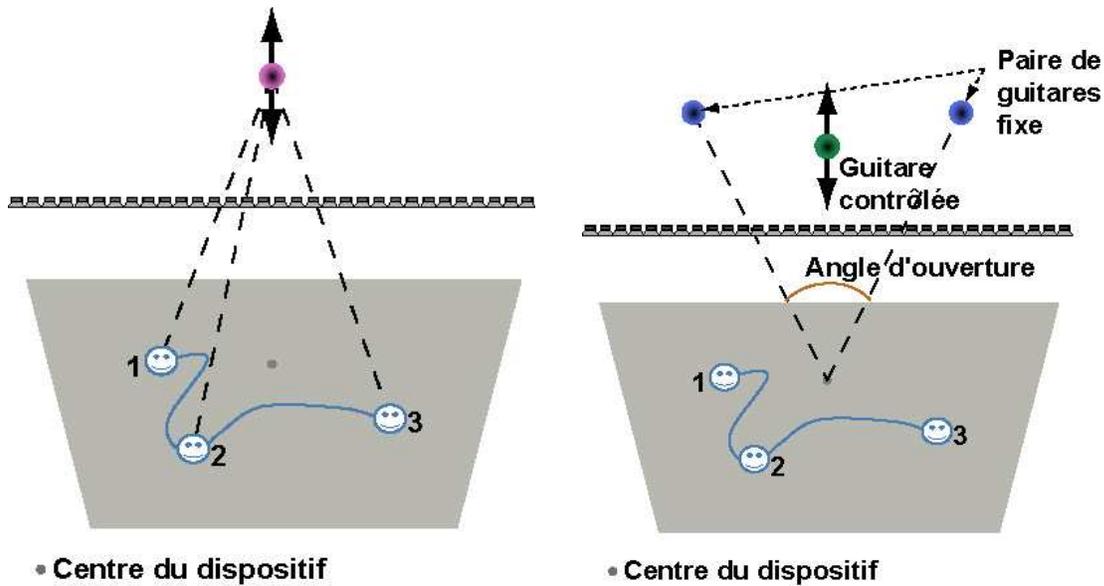
FIG. 7.4 – Création de perspective sonore par la distance holophonique pour un ensemble instrumental composé de trois guitares et d'une voix

Validation perceptive de la distance holophonique

Une première expérience perceptive a été réalisée dans le but de vérifier la validité et l'usage du paramètre de distance holophonique dans un contexte de mixage.

Le dispositif de reproduction est constitué de 4 MAPs (cf. partie 5.2.2) adjacents formant un banc de haut-parleurs de 5,3 m de long. Ce banc de haut-parleurs est installé dans le studio 4 de l'IRCAM dont l'acoustique est suffisamment "sèche" pour garantir une indépendance des effets observés vis-à-vis de l'effet de salle naturel. A ce dispositif correspond une zone d'écoute privilégiée de forme trapézoïdale délimitée au sol par un marquage (cf. partie 5.2.2). L'égalisation de ce dispositif est assurée par la procédure d'égalisation multicanal décrite dans la partie 5.1 réalisée pour un ensemble de sources ponctuelles omnidirectionnelles dont la discrétisation de l'espace répond aux critères définis dans 5.1.2. A partir d'une distance holophonique limite de 10 m, les sources virtuelles reproduites sont des "ondes planes" dont l'angle d'incidence vis à vis de la normale au banc de haut-parleurs est compris entre -30 et + 30 degrés avec un pas de 5 degrés. Toutes les sources virtuelles sont normalisées (temps d'arrivée, niveau) au centre du dispositif qui est à une position centrée et à 2 m du banc de haut-parleurs. La distance holophonique minimale admise est de 1 m.

Un exemple introductif est proposé pour lequel une unique source est reproduite, centrée par rapport au banc de haut-parleurs (cf. figure 7.5(a)). L'ingénieur du son peut manipuler la distance holophonique et se déplacer dans la zone de reproduction afin de vérifier les variations naturelles de localisation et se familiariser avec ce paramètre dans des situations de mixage "exocentrique". Une mixette MIDI portative lui permet de manipuler le paramètre cible en se déplaçant dans



(a) Manipulation préliminaire de la distance holophonique

(b) Configuration de tests

FIG. 7.5 – Validation de la distance holophonique pour la création de perspective sonore, test d'écoute réalisé

la zone d'écoute.

La première expérience vise à vérifier que la manipulation exclusive de la distance holophonique permet à l'ingénieur du son de gérer la cohérence de la perspective spatiale créée. L'organisation spatiale de la scène doit rester cohérente sur l'ensemble de la zone d'écoute.

Deux guitares sont reproduites à une distance holophonique fixe. La troisième est manipulée par l'ingénieur du son auquel on demande de la placer "entre les deux autres", à la même profondeur, en agissant exclusivement sur sa distance holophonique (cf. figure 7.5(b)).

L'ingénieur du son est invité à contrôler la perspective créée en se déplaçant à l'intérieur de la zone d'écoute. Neuf situations de test sont proposées dans un ordre aléatoire. Elles diffèrent la situation statique proposée : distance holophonique, angle d'ouverture, et "configuration" des deux guitares fixes (cf. tableau 7.2.1).

Pour la configuration "centrée", les deux guitares fixes sont situées de part et d'autre du centre du banc de haut-parleurs et la guitare manipulée est centrée. Pour la situation "excentrée" une des deux guitares fixe est centrée (angle par rapport à l'axe de 0 degré). L'autre guitare est située sur la gauche avec un angle θ , par rapport à l'axe. La guitare manipulée est alors située à une position angulaire intermédiaire avec un angle $\frac{\theta}{2}$.

La figure 7.2.1 représente la médiane ainsi que la répartition des quartiles pour chaque configuration de test. Les croix isolées représentent les réglages considérés comme non significatifs. Ils sont automatiquement exclus de l'analyse.

Situation	Distance des guitares fixes	Angle d'ouverture des guitares fixes	Configuration des guitares
1	1.5	30 degrés	centrée
2	1.5	60 degrés	centrée
3	1.5	30 degrés	excentrée
4	4	30 degrés	centrée
5	4	60 degrés	centrée
6	4	30 degrés	excentrée
7	9	30 degrés	centrée
8	9	60 degrés	centrée
9	9	30 degrés	excentrée

TAB. 7.1 – Situations proposées pour la première expérience

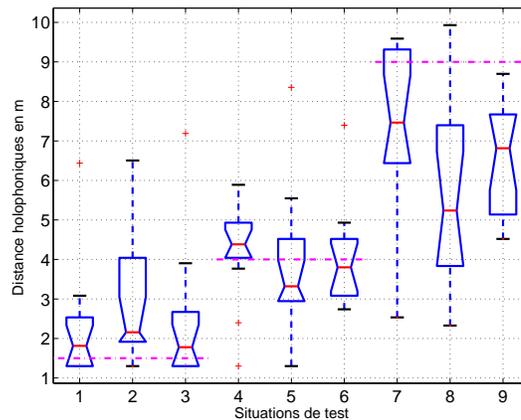


FIG. 7.6 – Résultats du premier test : deux guitares fixes, distance holophonique de la troisième manipulée

Les résultats montrent une forte corrélation entre le réglage de distance holophonique de la guitare centrale réalisé par les sujets et celle imposée aux deux autres guitares. Une analyse ANOVA, sur la dépendance des distances holophoniques (imposées/réglées) donne $p < 10^{-2}$ pour les configurations centrées avec un angle d'ouverture de 30 degrés, $p < 10^{-2}$ pour les configurations excentrées, $p < 10^{-2}$ pour les situations centrées avec un angle d'ouverture de 60 degrés. Aucune autre corrélation n'est significative ($p > 0.05$) pour les autres paramètres (ouverture, "configuration" des guitares).

On remarque que la distance holophonique choisie par les sujets se situe dans un intervalle de distance plus réduit que celui des guitares fixes. On peut avancer trois explications à ceci :

1. les distances holophoniques autorisées sont comprises entre 1 et 10 m et la conversion des valeurs de contrôleur MIDI en valeur de distance holophonique est effectuée suivant une échelle linéaire. Ainsi, pour les situations de test pour lesquelles les guitares fixes sont à 1,5 m ou 9 m, l'excursion de part et d'autre de la valeur de référence est faible. Les sujets choisissent rarement les valeurs extrêmes et l'excursion des résultats est donc plus faible. Lors d'entretiens informels après le test, certains sujets ont mentionnés une perception de

- type logarithmique du paramètre de distance holophonique ;
2. la synthèse des sources situées autour du banc de haut-parleurs (distance holophonique comprise entre 1,5 et 2,5 m) est susceptible d'introduire une coloration que la procédure d'égalisation multicanal, dans son état d'avancement à cette époque, n'a pas réussi à corriger. Cette zone reste critique pour l'égalisation car peu de haut-parleurs contribuent alors à la synthèse des sources virtuelles. On préfère, en général l'éviter à la reproduction. A la suite du test, certains sujets ont déclaré avoir préféré choisir une distance supérieure pour des questions de transparence de la reproduction ;
 3. les sources situées à des distances holophoniques importantes (entre 8 et 10 m, cf. figure 1.9(a)) sont potentiellement diffractées et certains sujets ont préféré les éviter pour préserver une bonne restitution de la perspective même pour de positions d'écoute très excentrées.

Les résultats de cette première expérience montrent clairement que les sujets ont une perception fiable de la distance holophonique et sont capables d'utiliser ce paramètre pour créer un "plan sonore" respectant les effets de parallaxe lors de déplacements dans la zone d'écoute.

7.2.2 Gestion de l'effet de distance en Wave Field Synthesis

La WFS offre la possibilité d'agir sur l'organisation spatiale des sources sonores en modifiant la position des sources sonores par un paramètre que l'on a nommé distance holophonique. Ce paramètre est, a priori, indépendant du réglage de l'impression d'éloignement subjectif. Cette impression d'éloignement subjectif est la "distance" classiquement utilisé par les ingénieurs du son. Elle lié à la synthèse d'un effet de salle synthétique, associé à la source. On considère classiquement que l'impression d'éloignement subjectif est liée aux niveaux relatifs du son direct et de la réverbération (rapport "Dir/Rev"). Cette notion n'est toutefois pas suffisante pour décrire complètement l'impression d'éloignement subjectif. On décrit ainsi dans cette partie, la notion de "présence de la source" qui donne une définition plus exacte de ce paramètre.

Après avoir redéfini cette deuxième notion, nous présentons les résultats d'une expérience perceptive qui permet de juger de l'indépendance relative de ces deux "distances".

Présence des sources sonores

La *présence de la source* a été définie à la suite d'études perceptives réalisées à l'IRCAM à la fin des années 80 et au début des années 90 sur "l'impression spatiale" (Lavandier, 1989) (Jullien *et al.*, 1992) (Jullien, 1995). L'impression spatiale est un terme générique qui rassemble tous les aspects de la perception de l'espace sonore. Elle a été identifiée comme appartenant à une dimension perceptive indépendante des autres notions qui décrivent l'impression spatiale. Elle est liée à l'éloignement subjectif de la source sonore.

Les ingénieurs du son ont une pratique ancienne de la manipulation de ce paramètre. Sa réalisa-

tion la plus simple consiste à régler le niveau du son direct en maintenant les envois auxiliaires vers un processeur de réverbération artificielle à un niveau constant.

Des études psychoacoustiques ont permis de mettre en évidence des mécanismes plus complexes faisant intervenir la distribution spatio-temporelle de réflexions précoces. Ces études ont abouti à la proposition de règles, correctes d'un point de vue perceptif, permettant de manipuler la notion d'éloignement subjectif de la source à l'aide d'un processeur de synthèse d'effet de salle (cf. partie 8.1.1).

Dans le standard MPEG-4 (MPEG, 2000), la présence de la source est objectivée par le paramètre E_s défini par :

$$E_s = \int_0^{40} h^2(t)dt + 0.18 \int_{40}^{80} h^2(t)dt, \quad (7.6)$$

$h(t)$ est la réponse impulsionnelle associée à l'effet de salle synthétique. A cette définition, on associe une description du rendu spatial des réflexions précoces (entre 20 et 40 ms). Celles-ci doivent être synthétisées dans deux directions proches de celle du son direct. L'influence des réflexions synthétisées entre 40 et 80 ms est plus faible. Celles-ci doivent être réparties tout autour de la zone d'écoute. Cette définition est conforme à celle adoptée par le spatialisateur (logiciel de spatialisation créé à l'IRCAM, (Jot & Warusfel, 1995)) qui est utilisé pour la synthèse de l'effet dans le dispositif de reproduction mis en place pour le test d'écoute.

Dans les situations naturelles, ces deux paramètres sont liés par les caractéristiques acoustiques de la pièce d'écoute. La WFS offre alors un degré de liberté supplémentaire par rapport aux situations naturelles puisqu'elle permet un réglage indépendant. On souhaite vérifier l'intérêt d'un tel réglage indépendant vis-à-vis du mixage et de la création de scènes sonores. Cette question importe notamment lorsque l'on veut créer des "outils auteur" spécifiques à la WFS.

Confrontation perceptive de la distance holophonique et de la présence des sources

Pour cette deuxième expérience, un effet de salle synthétique est créé par le spatialisateur. Celui-ci permet, pour chaque source virtuelle, le réglage de la présence de la source indépendamment de la distance holophonique. Pour le détail des modalités de synthèse de l'effet de salle sur le système WFS, on renvoie le lecteur à la partie 8.1.1.

Le trio de guitares est maintenant synthétisé à une distance holophonique et à un niveau de présence fixes. Les guitares forment un plan de référence pour la scène sonore proposée. On propose à l'ingénieur du son d'agir sur la voix soliste afin de créer une scène sonore cohérente (cf. figure 7.7). L'organisation globale du test consiste, pour la voix soliste, à :

- imposer différentes valeurs à la présence de la source et demander d'ajuster la distance holophonique ;
- imposer différentes valeurs à la distance holophonique et demander d'ajuster la présence de la source.

On présente les situations dans un ordre aléatoire sans informer le sujet du paramètre qu'il manipule (le sujet a à sa disposition une mixette MIDI dont un seul des faders est actif et qui

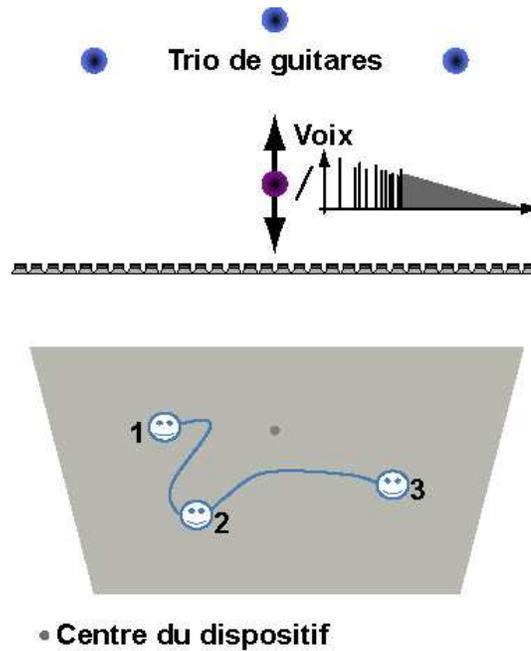


FIG. 7.7 – Situations proposées pour la deuxième expérience perception : évaluation de l'indépendance relative de la distance holophonique par rapport à la présence des sources

Situation	Distance de la voix (m)	Distance des guitares (m)	Présence des guitares (unités spécifiques)
1	4	4	78 (-14 dB)
2	8	4	78 (-14 dB)
3	1.5	4	78 (-14 dB)

TAB. 7.2 – Réglage de la présence de la voix pour une distance imposée, plan des guitares à 4m

est automatiquement assigné au paramètre utilisé pour le test). Douze configurations différentes sont ainsi définies (cf. tableaux 7.2, 7.3, 7.4, 7.5). Deux plans de référence sont proposés pour le trio de guitares. Ils se définissent par un couple distance holophonique/présence de la source identique pour les trois guitares. Les valeurs de la présence pour les deux plans de référence sont ajustées de manière à tenir compte du lien naturel entre distance holophonique et présence de la source. Ce plan de référence peut alors être considéré comme un *contexte* dans lequel on insère la voix de la chanteuse.

Afin de faciliter la compréhension des résultats, on regroupe les situations en fonction du para-

Situation	Distance de la voix (m)	Distance des guitares (m)	Présence des guitares (unités spécifiques)
4	7	8	63 (-19 dB)
5	10	8	63 (-19 dB)
6	3	8	63 (-19 dB)

TAB. 7.3 – Réglage de la présence de la voix pour une distance imposée, plan des guitares à 8m

Situation	Présence de la voix (unités spécifiques)	Distance des guitares (m)	Présence des guitares (unités spécifiques)
7	78 (-14 dB)	4	78 (-14 dB)
8	86 (-11.3 dB)	4	78 (-14 dB)
9	69 (-17 dB)	4	78 (-14 dB)

TAB. 7.4 – Réglage de la distance holophonique de la voix pour une distance imposée, plan des guitares à 4m

Situation	Présence de la voix (unités spécifiques)	Distance des guitares (m)	Présence des guitares (unités spécifiques)
10	63 (-19 dB)	8	63 (-19 dB)
11	71 (-16.3 dB)	8	63 (-19 dB)
12	54 (-22 dB)	8	63 (-19 dB)

TAB. 7.5 – Réglage de la distance holophonique de la voix pour une distance imposée, plan des guitares à 8m

mètre manipulé. Rappelons toutefois qu'au cours du test, les situations étaient présentées dans un ordre aléatoire.

Les figures 7.8(a) et 7.8(b) représentent la médiane et la répartition des quartiles de la *présence de la source* réglée par les sujets pour les deux contextes. Les résultats ne révèlent aucune corrélation significative entre la distance holophonique proposée et le niveau de présence réglé. L'analyse ANOVA sur ce paramètre donne $p = 0.97$. Par contre, les valeurs de présence de la voix choisies sont très fortement liées au niveau de présence imposé pour les guitares, l'analyse ANOVA donnant $p \approx 0$.

Les figures 7.9(a) et 7.9(b) représentent la médiane et la répartition des quartiles de la *distance holophonique* réglée par les sujets pour les deux contextes. Les résultats ne montrent pas de corrélation significative entre la présence de la source imposée et la distance holophonique réglée ($p = 0.64$). Au contraire, on remarque une forte corrélation entre la distance holophonique réglée par les sujets et la distance holophonique à laquelle est placée le trio de guitare ($p < 10^{-2}$).

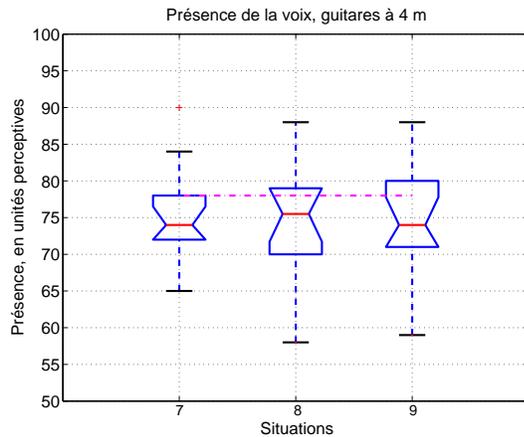
Cette expérience vise à identifier les liens possibles entre présence et distance holophoniques dans le cas où l'on peut manipuler ces deux paramètres de manière indépendante. On a ainsi proposé aux sujets deux différents "contextes" concernant le trio de guitare :

- un contexte "proche" pour lequel les guitares sont placés à une distance holophonique faible et ont un niveau de présence élevé (cf. tableaux 7.2 et 7.4)
- un contexte "lointain" pour lequel les guitares sont placés à une distance holophonique plus élevée et ont un niveau de présence plus faible (cf. tableaux 7.3 et 7.5)

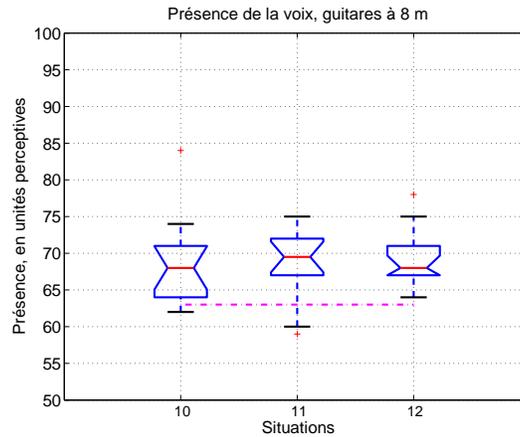
Les différences entre les deux contextes suivent une loi naturelle de variation de ces deux critères. Pour chacun de ces deux contextes, on propose des valeurs fixes de distance holophonique *ou* de présence et on leur propose de manipuler l'autre paramètre. La tâche consiste à assurer la

cohérence de la scène sur toute l'étendue de la zone d'écoute.

Les résultats de cette expérience montrent que les deux paramètres de "distance" (distance holophonique et présence de la source) sont perçus, dans ce contexte, comme deux paramètres indépendants. Lorsqu'on leur demande d'assurer une cohérence de la scène, les ingénieurs du son ont tendance à donner au paramètre manipulé une valeur qui est liée à celle prise par le contexte (guitares) sans interférence avec le paramètre figé.

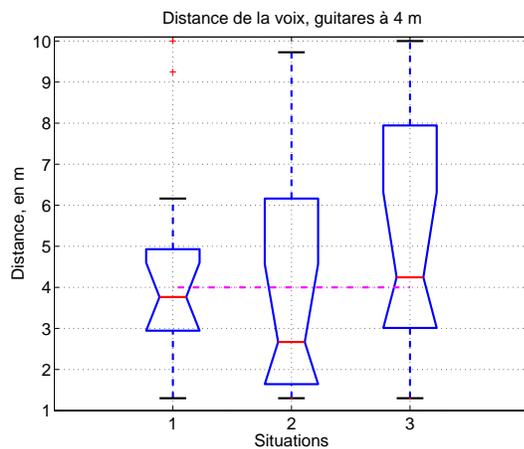


(a) Guitares à 4m, cf. tableau 7.2

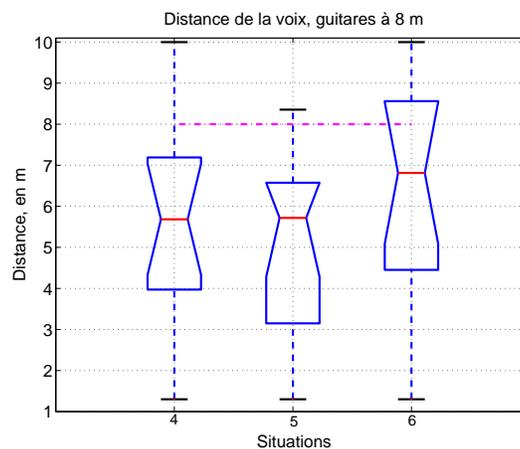


(b) Guitares à 8m, cf. tableau 7.3

FIG. 7.8 – Réglage de la présence de la voix pour une distance holophonique imposée, la ligne magenta correspond à la présence des guitares



(a) Guitares à 4m, cf. tableau 7.4



(b) Guitares à 8m, cf. tableau 7.5

FIG. 7.9 – Réglage de la distance holophonique de la voix pour une présence imposée, la ligne magenta correspond à la distance holophonique des guitares

7.2.3 Discussion : gestion de la distance en WFS

Dans cette partie, nous avons présenté les paramètres liés à la gestion de la "distance" en WFS. Nous avons mis en évidence un nouveau paramètre de contrôle, que nous avons nommé *distance holophonique*. Celui-ci est spécifique aux systèmes de reproduction sonore qui permettent une navigation de l'auditeur à l'intérieur d'un espace d'une taille "suffisante" (que nous ne saurions proprement décrire).

