



HAL
open science

Validation théorique de la correction des réflexions sur la base d'une représentation en harmoniques sphériques.

Romain Deprez, Rozenn Nicol, Emmanuel Friot

► To cite this version:

Romain Deprez, Rozenn Nicol, Emmanuel Friot. Validation théorique de la correction des réflexions sur la base d'une représentation en harmoniques sphériques.. 10ème Congrès Français d'Acoustique, Apr 2010, Lyon, France. hal-00550906

HAL Id: hal-00550906

<https://hal.science/hal-00550906>

Submitted on 31 Dec 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

10ème Congrès Français d'Acoustique

Lyon, 12-16 Avril 2010

Validation théorique de la correction des réflexions sur la base d'une représentation en harmoniques sphériques

Romain Deprez^{1,2}, Rozenn Nicol¹, Emmanuel Friot²

¹ Orange Labs, 2 av Pierre Marzin, F-22307 Lannion Cedex, {romain.deprez, rozenn.nicol}@orange-ftgroup.com

² Laboratoire de Mécanique et d'Acoustique, 31 chemin Joseph Aiguier, F-13402 Marseille Cedex 20, friot@lma.cnrs-mrs.fr

A l'heure de l'introduction dans le grand public de systèmes de spatialisation sonore offrant une expérience d'écoute plus riche que les systèmes stéréophoniques classiques, il apparaît que les perturbations dues au lieu d'écoute augmentent de façon non-linéaire avec le nombre de haut-parleurs. En effet, le jeu des interférences et des réflexions sur les murs engendre potentiellement des incohérences entre la localisation cible et la localisation effectivement perçue des sources virtuelles. La robustesse de la qualité d'écoute et de spatialisation au positionnement des haut-parleurs devient également critique. Les solutions classiques actuelles de correction de systèmes acoustiques consistent généralement en une inversion canal par canal des réponses impulsionnelles, réalisée sous la forme d'un filtrage. Des études ont aussi été menées afin de créer un environnement actif, mais cela nécessite le fonctionnement permanent de nombreux microphones. L'étude présentée vise, aux travers de simulations, à mettre en œuvre et valider une méthode basée sur une stratégie d'inversion appliquée dans le domaine spatio-fréquentiel. L'analyse spatiale repose sur le format Higher Order Ambisonics (HOA) qui offre la possibilité de créer des filtres spatiaux permettant une correction différenciée selon l'angle. L'étude est réalisée à partir de réponses impulsionnelles calculées par un modèle d'acoustique virtuelle. Différents dispositifs d'écoute sont examinés. L'évaluation des performances du traitement de correction est menée à travers un jeu de critères dérivés de mesures de la qualité acoustique d'une salle.

1 Introduction

Aujourd'hui les contenus audio spatialisés sont de plus en plus largement proposés au format multicanal 5.1. Dans le même temps, de nouveaux formats (tels que notamment les formats 6.1, 7.1, 10.2 ou 22.2) mettant en avant une spatialisation enrichie sont en voie d'émergence. Ces différentes solutions ont en commun de reposer sur un nombre croissant de haut-parleurs. Les progrès technologiques (miniaturisation des composants, nouvelles technologies de transducteurs, réseau de haut-parleurs...) rendent possible la mise en œuvre d'équipements multi haut-parleurs de plus en plus complexes, y compris pour le grand public. Ce dernier, de son côté, multiplie les occasions d'apprécier les attraits de la spatialisation sonore, devenant ainsi demandeur d'une expérience d'écoute plus riche, ce qui l'amène à intégrer dans ses installations des équipements audio multicanaux. Dès lors, ses attentes en termes de qualité sont aussi plus fortes, compte tenu du coût (à la fois économique et en termes de contrainte) de l'équipement.

La conséquence de cette évolution est qu'on se retrouve confronté à la généralisation d'un problème qui se manifestait hier de façon plus "marginale" : l'optimisation de la qualité de restitution sonore par un système multi haut-parleurs dans une salle dont les caractéristiques acoustiques ne sont pas contrôlées. La cause première est l'augmentation du nombre de haut-parleurs qui viennent solliciter et alimenter l'effet de salle dont l'interaction avec la scène sonore restituée n'est alors

