Diffusion de son 3D par synthèse de champs acoustiques binauraux

Adrien Vidal

UNIVERSITE D'AIX-MARSEILLE

ECOLE DOCTORALE SCIENCE POUR L'INGENIEUR : MECANIQUE, PHYSIQUE, MICRO ET NANOELECTRONIQUE (ED 353)

Thèse présentée pour obtenir le grade universitaire de docteur

Discipline : ACOUSTIQUE

Présentée et soutenue publiquement par

Adrien VIDAL

le 03 février 2017

DIFFUSION DE SON 3D PAR SYNTHESE DE CHAMPS ACOUSTIQUES BINAURAUX

Jury :

Rozenn NICOL	Orange Labs (Lannion)	Rapporteur
Olivier WARUSFEL	IRCAM (Paris)	Rapporteur
Cedric MAURY	ECM (Marseille)	Examinateur
Manuel MELON	Université du Maine (Le Mans)	Examinateur
Etienne PARIZET	INSA (Lyon)	Examinateur
Philippe HERZOG	LMA (Marseille)	Directeur de thèse
Christophe LAMBOURG	Genesis (Aix-En-Provence)	Co-Encadrant
Patrick BOUSSARD	Genesis (Aix-En-Provence)	Invité

Remerciements

Le travail de thèse n'est certainement pas un travail solitaire, et j'aimerais donc remercier toutes les personnes qui ont contribué de près ou de loin à l'aboutissement de ce projet.

Je souhaiterais tout d'abord remercier Philippe Herzog pour son encadrement bienveillant, sa pédagogie et son inépuisable vivier de connaissances. J'ai apprécié découvrir le travail de chercheur avec toi, même si maintenant j'ai des mouchoirs plein les poches ! Je remercie aussi vivement Christophe Lambourg pour l'encadrement côté Genesis, ses conseils pertinents ont su me guider au cours de ce travail. Ce fut un plaisir de travailler avec vous deux !

Je tiens à remercier Rozenn Nicol et Olivier Warusfel pour l'évaluation de ce travail et les rapports détaillés. Je remercie également Etienne Parizet pour avoir présidé le jury, ainsi que Cédric Maury et Manuel Melon pour avoir fait partie de ce jury.

J'adresse toute ma gratitude à Patrick Boussard pour m'avoir confié cette mission passionnante de trois ans. Je remercie également toute l'équipe Genesis : Martine pour la gestion administrative, Guillaume pour les parties de contrée, Antoine pour les conseils en psycho, Paul pour les sessions escalade, Sylvain pour la cause animale, ainsi que Hélène, Clément et Stéphane. Merci Hakim pour le rodage des Schtroumpfs !

J'adresse également tous mes remerciements aux chercheurs du LMA, en particuliers ceux des équipes GROINK et psychoacoustique : Jacques pour la mise en place des mesures d'HRTF, Guy pour les bricolages, Sophie et Sabine pour les discussions sur les tests perceptifs, Marc et Cédric pour toujours trouver le matos dont on a besoin, mais aussi Muriel, Pierre-Olivier, Sergio, Renaud. Je remercie chaleureusement Vincent de l'atelier pour la fabrication de la sphère et du Schtroumpf. Merci à tous les doctorants et post-docs du labo pour les cafés et autres réunions plus ou moins informelles : Pierre-Yvon pour les macarons, Pierre pour les délires en salle café, Lennie pour son calme absolu, ainsi que Volodimir, Lionel, Quentin, Marina(s), Gaëtan, Gaultier, Thomas.

Même si pour beaucoup c'est redondant avec les autres remerciements, je tiens à remercier toutes les oreilles qui ont participé aux tests d'écoute !

Mais aussi merci à Diane et Juliette pour les cours de salsa, Maxime pour les afterwork à la côte bleue, Maureen pour le Esquissé, Patou pour le covoiturage, Hugo pour ton super « salon » à la plaine.

Je souhaite remercier toute ma famille qui a toujours été là pour moi, en particulier mes parents qui m'ont permis de réaliser les études qui me tenaient à cœur. Merci Papa pour la relecture finale !

Enfin, grand merci à Sara pour son aide et son soutien dans tous les moments (surtout les derniers !)

Résumé

Ces travaux de thèse concernent la conception d'un dispositif de restitution sonore léger en salle usuelle, permettant la diffusion de signaux binauraux. La priorité du travail est d'assurer la précision dans la reproduction du niveau et du timbre, devant la spatialisation du son. Afin d'assurer la compatibilité avec les signaux binauraux existants et pour proposer un système à faible nombre de canaux, la technologie transaurale est prise comme point de départ. Pour limiter la coloration introduite par la salle d'écoute, particulièrement gênante, il est proposé de placer les sources du système à proximité de l'auditeur afin de maximiser le rapport champ direct sur champ diffus. Ce placement inhabituel a plusieurs effets, parmi lesquels quatre ont été étudiés séparément : les variations inter-individuelles de morphologies, l'influence des filtres transauraux sur les sources électro-acoustiques, l'effet de salle et le placement de l'auditeur. Des tests d'écoute ont été réalisés pour une sélection de configurations, et les résultats ont permis d'implémenter des indicateurs objectifs représentatifs des réponses des auditeurs. La synthèse de ces indicateurs a permis de proposer trois configurations considérées comme optimales, et dont la combinaison pourrait être envisagée.

Abstract

This work deals with the design of a 3D sound system involving a few number of loudspeakers and able to work inside any usual room, for reproducing binaural sounds. This system focuses on an accurate reproduction of perceived level and timbre, even before the sound spatialization. To ensure compatibility with binaural recordings and to achieve a system with a low number of loudspeakers, this work started from a transaural system. To avoid tone coloration induced by the listening room, the sound sources are placed close to the listener, thus maximizing the energy ratio between direct and diffuse fields. This has consequences on other aspects, four of which are considered separately: inter-individual morphological variations, demands on the electro-acoustic sources, room effect and misalignment of the listener. Some configurations have been evaluated though listening tests, and objective indicators are deduced from these results. The generalization allows to propose three configurations considered as optimum, which might also be combined.