La manipulation de la distance holophonique est rendue possible grâce à la variation cohérente des indices binauraux de localisation auditive perçue lors des mouvements de l'auditeur. Ceci permet de créer une véritable *perspective sonore* ce qui était jusqu'ici inaccessible avec les dispositifs de reproduction classiques.

Nous avons montré dans le cadre du test que ce paramètre est utilisable par les ingénieurs du son pour la création d'une perspective sonore. Ce cadre est certes restreint mais il a été choisi de façon à représenter une "perspective réaliste".

Les ingénieurs du son ont souvent perçu une relation logarithmique entre la distance holophonique et les effets de parallaxe restitués. C'est-à-dire qu'une modification de distance holophonique autour d'une valeur *faible* provoque une variation *rapide* des effets de parallaxe alors qu'une modification de même amplitude de distance holophonique autour d'une valeur *élevée* provoque une variation *lente* des effets de parallaxe. Ce point mériterait cependant des études complémentaires pour définir les relations précises entre le réglage de la distance holophonique et les effets de parallaxe qui en résulte afin d'aboutir à une perception linéaire lors de la manipulation.

En ce qui concerne les relations entre la distance holophonique et la notion d'éloignement subjectif (présence de la source), nous avons pu mettre en évidence une indépendance relative des deux paramètres proposés.

Il est clair que ces résultats sont à confirmer sur un plus grand nombre de sujets et d'exemples musicaux. Leur clarté est pourtant de bonne augure pour leur généralisation à d'autres situations d'écoute.

Un "outil auteur" pour la WFS doit ainsi permettre un réglage indépendant de la distance holophonique et de la présence de la source sonore mais proposer des modalités d'asservissement afin de simuler des variations d'impression de profondeur réaliste ou non lors des mouvements de la source sonore.

Un tel outil auteur doit être installé sur une interface portative afin de permettre au créateur de contenu de vérifier la cohérence de la scène dans toute la zone d'écoute et d'effectuer des corrections en divers positions. On passe, en effet, d'une notion de mixage "égocentré" à un mixage "exocentré" qui se doit d'être cohérent pour *toutes* les positions d'écoute.

Chapitre 8

Création/Manipulation/Reproduction de scènes sonores pour la WFS

La notion de scène sonore recouvre de nombreux aspects. Parmi ceux-ci, nous avons abordés principalement celui de la l'organisation spatiale des objets sonores de la scène, reproduits par des sources ponctuelles ou des "ondes planes" en WFS. Toutefois, cette notion est plus large et concerne aussi l'espace réel ou virtuel à l'intérieur duquel ces objets sont situées. Reste à savoir comment décrire et reproduire les caractéristiques *perceptives* de cet espace.

Contrairement aux systèmes traditionnels, la WFS vise une restitution de l'impression spatiale (son direct + effet de salle) dans une zone étendue de l'espace. La question est ici de vérifier dans quelle mesure l'effet de salle synthétisé est correctement restitué dans cette zone.

De plus, il n'existe aucun standard ni sur le nombre, ni sur la disposition des haut-parleurs pour la WFS et ainsi sur la taille de la zone d'écoute correspondante. Il est possible que la scène sonore soit réalisée puis jouée sur des dispositifs de reproduction de taille très différentes. Il pourrait arriver que l'on veuille créer une scène en studio, avec un dispositif de restitution d'une centaine de haut-parleurs et une zone d'écoute relativement réduite, puis qu'elle soit jouée dans une salle de concert sur un dispositif composé de plusieurs centaines de haut-parleurs pour une zone d'écoute associée de grande taille. Se pose alors la question de la persistance des effets identifiés dans le studio lors de la reproduction dans la salle de concert.

En conséquence, il est nécessaire de fournir aux créateurs de contenu des outils qui permettent la création et la manipulation en temps réel de la scène sonore. La description dynamique de la scène doit pouvoir être consignée. Il convient donc de définir un format de stockage.

Dans ce chapitre, nous tâchons de répondre ou tout au moins de préciser ces questions. Nous verrons dans un premier temps comment la WFS peut s'adapter aux approches classiques de description de l'effet de salle (géométriques, physiques, perceptives). Pour chacune d'entre elles, quelle peut être la validité de l'effet de salle restitué en n'importe quelle position de la zone d'écoute.

Nous intéressons à la reproduction de la scène pour différentes situations (taille et forme du

dispositif, acoustique de la salle d'écoute). Ceci portera, d'une part, sur la restitution de l'organisation spatiale de la scène et des effets de parallaxe en fonction de la taille de la zone d'écoute. D'autre part, nous verrons quelles sont les caractéristiques de l'effet de salle effectivement généré dans la pièce de restitution. Nous présenterons une méthode de type énergétique qui permet une compensation partielle des caractéristiques acoustiques résiduelles après réduction du niveau des réflexions précoces (cf. partie 3.2 et chapitre 6).

La troisième partie de ce chapitre concernera la description de la chaîne de production sonore mise en place à l'IRCAM en collaboration avec sonic emotion ag. La définition de cette chaîne de production fait suite aux travaux réalisés à l'IRCAM dans le cadre du projet européen Carrouso et se base sur des logiciels existant, adaptés, ou développés spécifiquement.

8.1 Description et synthèse d'une scène sonore en WFS

La description d'une scène sonore pour la WFS repose sur un ensemble de paramètres associés aux *sources* (position, orientation, directivité) et à une *qualité acoustique cible* (effet de salle). Dans cette partie, nous présentons les approches classiques de description de l'effet de salle et nous commentons leur utilisation pour la WFS. Notre analyse porte en particulier sur la validité de la restitution dans l'ensemble de la zone d'écoute.

Dans un deuxième temps nous présentons les modalités de synthèse de la scène sonore sur le système WFS. Nous présentons en particulier les structures de traitement de signal qui permettent la synthèse de sources virtuelles sur le système de reproduction en fonction du type d'égalisation mise en œuvre (aucune, individuelle, multicanal). Nous montrons notamment que ces structures restent similaires dans le cas où la méthode de compensation des réflexions précoces de la salle d'écoute est utilisée (cf. chapitre 6).

8.1.1 Description et synthèse de l'effet de salle

On distingue 3 types de description de l'effet de salle :

1. approche géométrique,
2. approche physique,
3. approche perceptive.

Dans le premier cas, il s'agit de décrire les *causes* (sources et caractéristiques acoustiques des parois). Dans le deuxième cas, on décrit directement le champ acoustique et sa traduction dans le domaine signal. Dans le troisième cas, on décrit *l'effet* sous forme de dimensions/catégories perceptives.

Remarquons que le passage par une description de type physique est indispensable pour la synthèse. Celle-ci est accessible soit

- par la modélisation, à partir de la description géométrique ou de la description perceptive par la création d'un effet de salle générique manipulable sur certaines dimensions ;

– par la mesure dans l'espace réel.

La WFS permet la synthèse de sources virtuelles ponctuelles ou d'ondes planes à partir d'une distribution de haut-parleurs. La synthèse de l'effet de salle va donc consister à créer des *canaux d'effet de salle* à partir du son direct qui seront diffusés par un ensemble de sources virtuelles supplémentaires sur le système de reproduction.

Rappelons toutefois que cette synthèse est limitée (reproduction 2D uniquement) et imparfaite (loi d'atténuation non conforme, aliasing spatial). Reste alors à savoir dans quelle mesure l'effet de salle restitué peut être conforme avec la description originale.

Dans cette partie, nous menons cette analyse de manière systématique pour chaque approche. Nous considérons ici que le banc de haut-parleur est placé en conditions anéchoïques ou tout au moins que l'effet de la salle d'écoute est négligeable dans l'impression spatiale restituée.

Approche physique

L'application de la description physique dans le cas de la WFS pose un problème de *changement d'échelle*. En effet, dans le cas des méthodes classiques de reproduction sonore (principes stéréophoniques), la description physique s'appuie sur l'existence d'une position d'écoute de référence, position unique à laquelle l'auditeur est sensé se situer. La relation entre la source et l'auditeur peut donc se concevoir en tant que canal acoustique unique (propagation dans la salle) que l'on caractérise en *un* point de l'espace en ajoutant, éventuellement, la notion de direction de provenance. La description physique est ainsi consigné sous forme d'une ou plusieurs réponses impulsionnelles associées à la direction de chaque haut-parleur du dispositif de reproduction.

Dans le cadre de la WFS, et plus généralement de toutes les méthodes de restitution qui permettent une synthèse de *champ*, ce canal acoustique n'existe plus et la description physique doit correspondre à une description spatiale du champ acoustique dans une zone étendue. Il s'agit alors de définir une "méta-réponse impulsionnelle" qui consigne le champ présent dans la salle sur une zone étendue.

La définition d'une telle description répond aux mêmes critères que ceux formulés dans le cadre de la caractérisation du système MIMO dans le cadre de l'égalisation multicanal ou de la procédure de compensation individuelle des réflexions (cf. parties 5.1.1 et 6.1.1). Il s'agit de décrire les caractéristiques spatiales du champ dans une zone étendue de l'espace. Toutefois, la description doit être suffisamment générique afin d'être applicable à différents dispositifs de reproduction et tailles de zone d'écoute.

Les solutions mises en œuvre se basent sur la mesure de réponses impulsionnelles sur des dispositifs microphoniques complexes tels que ceux décrits dans la partie 4.3.1. Le champ mesuré est ensuite décomposé sur une base de fonctions de propagation (harmoniques cylindriques, sphériques ou ondes planes). Dans le cadre de la Wave Field Analysis (dual de la WFS pour la capture), la caractérisation fournit une décomposition en ondes planes, soit directement à partir de bancs de microphones linéaires (Sonke, 2000), soit à partir d'un banc de microphones circulaire en réalisant une décomposition en harmoniques cylindriques puis en recomposant des

microphones équivalents hyper-directifs qui donnent la décomposition en ondes planes (Hulsebos *et al.*, 2001) (Hulsebos & de Vries, 2002). La description est donnée par une décomposition sur un nombre fini d'ondes planes qui permet a priori de décrire le rayonnement dans une zone de taille finie, par exemple, l'intérieur du banc de microphones circulaire.

Au niveau de la synthèse, on reproduit les ondes planes correspondant à la description. Ces sources virtuelles supplémentaires sont alimentées avec un signal formé par la convolution du "son direct" avec les réponses impulsionnelles mesurées dans chaque direction.

Dans le cas de la WFS, se posent un certain nombre de problèmes, à la fois d'un point de vue théorique mais aussi d'un point de vue pratique.

Premièrement, la WFS est un système de restitution 2D et ne permet ainsi la reproduction des composantes de l'effet de salle que dans le plan horizontal. La troisième dimension est alors apportée par la salle de restitution et n'est donc pas directement maîtrisable. Reste à savoir cependant ce qu'il faut faire avec les composantes décrites en dehors du plan horizontal, si on a pu les décrire.

Pour la WFS, les techniques de description physique de l'effet de salle se basent sur des dispositifs microphoniques 2D (linéaires, multi-linéaires (Sonke, 2000) ; ou circulaires (Hulsebos *et al.*, 2001) (Hulsebos & de Vries, 2002)). Le champ *capté* dans la dimension verticale est alors *automatiquement* réparti sur les composantes du plan horizontal.

Afin de décrire proprement le champ dans une zone étendue de l'espace, le nombre de composantes spatiale qui est lié au nombre de microphones est relativement important (éventuellement plusieurs centaines). La synthèse du champ sonore nécessite alors de reproduire simultanément l'ensemble de ces composantes sur le dispositif de reproduction et de réaliser la convolution du son direct avec toutes les réponses impulsionnelles mesurées dont la longueur peut atteindre plusieurs dizaines voir plusieurs centaines de milliers de coefficients. La puissance de calcul nécessaire en temps-réel est, encore à ce jour, colossale.

Dans le but de réduire le nombre de composantes nécessaires, Hulsebos (Hulsebos & de Vries, 2002) propose d'effectuer une paramétrisation énergétique de type temps/fréquence/espace pour un nombre réduit de direction (8/10). Cette procédure consiste à sélectionner les directions principales de provenance de l'énergie à partir de la décomposition 2D en ondes planes. Les autres directions sont alors "réparties" sur les 8/10 directions sélectionnées. La description réalisée n'est ainsi plus une description du champ mais une extraction des composantes principales.

Au niveau de la reproduction, il propose de reconstruire les réponses impulsionnelles à partir de séquences de bruit blanc, décorréélées entre elles, dans chaque canal. Les composantes directionnelles de l'effet de salle sont ensuite synthétisées par convolution du son direct avec les réponses impulsionnelles calculées. Celles-ci alimentent des ondes planes dans les directions correspondantes.

Rappelons que les "ondes planes" synthétisées en WFS ne sont pas réellement des ondes planes mais des ondes à symétrie cylindrique dont l'axe principal est la ligne constituée par le banc de

haut-parleurs. Dans le cas où l'on utilise plusieurs banc de haut-parleurs, la situation est encore plus complexe. Les fronts d'onde synthétisés présentent ainsi une loi d'atténuation qui décroît contrairement aux ondes planes dont le niveau devrait rester constant. Ainsi, même dans le cas où la description fournie est exacte, la synthèse du champ sonore par le système WFS ne peut être identique à l'original dans l'ensemble de la zone d'écoute.

La reproduction des réflexions précoces sur des ondes planes peut introduire des problèmes de localisation des réflexions par rapport au son direct en fonction de la position d'écoute. La figure 8.1 illustre ce type de problème dans le cas de la reproduction d'une source virtuelle proche *Dir* et de 2 réflexions *Ref₁* et *Ref₂* en ondes planes (situation 1) et à leur position réelle (situation 2). L'auditeur *L₁* étant situé à la position de référence, l'organisation spatiale des réflexion est conservée pour lui quelque soit la situation. Par contre, pour l'auditeur *L₂*, dans le cas de la synthèse des réflexions avec des ondes planes, l'organisation spatiale de l'ensemble est totalement modifiée.

Cette situation est relativement extrême mais montre bien le type de problème auxquels on peut être confronté. Afin de qualifier d'un point de vue perceptif l'importance de ce problème, il s'agirait de déterminer le flou de localisation pour les réflexions précoces. Malheureusement, ce sujet n'a pas fait l'objet d'un grand nombre de recherche dans la littérature et il nous est relativement impossible de conclure sur ce point.

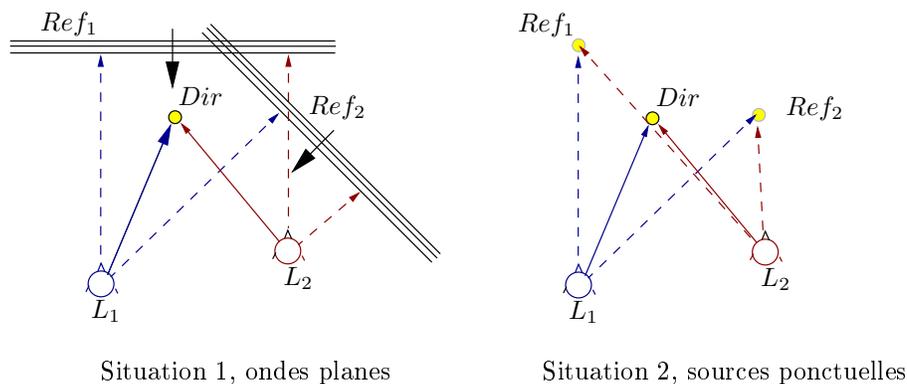


FIG. 8.1 – Synthèse des réflexions par le système WFS, mise en évidence des problèmes de l'utilisation d'ondes planes pour la synthèse des réflexions précoces

La synthèse de la réverbération tardive dans la zone d'écoute suppose de pouvoir créer un champ diffus dont le profil de décroissance suit la description temps/fréquence cible. D'après les études menées par Sonke à l'université de Delft, un nombre réduit de canaux décorrelés (8/10) diffusés sur un système WFS par des ondes planes réparties tout autour de la zone d'écoute s'avère suffisant pour synthétiser un champ dont les caractéristiques, d'un point de vue perceptif, sont proches de celles d'un champ diffus dans la pièce de restitution (Sonke, 2000).

Lors de la mesure de la réponse de la salle, les sources acoustiques utilisées ont classiquement des caractéristiques de directivité omnidirectionnelles ou qui se rapprochent de celles de la voix humaine (plus ou moins cardioïde). La directivité de la source conditionne l'"éclairage" de la

salle et modifie ainsi complètement l'effet de salle généré.

Il serait préférable d'utiliser des sources à directivité contrôlable, telles que la timée de l'IRCAM (Warusfel *et al.*, 2004b), afin d'extraire la réponse impulsionnelle de la salle pour des figures de directivité élémentaires de la source d'émission (harmoniques sphériques). Il serait alors possible de générer la réponse impulsionnelle associée à la directivité réelle de la source par recombinaison à partir des mesures réalisées sur les figures de directivité élémentaires.

Approche géométrique

L'approche géométrique repose sur un modèle de description de type architectural de la salle associée à une description du comportement acoustique des parois (coefficients de réflexion et de diffusion).

La description "signal" associée est obtenue par simulation de l'effet de salle créé par une source donnée. Ceci revient à simuler une mesure de réponse impulsionnelle ou bien faire un bilan énergétique (accès uniquement à la distribution énergétique temps/fréquence/espace mais pas au signal de pression) au niveau d'un ou plusieurs récepteurs. On se retrouve alors confronté au même type de problèmes que dans le cas de l'approche physique.

On peut, soit décider de réaliser une description à partir de la décomposition du champ synthétisé sur une base de fonctions de propagation ; soit essayer d'extraire les éléments importants d'un point de vue perceptif.

Si on souhaite réaliser une décomposition du champ sur un ensemble de composantes spatiales élémentaires, il est obligatoire d'avoir accès au signal de pression et donc de réaliser un calcul *exact* du champ. Les méthodes d'acoustiques prévisionnelles privilégiées sont alors les méthodes par éléments finis ou les méthodes modales. Celles-ci ont, cependant, un coût de calcul qui reste encore prohibitif et ne permettent généralement pas d'estimer, en un temps raisonnable, la réponse impulsionnelle sur toute la bande de fréquences audibles.

Il faut donc se rabattre sur les méthodes géométriques qui ne fournissent cependant qu'une description de type énergétique de l'effet de salle. Il devient alors impossible de le décrire sur une base des fonctions de rayonnement.

Parmi les méthodes géométriques classiques, la méthode des sources images se distingue car elle fournit une description de l'effet de salle relativement indépendante du récepteur (cf. partie 3.1.2). En effet, l'ensemble des sources images calculées ne dépendent pas de la position d'écoute considérée. Celles-ci pourraient ainsi être reproduites en tant que sources virtuelles sur le système WFS. Reste tout de même deux problèmes.

1. Le calcul des sources images est réalisé en trois dimensions. Il s'agit de déterminer, à l'aide de critères perceptifs, quelles sont les sources virtuelles à sélectionner et, idéalement, de définir une "situation équivalente" dans le plan horizontal. De tels critères n'ont pas été définis à ce jour et mériteraient une attention particulière.
2. Pour assurer la validité de la description, on doit vérifier la visibilité des sources images vis-à-vis du ou des récepteurs. Il convient alors d'éliminer les sources images qui ne sont pas

visibles dans l'ensemble de la zone d'écoute. Remarquons qu'un masquage partiel pourrait être si nécessaire reproduit en modifiant la directivité de la source virtuelle synthétisée. Reste à savoir si ceci est justifié d'un point de vue perceptif.

La méthode des sources images ne peut être appliquée que pour la description des réflexions spéculaires précoces. Pour l'énergie diffuse précoce et la diffusion tardive (réverbération), la méthode privilégiée est la méthode de radiosité.

La méthode de radiosité permet de simuler les échanges d'énergie entre surfaces en prenant l'hypothèse de réflexions diffuses.

Dans le cas classique pour lequel il existe une position d'écoute de référence, on collecte l'énergie des angles solides vus de la position du récepteur. Ce champ diffus est ensuite reproduit de manière local sous forme de réponses impulsionnelles décorellées dans chaque direction lors de la description.

Dans le cas de la WFS, le problème se pose de savoir comment reproduire les caractéristiques du champ diffus décrit par la méthode de radiosité. Il s'agit en effet de différencier :

- le champ diffus précoce issu des caractéristiques diffuses de la réflexion sur chaque paroi. Ce phénomène se décrit dans le domaine temporel par un étalement de la réponse de la paroi et dans le domaine spatial comme un champ issu d'une source étendue dont l'ensemble des points d'émission sont décorellés (hypothèse de réflexion diffuse) ;
- le champ diffus tardif qui est classiquement considéré comme un champ uniforme et isotrope, c'est-à-dire pour lequel l'énergie incidente en chacun des points est répartie de manière égale selon toutes les directions et à toutes les positions. De plus, on considère les phases 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.

La synthèse du champ diffus peut ainsi s'effectuer par un nombre réduit de composantes décorellées comme dans le cas physique. Il s'agit alors simplement de collecter l'énergie diffuse au niveau d'un unique récepteur.

La synthèse du champ diffus précoce est plus difficile à réaliser car la WFS ne permet que la synthèse de sources ponctuelles dans le plan horizontal. Pour une réflexion donnée, la composante temporelle du champ diffus peut simplement s'obtenir en associant à la source image correspondante une réponse impulsionnelle qui traduit l'étalement temporel (Pellegrini, 2001). Cependant cette approche n'assure pas la synthèse des caractéristiques *spatiales* de la diffusion.

Il n'est toutefois pas évident que les caractéristiques "directionnelles" du champ diffus précoce soient audibles. Les indices acoustiques fournis sont en effet très incohérents. On pourrait alors envisager de synthétiser un champ diffus précoce de la même manière que le champ diffus tardif. La validation de cette approche nécessiterait des études perceptives spécifiques qui dépassent le cadre de ce travail.