plus négligeable et d'autant plus difficile à contrôler, en comparaison d'un système stéréophonique. Du point de vue de la salle, les sources physiques sont les haut-parleurs, c'est à dire que les réflexions générées sont cohérentes avec la position des haut-parleurs et n'ont en revanche aucune relation avec les réflexions qu'auraient créées les sources virtuelles constituant la scène sonore à reproduire (cf figure 1). L'effet de salle du lieu d'écoute est ainsi incohérent avec l'information spatiale des ondes directes associées aux sources virtuelles. Dans le cas d'un système stéréophonique, ce problème peut être négligé car les sources virtuelles sont situées sur le segment constitué par les deux haut-parleurs [1]. Par contre, dans le cas d'un système multicanal, les sources virtuelles peuvent être recrées dans n'importe quelle direction et à n'importe quel distance dans le plan des haut-parleurs [2]. La disjonction devient alors potentiellement notable. Pour corriger ce problème, l'idéal serait d'annuler les réflexions induites par les haut-parleurs dans la salle. En pratique nous nous contenterons de chercher à les réduire sur une zone suffisamment étendue. Il s'agit de la première étape du traitement associé à la configuration d'écoute (dont la spécification inclut les caractéristiques de la salle et la disposition des haut-parleurs) qui constitue le sujet de la présente étude. La seconde étape concerne la synthèse d'un effet de salle (correspondant ou non à la salle d'écoute) cohérent avec les sources virtuelles. Un autre aspect porte sur la correction de l'onde directe afin de prendre en compte et de compenser d'éventuelles erreurs de positionnement

des haut-parleurs par rapport à une configuration idéale. Ces deux dernières questions ne seront pas abordées ici.

Le traitement de correction mis en œuvre se base sur une stratégie classique d'inversion [3, 4], dans laquelle on utilise l'ensemble des haut-parleurs pour corriger le rendu d'un haut-parleur donné. La spécificité de la solution présentée réside dans la représentation des ondes acoustiques au format Higher Order Ambisonics (HOA) dont les propriétés (décomposition de l'espace, hiérarchie de la représentation) offrent plusieurs atouts pour le traitement. La première partie de l'article rappelle les fondements théoriques et les propriétés de ce format. Ensuite la méthode de correction est décrite. La quatrième partie propose, pour évaluer l'efficacité de la correction, des critères inspirés des critères de la qualité acoustique d'une salle. Dans cette étude, la correction est mise en œuvre dans le cadre de simulations numériques de différents configurations d'écoute. Les réponses impulsionnelles des différents haut-parleurs sont ainsi calculées à l'aide d'un outil d'acoustique virtuelle se basant sur une méthode des sources images et qui est présenté en cinquième partie. Enfin, l'analyse des résultats des simulations sur la base des critères identifiés précédemment est donnée avant de conclure.

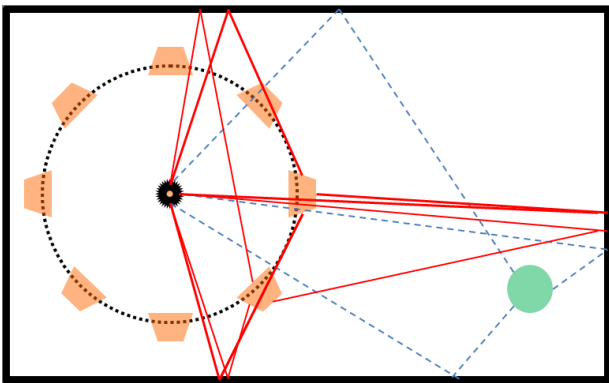


FIGURE 1 – Illustration de l'incohérence des réflexions.

En vert, une source virtuelle créée par le système orange, avec ses réflexions théoriques associées (tirets bleus). Les réflexions issues du système sont en rouge.

On note bien les différences d'angle et d'intensité (distance aux murs).

2 Harmoniques sphériques et format HOA

Le format HOA est un modèle de représentation d'une onde acoustique qui s'identifie au développement de l'onde sur la base des fonctions propres de l'équation des ondes acoustiques en coordonnées sphériques (r : rayon, φ : angle d'azimuth, θ : angle d'élévation) [5]. Ces fonctions propres combinent des fonctions de Bessel sphériques $j_m(kr)$ ¹ et $n_m(kr)$ ² et/ou des fonctions de

Hankel sphériques $h_m^+(kr)$ ³ et $h_m^-(kr)$ ⁴ qui décrivent les dépendances **radiales**, et des harmoniques sphériques $Y_{mn}^\sigma(\varphi, \theta)$ qui décrivent les dépendances **angulaires** de l'onde acoustique [6]. En général, la scène sonore à représenter ne comporte que des ondes convergentes et le développement de la pression acoustique $p(\vec{r}, \omega)$, où ω est la pulsation, se ramène à [5] :