Table des matières

Table des abréviations

ANOVA	Analyse of Variance
BEM	Boundary Element Method
ETC	Energy Time Curve
FIR	Finite Impulse Response
FRF	Fonction de Réponse en Fréquence
HOA	Higher Order Ambisonics
HRIR	Head-Related Impulse Response
HRTF	Head-Related Transfer Function
IIR	Infinite Impulse Response
ILD	Interaural Level Difference
ITD	Interaural Time Difference
JND	Just Noticeable Difference
MIMO	Multiple Input Multiple Output
OSD	Optimal Source Distribution
RI	Réponse Impulsionnelle
RMSE	Root Mean Square Error
SISO	Single Input Single Output
SRV	Source Virtuelle
SVD	Singular Value Decomposition
WFS	Wave Field Synthesis

Notations

θ	Angle d'azimut
ϕ	Angle d'élévation
r	Distance
f	Variable fréquentielle
t	Variable temporelle
λ	Longueur d'onde
ω	Pulsation
Q	Débit
c	Célérité du son dans l'air ($c = 344\ m.s^{-1}$ à 21°C)
ρ	Masse volumique de l'air ($\rho = 1.2\ kg.m^{-3}$ à 21°C)
A	Aire équivalente d'absorption
W	Puissance acoustique
Y_n^m	Harmonique sphérique d'ordre m et de degré n
$P_n^{\lvert m \rvert}$	Polynôme de Legendre d'ordre m et de degré n
h	Fonction de Hankel sphérique de 1ère espèce
R	Coefficient de corrélation de Bravais-Pearson
ρ_{Spear}	Coefficient de corrélation de rang de Spearman
β	Paramètre de régularisation
j	$\sqrt{-1}$
$Diag$	Matrice diagonale
$[.]$	Matrice
$\boldsymbol{a}$	Vecteur
$\lfloor . \rfloor$	Partie entière
$\langle . \rangle$	Moyenne arithmétique

Convention de repère adoptée

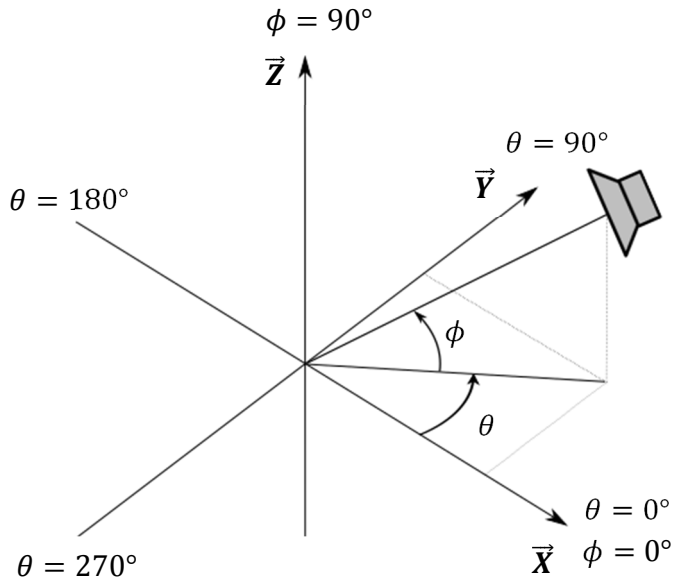


Figure 1 : convention de repère adoptée

Introduction

Ces dernières années, la diffusion sonore en trois dimensions connait un essor remarquable, se démocratisant notamment auprès du grand public. Par exemple, RadioFrance[1] propose un site internet dédié à la diffusion de créations sonores spatialisées, Facebook[2] et Youtube[3] proposent des outils pour créer des vidéos avec du son spatialisé. Par ailleurs, le nombre de publications scientifiques autour de la problématique a considérablement augmenté ces dernières années[4]. La restitution sonore spatialisée trouve également sa place dans le domaine des télécommunications, permettant par exemple d'accroître le réalisme et l'intelligibilité lors de téléconférences.

Le son spatialisé est aussi un enjeu industriel. Dans la phase de mise au point d'un produit, il peut être nécessaire de s'intéresser à son rendu sonore. Son évaluation n'est pas toujours possible : par exemple pour comparer le bruit dans l'habitacle de plusieurs véhicules en conditions de roulage, il est difficile d'assurer la reproductibilité du rendu sonore. Une évaluation alternative consiste à diffuser des signaux sonores *via* un dispositif de restitution : la qualité sonore peut alors être évaluée indépendamment d'autres composantes (confort de l'habitacle, image de marque, etc.), et les conditions de reproduction maîtrisées (le même signal est diffusé pour tous les auditeurs). Ce type d'approche peut également servir au stade de l'étude d'un produit, avant qu'il ne soit réellement fabriqué. Pour que cela soit possible, il faut cependant disposer d'outils de simulation réaliste. Cet exemple rejoint un autre type d'application : la réalité virtuelle. Ainsi des simulateurs sont depuis longtemps utilisés par les pilotes pour leur entraînement, effectué à moindre risque et moindre coût qu'en conditions réelles. Le réalisme de la reproduction est primordial pour que le pilote soit dans des conditions aussi proches que possible de la réalité.

La restitution sonore spatialisée n'est pas nouvelle, et il existe de nombreux systèmes et formats de restitution de complexités variées. Les systèmes généralisant la stéréophonie tels que le surround (5.1, 7.1, etc.) et le Vector Based Amplitude Panning (VBAP) font partie des systèmes précurseurs de la spatialisation sonore. Des approches plus complexes ont aussi été développées, visant notamment à reproduire un champ sonore dans une zone spatiale élargie : les Wave Field Synthesis (WFS) et Higher Order Ambisonics (HOA). Les performances de ce type de systèmes sont directement liées au nombre de canaux employés, généralement élevé. Au contraire, les approches basées sur les technologies binaurales, sont minimalistes et dédiées à une écoute individuelle. Pour ce dernier type de technologie, il s'agit de reproduire la pression acoustique au niveau des oreilles de l'auditeur telle qu'elle l'aurait été dans une situation réelle. La diffusion sonore binaurale peut être effectuée au casque d'écoute, celui-ci pouvant toutefois être gênant

1 http://nouvoson.radiofrance.fr/

2 https://facebook360.fb.com/spatial-workstation/

3 https://support.google.com/youtube/answer/6395969?hl=fr

4 Une sélection de références sera présentée dans le chapitre « Etat de l'art »

pour l'auditeur. Une alternative consiste à employer un dispositif de haut-parleurs : il s'agit alors d'un système transaural.