Remarquons que ces deux approches (physique et géométrique) ne permettent pas la synthèse d'un effet de salle exactement conforme à la spécification. Ces deux méthodes nécessitent la définition de critères perceptifs pour déterminer quels sont les éléments importants qui doivent

être synthétisés sur le système WFS. La description *signal* obtenue correspond à un ensemble de réponses impulsionnelles qui décrivent les caractéristiques du champ réfléchi dans un nombre fini de directions (approche physique) ou bien à un ensemble de sources virtuelles (sources images) et à une description du champ diffus sous forme de réponses impulsionnelles décorellées (approche géométrique).

La synthèse de l'effet de salle est réalisée par convolution du son direct avec les réponses impulsionnelles contenant la description de la diffusion. Les signaux obtenus forment des "canaux d'effet de salle diffus" qui alimentent des sources virtuelles synthétisées sur le système WFS a priori sous forme d'ondes planes. Les sources images peuvent alors être synthétisées séparément en tant que sources virtuelles sur le système WFS ou bien par répartition (panoramique d'intensité ou autre) sur les canaux d'effet de salle diffus.

L'avantage de cette dernière approche est de réduire très fortement le nombre de sources virtuelles à synthétiser sur le système WFS. En effet, dans une scène sonore composée de P sources, si pour chaque source on forme N canaux d'effet de salle diffus et qu'il y a Q sources images, on doit reproduire $P \times Q + N$ sources virtuelles sur le système WFS. Les directions associées aux canaux d'effet de salle diffus sont en effet les mêmes quelque soit la source virtuelle reproduite. Dans le cas où les réflexions précoces sont réparties sur les canaux d'effet de salle, le nombre de sources à synthétiser est de $P + N$ ce qui est bien inférieur à ce qui a été mentionné auparavant ... Nous avons cependant souligné les défauts potentiels de cette approche dans la mesure où la source virtuelle est proche et les canaux d'effet de salle sont synthétisés sous forme d'ondes planes. Il pourrait alors être utile de reproduire certaines réflexions précoces en tant que sources ponctuelles et/ou synthétiser les canaux d'effet de salle eux aussi en tant que sources ponctuelles. Reste à estimer si l'on reproduit alors correctement un champ diffus dans l'ensemble de la zone d'écoute.

Approche perceptive

Au contraire des approches géométriques et physiques, l'approche perceptive ne cherche pas à simuler l'acoustique d'un lieu réel à travers ses caractéristiques *acoustiques physiques*. L'approche perceptive propose au créateur de contenu d'agir sur des dimensions perceptives liées à l'impression spatiale. Ces paramètres de "haut-niveau" sont liés à une description de "bas-niveau" (physique) de l'effet de salle de type physique (ensemble de sources images/réflexions, description temps/fréquence de la réverbération tardive et éventuellement de la diffusion précoce) qui permet de *stimuler* les attributs perceptifs cibles.

Remarquons que toute approche perceptive est par définition égocentrée. Elle est définie de manière à donner à un auditeur situé à la position de référence une certaine impression spatiale qui est spécifiée par les paramètres perceptifs. Une dimension classique est celle de l'éloignement subjectif de la source. Dans la réalité, cette impression d'éloignement est lié à des critères objectifs et *dépend* évidemment de la position d'écoute.

Nous proposons d'illustrer ici deux approches perceptives distinctes qui ont été utilisées au cours

du projet Carrouso : l'approche du spatialisateur de l'IRCAM (Jot & Warusfel, 1995), et le modèle de Pellegrini (Pellegrini, 2001). Nous décrivons en particulier comment ces modèles peuvent être adaptés afin de les utiliser pour la WFS. Nous présenterons en particulier l'adaptation que nous avons réalisé pour le spatialisateur.

Au contraire des approches physiques et géométrique, les approches perceptives présentées se basent sur des modèles bas-niveau qui n'utilisent que des composantes réfléchies dans le plan horizontal. Aucune procédure d'élimination ou de répartition des composantes réfléchies en dehors du plan horizontal n'est alors nécessaire.

Ce dernier point aurait tendance à plaider en faveur de l'utilisation de l'approche perceptive pour la description et la synthèse de l'effet de salle en WFS. Dans ce cadre, on maîtrise en effet la qualité de l'impression spatiale rendue au moins au niveau du point d'écoute de référence auquel toutes les sources virtuelles WFS sont normalisées en temps et en niveau. Il s'agit ainsi de réaliser une implémentation qui permet une évolution progressive et réaliste de l'impression spatiale perçue lorsque l'on s'éloigne du point d'écoute de référence.

Approche du spatialisateur Le spatialisateur (Jot & Warusfel, 1995) se base sur une description perceptive de l'impression spatiale issue de recherches menées à l'IRCAM à la fin des années 80 et au début des années 90 sur les caractéristiques perceptives d'un effet de salle (Lavandier, 1989) (Jullien *et al.*, 1992) (Jullien, 1995) (Kahle, 1995). 9 paramètres indépendants ont été définis et forment les dimensions perceptives réglées par le créateur de contenu. Parmi ces 9 paramètres, 3 décrivent l'environnement de la source (niveau par bandes de fréquence son direct + réflexions précoces), 3 décrivent l'interaction de la source avec la salle, 3 autres décrivent les caractéristiques de la réverbération tardive. La "présence de la source" est un de ces paramètres (cf. partie 7.6).

La restitution/stimulation des dimensions perceptives réglées se base sur une description morphologique de l'effet de salle. Un ensemble de règles permet de modifier un modèle générique de réponse impulsionnelle en réglant un niveau énergétique dans quatre sections temporelles (R_0 , R_1 , R_2 , R_3) et trois bandes de fréquences. Ces sections temporelles sont

- R_0 : son direct ;
- R_1 : réflexions précoces (20-40ms) "directionnelles", réparties sur deux canaux L et R qui doivent être reproduits *autour* de la direction angulaire de provenance du son direct ;
- R_2 : réflexions précoces (40-100ms) diffuses, réparties sur 4 canaux S_1 , S_2 , S_3 , S_4 destinés à être diffusés tout autour de l'auditeur de manière statique ;
- R_3 : réverbération tardive formant 4 signaux décorrélés répartis sur S_1 , S_2 , S_3 , S_4 .

Afin de respecter l'organisation spatiale des réflexions suggérée par le modèle morphologique d'effet de salle du spatialisateur, il conviendrait de reproduire les canaux L et R sur deux sources virtuelles ayant la même distance holophonique, ou légèrement supérieure, que le son direct mais situées de part et d'autre en azimut. Ceci permettrait de garantir que ces composantes sont réparties de manière cohérente par rapport au son direct quelque soit la position d'écoute.

Les canaux S_1 , S_2 , S_3 , et S_4 devraient idéalement être diffusés par des ondes planes régulièrement

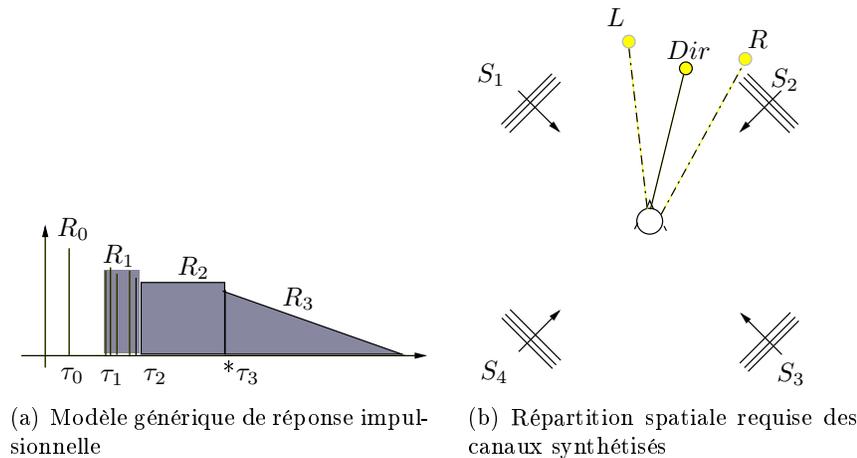


FIG. 8.2 – Modèle physique sous jacent au spatialisateur

réparties autour de la zone d'écoute. Remarquons que le fait de n'avoir que 4 canaux décorrelés est insuffisant vis-à-vis des critères définis à l'université de Delft pour créer un champ diffus (8/10 canaux décorrelés).

Dans l'implémentation que nous avons réalisée pour la WFS, l'ensemble des canaux d'effet de salle est distribué sur 8 haut-parleurs virtuels. Ceux-ci sont réalisés en tant que sources ponctuelles répartis autour de la zone d'écoute qui diffusent 8 canaux d'effet de salle.

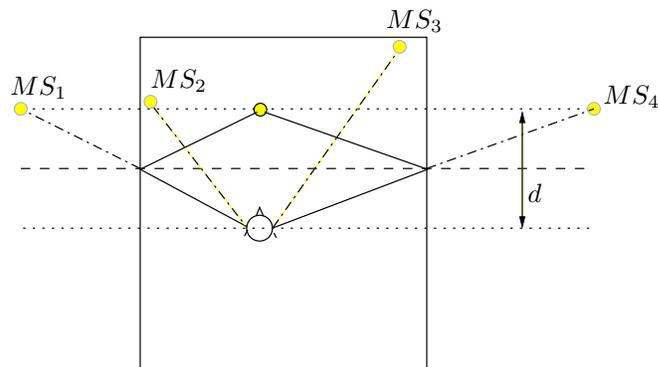
Les canaux L et R sont répartis par panoramique d'intensité sur les 8 canaux d'effet de salle. Ceci permet de limiter le nombre de sources WFS à reproduire sur le système. C'est dans cette optique que ceci a été réalisé. Le choix n'a été fait qu'en fonction de limitations d'ordre *technologique*. Reste à savoir si la différence par rapport au cas idéal est audible, point sur lequel nous n'avons pas encore de réponse claire à apporter.

Les canaux S_1 , S_2 , S_3 , et S_4 sont simplement distribués sur les 8 canaux d'effet de salle mais uniquement pour la partie des réflexions diffuses (40-100 ms). Pour la réverbération tardive, on récupère directement les 8 sorties décorrelées de la structure de réverbération utilisée dans le spatialisateur (FDN, Feedback Delay Network, (Jot, 1997)) que l'on assigne à chacun des 8 canaux d'effet de salle. La restitution de la réverbération diffuse n'est pas réalisée par des ondes planes mais par des sources ponctuelles. Dans la mesure où celles-ci sont relativement éloignées, il nous semble peu probable que ceci influe réellement sur la synthèse d'un champ dont les caractéristiques perceptives se rapprochent de celles d'un champ diffus dans l'ensemble de la zone d'écoute. Ceci resterait néanmoins à vérifier.

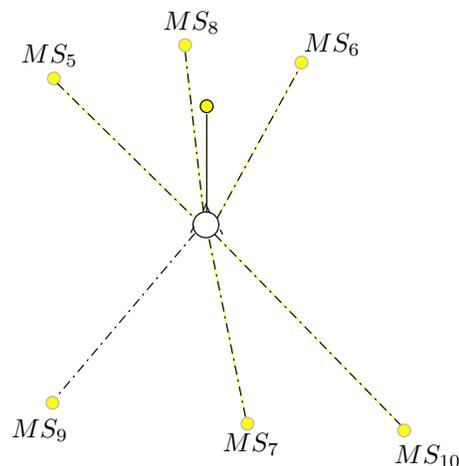
Ce type d'implémentation permet, a priori, de garantir une transition douce entre l'impression spatiale restituée au centre du dispositif (point de référence du système) et celle restituée lorsque l'on s'en éloigne.

Approche de Pellegrini Le modèle perceptif proposé par Pellegrini se base sur une description alternative de l'impression spatiale. Il se concentre sur deux attributs perceptifs : la

"distance" (éloignement subjectif, présence de la source) et la taille apparente de la salle. Pellegrini (Pellegrini, 2001) propose, pour la stimulation de la "distance", de contraindre le temps d'arrivée et l'angle d'ouverture de 4 réflexions. Il se base sur des critères géométriques simples liés à une pièce rectangulaire allongée pour un couple source récepteur légèrement excentré (cf. figure 8.3(a)). Cette asymétrie permet d'éviter un temps d'arrivé régulier des réflexions susceptible d'introduire de la coloration. Les réflexions MS_1 et MS_4 sont exactement les sources images de la source virtuelle par les parois latérale de la pièce choisie. Les réflexions MS_2 et MS_3 sont choisies telles que leur retard soit de l'ordre de $\frac{2}{3}$ de celui de, respectivement, MS_1 et MS_4 . Elles ont aussi un angle d'incidence par rapport à la source plus faible que celui de MS_1 et MS_4 . Ces réflexions sont alors utilisées afin de "remplir" à la fois en terme spatial et temporel le "vide" entre le son direct et les premières réflexions afin d'éviter de les percevoir en tant qu'écho. La stimulation de l'impression de taille de la salle est réalisée par un groupe de 6 réflexions réparties autour de l'auditeur (cf. figure 8.3(b)) dont la distance (et donc le retard et l'atténuation) dépend de la taille de la pièce. Là encore, une situation légèrement asymétrique est préférée afin d'éviter la coloration créée par des directions et des temps d'arrivée trop similaires.



(a) Réflexions stimulant l'impression de distance



(b) Réflexions stimulant l'impression de taille de la salle

FIG. 8.3 – Modèle physique sous jacent au modèle de Pellegrini

Le modèle de Pellegrini étant basé sur des critères à la fois géométriques et perceptifs, la solution correcte consisterait à synthétiser l'ensemble des réflexions considérées par des sources virtuelles WFS ce qui assurerait une transition douce entre les impressions spatiales perçues aux différentes positions d'écoute.

Remarquons toutefois que dans les critères proposés, la taille de la pièce est stimulée par des réflexions lointaines dont la propriété principale est que ces réflexions soient réparties tout autour de la zone d'écoute. Ces réflexions pourraient alors être diffusées sur les canaux d'effet de salle qui reproduisent la réverbération tardive.

Synthèse des canaux d'effet de salle

Quelque soit l'approche utilisée pour la description de l'effet de salle, la synthèse passe par une conversion dans le domaine physique ainsi obtenue (ensemble de réflexions + diffusion précoce et tardive ou réponses impulsionnelles affectées à des directions de provenance). La synthèse de l'effet de salle consiste alors à former, pour chaque source de la scène sonore, un ensemble de canaux supplémentaires à reproduire sur le système WFS. Cette opération s'effectue par filtrage du signal associé à une des sources.

Remarquons alors que cette étape peut être réalisée *indépendamment* de la synthèse des sources sur le système WFS. La structure de traitement de signal affectée à cette tâche est appelée *processeur d'effet de salle*.

Pour notre part, nous avons utilisée la structure de traitement classique de synthèse de l'effet de salle du spatialisateur (Jot & Warusfel, 1995).

Validité de l'effet de salle synthétisé

Comme on l'a vu au cours de cette partie, la Wave Field Synthesis ne permet pas une synthèse "exacte" d'un effet de salle donné dans l'ensemble de la salle d'écoute. Il est en particulier *impossible* de contrôler le champ dans la troisième dimension. La loi d'atténuation des fronts d'onde synthétisés en WFS est aussi problématique car le niveau des réflexions synthétisées ne peut être correct dans toute la zone d'écoute. Celui-ci est particulièrement élevé en proximité des haut-parleurs (cf. équation 2.20).

L'impression spatiale restituée ne peut ainsi être correcte dans toute la zone d'écoute. La compréhension et, éventuellement, la maîtrise de ces phénomènes passerait par une analyse détaillée des caractéristiques physiques de l'effet de salle restitué en fonction de la position d'écoute et sa traduction en terme d'impression spatiale. Une telle description nécessiterait des études complémentaires sur la perception du niveau, de la direction d'incidence et du temps d'arrivée de réflexions dans un champ sonore complexe qui dépasse largement le cadre de notre étude.

8.1.2 Reproduction sur le système WFS

Dans cette partie, nous illustrons la restitution d'une scène sonore sur un système WFS en temps-réel. On suppose que le processeur d'effet de salle délivre N canaux d'effet de salle à reproduire tout autour de la zone d'écoute ainsi que P canaux associés à des sources virtuelles WFS.

Nous considérons un système WFS incomplet constitué d'un banc linéaire de haut-parleurs situé dans la région frontale qui permet la synthèse de sources virtuelles en WFS dans une zone limitée de l'espace.

La figure 8.4 représente un tel système constitué d'un banc de haut-parleurs linéaire dans l'espace

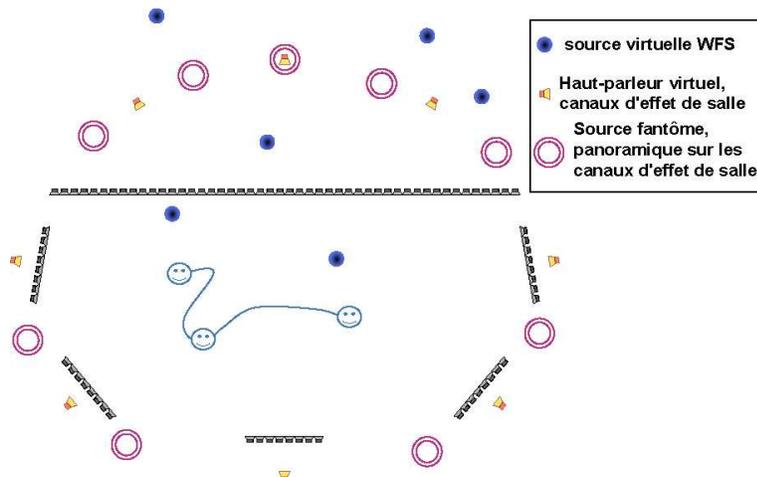


FIG. 8.4 – Système WFS considéré

frontal et d'un ensemble de bancs de haut-parleurs de taille plus petite et non continus sur les côtés et l'arrière.

Le système ne permet de synthétiser que des sources virtuelles en WFS dans le secteur frontal. Sur chacun des banc de haut-parleurs situés sur les côtés et à l'arrière, est synthétisée une source ponctuelle située derrière le banc de haut-parleurs. Celle-ci est placée légèrement derrière le banc de haut-parleurs. Ceci permet de faire en sorte que la zone de visibilité de cette source à travers le banc de haut-parleurs englobe une grande partie de la zone d'écoute.

De plus, on synthétise trois sources ponctuelles avec le banc de haut-parleurs frontal afin de former un système régulier de 8 *haut-parleurs virtuels* entourant la zone d'écoute et sur lesquels sont répartis les P canaux d'effet de salle.

Dans un premier temps, nous montrons comment les différents canaux peuvent être répartis sur le système. Nous donnerons ensuite les structures de traitement de signal temps-réel qui permettent de mettre en œuvre les différents techniques d'égalisations décrites dans la partie II de ce document.

Répartition des canaux sur le système WFS

Sur le type de système de reproduction considéré, la synthèse de sources WFS est limitée au secteur frontal. Le problème se pose de savoir que faire des sources virtuelles situées soit les côtés, à l'arrière ou bien, susceptibles de se déplacer tout autour de la zone d'écoute.

Nous proposons de distribuer ces sources sur les haut-parleurs virtuels destinés à diffuser les canaux d'effet de salle. La répartition s'effectue par panoramique stéréophonique. L'image n'est a priori correctement reproduite qu'au centre du dispositif. Cependant, les haut-parleurs virtuels sont peu écartés (30 à 50 degrés) ce qui a tendance à augmenter significativement la zone d'écoute dans laquelle la localisation reste correcte.

Il s'agit alors d'un mode de restitution que l'on peut qualifier de "dégradé" puisqu'il ne permet pas la restitution des effets de perspective sonore mais qui est extrêmement peu coûteux en terme de puissance de calcul temps réel (panoramique sur deux haut-parleurs virtuels déjà synthétisés par ailleurs). Cependant, il peut être utile de l'utiliser pour des sources qui n'ont pas besoin d'une précision de rendu spatial très fine ou qui ont une vocation plus immersive comme par exemple l'ambiance.

Cette technique est à rapprocher de la technique des Virtual Panning Spot (VPS) proposée par l'IRT (Institut für Rundfunktechnik) qui propose de synthétiser un ensemble de sources virtuelles sur le système WFS qui constituent autant de haut-parleurs virtuels sur lesquels on va répartir par panoramique stéréophonique les signaux associés aux sources (Theile *et al.*, 2003). Cette technique permet de réduire considérablement le nombre de canaux de transmission nécessaires, par exemple pour la restitution d'ensembles orchestraux, pour lesquels on ne peut distinguer chacun des instruments individuellement. Il est aussi possible par cette technique de tenter de reproduire des sources étendues (orgue, chœur, . . .) à l'aide plusieurs sources virtuelles sur lesquels on répartit la prise de son issue de plusieurs microphones. Ce type de technique ouvre aussi la voie à l'utilisation de techniques microphoniques "traditionnelles" pour la WFS telles que l'utilisation de microphones principaux utilisés couramment pour la prise de son "naturelle" (Hulsebos *et al.*, 2003) (Kuhn *et al.*, 2003). Il s'agit alors de capter une l'image globale de la scène dans laquelle on insère des éléments plus distincts (sources ponctuelles).

Dans le système mis en place à l'IRCAM, il appartient au créateur de contenu de choisir explicitement ce mode de rendu appelé *Pan* pour les sources susceptibles de passer à l'arrière ou sur les côtés. Le rendu des sources estampillées *WFS* est alors limité au seul secteur frontal. On laisse d'autre part la possibilité de choisir la distance holophonique des haut-parleurs virtuels synthétisés à l'avant afin de gérer un plan sonore supplémentaire avec les sources virtuelles situées à l'avant synthétisées en mode *Pan*.