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

Les coefficients B_{mn}^σ constituent la représentation HOA de l'onde acoustique. Le format Ambisonic proposé par Gerzon [7] est un cas particulier de cette représentation, dans lequel le développement de l'équation 1 est limité à l'ordre 1 et ne comporte donc que les 4 premières composantes. Le format HOA en est ainsi la généralisation aux ordres supérieurs. En pratique la représentation HOA doit être tronquée à un ordre M donné, ce qui conduit à représenter la scène audio 3D par $(M+1)^2$ composantes B_{mn}^σ ($m=0, 1, \dots, M$; $n=0, 1, \dots, m$; $\sigma = \pm 1$). Si on se restreint au plan horizontal (représentation dite "2D", par contraste avec la représentation "3D" précédente) et qu'on ne conserve que les harmoniques sphériques porteurs de l'information spatiale horizontale [5], la représentation ne comporte que $2M+1$ composantes.

Le format HOA est parfaitement universel au sens où il permet de décrire n'importe quelle onde acoustique (onde plane ou sphérique notamment) [2]. Mais surtout, cette représentation se caractérise par deux propriétés fondamentales. Il s'agit d'abord d'un format indépendant des systèmes de captation et de restitution, d'où une flexibilité dans son utilisation. Secundo il s'agit d'une description hiérarchique, c'est à dire que les composantes des premiers ordres (m) suffisent à représenter l'onde acoustique, les composantes des ordres supérieurs ne venant que préciser l'information spatiale. Cette propriété est très utile, car elle permet de faire évoluer la représentation, a posteriori de la captation, afin de s'adapter aux contraintes notamment de capacité en débit du réseau de transmission ou du système d'écoute disponible. Un autre atout est sa lisibilité, dans la mesure où cette représentation peut se lire directement en termes de structure spatiale de la scène sonore, offrant une analyse séparée des informations selon 2 axes : la distance (r) et la direction (φ, θ). Enfin le développement sur la base des fonctions propres n'est autre qu'une transformation du domaine des coordonnées d'espace (r, φ, θ) dans un domaine dual correspondant aux fréquences spatiales. Les coefficients B_{mn}^σ définissent ainsi le spectre spatial associé à l'onde acoustique, ce qui apporte une lisibilité supplémentaire à la représentation HOA en termes de fréquences (ou de variations) spatiales. Les coefficients B_{mn}^σ prennent ainsi le sens d'une représentation duale de la scène sonore, les deux étant liés par une transformation et sa réciproque. En outre, il a été montré que ce format était particulièrement pertinent pour l'analyse des premières réflexions d'une salle [8].

1. Fonctions de Bessel sphériques de première espèce.
2. Fonctions de Bessel sphériques de seconde espèce ou fonction de Neumann.

3. Fonctions de Hankel sphériques de première espèce : onde progressive convergente.

4. Fonctions de Hankel sphériques de seconde espèce : onde progressive divergente.

3 Principe du traitement de correction

Tous les signaux sont considérés dans le domaine spatio-fréquentiel, c'est à dire que la pression acoustique $p(\vec{r}, t)$ évaluée au point \vec{r} et à l'instant t est d'abord exprimée, via une transformée de Fourier, dans le domaine des pulsations ω . Le signal résultant $p(\vec{r}, \omega)$ est ensuite décomposé sur la base des harmoniques sphériques en $D = (M + 1)^2$ ou $D = 2M + 1$ composantes spatiales $B_{mn}^\sigma(\omega)$, respectivement pour une représentation 3D ou 2D, où M définit l'ordre de troncature de la décomposition (cf. Section 2).

La première étape consiste à caractériser la configuration d'écoute afin de décrire à la fois les propriétés acoustiques de la salle et le dispositif des haut-parleurs. Pour ce faire, les réponses impulsionnelles de chaque haut-parleur sont obtenues (par simulation ou mesure) pour une position de référence qu'on se donne dans la salle. Chaque réponse impulsionnelle est composée d'une onde directe décrivant les propriétés intrinsèques des haut-parleurs, suivie des réflexions générées par la salle et donc représentatives de cette dernière. Pour cette première étude, on a choisi de se focaliser sur la correction de l'effet de salle. On élimine donc l'onde directe par fenêtrage temporel pour ne retenir que les L échantillons suivants. Une Transformée de Fourier Discrète (DFT) est alors appliquée pour obtenir le spectre associé sur L bins fréquentiels ω_k , $k \in [1, \dots, L]$. Si le système de restitution se compose de N haut-parleurs, on réunit ainsi un ensemble de N fonctions de transfert $R_i(\omega_k)$ (l'indice i repérant le i ème haut-parleur) qui constituent le descriptif acoustique de la configuration d'écoute. Afin de séparer et d'analyser les différentes composantes spatiales, chaque fonction de transfert est représentée dans le format HOA. Pour la suite de l'étude on se limite à une représentation 2D, ce qui implique donc $D = 2M + 1$ composantes spatiales $R_{j,i}(\omega_k)$ où l'indice j parcourt les composantes HOA. Il en résulte pour chaque pulsation, une matrice $\mathbf{R}(\omega_k)$ de dimension $D \times N$ et dont les N colonnes sont formées par les $\mathbf{R}_i(\omega_k)$, $i \in [1, \dots, N]$, développées sur D composantes HOA.