Pour beaucoup d'applications industrielles évoquées précédemment, le réalisme du rendu sonore concerne prioritairement le respect du niveau et du timbre : le réalisme spatial est important mais moins critique. La majorité des systèmes de reproduction évoqués sont cependant sensibles à l'environnement d'écoute, qui apporte une « coloration », modifiant le timbre et le niveau du rendu. Un moyen de réduire cette influence est de placer le dispositif en milieu très peu réverbérant, s'approchant plus ou moins d'une chambre anéchoïque. Cependant, le coût d'un tel moyen d'essais augmente alors considérablement et n'est pas compatible avec beaucoup de contextes industriels : il est donc intéressant de trouver des solutions alternatives.

L'objectif des travaux présentés ici est de répondre à cette attente, c'est-à-dire de proposer une solution de reproduction sonore spatialisée respectant prioritairement le niveau et le timbre des signaux, à coût minimal, et avec la possibilité d'une installation dans une salle usuelle. Les besoins industriels sont au cœur de la démarche, et le système doit donc permettre la reproduction de signaux correspondant aux contenus pré-existants. De nombreux industriels utilisant les technologies binaurales de longue date, il est en particulier nécessaire que le système soit compatible avec leurs bibliothèques d'enregistrements existants. Cette contrainte est importante : l'information spatiale portée par de tels signaux est fortement liée au dispositif de prise de son, et le dispositif de restitution doit tenir compte de cette particularité.

Pour ces raisons, notre approche s'est focalisée initialement sur les systèmes transauraux qui permettent une reproduction spatialisée avec deux haut-parleurs. Ils ont l'inconvénient de modifier le timbre des signaux reproduits s'ils sont mal réglés, et la zone d'écoute optimale est limitée à une position relativement précise de l'auditeur. Comme tout système de diffusion sur haut-parleur, un système transaural est aussi sensible à l'effet de la salle d'écoute. Partant de l'hypothèse que cet effet joue un rôle prépondérant dans la modification du timbre de restitution, notre approche vise à le réduire *a priori*. Pour cela, nous proposons de rapprocher les sources jusqu'à les placer à proximité immédiate de l'auditeur, maximisant ainsi le rapport de l'énergie directe sur l'énergie renvoyée par les parois. Ce placement inhabituel des sources sonores peut alors rendre le système plus sensible à d'autres facteurs, tels que le placement incorrect de l'auditeur ou les variations morphologiques inter-individuelles.

L'effet d'une forte proximité des sources sur le rendu transaural est l'axe principal de ces travaux. Pour l'étudier, différents facteurs pouvant modifier le rendu aux oreilles de l'auditeur sont évalués séparément : influence de la morphologie, influence des filtres transauraux sur les sources électro-acoustiques, effet de salle et placement de l'auditeur. Pour tous ces facteurs, différents placements en distance, angle d'azimut et angle d'élévation des sources sonores ont été évalués.

Notre objectif est d'implémenter un système qui soit destiné à l'écoute de signaux, des tests perceptifs ont donc été à la base de notre démarche. En revanche, de multiples raisons pratiques font qu'une évaluation perceptive systématique n'est pas efficace pour mettre au point un nouveau dispositif de reproduction. Nous sommes donc partis de l'évaluation d'une sélection de configurations, et les résultats ont permis d'implémenter des indicateurs objectifs susceptibles

d'estimer les réponses des auditeurs à partir de l'information contenue dans les signaux. Ils ont par la suite permis d'étendre les résultats à des configurations qui ont été simulées mais n'ont pas été évaluées par les auditeurs.

Le premier chapitre est une présentation de l'état de l'art concernant la problématique de la restitution sonore spatialisée. Plusieurs aspects de cette problématique sont présentés : les technologies de diffusion spatialisée (en particulier le binaural et transaural), la diffraction par l'auditeur, l'influence de la salle et l'égalisation de systèmes audio.

Le principe du transaural consiste à compenser l'effet de diffraction par l'auditeur, caractérisé par les HRTF. Le deuxième chapitre permet d'étudier cet effet en fonction de la distance des sources. Des mesures ont été réalisées sur quatre mannequins à deux distances et comparées entre elles ainsi qu'à un modèle de sphère. La possibilité de propager des mesures obtenues en proximité vers des mesures à plus grande distance est étudiée. Ces travaux ont permis de valider l'utilisation d'une démarche simplifiée pour simuler la reproduction transaurale, afin d'évaluer l'influence de différents facteurs dans les autres chapitres.

L'utilisation de filtres transauraux implique une plus grande sollicitation des sources électro-acoustiques. Les sources utilisées en proximité de l'auditeur doivent être de petite taille, ce qui limite mécaniquement leurs performances aux basses fréquences. L'influence des filtres sur la sollicitation des sources est étudiée au chapitre III.

L'influence de la salle d'écoute est étudiée dans le chapitre IV, en cherchant à déterminer un « meilleur » moyen de limiter son influence, sous forme d'un compromis entre la proximité des sources et l'égalisation du système de diffusion. Ceci est évalué perceptivement dans une configuration aussi simple que possible : les signaux sont monophoniques et captés sans auditeur.

L'intuition suggère que le placement de sources en proximité peut rendre la diffusion plus sensible à un déplacement de l'auditeur. Cet aspect est étudié par le biais d'un test perceptif au chapitre V, suivant la même démarche qu'au chapitre IV.