Structure de traitement de signal temps-réel WFS

La figure 8.5 illustre les différentes structures de traitement de signal temps réel à mettre en œuvre pour la synthèse de sources virtuelles en temps réel par la WFS en fonction du type

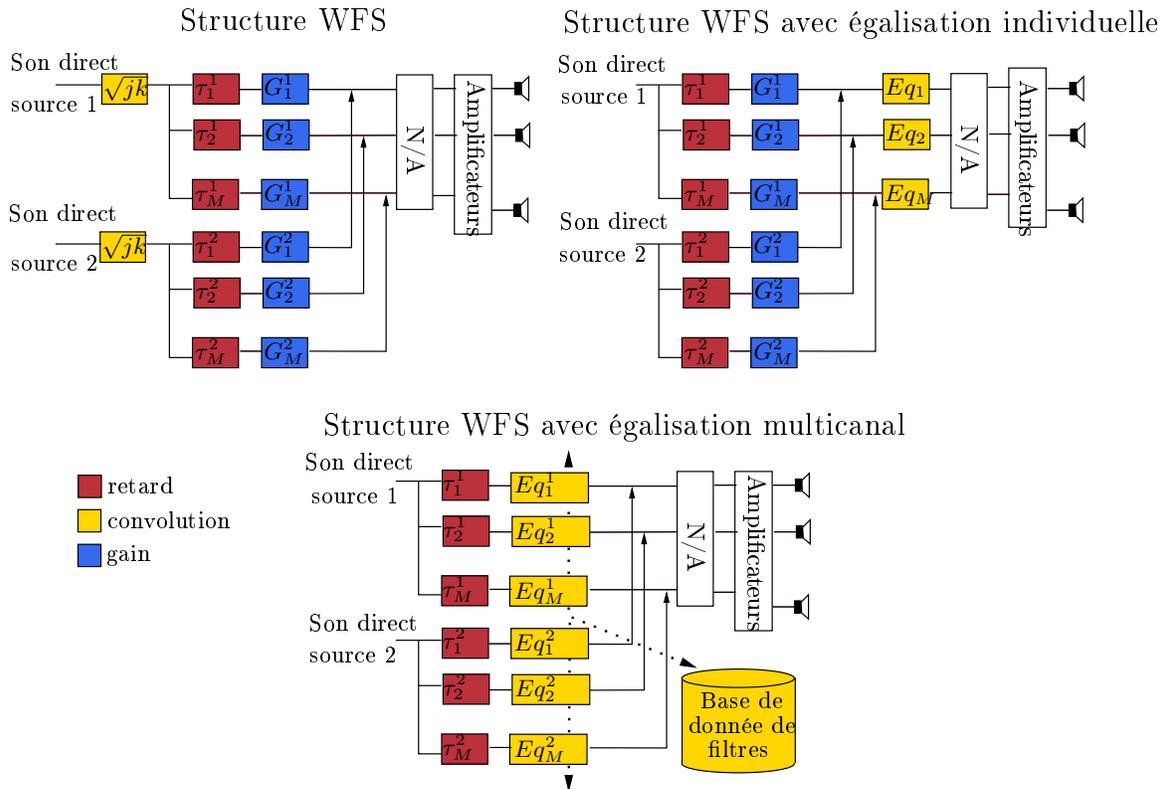


FIG. 8.5 – Structures de traitement de signal pour la synthèse de sources virtuelles en Wave Field Synthesis avec différentes méthodes d'égalisation

d'égalisation utilisée.

La synthèse de sources virtuelles en WFS sans égalisation s'effectue, pour un haut-parleur donné, en appliquant un retard et une atténuation, donnés par la position relative du haut-parleur et de la source en fonction de la vitesse du son dans l'air et par les caractéristiques de directivité, à partir d'un unique signal source filtré (filtre en \sqrt{jk} jusqu'à la fréquence d'aliasing). Les retards et les atténuations nécessaires sont soit calculés en temps réel, soit stockés dans une base de donnée pour tous les haut-parleurs et un ensemble de positions de sources et de figures de directivité élémentaires associées (cf. partie 5.1.2).

L'égalisation individuelle des haut-parleurs, nécessaire dans la plupart des cas, rend le coût de calcul associé plus important. Il est alors nécessaire d'appliquer un filtre d'égalisation par haut-parleur mais indépendamment de la source virtuelle synthétisée.

La mise en œuvre des procédures d'égalisation multicanal proposée dans ce document, de même que la procédure proposée par de Vries dans (de Vries, 1996), *dépendent* de la source virtuelle cible et nécessitent un filtre par haut-parleur *et* par source. Les filtres sont stockés dans une base de donnée pour l'ensemble des positions et des figures de directivité élémentaires calculées par égalisation multicanal. La synthèse d'une source virtuelle donnée s'effectue en chargeant les filtres nécessaires et en les convoluant au signal affecté à la source. Le passage d'un filtre à l'autre nécessite une procédure d'interpolation afin d'éviter les artefacts.

La synthèse d'une source virtuelle dont la directivité est complexe et ne correspond pas à une figure de directivité élémentaire nécessite une étape supplémentaire de composition des filtres. Celle-ci n'est pour le moment pas mise en œuvre.

La compensation des réflexions précoces de la pièce de restitution est réalisée simplement en utilisant les filtres calculés par la méthode de compensation individuelle des réflexions définie dans le chapitre 6.

Dans le système mis en place à l'IRCAM, ce traitement est assuré par un logiciel de convolution développé par sonic emotion ag. La base de donnée de filtres est calculée par égalisation multi-canal. Pour la synthèse d'une source virtuelle, un message est envoyé au logiciel de convolution indiquant la position de la source à synthétiser. Celui-ci choisit dans la base de donnée la source la plus proche figurant dans la base de donnée de la position cible.

Gestion des basses fréquences

Dans l'état de la technologie, les haut-parleurs étant espacés de 15 à 20 cm, on ne peut prétendre effectuer la Wave Field Synthesis sur l'ensemble de la bande audible (20-20000 Hz) avec des haut-parleurs électro-dynamiques. Toutefois de nouvelles approches permettent de reculer les limites de fonctionnement dans les basses fréquences de ces transducteurs (Polack *et al.*, 2002). Les MAP ne permettent pas non plus une synthèse efficace en dessous de 100 à 150 Hz. D'autre part, le champ synthétisé en basses fréquences pour de bancs de haut-parleurs finis présente certaines difficultés à cause de la diffraction (cf. partie 2.3.3).

La solution choisie pour le moment, qui ne se veut pas une solution définitive, consiste à utiliser un petit nombre de subwoofers (typiquement 3 ou 4) répartis autour de la zone d'écoute et destinés à reproduire la partie basses fréquences du champ. Pour chaque source synthétisée sur le système, la partie basse fréquence est distribuée sur chaque canal par panoramique d'intensité en fonction de la position angulaire de la source.

Ce point n'a cependant pas fait l'objet d'une attention particulière lors de cette étude. La gestion des basses fréquences dans une salle est un sujet complexe car il n'est en général plus possible de distinguer le son direct de l'acoustique de la salle d'écoute. Celle-ci doit être prise en compte de manière efficace par placement adéquat des subwoofers dans la salle et par égalisation (voir par exemple (Rubak & Johansen, 2000) ou (Pedersen, 2003)).

L'utilisation de la Wave Field Synthesis à ces fréquences ne requerrait "qu'un subwoofer tous les 1 à 2m". L'utilité d'une telle démarche reste à démontrer et nous n'avons pas de réponse précise à apporter sur ce point.

8.2 Validité de la scène reproduite

Dans cette partie, nous souhaitons aborder la question de la validité de la scène reproduite en fonction de la géométrie du dispositif de reproduction et de l'acoustique de la salle d'écoute.

Dans un premier temps, nous mettrons en évidence un ensemble de problèmes liés à la géométrie du dispositif de reproduction pour la synthèse de la perspective sonore. Nous illustrerons ces problèmes dans le cas de la synthèse de sources virtuelles par la WFS en combinaison avec une projection d'images 2D. Puis nous aborderons le problème plus général de la persistance de la perspective spatiale en fonction de la taille du dispositif utilisé.

A moins que la salle d'écoute soit une chambre anéchoïque ou que l'on obtienne une compensation "parfaite" de celle-ci, le résultat auditif de la synthèse d'une scène sonore *ne correspond pas*, tout au moins physiquement, à celui spécifié par le créateur de contenu. Une opération de compensation supplémentaire de la salle d'écoute est donc nécessaire et fera l'objet d'une deuxième partie.

8.2.1 Restitution de la perspective spatiale

Dans cette partie, on souhaite mettre en évidence la dépendance relative de la description de la scène sonore vis à vis du système de reproduction dans le cas de la WFS. Nous évoquerons ainsi les problèmes de cohérence d'une scène sonore reproduite en WFS avec une projection d'image 2D.

Dans un deuxième temps, nous verrons dans quelle mesure une *mise à l'échelle* de la scène sonore peut s'avérer nécessaire en fonction de la géométrie du dispositif de reproduction.

Cohérence image/son dans le cas de la projection 2D

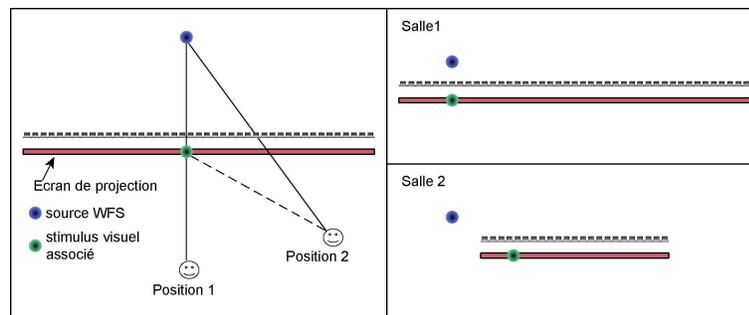


FIG. 8.6 – Problème lors de la reproduction sonore en WFS en combinaison avec une projection d'images en deux dimensions

Dans le cadre de la projection d'images en deux dimensions, une seule de ces deux dimensions est gérée en WFS, la largeur. Dans le but d'assurer une cohérence entre les stimuli visuels et auditifs, les sources sonores devraient être positionnée au niveau de l'écran à la même position que l'objet visuel auquel elles se réfèrent (Melchior *et al.*, 2003) (de Bruijn & Boone, 2003).

Ceci est d'ors et déjà impossible dans la direction verticale. Sur la largeur de l'écran, le positionnement des source est a priori contraint sur un unique plan, situé au niveau de l'écran. La partie gauche de la figure 8.6 montre le type de problème que l'on peut rencontrer dans le cas où la source virtuelle est située derrière l'écran. Si les stimuli auditifs et visuels apparaissent dans

la même direction à la position 1, ce n'est plus le cas à la position 2. Ce phénomène est lié aux effets naturels de parallaxe restitués par la Wave Field Synthesis.

Remarquons toutefois que, d'une part, la précision de localisation autorisée par le système auditif est limitée ; d'autre part, il existe un mécanisme de fusion des stimuli auditifs et visuels connu sous la nom d'"effet ventriloque". Ce phénomène se produit lorsque l'on présente simultanément un stimulus visuel et un stimulus auditif à des positions différentes. Le stimulus auditif est alors perçu à la position du stimulus visuel.

De Bruijn et Boone (de Bruijn & Boone, 2003) ont réalisé un ensemble d'expériences psychoacoustiques afin de quantifier la gêne associée à la synthèse par la WFS de la profondeur réelle de la source virtuelle dans le cadre d'une application de téléconférence. Ils utilisent une scène visuelle composée de plusieurs locuteurs à des profondeurs différentes et choisissent de reproduire ces sources à leur "vraie" profondeur (distance holophonique correspondante à la distance physique vis à vis de la caméra). Ils montrent ainsi que cette situation est vécue comme désagréable pour les positions excentrées, pour lesquelles les stimuli visuels et auditifs ne sont pas cohérents. Une expérience annexe menée sur l'intelligibilité des sources sonores reproduites sans projection d'image (Boone & de Bruijn, 2003) en comparant la reproduction sur haut-parleurs individuels et en WFS montre une augmentation significative de l'intelligibilité pour la WFS avec reproduction de profondeur. Ce résultat n'est pas expliqué précisément dans (Boone & de Bruijn, 2003).

Remarquons cependant que pour la WFS, l'interaction avec la salle d'écoute est réduite et dépend de la source virtuelle synthétisée (cf. partie 3.1). Les sources virtuelles sont reproduites à travers une fenêtre acoustique (banc de haut-parleurs) dont la taille est limitée. Ceci réduit le niveau des réflexions latérales ce qui pourrait expliquer ce résultat (cf. partie 3.1.4).

De Bruijn et Boone (de Bruijn & Boone, 2003) suggèrent d'appliquer une compression sur la profondeur de la scène sonore. Ceci permet de conserver une distribution en profondeur des sources tout en réduisant les erreurs de localisation pour l'ensemble des positions d'écoute.

Un problème supplémentaire apparaît lorsque l'on considère le cas classique de la projection des écrans de taille différente. La partie droite de la figure 8.6 illustre ce problème. La source WFS qui est bien positionnée dans la salle 1 est synthétisée en dehors de l'écran dans la salle 2.

Pour l'image, ceci est pris en compte par l'adaptation de la focale du projecteur. Dans le cas de la stéréophonie, il suffit d'écarter les haut-parleurs situés derrière l'écran.

Pour la WFS, Melchior *et al.* proposent d'appliquer un facteur d'échelle à la description de la scène (Melchior *et al.*, 2003) ce qui permet de "ramener" les sources à l'intérieur de l'écran.

Adaptation de la description de la scène au système de reproduction

Se pose, plus généralement, la question de la validité de la description de la scène sonore vis-à-vis du système de diffusion auquel on s'adresse.

Prenons pour cela l'exemple d'une source focalisée. Par définition, si cette source est focalisée, cela veut dire qu'elle est située à l'intérieur de l'espace de reproduction. Cependant cette source n'est *intérieure* qu'en regard de la disposition précise de haut-parleurs considérée. Sur un autre dispositif, une source synthétisée à cette même position "absolue" pourrait se retrouver soit en dehors de la pièce d'écoute, soit au milieu du public. Dans ce dernier cas, la moitié du public serait située dans la "zone d'ombre" de la source focalisée (entre les haut-parleurs et la position de la source) et ne serait donc pas perçue correctement.

Le positionnement d'une source virtuelle à l'intérieur de l'espace peut être considéré comme un effet à part entière voulu par le créateur de contenu. Ceci laisse en effet la possibilité à l'auditeur de se rapprocher indéfiniment de la source jusqu'à rentrer en contact direct avec elle. La source acquiert alors une "présence" toute particulière. Il est donc important de garantir que cette source sera bien située à l'intérieur du dispositif quelque soit la disposition géométrique des haut-parleurs.

Une étape de "mise à l'échelle" de la scène sonore apparaît donc nécessaire pour garantir que les effets produits seront similaires d'un dispositif à l'autre. Ceci peut être réalisé simplement en appliquant un facteur de compression ou d'expansion des distances holophoniques en fonction de la taille du dispositif.

Deux types de problèmes subsistent alors :

1. Cette mise à l'échelle ne fonctionne que si le dispositif de restitution est de forme simple. Idéalement, il devrait s'agir d'un cercle centré sur l'origine du repère par rapport auquel sont positionnées les sources. Si, comme dans la plupart des cas, le dispositif de reproduction est de forme rectangulaire ou trapézoïdal, le facteur de mise à l'échelle à appliquer revient à effectuer une homothétie de la scène dans deux dimensions. Reste à savoir si l'organisation spatiale de l'*ensemble* la scène est alors respectée et si les liens sémantiques entre les sources sonores sont toujours conservés.
2. En fonction du facteur de mise à l'échelle appliqué, les effets de parallaxe vont être modifiés. Typiquement, si la zone d'écoute est grande, les effets de parallaxe ne se ressentiront que par un déplacement important de l'auditeur ; si la zone d'écoute est petite, ceux-ci vont intervenir très rapidement. Ceci plaiderait plutôt en *défaveur* de l'utilisation de la mise à l'échelle, en particulier pour les dispositifs de petite taille qui n'autoriseront pas de toute façon de possibilité de déambulation suffisante. On pourrait alors choisir de reproduire une portion de la zone d'écoute. Ceci reviendrait, par exemple, à choisir un ensemble de sièges dans une salle de concert.

Nous n'avons pas pour le moment de réponse claire à ces problèmes. Ceux-ci sont dus à l'inexistence d'arrangement normatif de haut-parleurs pour la WFS qui, d'une part, permet une grande flexibilité d'installation, mais d'autre part apporte de nouveaux problèmes.

Parmi les problèmes mis en évidence, certains sont d'un ordre purement géométrique (source intérieure), d'autres mettent en jeu des fonctions sémantiques et/ou proprioceptives.

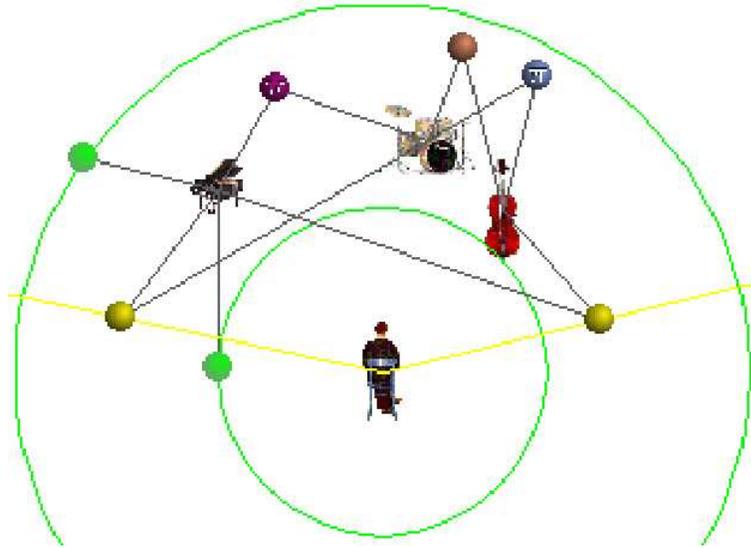


FIG. 8.7 – Système de contraintes utilisées dans le logiciel MusicSpace

Le respect de la volonté du créateur de contenu pourrait passer par l'utilisation de contraintes sur les positions des sources de la scène lors de l'adaptation au système de reproduction.

Une telle approche a été proposée par Olivier Delerue dans sa thèse (Delerue, 2004). Il y décrit le système "MusicSpace". Dans ce système, il s'agit de proposer au créateur de contenu un ensemble d'outils qui permettent d'appliquer des contraintes sur la position des sources sonores de la scène. Ces contraintes peuvent s'appliquer à plusieurs sources simultanément afin de conserver une unité dans un groupe instrumental, restreindre les positions que peuvent prendre certaines sources dans une zone de l'espace, etc... Pour plus de détails, nous renvoyons à la lecture de (Delerue, 2004). Ces contraintes permettent d'objectiver les liens sémantiques qui lient les sources entre elles. Le but de ce système est de garantir que lorsque l'on déplace une source dans la scène, l'organisation spatiale des sources et les liens sémantiques qui s'y rattachent ne sont pas totalement bouleversés. De plus, MusicSpace contient un algorithme de résolution de contraintes en temps réel qui permet une utilisation interactive du créateur de contenu ou de l'utilisateur. Le créateur de contenu a alors la possibilité de déplacer en une seule opération un groupe instrumental complet sans avoir à récupérer les sources une par une. Il conserve donc à tout moment de cette opération une *cohérence* de la scène. Il peut aussi contraindre les positions relatives des sources afin de fournir des modalités d'interaction à l'utilisateur qui garantissent une validité de la scène.

La figure 8.7 représente une scène telle qu'elle peut apparaître dans le système MusicSpace. La position des sources sonores (piano, batterie, contrebasse) est donnée par rapport à la position de l'auditeur. Les contraintes sont matérialisées par les boules et le type de contrainte par la

couleur de ces boules. C'est en liant une source à la contrainte que celle-ci lui sera appliquée. Les boules jaunes représentent une contrainte de limite de position angulaire, les boules vertes, une limite de position en distance. La boule marron, à laquelle sont reliées la contrebasse et la batterie impose un écart angulaire constant entre les deux sources quelque soit leur position.

Un tel système de contraintes n'a pas été mis en place pendant la thèse. Nous pensons cependant que cette approche permettrait de simplifier, d'une part, la manipulation de la scène sonore par le créateur de contenu et, d'autre part, de permettre une adaptation de la scène au système de restitution qui garantisse les liens sémantiques entre les sources sonores.

8.2.2 Impression spatiale restituée

En dehors de l'organisation spatiale des sources sonores, il convient de s'intéresser à l'effet de salle réellement synthétisé dans la pièce d'écoute. Celle-ci a en effet son acoustique propre qui va venir s'ajouter à l'effet de salle synthétique dans un processus de convolution énergétique naturel. Les procédures de compensation des réflexions précoces proposées dans ce document (cf. partie 3.2 et chapitre 6) n'ont pas l'ambition de transformer la salle d'écoute en une chambre anéchoïque. Ceci est en réalité *impossible* avec le type de système de reproduction utilisé. Nous avons mis l'accent sur le *possible*, c'est à dire la diminution du niveau des réflexions précoces du plan horizontal dans le plan horizontal, en dessous de la fréquence d'aliasing spatial.

Une compensation de l'effet de salle résiduel est alors nécessaire. Une telle procédure consiste à *modifier la description* de haut ou de bas niveau de l'effet de salle synthétique de manière à prendre en compte la présence l'effet de salle résiduel.

Jot (Jot, 1997) a proposé une telle technique dans le spatialisateur de l'IRCAM. Cette technique consiste en une déconvolution d'enveloppe énergétique. Son but est de "n'ajouter que ce qui manque". Nous en décrivons l'adaptation au cas de la WFS.

Cette procédure comprend trois étapes :

1. description de l'effet de salle résiduel ;
2. définition d'un objectif de reproduction ;
3. calcul de l'effet de salle à synthétiser pour obtenir le résultat souhaité.

La description de l'effet de salle résiduel (contexte) ainsi que de l'effet de salle cible est réalisée dans les quatre sections temporelles et les trois bandes de fréquence utilisées dans la description de bas-niveau du spatialisateur.

La synthèse d'un élément de la scène sonore en WFS implique de synthétiser le son direct par une source virtuelle et à diffuser les canaux d'effet de salle synthétiques issus du processeur d'effet de salle par un ensemble de haut-parleurs virtuels. Pour décrire le contexte, il convient de caractériser, par mesures de réponses impulsionnelles au niveau du point d'écoute de référence, le champ synthétisé dans la salle pour *toutes* les sources virtuelles ainsi que l'ensemble des haut-parleurs virtuels. A partir de ces mesures, on déduit la paramétrisation énergétique. Dans le cas de la WFS, le contexte *dépend* de la source virtuelle synthétisée (cf. partie 3.1).