La matrice $\mathbf{R}(\omega_k)$ doit être comparée à une matrice "cible" $\mathbf{T}(\omega_k)$ définissant l'effet de salle qu'on souhaite obtenir, à la place de celui décrit par $\mathbf{R}(\omega_k)$. La matrice $\mathbf{T}(\omega_k)$ est obtenue par une processus analogue à celui qu'on vient d'appliquer pour exprimer $\mathbf{R}(\omega_k)$. Comme cette dernière, elle peut être issue de mesures dans une salle réelle ou de simulations. Elle peut être aussi imposée arbitrairement. Dans le cadre de ces simulations par exemple, les réponses cibles sont constituées des réponses directes des haut-parleurs, c'est-à-dire sans aucune réverbération.

Le processus de correction consiste à appliquer en amont des haut-parleurs une matrice de filtres. Le signal alimentant un haut-parleur n'est alors plus alimenté indépendamment des autres haut-parleurs, mais prend en compte leurs signaux pour compenser les réflexions associées. Ainsi les réflexions générées par un haut-parleur sont corrigées par l'ensemble des autres haut-parleurs. La matrice de correction $\mathbf{c}(\omega_k)$ réalise à la fois le mélange et la pondération des signaux des haut-parleurs. On reconnaît une sorte de généralisation du procédé d'annu-

lation des trajets croisés utilisés pour la reproduction de signaux binauraux [9]. Pour en illustrer le principe, considérons la correction de la contribution d'un seul haut-parleur. La fonction de transfert cible associée à ce haut-parleur est représentée par la colonne $\mathbf{T}_i(\omega_k)$. Les N haut-parleurs du dispositif sont sollicités pour atteindre cette cible. Leurs fonctions de transfert effectives sont contenues dans la matrice $\mathbf{R}(\omega_k)$ et sont alors superposées après pondération par le vecteur $\mathbf{c}_i(\omega_k)$, de sorte que :

$$\mathbf{R}(\omega_k)\mathbf{c}_i(\omega_k) = \mathbf{T}_i(\omega_k) \quad (2)$$

Le vecteur $\mathbf{c}_i(\omega_k)$ se compose des N pondérations associées aux différents haut-parleurs pour corriger le i ème haut-parleur. Pour décrire les corrections de l'ensemble du dispositif, on obtient au final une matrice $\mathbf{c}(\omega_k)$ de dimension $N \times N$. Chaque vecteur de correction $\mathbf{c}_i(\omega_k)$ est calculé en utilisant la pseudo-inverse de $\mathbf{R}(\omega_k)$, ce qui donne, en soulageant la notation des ω_k :

$$\mathbf{c}_i = (\mathbf{R}^*\mathbf{R})^{-1} \cdot \mathbf{R}^*\mathbf{T}_i \quad (3)$$

Dans cette expression, l'opérateur $*$ définit une transposée hermitienne (matrices complexes). On note que le procédé permet potentiellement d'adapter une correction différente selon la composante HOA. Cette possibilité sera examinée ultérieurement. Les réponses corrigées sont calculées en recombinaison des composantes HOA après convolution des réponses originelles par les filtres de correction.

Cette méthode convient bien pour des simulations, dont les données ne sont pas bruitées. Dans le cas de mesures dans une salle réelle, il serait plus judicieux d'utiliser une opération de régularisation, de type Tikhonov, pour effectuer l'inversion. De plus, un autre obstacle au passage à des conditions réelles est qu'il sera alors nécessaire d'effectuer un pré-traitement spatial sur les réponses impulsionnelles, afin de réduire voire supprimer les réflexions ne venant pas du plan défini par le cercle de haut-parleurs. Ces réflexions ne peuvent en effet physiquement pas être contrôlées. Pour ce faire, l'emploi de gabarit de filtre spatiaux de forme toroïdale est envisagé, leur construction et leur application ne posant pas de problèmes dans le domaine HOA.