Le dernier chapitre tente alors de généraliser les résultats des chapitres IV et V à un grand nombre de configurations qui n'ont pas toutes été effectivement évaluées par des auditeurs. Des indicateurs issus des tests d'écoute permettent alors de déduire leurs résultats. Ce dernier chapitre permet alors de faire la synthèse des différents phénomènes influant le rendu sonore, et de les mettre en relation. Cette synthèse permet ainsi de proposer quelques configurations optimales selon les critères identifiés, pour répondre au cahier des charges initial.

Chapitre I Etat de l'art

Table des matières

I.1 Introduction

L'objectif des travaux de la thèse est d'optimiser un système de diffusion sonore dans une salle d'écoute, et ceci pose plusieurs problèmes qui ont été abordés dans la littérature. Tout d'abord, la salle d'écoute modifie le rendu sonore, et plusieurs aspects liés à cette influence sont présentés. La reproduction sonore dépend aussi des caractéristiques des sources, qu'il est possible de contrôler dans une certaine mesure par une égalisation fréquentielle. Cette égalisation n'est d'ailleurs pas limitée aux sources et des méthodes ont été proposées pour compenser aussi l'effet de la salle. Le système de reproduction est destiné à un auditeur : il faut prendre en compte sa présence au centre du dispositif. Elle particularise significativement le champ de pression au niveau de ses oreilles, et ces particularités lui permettent de localiser des sources dans l'espace. Les principaux mécanismes de perception spatiale sont présentés, ainsi que des travaux récents concernant la caractérisation de la diffraction par l'auditeur.

Plusieurs techniques de systèmes spatialisés ont déjà été développées et sont succinctement présentées, l'accent étant mis sur les systèmes transauraux qui correspondent au contexte du travail présenté ici.

I.2 Influence de la salle d'écoute sur le rendu sonore

Les concepts expliqués dans la suite sont principalement issus de livres : [Gade, 2007], [Everest & Pohlman, 2009]. Lorsqu'une source sonore rayonne, l'énergie est réfléchie et diffusée par les différentes parois et objets jusqu'à être dissipée. La répartition temporelle de ces réflexions dépend des caractéristiques de la salle. Trois composantes sont généralement considérées comme l'illustre la Figure 2 : le champ direct, les premières réflexions et la réverbération tardive. Le champ direct correspond au premier front d'onde, associé au rayonnement direct de la source. Les premières réflexions sont généralement assez bien identifiables, et la réverbération correspond à une multitude de réflexions difficilement discernables. Cette réverbération est souvent associée à un champ diffus, c'est-à-dire un champ de pression équivalent en tout point de la salle avec une phase aléatoire. La limite entre les premières réflexions et la réverbération n'est pas toujours clairement identifiable, une méthode pour l'estimer est par exemple proposée par [Bidondo et al., 2016].

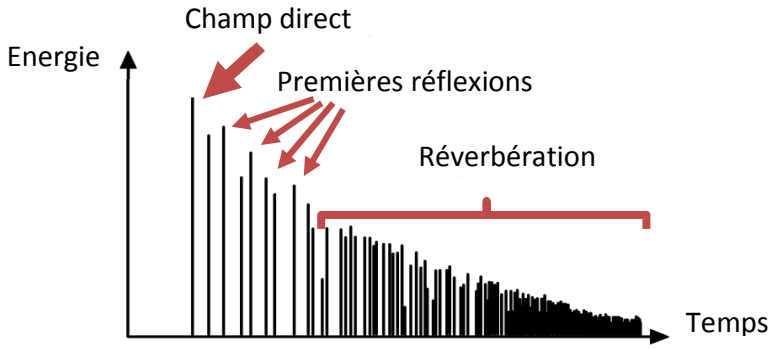


Figure 2 : représentation d'une Energy Time Curve (ETC) de salle

Cette répartition de l'énergie a plusieurs effets sur le rendu sonore : des effets temporels, spectraux et spatiaux. L'effet le plus connu d'une salle d'écoute est sans doute la traînée de réverbération, qui correspond à la réverbération tardive. Un exemple extrême est le jeu de l'organiste dans une église : le son peut se propager plusieurs secondes après l'intervention du musicien [Friot et al., 2016]. Dans le cas d'une salle usuelle le phénomène est moins prononcé mais peut modifier le rendu. Le temps de réverbération désigné par TR_{60} est un indicateur de ce phénomène, correspondant au temps nécessaire pour que le niveau de pression décroisse de 60 dB après interruption d'une excitation par un bruit stationnaire.

I.2.A. Cas d'une réflexion

Si les premières réflexions apparaissent suffisamment tardivement, un phénomène d'écho est perçu, éventuellement plusieurs fois. La perception de cet écho dépend de l'amplitude de la réflexion ainsi que du retard associé. Par exemple, Olive et Toole [Olive & Toole, 1989] ont investigué la perception d'une réflexion latérale en environnement anéchoïque, pour une incidence frontale du champ direct. Ils ont étudié l'effet de deux paramètres : le délai entre la réflexion et le champ direct, et l'amplitude relative de la réflexion. La courbe de gauche de la

Figure 3 représente les résultats de l'expérience pour un signal de parole. Un écho est audible si la réflexion est de même amplitude que le champ direct et retardée de plus de 40 ms. Si au contraire la réflexion intervient peu de temps après le champ direct, la réflexion n'est pas perçue comme un écho mais comme une modification du champ direct. Pour la zone comprise entre les seuils B et C, la réflexion introduit un changement de la perception de la localisation, ainsi qu'une sensation d'espace. Le signal a une influence significative sur la perception d'une réflexion. La courbe de droite de la Figure 3 illustre ce phénomène : par exemple pour une réflexion retardée de 10 ms, le seuil de détection est à -25 dB pour des clics, alors qu'il est à -15 dB pour de la parole.