La figure 8.8 illustre la convolution d'enveloppe énergétique naturelle pour la section temporelle R_1 de l'effet de salle cible avec la paramétrisation du contexte. Pour chaque bande de fréquence,

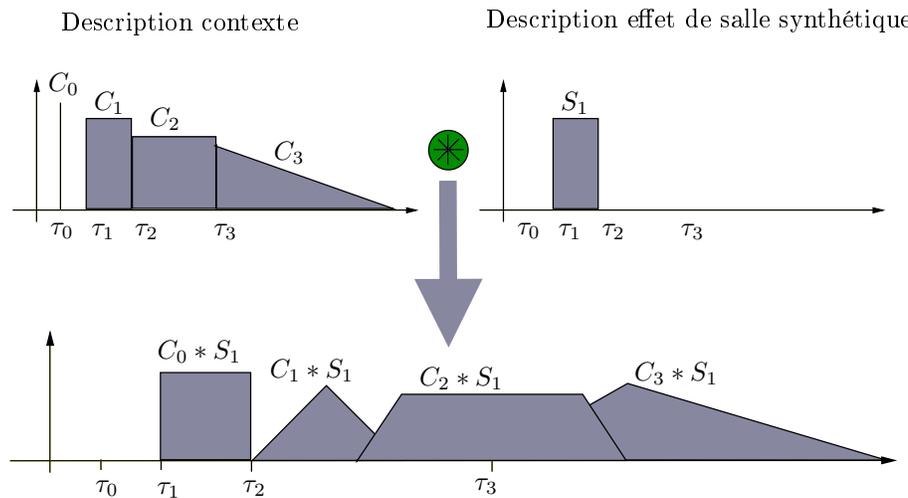


FIG. 8.8 – Transfert d'énergie par convolution d'enveloppe énergétique

il s'agit de résoudre l'équation

$$T = \mathbf{C}S, \quad (8.1)$$

où, T est composé des T_k énergies de la section temporelle k ($k = 0, 1, 2, 3$) de l'effet de salle cible ; la matrice \mathbf{C} est composée des $C_{j,k}$ contributions énergétiques de S_j pour T_k ; S est composé des S_j énergies de la section temporelle j ($j = 0, 1, 2, 3$). S correspond donc à la distribution énergétique temps/fréquence à synthétiser (Spat) pour réaliser à travers la pièce de restitution la qualité acoustique décrite par T .

La matrice \mathbf{C} correspond au transfert d'énergie réalisé dans la pièce de restitution entre la section j du Spat et la section k réalisée dans la salle. Le calcul de cette matrice s'effectue à partir de la caractérisation du contexte et de la modélisation des transferts d'énergie entre les sections temporelles par le processus de convolution énergétique. Les coefficients $C_{j,k}$ de la matrice \mathbf{C} sont donc nuls pour $j > k$ par principe de causalité. Le problème est alors équivalent à l'inversion d'une matrice triangulaire inférieure.

Les valeurs de S_j se référant à des niveaux énergétique, elles doivent être nécessairement positives. Ceci nécessite de contrôler l'inversion.

Cette procédure est intéressante par sa simplicité et l'approche utilisée est raisonnable. Cependant, on peut douter de son efficacité pour l'adaptation de la distribution de réflexions précoces à synthétiser. Ce problème s'avère bien plus délicat et nécessiterait une compréhension plus approfondie de la perception individuelle des réflexions dans une réponse impulsionnelle complexe. On pourrait utiliser des critères de masquage temporel et spatial afin de synthétiser uniquement les réflexions *nécessaires* à la synthèse de l'effet de salle cible et masquer les réflexions gênantes

qui pourraient subsister, notamment en dehors du plan horizontal.

Cette méthode n'a pas, pour le moment, été mise en œuvre pour la WFS.

8.3 Chaîne de production

Dans cette partie nous présentons la chaîne de production mise en place à l'IRCAM en collaboration avec sonic emotion ag. Cette chaîne de production a été développée en reprenant et éventuellement en adaptant des éléments existants à l'IRCAM, développés au cours du projet CARROUSO ou fournis par sonic emotion.

Nous donnons d'abord une description générale de l'architecture de cette chaîne de production avant d'insister plus particulièrement sur la partie concernant les interfaces de création et de manipulation de la scène (Corteel, 2004).

Cette chaîne de production sonore a ainsi permis de travailler avec plusieurs compositeurs (Nicolas Vérin, Rémy Gallichet, Philippe Schoeller, et Gilles Grand) pour la création d'esquisses sur le système WFS de l'IRCAM à partir de matériaux préexistants qui ont ainsi été "remixés" sur un système WFS.

8.3.1 Chaîne de production sonore

Dans cette partie, nous proposons une chaîne de production destinée à la création de scène sonore pour la WFS mais dont la généralité permettrait de la destiner à tout système fonctionnant sur le principe du codage de contenu. Dans notre cas, la WFS constitue le système de monitoring utilisé.

Dans un premier temps, nous montrons la structure générale de cette chaîne de production. Nous présentons, ensuite, les différentes interfaces utilisateur qui permettent de créer et de manipuler une scène sonore en WFS.

8.3.2 Structure de la chaîne de production

La figure 8.9 présente la structure générale de la chaîne de production. Elle est constituée de trois blocs qui composent les trois éléments de base de la synthèse et de la manipulation d'une scène sonore en WFS :

1. le bloc création du son direct, manipulation de la description de la scène sonore. Ce bloc, comme son nom l'indique, est chargé des opérations de synthèse/lecture/"transformations autres que spatiales" des flux audio associés aux sources qui composent la scène sonore. Il permet la création, l'édition et la manipulation des paramètres de description des caractéristiques spatiales de la scène sonore ;
2. le bloc synthèse de l'effet de salle, répartition sur canaux d'effet de salle ;
3. le bloc système WFS qui ne se charge que de la synthèse des sources et des haut-parleurs virtuels sur lesquels sont diffusés les canaux d'effet de salle.

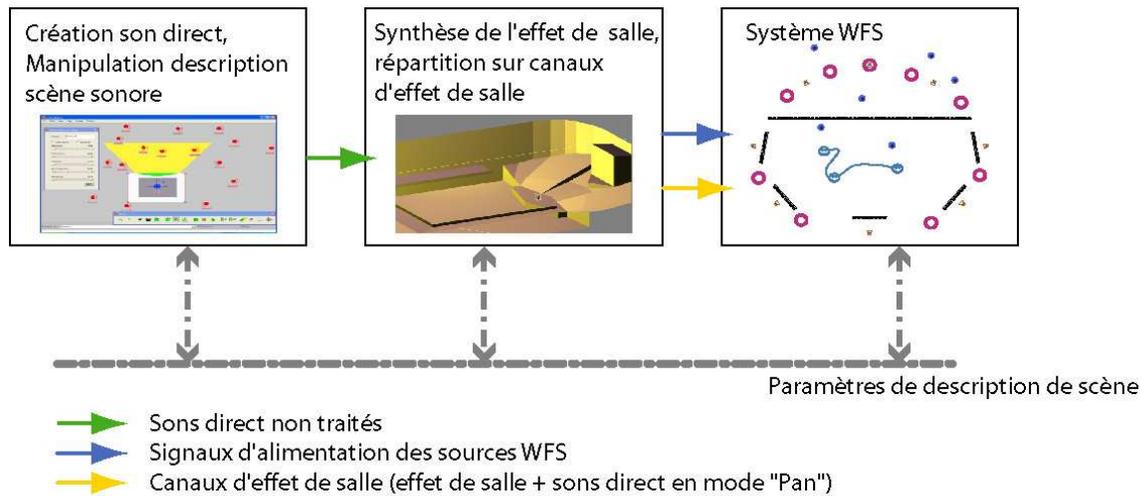


FIG. 8.9 – Structure générale de la chaîne de production pour la Wave Field Synthesis

Les trois blocs sont reliés au même réseau afin de communiquer les paramètres liés à la description de la scène sonore entre les différents blocs. Le réseau utilisé est le *zsonicNet* développé par *sonic emotion ag*. Il fonctionne comme une base de données répartie à laquelle l'ensemble des éléments de la chaîne accèdent. Ce réseau permet à l'ensemble des applications connectées d'accéder à l'ensemble des paramètres de description de la scène sonore, de modifier la valeur de ces paramètres, d'être informé de la modification de la valeur de ces paramètres par un système d'abonnement et éventuellement de résilier cet abonnement.

Cette forme s'avère particulièrement pertinente et bien plus efficace que l'utilisation d'une communication point à point. Ce type de communication nécessite en effet d'établir une communication directe entre les machines distantes. C'est le cas du protocole *OSC* (Wright *et al.*, 2003) qui permet la communication de messages par liaison *UDP* entre deux ordinateurs distants.

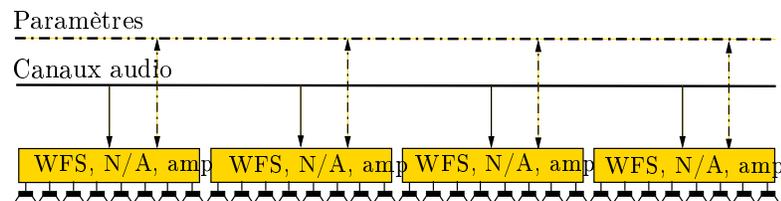


FIG. 8.10 – Architecture du système WFS

Le système WFS est constitué de haut-parleurs MAP. A chaque haut-parleur MAP est associé un ensemble de 8 amplificateurs et convertisseurs Numérique/Analogique. La convolution des filtres issus de l'égalisation multicanal 5 est assurée par des PC sur lesquels tourne le logiciel développé par *sonic emotion*. Chaque PC permet de gérer 2 haut-parleurs MAP (16 canaux) et 9 sources WFS avec égalisation multicanal.

Le système WFS possède une architecture distribuée. Chaque machine n'effectue que des traitements *locaux*, liés aux haut-parleurs auxquels elle s'adresse (cf. figure 8.10).

Le système de synthèse de l'effet de salle est assuré par une version spécifique du spatialisateur développée à l'IRCAM sous Max/MSP. En effet, dans la structure classique, les sorties du spatialisateur sont ainsi destinées, a priori, à alimenter *directement* les haut-parleurs. Nous avons ainsi "démonté" la structure afin d'alimenter les canaux d'effet de salle destinés aux haut-parleurs virtuelles.

Dans la version actuelle, la synthèse des sources en mode *Pan* est réalisée par panoramique d'intensité sur les différents canaux d'effet de salle.

8.3.3 Manipulation de la description de la scène sonore

Dans cette partie, on décrit le bloc de manipulation de la scène sonore et de synthèse du son direct. La figure 8.11 est une description détaillée de ce bloc. Il comporte :

1. un système de manipulation de l'organisation spatiale qui permet d'afficher et de modifier l'organisation spatiale de la scène sonore. Ce système doit pouvoir être installé sur une interface portable, par exemple un Tablet PC afin de permettre à l'auteur de manipuler l'organisation spatiale de la scène depuis n'importe quel emplacement de l'installation ;
2. un séquenceur qui permet de gérer l'aspect temporel de la scène sonore aussi bien au niveau des flux sonores que des paramètres de description de la scène sonore (position/ directivité/ orientation des sources, paramètres perceptifs du spatialisateur) . Ceux-ci sont accessibles en tant que paramètres d'automatisation ce qui permet d'enregistrer leurs variations temporelles. Le créateur de contenu peut modifier les paramètres en temps-réel et enregistrer leur évolution temporelle. Il peut ensuite les éditer et les modifier en temps différé afin d'assurer leur cohérence avec les événements sonores ;
3. un ensemble de traitements sonores autres que spatiaux peut être effectué, soit en partie dans le séquenceur (égalisation, compression, et autres effets), soit dans un module séparé, par exemple sous Max/MSP. Les traitements peuvent alors être aussi nombreux que le permet l'imagination du créateur de contenu. Dans ce bloc, un ensemble de canaux de son direct supplémentaires peuvent être ajoutés par traitement temps-réel de flux captés ou par synthèse ;
4. un ensemble d'interfaces haptiques pourraient être ajoutés afin de permettre une interaction gestuelle avec les paramètres spatiaux de la scène sonores. Les correspondances entre les paramètres gestuels et les paramètres spatiaux à mettre en œuvre sont à définir et dépassent largement le cadre de cette étude.

L'interface de manipulation de l'organisation spatiale de la scène sonore utilisée est l'interface ListenSpace (cf. figure 8.12(a)) développée par Olivier Delerue et Riita Väänänen dans le cadre du projet Carrouso. Cette interface offre une visualisation en deux dimensions de l'organisation spatiale (position des sources) de la scène sonore. Elle permet d'agir sur les paramètres perceptifs MPEG-4 (spatialisateur) pour chacune des sources. La description de la scène sonore peut être codée sous forme d'audio BIFS pour être exportée au format MPEG-4. Dans ce cadre, ListenSpace

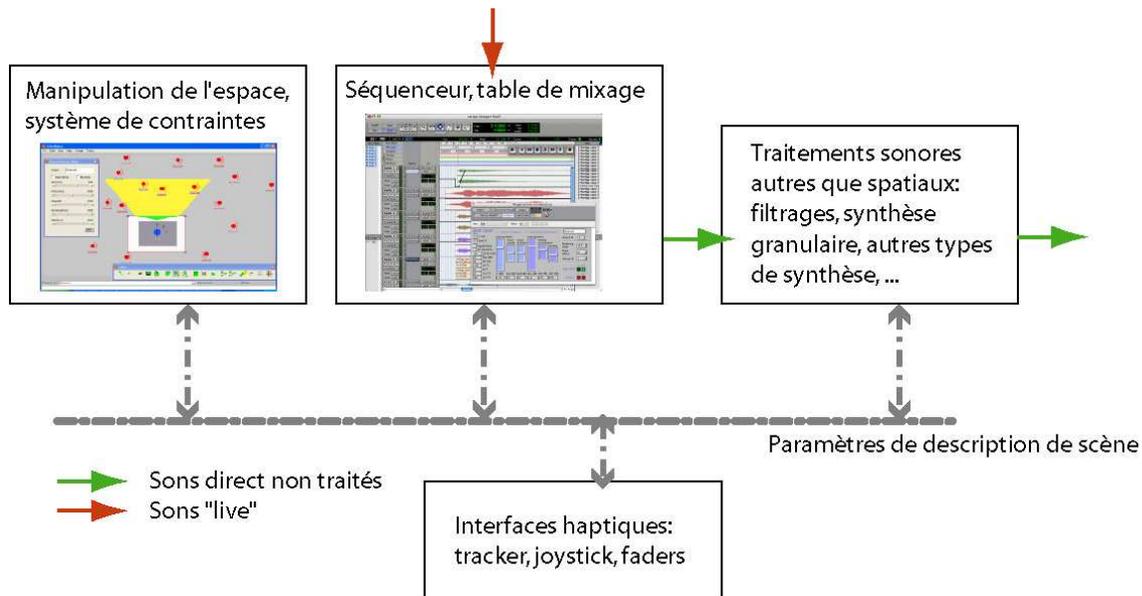


FIG. 8.11 – Bloc manipulation de la scène sonore et synthèse du son direct de la chaîne de production

génère automatiquement une interface utilisateur 2D compatible avec le format MPEG-4 qui est véhiculée avec le contenu (cf. figure 8.12(b)). ListenSpace, installé sur un Tablet PC, permet à l'utilisateur de se déplacer dans l'installation sonore et d'agir en temps réel sur la position des sources sonores.

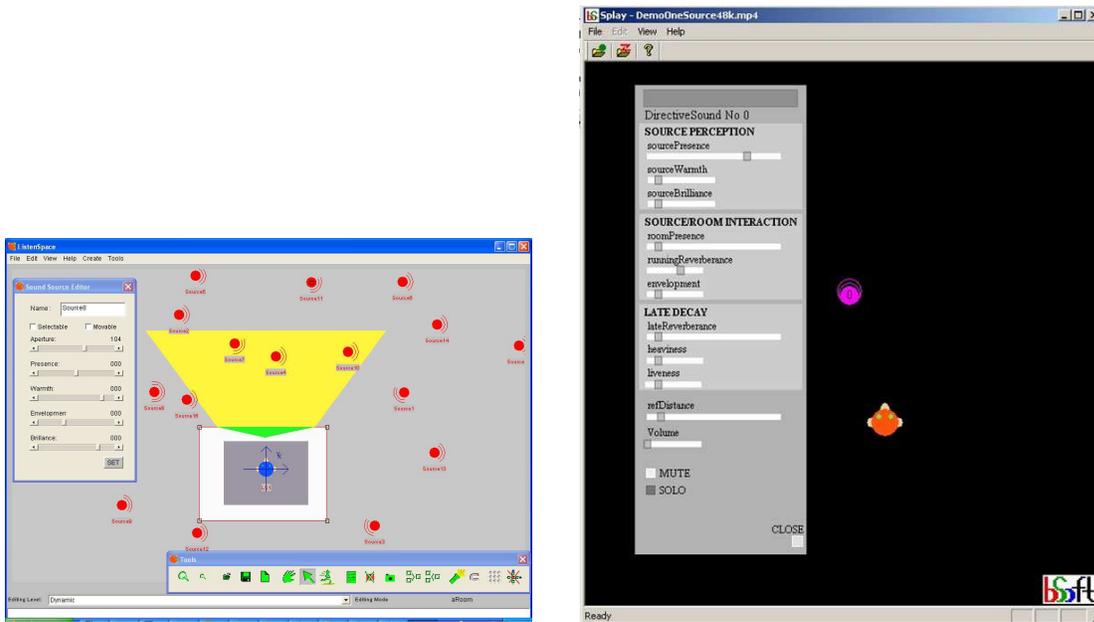
L'interface entre les paramètres de description de la scène sonore et l'automatisation du séquenceur est assuré par un ensemble de plugins développés par Manuel Poletti à l'IRCAM et repris par l'auteur.

Trois plugins ont été développés :

- le plugin SourceCtrl permet de manipuler l'ensemble des paramètres perceptifs attachés à chacune des sources sonores de la scène. En plus de la position des sources et des paramètres d'effet de salle, ce plugin permet le contrôle du mode de rendu ("WFS" ou "Pan") ;
- le plugin ReverbCtrl permet de régler les paramètres perceptifs liés à une structure de réverbération particulière ;
- le plugin OSC/zsonicNet permet la communication des paramètres par le réseau (soit lecture/écriture/abonnement/désabonnement sur le zsonicNet, soit envoi et réception de messages par OSC).

Le format de stockage choisi pour le moment n'est pas MPEG-4. Ceci est dû à l'indisponibilité de décodeur MPEG-4 qui permette la lecture des scènes sonores produites.

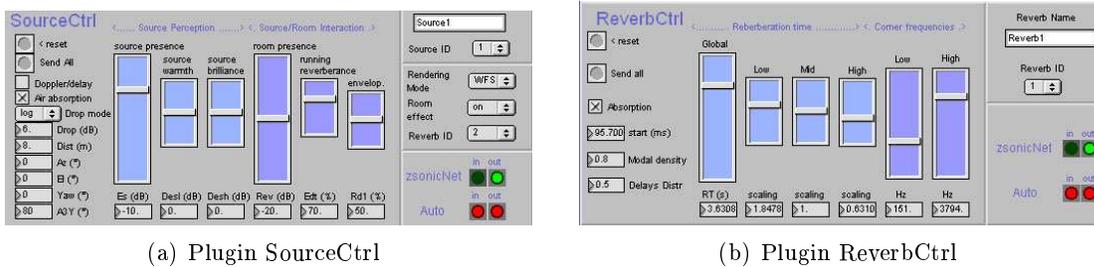
Une étape de "post-production" permet de stocker l'ensemble des fichiers sonores ainsi que l'évolution de la description de la scène sonore sous forme d'un fichier multi-piste associé à un fichier MIDI. Chaque paramètre est stocké comme une valeur de contrôleur. Un lecteur développé dans Max/MSP permet la conversion des valeurs MIDI en paramètres de description ainsi que



(a) ListenSpace

(b) Interface scène sonore MPEG-4 générée par ListenSpace rendus sur un décodeur MPEG-4 (BSoft)

FIG. 8.12 – Interface ListenSpace



(a) Plugin SourceCtrl

(b) Plugin ReverbCtrl

FIG. 8.13 – Plugins permettant le réglage des paramètres de description de la scène sonore depuis le séquenceur

la lecture du ou des fichiers sons ce qui permet une restitution des scènes sonores dans un mode sans séquenceur.

Conclusion

La Wave Field Synthesis est une technique de reproduction sonore holophonique qui a l'ambition de synthétiser les caractéristiques physiques d'un champ sonore dans une zone étendue de l'espace. Cette technique se base sur un ensemble d'approximations des principes physiques qui garantissent une telle reproduction en champ libre. La formulation originale de la WFS se réfère à la synthèse de sources virtuelles omnidirectionnelles par une distribution linéaire de haut-parleurs omnidirectionnels de taille finie.

En pratique, les haut-parleurs ont des caractéristiques de directivité qui s'écartent de ce modèle idéal et sont placés dans une pièce d'écoute. Les questions que nous nous sommes posé ont été de savoir quelle l'influence de l'ensemble de ces approximations, qu'elles soient théoriques ou pratiques, sur la qualité du champ acoustique synthétisé vis-à-vis de la perception ; comment peut-on compenser les "défauts" mis en évidence pour garantir la transparence de la reproduction dans une zone étendue de l'espace ; peut-on garantir la restitution et la manipulation d'une perspective sonore ; quels sont les paramètres qui permettent de la manipuler ; et enfin comment les mettre en œuvre dans un contexte de création musicale ?

L'approche que nous avons suivi a consisté à analyser, dans la partie I, la validité des approximations utilisées dans un cadre purement théorique et à compléter le formalisme de la WFS par deux extensions (sources à directivité arbitraire, chapitre 1, compensation des réflexions précoces du plan horizontal dans le plan horizontal, chapitre 3). Cette analyse nous a permis de mettre en évidence une évolution de la qualité des fronts d'onde au cours de la propagation dans ce cadre "idéal". Nous avons notamment proposé une définition de la fréquence d'aliasing spatial qui tient compte de la modification de cette fréquence limite en fonction de la position d'écoute. De plus, nous avons décrit de manière qualitative l'interaction du banc de haut-parleurs avec la salle de restitution. Nous avons montré que la distribution de réflexions précoces générée dépend de la source virtuelle synthétisée.

Ce dernier point mériterait d'être étendue à la description de l'énergie communiquée à la salle par le banc de haut-parleurs qui va notamment conditionner la réverbération tardive. Ce point, ainsi que l'utilisation de sources à directivités arbitraires, fait partie du travail de thèse de Terence Caulkins dans l'équipe acoustique des salles de l'IRCAM (Caulkins *et al.*, 2004).

Nous avons aussi proposé une méthode de compensation des réflexions précoces afin d'améliorer la transparence de la reproduction (chapitre 3) pour chaque source virtuelle. Nous avons vu que cette méthode pouvait se formuler sous forme d'extension de la Wave Field Synthesis. Nous avons aussi souligné la nécessité de réaliser une évaluation perceptive approfondie de cette technique ainsi que d'analyser son influence sur la réverbération tardive de la salle.

Une deuxième partie de notre travail a consisté à mettre en place des méthodes d'égalisation qui permettent de prendre en compte et de compenser les caractéristiques *réelles* de la reproduction dans l'ensemble de la zone d'écoute. Ces méthodes se basent sur des techniques d'inversion multicanal qui permettent d'assurer que le champ synthétisé est correct en un nombre fini de point de contrôles (chapitre 4). L'approche que nous avons suivie a consisté à la mise en place d'un

protocole de caractérisation et de mise en forme du problème afin de garantir, dans une certaine mesure, que la reproduction reste correcte dans l'ensemble de la zone d'écoute.