4 Critères d'évaluation

La description de la qualité acoustique d'une salle fournit une palette de critères objectifs capables de quantifier les propriétés d'un effet de salle [10, 11]. Cependant la plupart de ces critères n'est pas transposable directement à notre problème. Le traitement vise à réduire l'effet de salle en annulant les réflexions engendrées par les haut-parleurs. L'objectif est de minimiser l'incohérence entre les réflexions et les sources sonores virtuelles, afin de préserver la qualité de la spatialisation. En pratique, le traitement ne permet pas d'annuler les réflexions, mais va plutôt "sculpter" la réverbération pour la rendre plus discrète. Il travaille notamment à réduire l'émergence des réflexions. Or cette propriété n'est pas reflétée par les critères classiques. Quant à la qualité de spatialisation des sources sonores, cet aspect est as-

sez peu pris en compte. Un premier jeu de critères pour pallier ces lacunes est proposé ici.

4.1 Émergence des réflexions

La présence de réflexions importantes autour du son direct apparaît comme une première valeur permettant de juger l'apport de la correction. Comme nous travaillons ici sur des réponses impulsionnelles simulées, chaque réflexion se traduit par une impulsion idéale caractérisée par ses positions dans le temps et dans l'espace, ainsi que par son amplitude.

Des mesures d'émergence d'échos comme le coefficient d'écho ou la tangente ajustée, que l'on peut trouver dans [10], ne semblent pas adaptées. Elles s'intéressent en effet à évaluer si des échos sont bel et bien perçus par l'auditeur. Or, dans le cas présent, la question n'est pas de savoir si ces réflexions sont individuellement perçus. On suppose simplement qu'elles jouent un rôle dans l'impression spatiale, et que c'est leur incohérence avec le son direct qui est ressentie.

On préfère comptabiliser, en s'inspirant des recommandations de l'ITU [12] pour la composante omnidirectionnelle du signal HOA (composante W dans le formalisme Ambisonics) et sur une fenêtre de 50ms (5ms avant le son direct, 45ms après), le nombre de pics $nbPeak$ dont l'énergie est inférieure d'au plus 12dB à l'énergie de l'onde directe. Ces valeurs sont plus exigeantes que celles proposées par l'ITU (15ms et 10dB). L'énergie moyenne $salPeak$ de ces pics a également été calculée. Une baisse de ces 2 valeurs suggère que les perturbations liées aux réflexions sont potentiellement réduites.

4.2 Corrélation croisée inter-aurale

Les critères précédents ne permettant pas de rendre compte de la localisation des sources, une troisième critère est identifié : il s'agit de l'Inter-Aural Cross Correlation ou IACC, dont l'intérêt est rappelé dans [10]. De plus, une étude récente met en évidence sa pertinence en relation avec le format HOA [13]. Si $p^R(t)$ et $p^L(t)$ représentent respectivement les pressions induites au niveau des oreilles droite et gauche, la fonction normalisée de l'IACC est définie entre les bornes temporelles t_1 et t_2 par :

$$\rho_{t_1, t_2}(\tau) = \frac{\int_{t_1}^{t_2} p^R(t)p^L(t + \tau)dt}{\sqrt{\int_{t_1}^{t_2} [p^R(t)]^2 dt \cdot \int_{t_1}^{t_2} [p^L(t)]^2 dt}} \quad (4)$$

Le coefficient d'IACC s'obtient ensuite en trouvant la valeur maximum selon τ ,

$$IACC_{t_1, t_2} = \max_{\tau} |\rho_{t_1, t_2}(\tau)|, \tau \in (-1, 1)ms. \quad (5)$$

Dans notre cas, les signaux $p^R(t)$ et $p^L(t)$ sont obtenus grâce à un outil développé à Orange Labs [14] et destiné à adapter des signaux HOA à une écoute sur casque. Il consiste à simuler des haut-parleurs virtuels autour de l'auditeur et à synthétiser les signaux au niveau de ses oreilles, en utilisant les fonctions de transfert correspondantes définies comme les Head Related Transfer

Function (HRTF). Les HRTF mises en œuvre ici sont celles d'un sujet de la base privée d'Orange Labs.

Cet indice mesure le degré de cohérence de l'information spatiale perçue entre les deux oreilles, et par suite donne une idée de la précision de localisation de la source sonore en présence d'un effet de salle. Dans un champ diffus, il vaut 0, car l'énergie acoustique provient uniformément de toutes les directions de façon décorrélée. En présence d'une source unique il est en revanche proche de 1, car les deux oreilles perçoivent la même onde acoustique moyennant les différences interaurales de temps et d'intensité et la diffraction par le corps de l'auditeur. Ainsi une valeur proche de 1 traduit une impression de localisation plus précise.