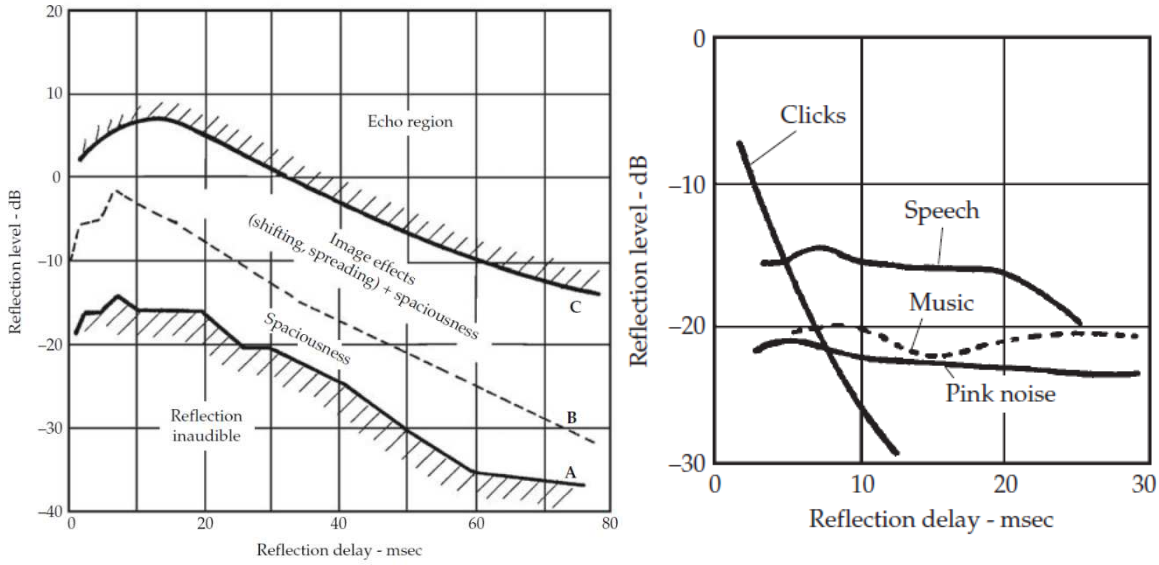


Figure 3 : Courbes de gauche : perception d'une réflexion latérale en fonction du retard et de l'amplitude de la réflexion pour un signal de parole. La courbe A correspond au seuil de détection de la réflexion, la courbe B est le seuil de perception d'un décalage de l'incidence perçue et la courbe C est le seuil à partir duquel la réflexion est perçue comme un écho. La figure de droite correspond au seuil de détection de la réflexion pour différents stimuli. Figures extraites de [Everest & Pohlman, 2009] pp104 à partir de travaux de [Olive & Toole, 1989]

Pour les réflexions intervenant peu de temps après le champ direct, une modification spectrale apparaît sous forme de « filtrage en peigne » : une réflexion peut être modélisée comme une version retardée et atténuée du champ direct, et les deux contributions se somment au point récepteur, et sur l'ensemble du spectre ce filtre a l'allure d'un peigne.

L'influence d'une réflexion est toujours un sujet d'étude, par exemple des auteurs [Robotham et al., 2016] ont évalué l'influence de la présence ou non d'une réflexion sur le plafond pour la diffusion de signaux musicaux. Certains auditeurs ont préféré la restitution avec la réflexion au plafond : selon le contexte d'étude, l'effet de salle n'est donc pas systématiquement négatif.

I.2.B. Modes acoustiques

Aux basses fréquences, le comportement modal de la salle d'écoute domine, caractérisé par la présence de résonances et anti-résonances. A chaque mode est associée une « fréquence propre » liée aux dimensions de la pièce. Par exemple, dans le cas d'une salle rectangulaire aux parois rigides, les fréquences f_{propre} pour lesquelles des résonances apparaissent sont données par la relation suivante [Everest & Pohlman, 2009] :

$$f_{propre} = \frac{c}{2}\sqrt{\left(\frac{n_x}{x}\right)^2 + \left(\frac{n_y}{y}\right)^2 + \left(\frac{n_z}{z}\right)^2} \quad (1)$$

Avec (x, y, z) les dimensions de la salle, et n_x, n_y, n_z des entiers supérieurs ou égaux à 0. Les premières résonances apparaissent donc à des fréquences plus élevées pour les salles de petites dimensions. Pour des salles de géométries plus complexes, ces résonances existent mais l'estimation des fréquences propres n'est pas aussi simple.

Aux basses fréquences, les modes de la salle sont peu denses : la distribution fréquentielle et spatiale de l'énergie est alors hétérogène. La densité modale augmente avec le carré de la fréquence. La fréquence à partir de laquelle le champ est considéré comme « diffus » f_{diffus} peut être estimée à partir du volume V de la salle et du temps de réverbération TR_{60} [Everest & Pohlman, 2009] :

$$f_{diffus} = 2000\sqrt{\frac{V}{TR_{60}}} \quad (2)$$

Des auteurs se sont intéressés à la perception des modes acoustiques, en étudiant notamment leur seuil d'audibilité en fonction de leur durée d'extinction [Fazenda et al., 2015]. Ce seuil est plus bas pour des signaux artificiels que pour des signaux musicaux, et d'autant plus élevé que la fréquence est basse. Par exemple pour des signaux artificiels, à 32 Hz un mode est audible s'il résonne pendant plus de 0.9 s, alors qu'à 100 Hz il est audible si sa durée dépasse 0.18 s.

I.2.C. Recommandations

Des recommandations ont été établies pour le dimensionnement d'une salle d'écoute destinée à des tests perceptifs.

Pour l'évaluation de petits défauts dans les systèmes audio multicanaux, l'ITU recommande l'utilisation de salle d'écoute d'une surface comprise entre 30 m² et 70 m² [ITU-R BS.1116-3, 2015]. Pour les systèmes stéréo, la surface peut être comprise entre 20 m² et 60 m². La recommandation détaille également les proportions que doit respecter la salle d'écoute, et le temps de réverbération entre 200 Hz et 4 kHz doit être de $TR_{60} = 0.25\left(\frac{V}{V_0}\right)^{\frac{1}{3}}$ avec une tolérance de +/- 50 ms, $V_0 = 100\ m^3$ étant le volume de référence. Ce dernier point est le plus important à satisfaire d'après l'ITU : une salle aux dimensions plus petites peut être acceptable si le temps de réverbération est conforme.