Ce type de méthode a été appliquée au contrôle du champ acoustique synthétisé en champ libre (chapitre 5) et, de manière distincte, à la compensation des réflexions sur les parois de la salle d'écoute (chapitre 6). Nous avons pu vérifier la validité de l'approche dans les deux cas mentionnés, d'abord dans un cadre idéal puis dans un cadre pratique. La méthode d'égalisation multicanal en champ libre a été présentée à de nombreuses reprises lors de démonstrations organisées dans le cadre de manifestations scientifiques et technologiques pendant et après le projet Carrouso.

Cette méthode semble pouvoir compenser une partie des défauts des caractéristiques de directivité des haut-parleurs. Il serait intéressant de déterminer précisément quelles sont les caractéristiques que cette méthode peut compenser et quelles sont celles qu'elle ne *peut pas* compenser.

Enfin, nous nous sommes intéressé à la validation perceptive des modalités d'utilisation de la perspective sonore (chapitre 7) et à la mise en place d'une chaîne de production qui permet de créer, de manipuler et de stocker des scènes sonores pour la WFS (chapitre 8).

En-dessous de la fréquence d'aliasing, nous avons vu que les indices binauraux de localisation de l'événement auditif correspondent à ceux qu'aurait fourni la source cible quelque soit la position de l'auditeur. Nous avons alors mis en évidence des problèmes potentiels pour la localisation des sources virtuelles WFS au-dessus de la fréquence d'aliasing. Nous avons défini un nouveau paramètre appelé *distance holophonique* (distance de la source virtuelle par rapport au centre de la zone d'écoute) dont nous avons vu qu'il permettait la gestion de la perspective sonore et dont la manipulation est relativement indépendante de la notion classique de distance (éloignement subjectif ou présence de la source) lié à l'effet de salle réel ou synthétique. La distance holophonique est un paramètre original de manipulation de la scène sonore *spécifique* à la WFS. Il conviendrait de mener des études plus approfondies sur la perception de ce paramètre et de voir de quelle mesure les impressions restituées résistent au passage d'une installation sonore à l'autre.

Nous avons aussi analysé de manière systématique les modalités de synthèse d'un effet de salle synthétique à partir des trois approches classiques (géométriques, physiques, perceptives). La loi d'atténuation des fronts d'onde synthétisés, le repliement de l'élévation dans le plan horizontal ainsi que l'acoustique résiduelle de la pièce ne permettent pas a priori d'avoir une restitution physique exacte de l'effet de salle dans l'ensemble de la pièce d'écoute. Nous suggérons alors de se baser sur une approche de type perceptive qui devrait s'attacher à recréer une impression spatiale similaire. Il conviendrait alors de réaliser une vérification systématique de la qualité acoustique restituée dans l'ensemble de la zone en fonction de la distribution spatio/temporelle de l'effet de salle. Ce point reste à notre avis à explorer au regard d'une compréhension plus approfondie de la perception de l'effet de salle par le système auditif.

Ces analyses ont conduit à la mise en place d'une chaîne de production sonore pour la WFS qui repose sur un réseau permettant d'accéder et/ou de modifier la description de la scène sonore depuis n'importe quel point de l'installation. Nous avons alors défini, adapté ou tout simplement

utilisé un ensemble d'outils qui permet au créateur de contenu de créer une scène sonore et d'agir sur l'ensemble des paramètres qui la décrivent. L'évolution temporelle de ces paramètres peut être éditée et modifiée sous forme de paramètres d'automatisation dans un séquenceur.

Les résultats présentés constituent les premiers travaux réalisés à l'IRCAM sur la Wave Field Synthesis. Nous avons pu bénéficier du contexte du projet européen Carrouso pour améliorer notre compréhension de la WFS et, proposer et implémenter des solutions originales dans un contexte pratique. Ceci a abouti à la mise en place d'un système installé dans le studio 4 de l'IRCAM.

Cette installation permanente a permis de présenter cette technique à nombre croissant de scientifiques, compositeurs, ingénieurs du son, ou même au grand public. Les réactions sont enthousiastes et nombre de compositeurs se sont montrés intéressés par cette technique pour la réalisation d'installations sonores ou par son utilisation en concert.

A notre avis, cette technique ouvre des champs d'investigation scientifiques et musicaux neufs qui se doivent d'être abordés dans un futur proche.

Annexe A

Validité de l'approximation de la phase stationnaire

A.1 Approximation de la phase stationnaire

L'approximation de la phase stationnaire est utilisée dans ce document ainsi que dans la littérature sur la Wave Field Synthesis de manière "intense". Il nous a ainsi semblé nécessaire d'examiner les conditions de validité de cette *approximation*.

A.1.1 Formulation

L'approximation de la phase stationnaire est utilisée afin d'estimer des intégrales du type :

$$I(\lambda) = \int_{-\infty}^{+\infty} f(x)e^{j\lambda\chi(x)} dx, \quad (\text{A.1})$$

pour lesquelles les fonctions f et χ prennent des valeurs des valeurs réelles. Dans la suite, on appellera f la fonction d'enveloppe et χ la fonction de phase.

On considère une fonction de phase dont la dérivé $\chi'(x)$ par rapport à x ne s'annule qu'en un unique point $x = x_0$ mais pour laquelle $\chi''(x_0) \neq 0$. La fonction de phase admet alors un extremum unique en $x = x_0$. C'est un maximum si $\chi''(x_0) < 0$, un minimum si $\chi''(x_0) > 0$. De chaque côté de $x = x_0$, la fonction de phase est ainsi strictement monotone et le terme $e^{j\lambda\chi(x)}$ est oscillatoire. Le point $x = x_0$ est alors appelé *point de phase stationnaire*. λ apparaît comme un facteur d'échelle qui conditionne la vitesse des oscillations.

Si f' et f'' sont continues dans $] -\infty + \infty[$, $f(x)$ est monotone autour de x_0 , et $f(x_0) \neq 0$, le développement limité autour de x_0 de l'expression sous l'intégrale donne :

$$f(x)e^{j\lambda\chi(x)} \approx f(x_0)e^{j\lambda\left[\chi(x_0) + \frac{(x-x_0)^2}{2}\chi''(x_0)\right]}. \quad (\text{A.2})$$

En évaluant

$$\hat{I}(\lambda) = \int_{-\infty}^{+\infty} f(x_0)e^{j\lambda\left[\chi(x_0) + \frac{(x-x_0)^2}{2}\chi''(x_0)\right]} dx, \quad (\text{A.3})$$

on obtient

$$\hat{I}(\lambda) = f(x_0)e^{j\lambda\chi(x_0)} \int_{-\infty}^{+\infty} e^{j\lambda\frac{(x-x_0)^2}{2}\chi''(x_0)} dx. \quad (\text{A.4})$$

En effectuant le changement de variable

$$\kappa = (x - x_0) \sqrt{\frac{|\lambda\chi''(x_0)|}{2}}, \quad (\text{A.5})$$

et à la condition que

$$\chi''(x_0) \neq 0, \quad (\text{A.6})$$

l'équation A.4 s'écrit

$$\hat{I}(\lambda) = f(x_0) \sqrt{\frac{2}{|\lambda\chi''(x_0)|}} e^{j\lambda\chi(x_0)} \int_{-\infty}^{+\infty} e^{j\text{sign}(\lambda\chi''(x_0))\kappa^2} d\kappa. \quad (\text{A.7})$$

Remarquons que

$$\int_{-\infty}^{+\infty} e^{j\pm\kappa^2} d\kappa = \sqrt{\pi} e^{\pm\frac{\pi}{4}}, \quad (\text{A.8})$$

on a ainsi,

$$\hat{I}(\lambda) = f(x_0) \sqrt{\frac{2\pi}{|\lambda\chi''(x_0)|}} e^{j[\lambda\chi(x_0) + \text{sign}(\lambda\chi''(x_0))\frac{\pi}{4}]}. \quad (\text{A.9})$$

On peut montrer (Guillemin & Sternberg, 1977) qu'à la condition unique que la fonction f soit C_∞ :

$$I(\lambda) = \hat{I}(\lambda) + O\left(\lambda^{-\frac{3}{2}}\right), \quad (\text{A.10})$$

ce qui établit la formule habituelle.

A.1.2 Domaine de validité

La validité de l'approximation de la phase stationnaire est établie pour de grandes valeurs de λ . Ceci est aussi le cas si les fonctions f et χ sont correctement approchées pour n'importe quelle valeur de x par l'expression de l'équation A.2 (cf. équation A.9). La fonction d'enveloppe doit ainsi être pratiquement constante *et* la fonction de phase est correctement estimée par un développement limité à l'ordre 2 aux alentours du point de phase stationnaire.

On considère habituellement que si la fonction d'enveloppe varie lentement en comparaison de la fonction χ , l'approximation de la phase stationnaire reste valable quelle que soit la valeur de λ ce qui étend le domaine de validité précédent.

L'idée sous-jacente est de considérer que les oscillations de la fonction $e^{j\lambda\chi(x)}$ s'annulent par intégration et il s'agit donc d'évaluer correctement l'expression $f(x)e^{j\lambda\chi(x)}$ autour du point de phase stationnaire où la fonction de phase devient localement "constante". Afin d'évaluer ce voisinage, considérons tout d'abord que la fonction de phase est correctement approchée par un développement limité d'ordre 2 autour de x_0 . On peut alors écrire,

$$I(\lambda) \approx \int_{-\infty}^{+\infty} f(x) e^{j\lambda\left[\chi(x_0) + \frac{(x-x_0)^2}{2}\chi''(x_0)\right]} dx, \quad (\text{A.11})$$

d'où,

$$I(\lambda) \approx e^{j\lambda\chi(x_0)} \int_{-\infty}^{+\infty} f(x) e^{j\lambda\frac{(x-x_0)^2}{2}\chi''(x_0)} dx, \quad (\text{A.12})$$

et,

$$I(\lambda) \approx e^{j\lambda\chi(x_0)} \sqrt{\frac{2\pi}{|\lambda\chi''(x_0)|}} \int_{-\infty}^{+\infty} f\left(\kappa\sqrt{\frac{2}{|\lambda\chi''(x_0)|}} + x_0\right) e^{j\kappa^2} d\kappa, \quad (\text{A.13})$$

où, κ est introduit par le changement de variable A.5. La taille de l'intervalle sur lequel la fonction f doit être pratiquement constante autour de x_0 dépend de λ et de $\chi''(x_0)$.

On donne une idée de la taille $\Delta x(\lambda, x_0)$ de cet intervalle en considérant les valeurs de x pour lesquelles $\kappa^2 = \frac{\pi}{2}$, le régime d'oscillations démarrant au delà de cette valeur,

$$\Delta x(\lambda, x_0) = 2\sqrt{\frac{\pi}{|\lambda\chi''(x_0)|}}. \quad (\text{A.14})$$

Ainsi, plus les valeurs de λ et de $\chi''(x_0)$ sont grandes et plus l'intervalle à considérer est petit.

En résumé, on peut considérer l'approximation de la phase stationnaire comme justifiée si $\forall x \in \left[x_0 - \sqrt{\frac{\pi}{|\lambda\chi''(x_0)|}}, x_0 + \sqrt{\frac{\pi}{|\lambda\chi''(x_0)|}} \right]$

$$\frac{\chi(x) - \left(\chi(x_0) + \chi''(x_0) * \frac{(x-x_0)^2}{2} \right)}{\left(\chi(x_0) + \chi''(x_0) * \frac{(x-x_0)^2}{2} \right)} \ll 1, \quad (\text{A.15})$$

et

$$\frac{f(x) - f(x_0)}{f(x_0)} \ll 1. \quad (\text{A.16})$$

A.2 Phase stationnaire et Wave Field Synthesis

Dans ce document, l'approximation de la phase stationnaire est utilisée à deux endroits afin de simplifier l'expression du champ rayonné par la distribution de sources secondaires et estimer le champ rayonné par :

- une colonne afin de passer d'une distribution planaire (Rayleigh) à une distribution linéaire (WFS) de sources secondaires (dimension *verticale*, dans les parties 1.2.1 et 1.3).
- la distribution linéaire de sources secondaires afin de définir la loi d'atténuation du champ sonore effectivement synthétisé, pour une source virtuelle donnée (dimension *horizontale*, dans la partie 2.2.1).

A.2.1 Identification des fonctions pour la Wave Field Synthesis

Dans les deux cas mentionnés (cf. figures 1.2 et 2.1), la fonction de phase

$$\chi_R^\psi(x, z) = -(\Delta r \pm r). \quad (\text{A.17})$$

χ_R^ψ présente une dépendance non seulement vis-à-vis de la position de la source ψ mais aussi par rapport à la position d'écoute R . Si ψ est dans Ω_ψ , il s'agit d'une somme ; si ψ est dans Ω_R (source focalisée), il s'agit d'une soustraction.

Dans le premier cas, x est fixé ($x = x_C$) et z varie entre $-\infty$ et $+\infty$. Dans le deuxième cas, z est fixé ($z = z_L$) et x varie entre $-\infty$ et $+\infty$. On doit donc évaluer la valeur de la dérivée seconde

de χ_R^ψ par rapport à x et à z aux points de phase stationnaire respectifs (dimension *verticale* : $z = z_L$; dimension *horizontale* : $x = x_0$, intersection de la ligne joignant la source et la position d'écoute avec la ligne de sources secondaires L). Ceci donne

$$\frac{d^2 \chi_R^\psi(x_C, z_L)}{dz^2} = - \left(\frac{1}{r_0} \pm \frac{1}{\Delta r_0} \right), \quad (\text{A.18})$$

et

$$\frac{d^2 \chi_R^\psi(x_0, z_L)}{dx^2} = - \cos^2 \theta_0 \left(\frac{1}{r_0} \pm \frac{1}{\Delta r_0} \right). \quad (\text{A.19})$$

Dans le cas 1 (dimension verticale), la fonction d'enveloppe f_1 est donnée par,

$$f_1(z) = S(\omega) \frac{jk \cos \theta_\psi}{16\pi^2 r \Delta r}. \quad (\text{A.20})$$

Pour la dimension horizontale, f_2 vaut

$$f_2(x) = S(\omega) \sqrt{\frac{jk}{2\pi}} g_\Psi(y_{R_{av}}) \frac{\cos \theta_\psi}{4\pi \sqrt{r} \Delta r}. \quad (\text{A.21})$$

A.2.2 Domaine de validité

La validité de l'approximation de la phase stationnaire dans le cas de la WFS s'exprime grâce aux équations A.15 et A.16. La réécriture de A.15 ne donne pas d'indications supplémentaires sur cette condition, par contre, A.16 se simplifie fortement :

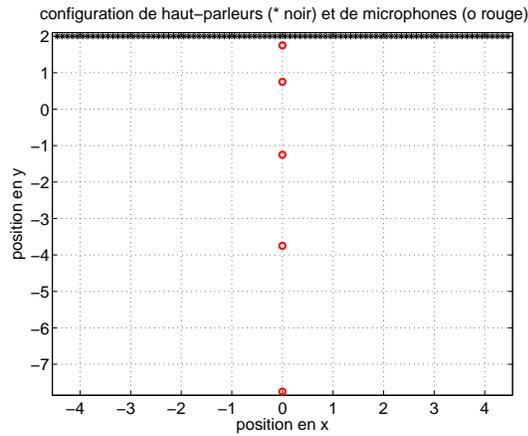
$$\begin{cases} \frac{r_0}{r} \frac{\Delta r_0}{\Delta r} \frac{\cos \theta_\psi}{\cos \theta_0} \ll 1, & \text{dimension verticale} \\ \sqrt{\frac{r_0}{r}} \frac{\Delta r_0}{\Delta r} \frac{\cos \theta_\psi}{\cos \theta_0} \ll 1, & \text{dimension horizontale} \end{cases}. \quad (\text{A.22})$$

Afin d'évaluer l'importance des erreurs commises par l'utilisation de l'approximation de la phase stationnaire, on réalise des simulations de l'erreur commise, sur la phase ainsi que sur l'enveloppe, pour un ensemble de points d'écoute (figure A.1). Ces simulations sont réalisées pour des sources virtuelles tests (source omnidirectionnelle focalisée en (0, 1), source omnidirectionnelle en (0, 3), source omnidirectionnelle en (3, 5), source omnidirectionnelle en (0, 20)).

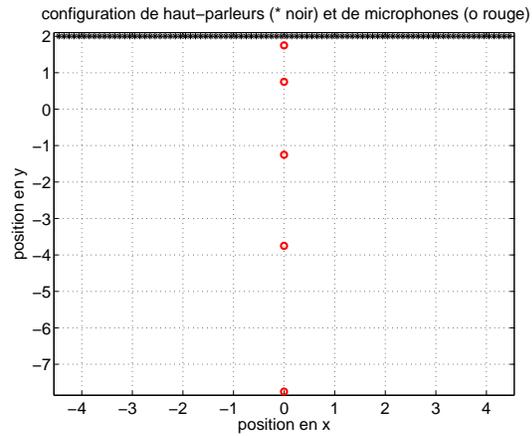
Les figures A.2.2 et A.2.2 donnent, respectivement, les erreurs commises sur l'estimation de la phase et de la fonction d'enveloppe, dans la dimension *horizontale*. Des résultats similaires sont obtenus pour la dimension *verticale*. Le paramètre de variation entre les différentes courbes d'une même figure est la position en y du point d'écoute. L'intervalle de confiance est donné pour chacune des courbes (* noir, 100 Hz ; o rouge, 500 Hz ; + bleu, 2000 Hz).

De manière générale, on remarque que l'estimation sur la fonction de phase reste robuste dans l'intervalle de confiance pour l'ensemble des sources omnidirectionnelles situées en dehors du sous-espace de reproduction, y compris aux basses fréquences.

Pour la source focalisée, l'estimation s'écarte de manière significative de la valeur réelle dans



(a) Configuration pour les sources extérieures à l'espace de reproduction



(b) Configuration pour les sources focalisées

FIG. A.1 – Configurations de tests pour l'estimation des erreurs introduites par l'approximation de la phase stationnaire.

l'intervalle de confiance pour pratiquement l'ensemble des positions d'écoute, particulièrement en basse fréquence (cf. figure). On s'attend donc à une mauvaise estimation du champ sonore par l'approximation de la phase stationnaire pour ce type de source.

L'estimation sur la fonction d'enveloppe s'avère beaucoup moins précise et robuste que celle de la fonction de phase. Elle est particulièrement mauvaise en grande proximité du banc de haut-parleurs et en basses fréquences mais devient plus précise lorsque la source s'éloigne et/ou que la fréquence augmente. Elle ne semble exacte, à l'intérieur de l'intervalle de confiance, que pour les hautes fréquences et les distances importantes. Pour la source focalisée, elle semble également mauvaise quelque soit la distance du point d'écoute. Seul l'intervalle de confiance se rétrécit, tendant vers une approximation plus juste.

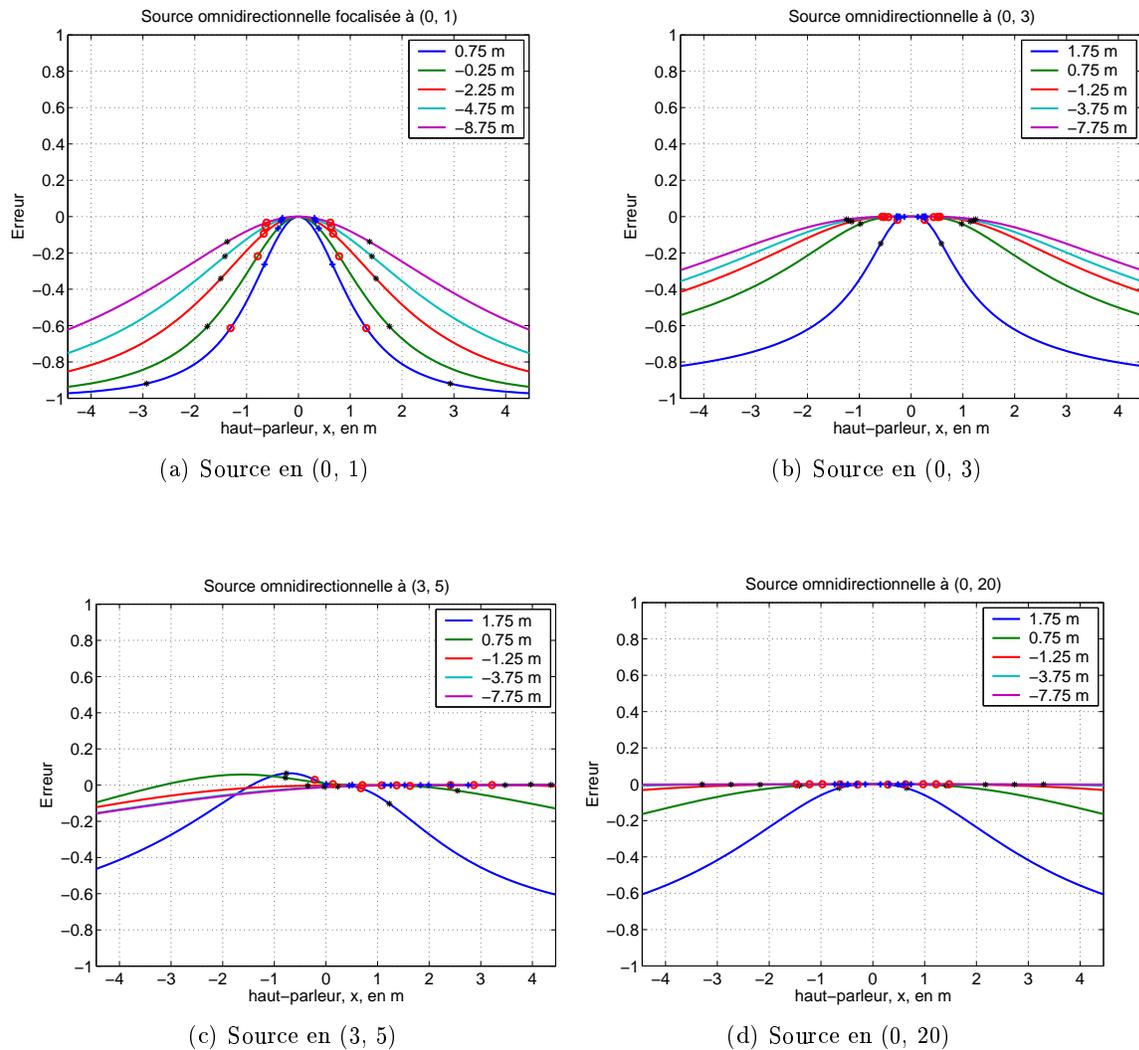


FIG. A.2 – Erreur de l'évaluation de la phase pour la dimension horizontale. Le paramètre de variation entre les différentes courbes est la position en y du point d'écoute. L'intervalle de confiance est donné pour chacune des courbes : * noir, 100 Hz ; o rouge, 500 Hz ; + bleu, 2000 Hz