5 Simulations

L'outil RoomSim a été utilisé pour simuler les réponses impulsionnelles $R_i(\omega_k)$. Celui-ci provient de l'adaptation pour Matlab [15] par Campbell *et al.* [16] d'un algorithme de calcul de réponses impulsionnelles de salles parallélépipédiques par la méthode des sources images proposé initialement par Allen et Berkeley [17]. Pour adapter RoomSim au formalisme HOA, le microphone associé à la captation de la réponse impulsionnelle est remplacé par un jeu de capteurs reproduisant les directivités des harmoniques sphériques afin d'extraire les composantes HOA. Un ensemble de directivités a été défini, afin de représenter les 25 composantes du développement HOA jusqu'à l'ordre 4. Comme cette étude se réfère à des systèmes réels existants, il a été choisi de ne pas s'intéresser aux ordres supérieurs, car il n'existe pas aujourd'hui de système de captation pour ces composantes.

La salle simulée est censée représenter un salon domestique "standard". Cependant, du fait qu'on se limite à une représentation 2D, le sol et le plafond ont été considérés parfaitement absorbants, c'est à dire qu'ils ne génèrent pas de réflexions.

Trois des murs sont relativement absorbants, le dernier a les propriétés de réflexion classique du verre. Deux systèmes d'écoute différents sont considérés, sur la base d'un cercle de haut-parleurs de rayon $r = 2m$. Le premier est un dispositif régulier de 8 haut-parleurs, le second de 5 disposés selon la configuration standard ITU. Dans tous les cas, ils sont considérés comme des sources omnidirectionnelles. Les paramètres variant lors de ces simulations sont

- L'ordre ambisonique du domaine de traitement, compris entre 1 et 4, aisément accessible par troncature grâce à la scalabilité de HOA.
- L'ordre maximal des réflexions considérées, entre 1 et 13. Néanmoins, dans une salle de cette dimension, les réflexions d'ordre supérieur à 4 arrivent déjà quasiment toutes avec plus de 100ms de retard par rapport à l'onde directe.
- La longueur des filtres utilisés, entre 512 et 8192. Ce paramètre est bien entendu déterminant dans le temps de calcul de l'algorithme, et un filtre de correction trop long pourrait être à exclure dans le cadre d'écoute en temps réel.

6 Résultats

6.1 Auralisation

Afin d'évaluer qualitativement l'apport de la méthode sur la restitution sonore, une scène sonore HOA virtuelle d'ordre 4 a été encodée. Elle contient un son d'ambiance à l'arrière, de la musique diffusée en proximité sur une configuration stéréophonique ainsi qu'un bruit percussif répétitif au loin. Cette scène virtuelle est ensuite décodée sur le cercle de 8 haut-parleurs, ce qui permet d'obtenir les signaux d'alimentation du système. Ces signaux sont ensuite convolués dans le domaine spatio-fréquentiel avec les réponses cibles, originelles et corrigées, ce qui permet d'obtenir le champ HOA global restitué, qui est finalement transformé en signaux binauraux grâce à la méthode décrite dans [14]. L'écoute binaurale des versions corrigées et originales fait ressortir assez nettement une plus grande précision spatiale de la scène. La coloration due à la salle est également très réduite. Ces observations restent à confirmer par des tests subjectifs dans le futur.

6.2 Critères objectifs

Très vite, l'ordre maximum des réflexions considérées dans les réponses impulsionnelles est apparu comme un paramètre non pertinent de simulation. Trop bas, il ne conduisait pas à des données réalistes, et à partir d'une certaine valeur, toutes les réflexions étaient situées dans la réponse tardive de la réverbération, bien après 100ms du son direct. Il a donc été choisi de le fixer à 10.

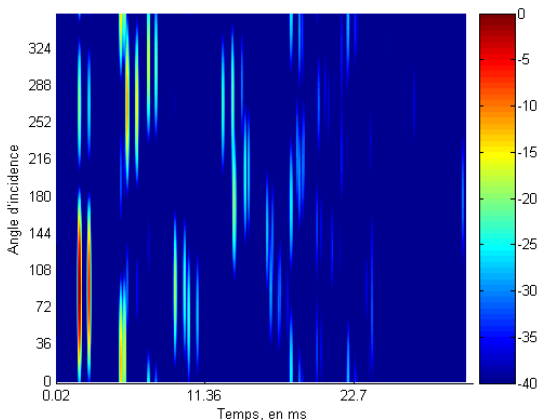


FIGURE 2 – Réponse impulsionnelle de base.