Par ailleurs, l'Audio Engineering Society (AES) recommande l'utilisation de salles d'écoute d'au moins 20 m², avec une hauteur sous plafond d'au moins 2,1 m [AES, 1996]. Le temps de réverbération est également spécifié, il doit par exemple être de 450 ms +/- 150 ms aux fréquences moyennes.

I.3 Egalisation de systèmes d'écoute

Nous avons vu qu'une salle d'écoute peut modifier significativement le signal perçu. La chaîne électro-acoustique peut également introduire des distorsions linéaires et non-linéaires,

principalement liées aux propriétés des haut-parleurs et de leurs enceintes. Des travaux se sont intéressés à la perception de ces distorsions. Notamment, Bucklein [Bucklein, 1981] a montré que les pics d'amplitude de la réponse linéaire sont plus audibles que ses creux. Fryer [Fryer, 1977] a montré que les résonances couvrant un large domaine spectral sont plus audibles que les résonances étroites. Toole et Olive [Toole & Olive, 1988] ont confirmé les constats précédents, ont montré l'influence du signal à évaluer, et trouvé que les résonances pouvaient être moins perceptibles en conditions réverbérantes. Lavandier [Lavandier, 2005] s'est intéressé aux dissemblances perçues entre différentes enceintes acoustiques et a identifié des dimensions perceptives liées à ces dissemblances. Il a également montré que les dissemblances perçues sont bien corrélées avec la densité de sonie calculée à partir des signaux. Pierre-Yohan Michaud a poursuivi ces investigations, en s'intéressant plus particulièrement aux distorsions non linéaires [Michaud, 2012].

Pour limiter les distorsions linéaires, un certain nombre de méthodes ont été proposées pour compenser les irrégularités de la réponse d'un système par filtrage du signal à reproduire. Les méthodes d'égalisation sont aujourd'hui implémentées sous forme numérique, où deux grandes familles peuvent être distinguées : les filtres à réponse impulsionnelle finie (Finite Impulse Response, FIR) et les filtres à réponse impulsionnelle infinie (Infinite Impulse Response, IIR) [Oppenheim & Schafer, 1975] pp 195-271. Les filtres FIR ont l'avantage d'être toujours stables, plus simples à construire et peuvent avoir un contrôle très fin sur l'amplitude et la phase. Les filtres IIR ont l'avantage d'être plus efficaces pour un nombre donné de coefficients.

I.3.A. Méthodes d'égalisation « générique »

Les méthodes « génériques » sont celles qui sont usuelles dans la communauté audio. Une revue très complète des méthodes d'égalisation a été proposée dans [Valimaki & Reiss, 2016], synthétisée ici. Trois approches différentes de calcul de filtre d'égalisation peuvent être distinguées : égalisation paramétrique, égalisation graphique et égalisation globale.

I.3.A.a Egalisation paramétrique

Les filtres d'égalisation paramétrique sont définis à partir de trois paramètres : le gain, la fréquence centrale et le facteur de qualité. Ces filtres amplifient ou atténuent une bande de fréquences, sans modifier significativement le reste du spectre. Ces filtres sont de type « plateau » (« shelf »), permettant de modifier le gain d'une large bande de fréquences, ou de type « peak / notch » permettant d'ajouter des pics ou creux dans la réponse. Ces filtres ont initialement été conçus sous forme analogique, et leur implémentation numérique est généralement réalisée par des filtres IIR.

Il est souvent nécessaire de combiner plusieurs filtres paramétriques pour corriger une large bande de fréquence. Ainsi Ramos et Lopez [Ramos & Lopez, 2006] ont proposé une méthode pour ajuster une série de filtres IIR d'ordre 2 de type « peak / notch », où chacun d'entre eux vise à corriger un défaut particulier de la réponse fréquentielle à égaliser.

I.3.A.b Egalisation graphique

Les filtres d'égalisation graphique permettent de corriger individuellement le gain de bandes de fréquences adjacentes. Ces bandes de fréquences correspondent en général aux 31 bandes de

tiers d'octave de fréquences centrales comprises entre 20 Hz et 20 kHz. Ce type d'égalisation est moins flexible que l'égalisation paramétrique car les bandes de fréquences sont fixes, mais son utilisation est considérée comme plus intuitive.

I.3.A.c Egalisation globale

Les égalisations paramétriques et graphiques sont des méthodes facilitant l'interaction avec l'utilisateur. Elles sont typiquement utilisées en sonorisation, où le dispositif de reproduction sonore est installé de manière temporaire. L'ingénieur du son dispose généralement d'un temps réduit pour égaliser le dispositif de restitution et doit parfois adapter les réglages pendant la représentation : il a besoin d'outils souples et efficaces.

D'autres méthodes existent, moins adaptées à l'interaction avec l'utilisateur. Ces méthodes permettent notamment de calculer le filtre d'égalisation conduisant à un gabarit cible. Kirkeby et Nelson [O. Kirkeby & Nelson, 1999] ont proposé une méthode de calcul de filtre FIR à partir d'une minimisation au sens des moindres carrés, en appliquant une régularisation de Tikhonov.

Pour le calcul de tels filtres d'égalisation, certains auteurs ont proposé de déformer l'échelle des fréquences, pour optimiser la répartition des coefficients du filtre selon des critères psychoacoustiques [Karjalainen et al., 1996] [Harma et al., 2000]. Les filtres associés sont désignés comme Warped Filters, WFIR ou WIIR selon le cas.

I.3.B. Egalisation spécifique aux salles d'écoute

Pour minimiser l'influence de la salle d'écoute, deux approches sont classiquement employées : soit par traitement acoustique de la salle d'écoute, soit par égalisation du signal à reproduire.