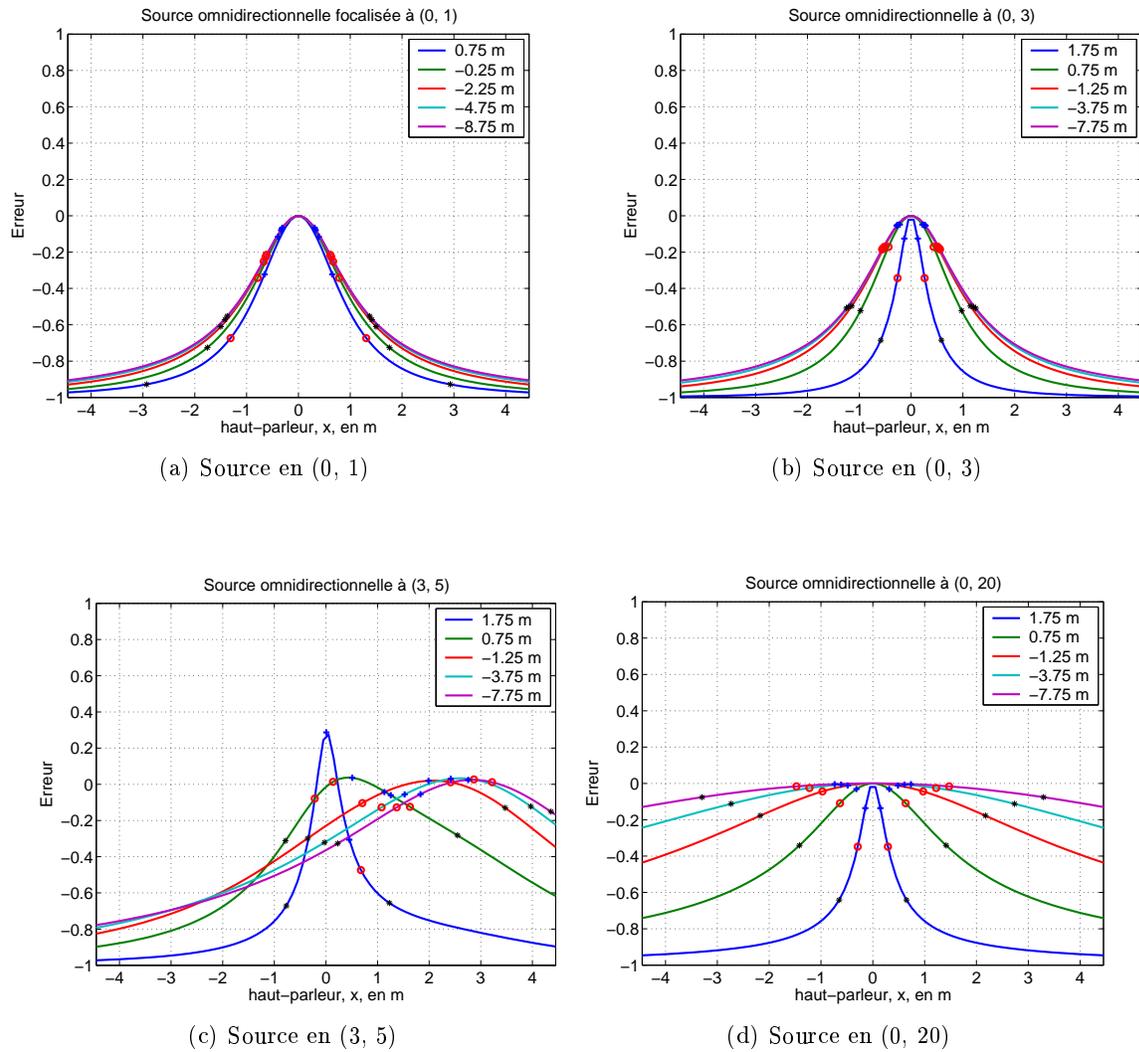


FIG. A.3 – Erreur de l'évaluation de la fonction d'enveloppe pour la dimension horizontale. Le paramètre de variation entre les différentes courbes est la position en y du point d'écoute. L'intervalle de confiance est donné pour chacune des courbes : * noir, 100 Hz ; o rouge, 500 Hz ; + bleu, 2000 Hz

Annexe B

Description de l'algorithme MFAP

B.1 Algorithme adaptatif MFAP pour l'inversion multicanal

Nous présentons ici les détails de l'algorithme MFAP utilisé pour le calcul des filtres pour les méthodes d'égalisation multicanal proposées. On donne dans un premier temps le détail de l'algorithme MFAP monocanal puis on en donne une version multicanal.

B.1.1 Détail de l'algorithme MFAP simple

La version de l'algorithme MFAP que nous proposons d'utiliser a été développée par Liu *et al.* (Liu *et al.*, 1996) pour une application d'annulation d'écho. Nous en proposons une adaptation liée à la mise en forme particulière de notre problème.

Cet algorithme est dérivé de la classe APA et permet de calculer itérativement le filtre d'égalisation. La particularité de cet algorithme est d'utiliser les N valeurs précédentes du signal d'erreur afin d'améliorer la vitesse de convergence de la solution vers un minimum d'erreur global. Le vecteur \mathbf{h} de taille I représente le filtre que l'on calcule de manière itérative. En définissant :

$$\mathbf{s}_n = \begin{bmatrix} s(n) & s(n-1) & \dots & s(n-I+1) \end{bmatrix}^T, \quad (\text{B.1})$$

le signal d'erreur à minimiser donné par

$$e(n) = d(n) - \mathbf{s}_n^T \mathbf{h}_{n-1}, \quad (\text{B.2})$$

où, $d(n)$ est le signal cible à l'instant n . On introduit la matrice $\mathbf{S}_n = \begin{bmatrix} \mathbf{s}_n & \mathbf{s}_{n-1} & \dots & \mathbf{s}_{n-N+1} \end{bmatrix}$ de taille $I * N$ et le vecteur $\mathbf{d}_n = \begin{bmatrix} d(n) & d(n-1) & \dots & d(n-N+1) \end{bmatrix}^T$. Les différentes étapes de l'algorithme sont alors :

$$\mathbf{e}(n) = \mathbf{d}(n) - \mathbf{S}_n^T \mathbf{h}_{n-1}, \quad (\text{B.3})$$

$$\mathbf{P}(n) = [\mathbf{S}_n^T \mathbf{S}_n + \delta \mathbf{I}_N]^{-1}, \quad (\text{B.4})$$

$$\boldsymbol{\epsilon}_n = \mu \mathbf{P}_n \mathbf{e}_n, \quad (\text{B.5})$$

$$\mathbf{h}_n = \mathbf{h}_{n-1} + \mathbf{S}_n \boldsymbol{\epsilon}_n, \quad (\text{B.6})$$

où, δ est un paramètre de régularisation pour le calcul de l'inverse de la matrice d'autocorrélation du signal s , μ est un scalaire appelé facteur de relaxation qui devra être compris entre 0 et 2 pour que l'algorithme converge.

La complexité de l'algorithme est de $2IN + O(N^2)$ multiplications par itération. Dans (Gay & Tavathia, 1995), une version rapide de cet algorithme est proposée sous le nom de FAP (Fast Affine Projection) permettant de réduire la complexité de l'algorithme à $2I + 21N$.

La première idée importante est d'actualiser \mathbf{e}_n en utilisant l'approximation suivante :

$$\mathbf{e}_n = \begin{bmatrix} e(n) \\ (1 - \mu) \bar{\mathbf{e}}_{n-1} \end{bmatrix}, \quad (\text{B.7})$$

où, $\bar{\mathbf{e}}_{n-1}$ correspondant aux $N - 1$ premiers éléments de \mathbf{e}_{n-1} .

La seconde modification concerne le calcul de la matrice \mathbf{P}_n . Ce calcul est réalisé par des techniques de prédictions forward et backward inspirées de l'algorithme RLS à fenêtre glissante pour calculer efficacement ϵ_n . Cette méthode est toutefois complexe et on recommande un calcul itératif basé sur le lemme d'inversion de matrice.

Enfin, le vecteur \mathbf{h}_n n'est pas calculé exactement. On utilise au contraire le vecteur $\hat{\mathbf{h}}_n$ qui est lié à \mathbf{h}_n de la manière suivante :

$$\mathbf{h}_n = \hat{\mathbf{h}}_n + \mu \bar{\mathbf{S}}_n \bar{\eta}_n, \quad (\text{B.8})$$

$\bar{\mathbf{S}}_n$ correspond aux $N - 1$ premières colonnes de \mathbf{S}_n et $\bar{\eta}_n$ aux $N - 1$ premiers éléments du vecteur η_n dont la mise à jour est réalisée avec,

$$\eta_n = \begin{bmatrix} 0 \\ \bar{\eta}_{n-1} \end{bmatrix} + \epsilon_n. \quad (\text{B.9})$$

Le vecteur $\hat{\mathbf{h}}_n$ peut alors aussi être calculé par façon récursive

$$\hat{\mathbf{h}}_n = \hat{\mathbf{h}}_{n-1} + \mu \eta_{n,N} \mathbf{S}_{n-(N-1)}, \quad (\text{B.10})$$

où, $\eta_{n,N}$ est le $N^{\text{ème}}$ élément de η_n .

On se rend compte que l'utilisation de $\hat{\mathbf{h}}_n$ est tout à fait satisfaisante et il n'est donc pas utile de tenter d'accéder à nouveau à \mathbf{h}_n .

Enfin, le calcul du signal d'erreur $e(n)$ peut être obtenue en substituant B.8 dans B.2 :

$$e(n) = d(n) - \hat{\mathbf{h}}_{n-1}^T \mathbf{s}_n - \mu \bar{\eta}_{n-1}^T \bar{\mathbf{S}}_{n-1}^T \mathbf{s}_n \quad (\text{B.11})$$

D'autres simplifications peuvent être apportées pour le calcul du dernier terme de B.11 en introduisant le vecteur $\mathbf{r}_n = \bar{\mathbf{S}}_n^T \mathbf{s}_n$ qui est calculé de manière récursive

$$\mathbf{r}_n = \mathbf{r}_{n-1} + \bar{\xi}_{n-1}^T \mathbf{s}_n - \bar{\xi}_{n-I}^T \mathbf{s}_{n-I-1}, \quad (\text{B.12})$$

où, $\bar{\xi}_n$ correspond aux $N - 1$ premiers éléments du vecteur ξ_n défini par :

$$\xi_n = \begin{bmatrix} s(n) & s(n-1) & \dots & s(n-N+1) \end{bmatrix}^T. \quad (\text{B.13})$$

Les modifications apportées à cet algorithme concernent le calcul récursif de \mathbf{P}_n et de ϵ_n . Ceci permet d'éviter d'utiliser les solutions proposées dans les versions rapides des algorithmes RLS qui présentent des problèmes d'instabilité. La solution proposée consiste à utiliser le lemme d'inversion de matrice pour à jour ces données.

On exprime l'inverse de \mathbf{P}_n sous la forme :

$$\mathbf{P}_n^{-1} = \sum_{i=n-(I-1)}^n \xi_i \xi_i^T + \delta \mathbf{I}_N. \quad (\text{B.14})$$

Un calcul récursif de \mathbf{P}_n^{-1} peut alors être donné

$$\mathbf{P}_n^{-1} = \mathbf{Q}_n - \xi_{n-I} \xi_{n-I}^T, \quad (\text{B.15})$$

$$\mathbf{Q}_n = \mathbf{P}_{n-1}^{-1} + \xi_n \xi_n^T. \quad (\text{B.16})$$

Le lemme d'inversion de matrice est donné pour deux matrices A et B définies positives de taille $N * N$, C , une matrice de taille $N * M$, et D une matrice définie positive de taille $M * M$. Si,

$$A = B + CD^{-1}C^T, \quad (\text{B.17})$$

alors,

$$A^{-1} = B^{-1} - B^{-1}C(D + C^T B^{-1}C)^{-1}C^T B^{-1} \quad (\text{B.18})$$

On peut alors l'utiliser pour le calcul de \mathbf{P}_n à partir de B.15. On obtient

$$\mathbf{P}_n = \mathbf{Q}_n^{-1} - \beta \mathbf{b}_n \mathbf{b}_n^T, \quad (\text{B.19})$$

où, $\mathbf{b}_n = \mathbf{Q}_n^{-1} \xi_{n-I}$, $\beta = (-1 + \xi_{n-I}^T \mathbf{b}_n)^{-1}$ et \mathbf{Q}_n^{-1} peut être obtenu en utilisant le lemme d'inversion de matrice dans B.16 :

$$\mathbf{Q}_n^{-1} = \mathbf{P}_{n-1} - \alpha \mathbf{a}_n \mathbf{a}_n^T \quad (\text{B.20})$$

où, $\mathbf{a}_n = \mathbf{P}_{n-1} \xi_n$, $\alpha = (1 + \xi_n^T \mathbf{a}_n)^{-1}$.

La complexité de l'approche présentée pour le calcul de \mathbf{P}_n est de $3N^2 + 7N$ multiplications par itération. Cependant, les valeurs de N utilisées seront bien inférieures à la longueur I des filtres que l'on cherche à synthétiser et qui conditionneront de façon bien plus importante la complexité globale de l'algorithme.

Une fois \mathbf{P}_n obtenue, ϵ_n peut être calculé comme dans l'équation B.5. Si μ a pour valeur 1, l'équation B.5 s'écrit

$$\epsilon_n = e(n) \mathbf{P}_{n,1}, \quad (\text{B.21})$$

où, $\mathbf{P}_{n,1}$ est la première colonne de \mathbf{P}_n . Le calcul de ϵ_n avec B.21 pour des valeurs de μ légèrement différentes de 1 (ex. 0.7) n'a que peu d'effet sur les résultats obtenus.

L'algorithme MFAP que nous proposons d'utiliser peut alors être résumé comme suit :

0. Initialisation :

$$\mathbf{P}_0 = \delta^{-1} \mathbf{I}_N, \quad \mathbf{r}_0 = \mathbf{0}, \quad \eta_0 = \mathbf{0}. \quad (\text{B.22})$$

1. Mise à jour de \mathbf{P}_n :

$$\mathbf{a}_n = \mathbf{P}_{n-1}\xi_n, \quad (\text{B.23})$$

$$\alpha = (1 + \xi_n^T \mathbf{a}_n)^{-1}, \quad (\text{B.24})$$

$$\mathbf{q}_n = \mathbf{P}_{n-1}\xi_{n-I}, \quad (\text{B.25})$$

$$\mathbf{b}_n = \mathbf{q}_n - \alpha (\mathbf{a}_n^T \xi_{n-I}) \mathbf{a}_n \quad (\text{B.26})$$

$$\beta = (-1 + \xi_{n-I}^T \mathbf{b}_n)^{-1} \quad (\text{B.27})$$

$$\mathbf{P}_n = \mathbf{P}_{n-1} - \alpha \mathbf{a}_n \mathbf{a}_n^T - \beta \mathbf{b}_n \mathbf{b}_n^T \quad (\text{B.28})$$

2. Calcul de $e(n)$:

$$\mathbf{r}_n = \mathbf{r}_{n-1} + \bar{\xi}_{n-1}^T \mathbf{s}_n - \bar{\xi}_{n-I}^T \mathbf{s}_{n-I-1} \quad (\text{B.29})$$

$$e(n) = d(n) - \hat{\mathbf{h}}_{n-1}^T \mathbf{s}_n - \mu \bar{\eta}_{n-1}^T \mathbf{r}_{n-1} \quad (\text{B.30})$$

3. Mise à jour de $\hat{\mathbf{h}}_n$ et de $\bar{\eta}_n$:

$$\epsilon_n = \mu e(n) \mathbf{P}_{n,1} \quad (\text{B.31})$$

$$\eta_n = \begin{bmatrix} 0 \\ \bar{\eta}_{n-1} \end{bmatrix} + \epsilon_n \quad (\text{B.32})$$

$$\hat{\mathbf{h}}_n = \hat{\mathbf{h}}_{n-1} + \mu \eta_{n,N} \mathbf{s}_{n-(N-1)} \quad (\text{B.33})$$

La complexité totale de cet algorithme est de $2L + 3N^2 + 12N$ multiplications par itérations.

B.1.2 Dérivation de l'algorithme MFAP multicanal

Afin de dériver une version multicanal de l'algorithme MFAP nous allons utiliser la mise en forme des données proposée au paragraphe 4.2.2.

Il s'agit donc maintenant de trouver les filtres $\hat{\mathbf{h}}^i$ ($i = [1 \dots M]$) qui permettent d'obtenir à l'instant n les L signaux désirés $d_j(n)$ ($j = [1 \dots L]$) qui forment le vecteur $\mathbf{d}_n^{multi} = [d_1(n) \ d_2(n) \ \dots \ d_L(n)]^T$ à partir d'un signal d'entrée unique $x(n)$ filtré par les M $\hat{\mathbf{h}}^i$ puis la matrice C de fonctions de transfert du système de taille $L * M$.

Tout d'abord, il convient de remplacer le vecteur $\hat{\mathbf{h}}$ de taille I par le vecteur $\hat{\mathbf{w}}$ de taille $M * I$ tel que

$$\hat{\mathbf{h}}^i(n) = \hat{\mathbf{w}}((n-1) * M + i) \quad i = [1 \dots M]. \quad (\text{B.34})$$

On définit de même $\mathbf{s}_n^{multi} = [C(n) \ C(n-1) \ \dots \ C(n-I+1)]^T$ qui correspond à la matrice de taille $(M * I) * L$ des signaux de sortie du système pendant un intervalle de temps de taille I dans le cas où $x(n)$ est une impulsion de dirac.

La matrice des signaux d'erreur $\mathbf{e}_n^{multi} = [e_1(n) \ e_2(n) \ \dots \ e_L(n)]^T$, avec $e_i(n) \ j =$

$\begin{bmatrix} 1 & \dots & L \end{bmatrix}$ les signaux d'erreur obtenus aux L positions de microphones, est donnée par :

$$\mathbf{e}_n^{multi} = \mathbf{d}_n^{multi} - \mathbf{s}_n^{multi,T} \hat{\mathbf{w}}_{n-1}, \quad (\text{B.35})$$

où, $\mathbf{s}_n^{multi,T}$ est la matrice transposée de \mathbf{s}_n^{multi} .

On définit alors, en référence au cas simple, $\mathbf{S}_n^{multi} = \begin{bmatrix} \mathbf{s}_n & \mathbf{s}_{n-1} & \dots & \mathbf{s}_{n-N+1} \end{bmatrix}$ de taille $(M * I) * (L * N)$ qui donne les N erreurs passées $\mathbf{E}^{multi}(n) = \begin{bmatrix} \mathbf{e}_n^{multi,T} & \mathbf{e}_{n-1}^{multi,T} & \dots & \mathbf{e}_{n-N+1}^{multi,T} \end{bmatrix}^T$ par

$$\mathbf{E}_n^{multi} = \mathbf{d}_n^{multi} - \mathbf{S}_n^{multi,T} \hat{\mathbf{w}}_{n-1}. \quad (\text{B.36})$$

Afin de dériver l'algorithme complet, il ne reste plus qu'à introduire

$$\xi_n^{multi} = \begin{bmatrix} C(n) & C(n-1) & \dots & C(n-N+1) \end{bmatrix}^T. \quad (\text{B.37})$$

L'algorithme complet peut alors être résumé de la manière suivante :

0. Initialisation :

$$\mathbf{P}_0^{multi} = \delta^{-1} \mathbf{I}_{L*N}, \quad \mathbf{r}_0^{multi} = \mathbf{0}, \quad \eta_0^{multi} = \mathbf{0} \quad (\text{B.38})$$

1. Mise à jour de \mathbf{P}_n^{multi} :

$$\mathbf{a}_n^{multi} = \mathbf{P}_{n-1}^{multi} \xi_n^{multi} \quad (\text{B.39})$$

$$\alpha^{multi} = \left(\mathbf{I}_L + \xi_n^{multi,T} \mathbf{a}_n^{multi} \right)^{-1} \quad (\text{B.40})$$

$$\mathbf{q}_n^{multi} = \mathbf{P}_{n-1}^{multi} \xi_{n-I}^{multi} \quad (\text{B.41})$$

$$\mathbf{b}_n^{multi} = \mathbf{q}_n^{multi} - \alpha^{multi} \left(\mathbf{a}_n^{multi,T} \xi_{n-I}^{multi} \right) \mathbf{a}_n^{multi} \quad (\text{B.42})$$

$$\beta^{multi} = \left(-\mathbf{I}_L + \xi_{n-I}^{multi,T} \mathbf{b}_n^{multi} \right)^{-1} \quad (\text{B.43})$$

$$\mathbf{P}_n^{multi} = \mathbf{P}_{n-1}^{multi} - \alpha^{multi} \mathbf{a}_n^{multi} \mathbf{a}_n^{multi,T} - \beta^{multi} \mathbf{b}_n^{multi} \mathbf{b}_n^{multi,T} \quad (\text{B.44})$$

2. Calcul de $e(n)$:

$$\mathbf{r}_n^{multi} = \mathbf{r}_{n-1}^{multi} + \bar{\xi}_{n-1}^{multi,T} \mathbf{s}_n^{multi} - \bar{\xi}_{n-I}^{multi,T} \mathbf{s}_{n-I-1}^{multi} \quad (\text{B.45})$$

$$\mathbf{e}_n^{multi} = \mathbf{d}_n^{multi} - \hat{\mathbf{w}}_{n-1}^T \mathbf{s}_n^{multi} - \mu \bar{\eta}_{n-1}^{multi,T} \mathbf{r}_{n-1} \quad (\text{B.46})$$

3. Mise à jour de $\hat{\mathbf{w}}_n$ et de $\bar{\eta}_n^{multi}$:

$$\epsilon_n^{multi} = \mu \mathbf{e}_n^{multi} \mathbf{P}_{n,1}^{multi} \quad (\text{B.47})$$

$$\eta_n^{multi} = \begin{bmatrix} 0 \\ \bar{\eta}_{n-1}^{multi} \end{bmatrix} + \epsilon_n^{multi} \quad (\text{B.48})$$

$$\hat{\mathbf{w}}_n = \hat{\mathbf{x}}_{n-1} + \mu \eta_{n,N}^{multi} \mathbf{s}_{n-(N-1)} \quad (\text{B.49})$$

Remarquons que $\bar{\xi}^{multi}$ ne correspond plus au $N - 1$ premiers éléments de ξ_n^{multi} mais aux $(N-1)*L$. De même, $\eta_{n,N}^{multi}$ correspond aux L derniers éléments de η_n^{multi} . Enfin $\mathbf{P}_{n,1}^{multi}$ correspond aux L premières colonnes de \mathbf{P}_n^{multi} .