Un des résultats de ces simulations est illustré sur les figures 2 et 3. La représentation utilisée représente l'énergie du signal en dB sur une échelle de couleur, en fonction du temps (abscisse) et de l'angle d'incidence (ordonnée). La dynamique d'affichage est de 40 dB. Ces figures sont obtenues pour un haut-parleur du système en comprenant 8, avec une taille de filtre de 4096 et à l'ordre 4.

L'efficacité de la correction proposée apparaît assez clairement de façon visuelle sur ces représentations. En plus d'une augmentation légère de l'énergie moyenne du

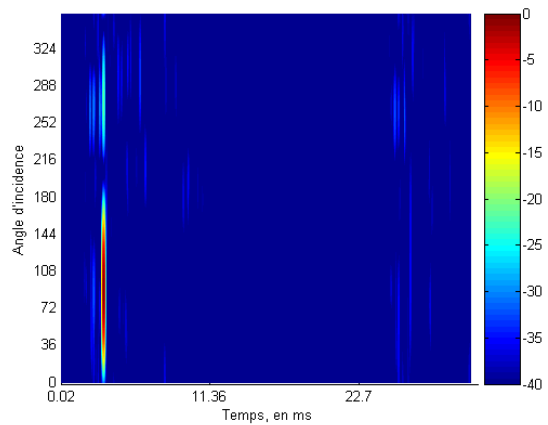


FIGURE 3 – Réponse impulsionnelle corrigée.

signal, les résultats objectifs selon les critères exposés à la partie 4 sont les suivants :

- Dans le cas du système à 8 haut-parleurs (figure 4), on observe une évolution des résultats globalement prévisible. Plus le filtre de correction est long, moins on compte de pics et plus l'IACC est élevée. Ainsi, avec des filtres de correction à 4096 coefficients et à l'ordre 4, on peut proposer une correction objectivement efficace.
- D'autre part, le cas à 5 haut-parleurs (figure 5) est plus problématique. Si les réflexions sont globalement réduites avec des filtres suffisants, il n'apparaît pas possible de les éliminer toutes avec cette méthode simple. Cela provient probablement du fait de l'irrégularité de la distribution des haut-parleurs [5].
- Dans tous les cas, l'évolution selon l'ordre HOA se révèle plus délicate à interpréter, car si on peut noter une amélioration sur les pics, l'IACC a tendance à baisser lorsque l'ordre augmente. Pour en comprendre les raisons, il conviendrait d'analyser plus en détail l'impact de l'ordre sur la distribution spatiale de l'énergie.

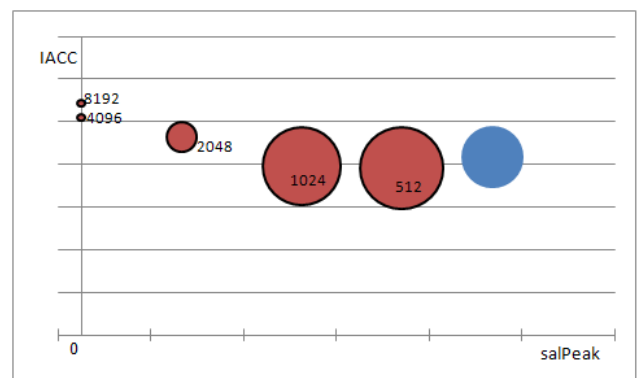


FIGURE 4 – Évolution des résultats selon la taille des filtres de correction : système à 8 haut-parleurs.

Les figures 4 et 5 doivent être interprétées comme ceci : les cercles représentent les valeurs moyennes sur les différents haut-parleurs des critères énoncés à la partie 4, selon le paramètre de simulation, ici la taille des filtres de correction. La taille de chaque cercle est directement associée au nombre de pics détectés $nbPeak$. L'axe des ab-

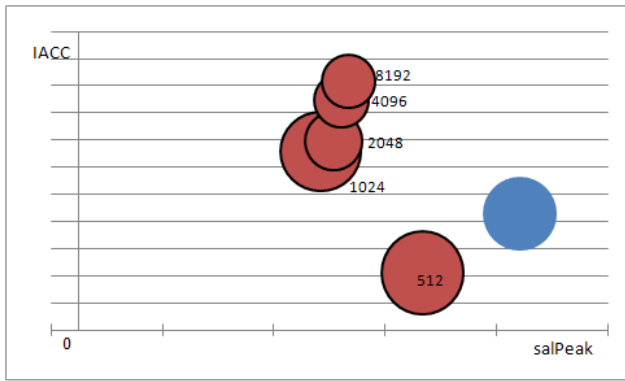


FIGURE 5 – Évolution des résultats selon la taille des filtres de correction, système à 5 haut-parleurs.

sisées est *salPeak*, celui des ordonnées l'*IACC*. Le cercle bleu sans bordure est constitué des valeurs de référence prises dans le cas du système non corrigé, les cercles rouges cerclés sont les issus de différentes corrections, avec les longueurs de filtres inscrites en étiquette.