Les quantités de matériaux absorbants à utiliser pour absorber l'énergie aux parois peuvent être très importantes notamment aux basses fréquences où les longueurs d'onde mises en jeu peuvent atteindre plusieurs mètres. Une alternative moins coûteuse que l'ajout de matériaux absorbants peut se faire par le contrôle des signaux à reproduire. Ce procédé n'est pas trivial, notamment car les réponses impulsionnelles mesurées dans des salles d'écoute se modélisent rarement comme une réponse impulsionnelle à phase minimale et un retard pur [Neely & Allen, 1979]. Cela se traduit par une difficulté à égaliser la phase d'un système de reproduction en salle d'écoute. Les méthodes génériques peuvent être employées, mais elles ne compensent généralement que l'amplitude spectrale. Des solutions spécifiques à la reproduction en salle d'écoute ont donc été proposées [Elliott & Nelson, 1989] [Craven & Gerzon, 1992] [Bharitkar et al., 2004] [Mertins et al., 2010] [Bank, 2013]

Certains auteurs proposent d'égaliser dans un premier temps les sources à partir de mesures anéchoïques, puis de calculer ensuite un filtre d'égalisation ciblant la salle de restitution [Craven & Gerzon, 1992]. Une alternative évitant les mesures anéchoïques est proposée par Bank [Bank, 2013] : elle consiste à tronquer la réponse impulsionnelle mesurée dans la salle pour calculer un filtre d'égalisation ciblant les fréquences moyennes et hautes. Une mesure complémentaire est réalisée après application de ce filtre, et les basses fréquences sont alors égalisées par bandes de fréquences.

Des auteurs [Mertins et al., 2010] ont proposé une méthode visant à réduire la durée de la réverbération perçue. Le principe est de réduire l'amplitude des réflexions, de sorte que les phénomènes de masquage temporel les rendent moins perceptibles.

Le champ acoustique dans une salle d'écoute varie d'un point à l'autre, et il est donc impossible en toute rigueur d'utiliser un filtre d'égalisation unique pour une zone d'écoute étendue. Une solution consiste à cibler une moyenne sur plusieurs points de mesure, et le calcul du filtre d'égalisation est alors un problème d'optimisation [Elliott & Nelson, 1989]. Une alternative plus simple est de calculer un filtre d'égalisation à partir de la moyenne des transferts mesurés [Bharitkar et al., 2004].

Nous avons évoqué le fait que la densité modale est faible en basses fréquences. Les méthodes génériques peuvent s'appliquer aux basses fréquences, mais posent plusieurs problèmes pratiques. Premièrement, l'effet des modes peut introduire des creux d'amplitude marqués, qui nécessitent alors d'employer un filtre avec un fort gain limitant la dynamique du système. De plus, les modes de salle ont généralement un support temporel très long, délicat à compenser avec les méthodes génériques. Des méthodes ont été proposées visant à réduire le support temporel des résonances aux fréquences propres [Makivirta et al., 2001]. Par ailleurs, la variation spatiale des modes de salles empêche de les contrôler rigoureusement sur une zone étendue au moyen d'une seule source. Des solutions spécifiques ont été proposées et certaines ont été explorées dans le cadre du stage de Pierre-Yvon Bryk [Bryk, 2014]. Parmi elles, une méthode consiste à simuler la propagation d'une onde plane à partir d'un réseau de haut-parleurs [Santillan et al., 2007]. Cette méthode ne s'applique que pour des salles rectangulaires, et ne permet de contrôler que les modes axiaux. Le principal avantage de cette méthode est qu'elle permet d'avoir un contrôle dans la quasi-totalité de la salle d'écoute. Au contraire, une autre méthode vise à ne réduire l'influence des modes acoustiques que dans la zone de l'espace où se trouvent les auditeurs [Herzog, 2005].

Des auteurs [Welti & Devantier, 2006] ont proposé une méthode visant à optimiser le placement des sources pour minimiser la variance d'amplitude entre différents points d'écoute. Pour chacune des sources un filtre paramétrique est appliqué. Les paramètres optimaux de positions des sources et les filtres associés sont déterminés par un algorithme d'optimisation. L'inconvénient de cette méthode est qu'elle nécessite de connaître toutes les fonctions de transfert entre les points d'écoute et les positions des sources à tester, ce qui nécessite de nombreuses mesures.

Pour le problème spécifique des basses fréquences, des solutions hybrides entre contrôle actif et passif ont enfin été proposées. Elles consistent à placer des absorbeurs électro-acoustiques en des points particuliers de la salle d'écoute [Meynial, 1999] [Boulandet, 2012] [Rivet et al., 2016].

I.3.C. Egalisation de systèmes non-linéaires

S'il est éventuellement possible de corriger des effets linéaires jusqu'à une certaine limite, la compensation des non-linéarités est bien plus complexe. En effet, les distorsions non-linéaires dépendent du contenu spectral et du niveau, le traitement d'égalisation devant alors s'adapter à leurs évolutions. Des méthodes de compensation de non-linéarité ont néanmoins été présentées dans la littérature, par exemple [Klippel, 1996],[Shi et al., 2007],[Defraene et al., 2012].

La perception de la distorsion non-linéaire n'est pas encore complétement maîtrisée, même si Pierre-Yohan Michaud s'est intéressé dans sa thèse à l'identification des dimensions perceptives associées à des distorsions non-linéaires [Michaud, 2012]. Pour l'évaluation perceptive de petits défauts dans les systèmes audio, l'ITU recommande d'utiliser des sources dont le taux de distorsion harmonique n'excède pas 3 % pour les fréquences inférieures à 250 Hz et 1 % pour les fréquences supérieures pour un signal de 90 dB_{SPL} à 1 m du haut-parleur [ITU-R BS.1116-3, 2015].

I.4 Influence de l'auditeur

L'objectif est de réaliser un système d'écoute, ce qui implique qu'il faut tenir compte de la présence d'un auditeur au sein du dispositif. La présence de l'auditeur modifie significativement le champ de pression, en particulier au niveau de ses oreilles, et ces modifications sont essentielles car elles lui permettent de localiser des sources dans l'espace. Les mécanismes permettant à l'auditeur de percevoir l'espace sonore sont donc brièvement rappelés, suivis de la caractérisation de la diffraction par l'auditeur.