Le point sensible de l'algorithme, en terme de complexité, est l'inversion des matrices \mathbf{a}_n^{multi} et \mathbf{b}_n^{multi} qui sont maintenant de taille $L * L$. Toutefois, la complexité globale de l'algorithme reste similaire, pour les tailles de filtre considérées, à celle de l'algorithme LMS multicanal.

References

- Abhayapala, T. D. 1999. *Modal Analysis et Synthesis of Broadband Nearfield Beamforming Arrays*. Ph.D. thesis, The Australian National University, Canberra, Australie.
- Abhayapala, T. D., & Ward, D. B. 2002 (Mai). Theory et Design of High Order Sound Field Microphones Using Spherical Microphone Array. *In : IEEE ICASSP-02*.
- Angus, J. A. S. 2000 (Février). Distributed Mode Loudspeaker Radiation Mechanisms. *In : 108th Convention of the Audio Engineering Society*. Preprint Number 5164.
- Barron, M. 1974. *The effects on early reflections on subjective acoustical quality in concert halls*. Ph.D. thesis, University of Southampton, Royaume Uni.
- Barron, M., & Marshall, A. H. 1981. Spatial Impression Due to Early Lateral Reflections in Concert Halls : the Derivation of a Physical Measure. *Journal of Sound and Vibration*, **70**(2), 211–231.
- Baskind, A. 2003. *Modèles et Méthodes de Description Spatiale des Scènes Sonores, application aux enregistrements binauraux*. Ph.D. thesis, Université de Paris VI, Paris, France.
- Bauck, J. L., & Cooper, D. H. 1996. Generalized Transaural Stereo and Applications. *Journal of the Audio Engineering Society*, **44**(9), 683–705.
- Bech, S. 1995. Timbral aspects of reproduced sound in small rooms I. *Journal of the Acoustical Society of America*, **97**(3).
- Bech, S. 1996. Timbral aspects of reproduced sound in small rooms II. *Journal of the Acoustical Society of America*, **99**(6).
- Bech, S. 1998. Spatial aspects of reproduced sound in small rooms. *Journal of the Acoustical Society of America*, **103**, 434–445.
- 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**, 2764–2778.
- Berkhout, A. J., de Vries, D., & Sonke, J. J. 1997. Array technology for acoustic wave field analysis in enclosures. *Journal of the Acoustical Society of America*, **102**, 2757–2770.
- Blauert, J. 1999. *Spatial Hearing, The Psychophysics of Human Sound Localization*. MIT Press.
- Blauert, J., & Cobben, W. 1978. Some Considerations on Binaural Cross Correlation Analysis. *Acustica*, **32**(2), 96–104.
- Boone, M. M., & Verheijen, E. N. G., & van Tol, P. F. 1995. Spatial sound field reproduction by wave field synthesis. *Journal of the Audio Engineering Society*, **43**, 1003–1012.
- Boone, M. M., & de Bruijn, W. P. J. 2000 (Février). On the Applicability of Distributed Mode Loudspeakers Panels for Wave Field Synthesis Sound Reproduction. *In : 108th Convention of the Audio Engineering Society*. Preprint Number 5165.

- Boone, M. M., & de Bruijn, W. P. J. 2003 (Mars). Improving speech intelligibility in teleconferencing by using Wave Field Synthesis. *In : 114th Convention of the Audio Engineering Society*. Preprint Number 5800.
- Boone, M.M., & Verheijen, E.N.G. 1998. Sound reproduction applications with wave field synthesis. *In : 104th Convention of the Audio Engineering Society*. Preprint Number 4689.
- Brix, S., Sporer, T., & Plogsties, J. 2001 (Mai). CARROUSO, A European approach to 3D audio. *In : 110th Convention of the Audio Engineering Society*. Preprint Number 5314.
- Bruneau, M. 1998. *Manuel d'Acoustique Fondamentale*. Hermes.
- Bucklein, R. 1985. The Audibility of Frequency Response Irregularities. *Journal of the Audio Engineering Society*, **29**, 126–131. Traduction anglaise d'un article paru en 1962.
- Burgtof, W. 1961. Untersuchung zur Wahnehmbarkeit verzögerter Schallsignale. *Acustica*, **11**, 97–111.
- Caulkins, T. 2003 (Juillet). *Restitution du effet de salle cohérent avec le lieu d'écoute dans le cadre de la Wave Field Synthesis*. M.Phil. thesis, Dea Atiam, Université Paris 6.
- 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*.
- Cerveau, L. 1999. *Couplage temps réel d'outils acoustique prévisionnelle et de dispositifs d'auralisation*. Ph.D. thesis, Université de Paris VI, Paris, France.
- Clarkson, P. M., Mourjopoulos, J., & Hammond, J. K. 1985. Spectral, Phase and Transient Equalization for Audio Systems. *Journal of the Audio Engineering Society*, **33**(3), 127–132.
- Corteel, E. 2000 (Juin). *Wave Field Synthesis (Hologophonie) : Possibilités offertes et limitations*. M.Phil. thesis, Ecole National Supérieur des Télécommunications.
- Corteel, E. 2001 (Juillet). *Egalisation de bancs de haut parleurs par techniques adaptatives multi voies, application à la reproduction de sources virtuelles en zone étendue*. M.Phil. thesis, DEA ATIAM.
- Corteel, E. 2004. Création et manipulation de scènes sonores pour la wave field synthesis. *Cahiers de Louis Lumière*, **2**.
- 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.
- Cotterell, P. 2002. *On The Theory of the Second-Order Soundfield Microphone*. Ph.D. thesis, University of Reading, UK.
- Craven, P.G., & Gerzon, M. A. 1977. *Coincident Microphone Simulation Covering Three Dimensional Space and Yielding Various Directional Outputs*. U.S. Patent number 4,042,779.
- Damaske, P. 1971. Head-related two-channel stereophony with loudspeaker reproduction. *Journal of the Acoustical Society of America*, **50**, 1109–1115.
- Daniel, J. 2000. *Représentation de Champs Acoustiques, Application à la Transmission et à la Reproduction de Scènes Sonores Complexes dans un Contexte Multimédia*. Ph.D. thesis, Université de Paris VI, Paris, France.

- Daniel, J. 2003 (Juin). Spatial Sound Encoding Including Near Field Effect : Introducing Distance Coding Filters and a Viable, New Ambisonic Format. *In : 23th International Conference of the Audio Engineering Society*.
- Daniel, J., & Moreau, S. 2004 (Mars). Design refinement of high order ambisonics microphones - Experiments with a 4th order prototype. *In : CFA/DAGA 04*.
- de Bruijn, W. P. J., & Boone, M. M. 2003 (Mars). Application of Wave Field Synthesis in life-size videoconferencing. *In : 114th Convention of the Audio Engineering Society*. Preprint Number 5801.
- de Vries, D. 1996. Sound Reinforcement by Wave Field Synthesis : Adaptation of the Synthesis Operator to the Loudspeakers Directivity Characteristics. *Journal of the Audio Engineering Society*, **44**(12), 1120–1131.
- Deer, J. A., Bloom, P. J., & Preis, D. 1985. Perception of Phase Distortion of All-pass Filters. *Journal of the Audio Engineering Society*, **33**(10), 782–786.
- Delerue, O. 2004. *Spatialisation du son et programmation par contrainte : le système MusicSpace*. Ph.D. thesis, Université de Paris VI, Paris, France.
- Elliott, S. J., & Nelson, P. A. 1989. Multiple-Point Equalization in a Room Using Adaptive Digital Filters. *Journal of the Audio Engineering Society*, **37**(11), 899–907.
- Elliott, S. J., Bhatia, L. P., Deghan, F. S., Fu, A. H., Stewart, M. S., & Wilson, D. W. 1994. Practical Implementation of Low-Frequency Equalization Using Adaptive Digital Filters. *Journal of the Audio Engineering Society*, **42**(12), 988–998.
- Fielder, L. D. 2003. Analysis of Traditional end Reverberation-Reducing Methods of Room Equalization. *Journal of the Audio Engineering Society*, **51**(1/2), 912–932.
- Flanagan, J. L., Johnston, J. D., Zahn, R., & Elkow, G. W. 1985. Computer-steered microphone arrays for sound transduction in large rooms. *Journal of the Acoustical Society of America*, **78**(5), 1508–1518.
- Flikkema, P. G. 2002 (Juin). An Algebraic Theory of 3D Sound Synthesis with Loudspeakers. *In : 22nd International Conference of the Audio Engineering Society*.
- Gansler, T., & Benesty, J. 2000. Stereophonic acoustic echo cancellation et two-channel adaptive filtering : an overview. *International Journal of Adaptive Control et Signal Processing*, **14**, 565–586.
- Gardner, B., & Martin, K. D. 1994. *HRTF Measurements of a KEMAR Dummy-Head Microphone*. MIT Media Lab Perceptual Computing - Technical Report number 280.
- Gardner, W. G. 1995. Efficient convolution without input-output delay. *Journal of the Audio Engineering Society*, **43**(3).
- Gay, S. L., & Tavathia, S. 1995. The Fast Affine Projection Algorithm. *Pages 3023–3026 of : IEEE International Conference on Acoustics et Speech Signal Processing*.
- Gerzon, M. A. 1973. Periphony : With-Height Sound Reproduction. *Journal of the Audio Engineering Society*, **21**(1), 2–10.
- Gerzon, M. A. 1985. Ambisonics in Multichannel Broadcasting et Video. *Journal of the Audio Engineering Society*, **33**(11), 859–871.
- Glasgal, R. 2001 (Septembre). AMBIOPHONICS, Achieving Physiological Realism in Music Recording and Reproduction. *In : 111th Convention of the Audio Engineering Society*.
- Guillemin, V., & Sternberg, S. 1977. *Geometric Asymptotics*. AMS.

- Harris, N., & Hawksford, M. O. J. 1997 (September). The Distributed Mode Loudspeaker as a Broad Band Acoustic Radiator. *In : 103rd Convention of the Audio Engineering Society*. Preprint Number 4526.
- Hatziantoniou, P. D., & Mourjopoulos, J. N. 2003 (Mars). Results for Room Acoustics Equalisation Based on Smoothed Responses. *In : 114th Convention of the Audio Engineering Society*. Preprint Number 5779.
- Hershkowitz, R., & Durlach, N. 1962. Frequency-correlation functions of frequency responses in rooms. *Journal of the Acoustical Society of America*, **34**(12), 1819–1823.
- Horbach, U., & Corteel, E. 2003. *Array Loudspeaker System for Sound Synthesis*. U.S. Patent. en cours d'examen.
- Horbach, U., Corteel, E., & de Vries, D. 2002 (Juin). Spatial Audio Reproduction using Distributed Mode Loudspeaker Arrays. *In : 21st International Conference of the Audio Engineering Society*.
- Horbach, U., Corteel, E., Pellegrini, R. S., & Hulsebos, E. 2002 (Juillet). Real-Time rendering of Dynamic Scenes using Wave Field Synthesis. *In : IEEE International Conference on Multimedia, ICME*.
- Howe, R. M., & Hawksford, M. O. J. 1991 (Octobre). Methods of Local Room Equalization et Their Effect over the Listening Area. *In : 91st Convention of the Audio Engineering Society*. Preprint Number 3138.
- 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.
- Hulsebos, E., Schuurmans, T., de Vries, D., & Boone, M. 2003 (Mars). Circular microphone array for discrete multichannel audio recording. *In : 114th Convention of the Audio Engineering Society*. Preprint Number 5716.
- Ise, S. 1999. A principle of Sound Field Control Based on the Kirchhoff-Helmholtz Integral Equation and the Theory of Inverse System. *ACUSTICA*, **85**, 78–87.
- Jessel, M. 1973. *Acoustique théorique : propagation et holophonie*. Masson.
- Johansen, L. G., & Rubak, P. 1996 (April). The Excess Phase in Loudspeaker/Room Transfer Functions : Can It Be Ignored in Equalization Tasks ? *In : 100th Convention of the Audio Engineering Society*. Preprint Number 4181.
- Jot, J. M. 1997 (Septembre). Efficient Models for Reverberation and Distance Rendering in Computer Music and Virtual Audio Reality. *In : ICMC 97*.
- Jot, J. M., & Warusfel, O. 1995. Spat : A Spatial Processor for Musicians and Sound Engineers. *In : CIARM 95*.
- Jot, J. M., Larcher, V., & Warusfel, O. 1995 (Mars). Digital Signal Processing Issues in the Context of Binaural and Transaural Stereophony. *In : 98th Convention of the Audio Engineering Society*. Preprint Number 3980.
- 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, E. 1995. *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, France.
- Karjalainen, M., Piirilä, E., Jäärvinen, A., & Huopaniemi, J. 1999. Comparison of Loudspeaker Equalization Methods Based on DSP Techniques. *Journal of the Audio Engineering Society*, **47**(1/2).
- Kirkeby, O., & Nelson, P. A. 1998. The Šstereo dipoleŤ - a virtual source imaging system using two closely spaced loudspeakers. *Journal of the Audio Engineering Society*, **46**(5).
- Kirkeby, O., & Nelson, P. A. 1999. Digital Filter Design for Inversion Problems in Sound Reproduction. *Journal of the Audio Engineering Society*, **47**(7/8).
- Kirkeby, O., Nelson, P. A., Hamada, H, & et Orduña Bustamente F. 1998. Fast deconvolution of multichannel systems using regularization. *IEEE Transactions on Speech and Audio Processing*, **6**, 189–194.
- Koya, D. 2000 (Mai). *Aural Phase Distortion Detection*. M.Phil. thesis, Université de Miami.
- Kuhn, C., Pellegrini, R. S., Leckschat, D., & Corteel, E. 2003 (Octobre). An approach to mixing and Miking of Music Ensembles Using Wave Field Synthesis. *In : 115th Convention of the Audio Engineering Society*.
- Kuhn, G. F. 1977. Model for the interaural time differences in the azimuthal plane. *Journal of the Acoustical Society of America*, **62**(1), 157–167.
- Kuo, S. M., & Morgan, D. R. 1999. Active Noise Control : a tutorial review. *Proceedings of the IEEE*, **87**(6), 943–972.
- 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 National Supérieur 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. 2004 (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 réalité virtuelle*. Ph.D. thesis, Université de Paris VI, Paris, France.
- Lavandier, C. 1989. *Validation perceptive d'un modèle objectif de caractérisation de la qualité acoustique des salles*. Ph.D. thesis, Université du Maine, Le Mans, France.
- Liu, Q. G., Champagne, B., & Ho, K. C. 1996. The Fast Affine Projection Algorithm. *Pages 354–347 of : IEEE Digital Signal Processing Workshop*.
- Martin, K. D. 1995 (June). *A Computational Model os Spatial Hearing*. M.Phil. thesis, Massachusetts Institue of Technology.
- Melchior, F., Brix, S., Sporer, T., Röder, T., & Klehs, B. 2003 (Juin). Wave Field Synthesis in combination with 2D Video Projection. *In : 24th International Conference of the Audio Engineering Society*.

- Meyer, J., & Agnello, T. 2003 (Octobre). Spherical Microphone Array for Spatial Sound Recording. *In : 115th Convention of the Audio Engineering Society*. Preprint Number 5975.
- Meyer, J., & Elko, G. W. 2002 (Mai). A Highly Scalable Spherical Microphone Array Based on an Orthonormal Decomposition of the Soundfield. *In : IEEE ICASSP-02*.
- Mitra, S. K., & Kaiser, J. F. 1993. *Handbook for Digital Signal Processing*. Wiley-Interscience.
- Miyoshi, M., & Kaneda, Y. 1988. Inverse filtering of room acoustics. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **36**(2), 145–152.
- Mourjopoulos, J. N. 1994. Digital Equalization of Room Acoustics. *Journal of the Audio Engineering Society*, **42**(11), 884–900.
- MPEG, Consortium. 2000. *MPEG-4 Systems standard, 2nd edition*. ISO/IEC 14496-1 :2000.
- Nelson, P. A., Hamada, H., & Elliott, S. J. 1992. Adaptive Inverse Filters for Stereophonic Sound Reproduction. *IEEE Transactions on Signal Processing*, **40**(7), 1621–1632.
- Nelson, P. A., Orduña Bustamente, F., & Hamada, H. 1995. Inverse Filter Design and Equalization Zones in Multichannel Sound Reproduction. *IEEE transactions on speech and audio processing*, **3**(3), 185–192.
- Nelson, P. A., Orduña Bustamente, F., & Hamada, H. 1996. Multichannel Signal Processing Techniques in the Reproduction of Sound. *Journal of the Audio Engineering Society*, **44**(11), 973–989.
- 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., & et Depollier, C. 1998. Lending an ear to the dimensionality of space. *In : ICA/ASA '98*.
- Noguès, 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 Investigations of Inverse Filtering. *Journal of the Audio Engineering Society*, **52**(10), 1003–1024.
- Olive, S. E., & Toole, F. E. 1989. The Detection of Reflections in Typical Rooms. *Journal of the Audio Engineering Society*, **37**(6/7).
- Olive, S. E., Schuck, P. L., Ryan, J. G., Sally, S. L., & Bonneville, M. E. 1997. The Detection Thresholds of Resonances at Low Frequencies. *Journal of the Audio Engineering Society*, **45**(3), 116–128.
- Patterson, R. D. 1994. The sound of a Sinusoid : Spectral Models. *Journal of the Acoustical Society of America*, **96**, 1409–1418.
- Pedersen, J. A. 2003 (Juin). Adaptive Bass Control - the ABC Room adaptation system. *In : 23th International Conference of the Audio Engineering Society*.
- Pellegrini, R. S. 2001. *A Virtual Listening Room as an Application of Auditory Virtual Environments*. Ph.D. thesis, Ruhr-Universität, Bochum, Allemagne.
- Perrot, D. R., Briggs, R., & Perrot, S. 1970. Binaural fusion : its limits as defined by signal duration and signal onset. *Journal of the Acoustical Society of America*, **47**(2), 565–568.
- Polack, J.-D. 1988. *La transmission de l'énergie sonore dans les salles*. Ph.D. thesis, Université du Maine, Le Mans, France.
- Polack, J.-D., Morckerken, J.-P., Pellerin, G., & Parzy, B. 2002 (Mai). Vented Box Geometry and Low-Frequency Reproduction : The Aerodynamical Approach. *In : 112th Convention of the Audio Engineering Society*. Preprint Number 5522.

- Poletti, M. A. 2000. A Unified Theory of Horizontal Holographic Sound Systems. *Journal of the Audio Engineering Society*, **48**(12), 1155–1182.
- Preis, D. 1982. Phase Distortion and Phase Equalization in Audio Signal Processing - A Tutorial Overview. *Journal of the Audio Engineering Society*, **30**(11), 774–794.
- Preis, D., & Bloom, P. J. 1984. Perception of Phase Distortion in Anti-Aliasing Filters. *Journal of the Audio Engineering Society*, **32**(11), 842–848.
- Prokofieva, E. 2003 (Mars). Effect of porous material on the diffusivity of an un baffled DML panel. In : *114th Convention of the Audio Engineering Society*. Preprint Number 5735.
- Pulkki, V., Karjalainen, M., & Huopaniemi, J. 1999. Analysing Virtual Sound Source Attributes Using a Binaural Auditory Model. *Journal of the Audio Engineering Society*, **47**(4).
- Putnam, W., Rocchesso, D., & Smith, J. O. 1995 (Octobre). A Numerical Investigation of the Invertibility of Room Transfer Functions. In : *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*.
- Rubak, P. 2004 (Juin). Coloration in Room Impulse Response. In : *Joint Baltic-Nordic Acoustics Meeting 2004, BNAM2004*.
- Rubak, P., & Johansen, L. G. 2000 (Février). Design and Evaluation of Digital Filters Applied to Loudspeaker/Room Equalization. In : *108th Convention of the Audio Engineering Society*. Preprint Number 5172.
- Schroeder, M. 1996. The Schroeder frequency revisited. *Journal of the Acoustical Society of America*, **99**, 3240–3241.
- Slaney, M. 1993. *An Efficient Implementation of the Patterson-Holdsworth Filter Bank*. Apple Computer Inc. - Technical Report number 35.
- Sommerfeld, A. 1964. *Optics*. New York : Academic Press.
- Sonke, J. J. 2000. *Variable Acoustics by Wave Field Synthesis*. Ph.D. thesis, TU Delft, Delft, Pays Bas.
- 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. 2004 (Mars). Efficient Active Listening Room Compensation for Wave Field Synthesis. In : *116th Convention of the Audio Engineering Society*.
- Start, E. W. 1997. *Direct Sound Enhancement by Wave Field Synthesis*. Ph.D. thesis, TU Delft, Delft, Pays Bas.
- 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*.
- Toole, F. E., & Olive, S. E. 1988. The modification of Timbre by Resonances : Perception and Measurements. *Journal of the Audio Engineering Society*, **36**, 122–142.
- Urban, M., Heil, C., & Bauman, P. 2003. Wavefront Sculpture Technology. *Journal of the Audio Engineering Society*, **51**(10), 912–932.
- Van Veen, B. D., & Buckley, K. M. 1988. Beamforming : A Versatile Approach to Spatial Filtering. *IEEE ASSP*, Avril, 4–24.

- van Zon, R. 2003 (Octobre). *Room compensation for Wave Field Synthesis using MAP loudspeakers*. M.Phil. thesis, TU Delft.
- van Zon, R., Corteel, E., de Vries, D., & Warusfel, O. 2004 (Mars). Multi-Actuator Panel (MAP) loudspeakers : how to compensate for their mutual reflections? *In : 116th Convention of the Audio Engineering Society*.
- Verheijen, E. N. G. 1997. *Sound Reproduction by Wave Field Synthesis*. Ph.D. thesis, TU Delft, Delft, Pays Bas.
- Väänänen, 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*.
- Vogel, P. 1993. *Application of Wave Field Synthesis in room acoustics*. Ph.D. thesis, TU Delft, Delft, Pays Bas.
- Ward, D. B., & Williamson, R. C. 1999 (Août). Beamforming for a source located in the interior of a sensor array. *In : 5th International Symposium on Signal Processing and Applications, ISSPA99*, vol. 2.
- Warusfel, O., & Viaud-Delmon, I. 2003 (Octobre). Contribution of Interactive 3D Sound to Presence. *In : Presence 2003, 6th International Workshop on Presence*.
- 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*.
- Wilson, R. 1989. Equalization of Loudspeaker Drive Units Considering Both On- and Off-Axis Responses. *Journal of the Audio Engineering Society*, **39**(3), 127–139.
- Wittek, H. 2004 (Mars). Spatial perception in Wave Field Synthesis rendered sound fields : Distance of real and virtual nearby sources. *In : CFA/DAGA 04*.
- Wright, M., Freed, A., & Momeni, A. 2003. OpenSound Control : State of the Art 2003. *In : Conference on New Interfaces for Musical Expression (NIME-03)*.