7 Conclusion

Cette étude a permis de montrer l'efficacité du traitement sur des données de simulation pour réduire de façon notable les réflexions d'un système multicanal dans une salle d'écoute domestique, et la similarité des simulations à la réalité (cf. figure 6) permet d'envisager le passage à des conditions réelles. Il faudra pour ce faire mettre au point une méthode séparant, au sein des réponses impulsionnelles, les réflexions provenant du sol et du plafond du reste, ainsi que proposer un algorithme plus robuste pour l'inversion.

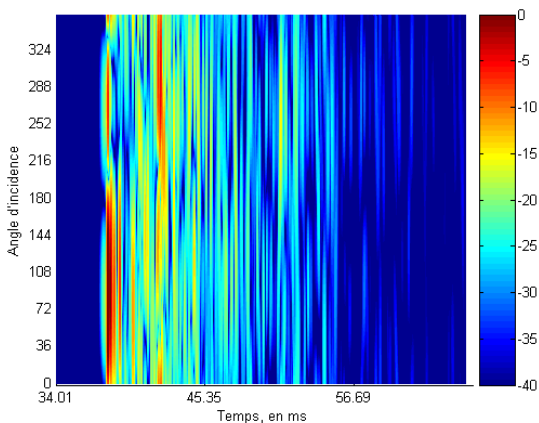


FIGURE 6 – Mesure d'une salle réelle.

Remerciements

Cette étude a été réalisée au sein des laboratoires d'Orange Labs - Lannion.

Références

- [1] Blauert, J., "Spatial Hearing, The Psychophysics of Human Sound Localization".
- [2] Daniel, J., Nicol, R. and Moreau, S. "Further investigations of high order ambisonics and wave-field synthesis for holophonic sound imaging", *114th AES Convention* (2003).
- [3] Guillaume, M., Grenier, Y. and Richard, G. "Iterative algorithms for multichannel equalization in sound reproduction systems", *ICASSP* (2005).
- [4] Corteel, E., "Adaptations de la Wave Field Synthesis aux conditions réelles", Thèse de Doctorat, Université de Paris 6 (2004).
- [5] Daniel, J. "Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia", Thèse de Doctorat, Université de Paris 6 (2000).
- [6] Bruneau M., "Introduction aux théories de l'acoustique", Université du Maine (1983).
- [7] Gerzon, M. A., "General metatheory of auditory localization", *92nd AES Convention* (1992).
- [8] Caulkins, T. *et al.*, "Use of a high spatial resolution microphone to characterize the early reflections generated by a WFS loudspeaker array", *AES 28th Int. Conf.*, Piteå(2006).
- [9] Gardner, W. G., "3-D audio using loudspeakers", Kluwer Academic Publisher (1997).
- [10] Jouhaneau, J., "Acoustique des salles et sonorisation".
- [11] Kahle, E., "Validation d'un modèle objectif de la perception de la qualité acoustique dans un ensemble de salles de concerts et d'opéras", Thèse de Doctorat, Université du Maine (1995).
- [12] ITU BS.1116, "Methods for the subject assessment of small impairments in audio systems, including multichannel sound systems." *ITU*, Geneva (1998)
- [13] Rafaely, B. and Avni, A., "Interaural cross correlation in a sound field represented by spherical harmonics", *J. Acoust. Soc. Am.* 127 (2), 823-828 (2009).
- [14] Faure, J. , Daniel, J. , Emerit, M. , "Optimisation d'une spatialisation sonore binaurale à partir d'un encodage multicanal", **WO 2007101958**.
- [15] The Math Works Inc. , *Matlab*, <http://www.mathworks.com>
- [16] Campbell, D.R., Palomäki, K.J. and Brown, G.J., "Roomsim, a MATLAB Simulation of "Shoebox" Room Acoustics for use in Teaching and Research", http://www.dcs.shef.ac.uk/~guy/pdf/Campbell_Palomaki_Brown.PDF
- [17] Allen, J. B. and Berkeley, D. A., "Image method for efficiently simulating small-room acoustics", *J. Acoust. Soc. Am.* 65 (4), 943-950 (1979).