I.4.A. Mécanismes de la perception spatiale

Les « capteurs » naturels d'une scène sonore sont les deux oreilles, utilisés pour identifier l'origine d'un champ sonore dans l'espace en trois dimensions avec une précision suffisante. L'utilisation de deux capteurs pour séparer trois dimensions sous-entend une interprétation complexe, probablement gérée par le cerveau. Cette interprétation serait basée sur plusieurs indices contenus dans l'information sonore, dont les plus usuellement décrits sont la différence interaurale de temps (ITD), la différence interaurale de niveau (ILD) et les indices spectraux [Blauert, 1997].

I.4.A.a Différence Interaurale de Temps (ITD)

Pour une source sonore qui n'est pas située dans le plan médian de l'auditeur, le front d'onde arrive avec un certain décalage temporel entre les deux oreilles. Ce décalage est nommé Différence Intéraurale de Temps ou Interaural Time Difference (ITD). Selon la longueur d'onde mise en jeu, deux mécanismes sont invoqués. Pour les longueurs d'onde supérieures au diamètre de la tête (environ 18 cm, correspondant à environ 1800 Hz), l'ITD se traduit par une différence de phase. Aux fréquences plus élevées, une ambiguïté peut apparaitre : les décalages temporels peuvent être supérieurs à une période. Un autre indice temporel est alors proposé, basé sur les différences d'arrivée de l'énergie.

I.4.A.b Différence Interaurale de Niveau (ILD)

La position d'une source sonore induit une différence de niveau entre les deux oreilles. Cette Différence Interaurale de Niveau ou Interaural Level Difference (ILD) est principalement liée à la diffraction par l'auditeur : la tête fait « écran » et atténue l'énergie selon l'incidence de la source. L'ILD est plus marqué aux hautes fréquences, alors que cette différence devient faible lorsque la longueur d'onde est grande devant les dimensions de la tête.

I.4.A.c Les indices spectraux

Les deux mécanismes décrits précédemment sont complémentaires sur la bande des fréquences audibles, et forment ce que l'on nomme la théorie duplex [Rayleigh, 1907]. Ils ne sont cependant pas suffisant pour permettre une localisation parfaite : selon cette modélisation il existe des points

dans l'espace pour lesquels les couples de valeur (ITD, ILD) sont identiques. En outre, ces mécanismes ne permettent pas de localiser une source dans le plan médian.

En complément des indices binauraux, le cerveau humain se baserait également sur des indices monauraux liés au spectre du signal perçu. Lorsqu'une onde sonore parvient à nos oreilles, elle est diffractée en fonction de notre morphologie : certaines fréquences sont amplifiées alors que d'autres sont atténuées, et il en résulte un signal filtré aux oreilles de l'auditeur. Chaque position de l'espace conduit à un filtrage spécifique résultant des formes des oreilles, de la tête et du torse. Un apprentissage individuel de ce filtrage pourrait expliquer l'importance de la morphologie individuelle pour la perception spatiale [Andeol, 2012].

I.4.A.d Indices multi-sensoriels

La localisation de sources sonores peut être facilitée par d'autres indices. Notamment, les mouvements de tête, même inconscients, aident à la résolution d'ambigüités de localisation [Moller, 1992] [Begault, 1994] [Bronkhorst, 1995]. L'association d'un autre sens perceptif peut également contribuer à améliorer la précision de la localisation. Il semblerait en particulier que la contribution visuelle soit prépondérante par rapport à la contribution sonore : l'effet « ventriloque » mis en évidence dans [Recanzone, 1998] est l'illusion de percevoir une source sonore à l'emplacement d'une source visuelle alors que la source sonore provient d'une direction différente.

I.4.A.e Performances de la localisation humaine

Les capacités de l'humain à localiser des sons ont été étudiées par de nombreux auteurs, notamment pour les mettre en relation avec les performances de systèmes de reproduction [Bertet, 2009; Blauert, 1997; Carlile et al., 1997; Pulkki & Hirvonen, 2005]. En particulier, la localisation est plus précise pour des sources face à l'auditeur et dans le plan médian. Blauert [Blauert, 1997] a reporté les résultats de deux tests d'écoute réalisés sur 600 et 900 auditeurs, et l'erreur moyenne de localisation en azimuth est de 1° pour les incidences frontales, alors que pour les positions latérales l'erreur moyenne de localisation est de l'ordre de la dizaine de degrés et le flou de localisation autour de ±10°. Blauert a également reporté dans son ouvrage la précision en localisation dans le plan médian, qui est encore plus approximative que dans le plan horizontal. En outre, il est fréquemment fait allusion dans la littérature au phénomène de « confusion avant/arrière » qui consiste à percevoir un son derrière soi alors qu'il provient de devant (ou inversement) [Carlile et al., 1997].

I.4.B. Head-Related Transfer Function (HRTF)

Pour décrire la diffraction par l'auditeur, une modélisation sous forme de fonctions de transfert est couramment adoptée. Ces fonctions de transfert désignées par « Head-Related Transfer Functions » (HRTF) caractérisent la diffraction d'une onde incidente par la tête d'un auditeur.

La définition des HRTF adoptée dans ce mémoire est le rapport de la pression à l'entrée du conduit auditif de l'oreille de l'auditeur $P_{oreille}$, sur celle au centre de la tête (sans l'auditeur) $P_{champ\ libre}$ pour une source de coordonnées (r, θ, ϕ) :

$$HRTF(r,\theta,\phi) = \frac{P_{oreille}}{P_{champ\ libre}} \quad (3)$$

Les HRTF dépendent des coordonnées r, θ, ϕ de la source à faible distance, alors qu'elles ne dépendent plus de r en champ lointain (lorsque la distance à la source est supérieure aux plus grandes dimensions de l'objet diffractant, elle n'influence plus significativement la diffraction) [Duda & Martens, 1998].

I.4.B.a Mesures de HRTF

Les HRTF résultent de la morphologie de l'auditeur donc pour régler un système de diffusion il est utile de prendre en compte les HRTF de l'auditeur, qu'il est nécessaire de caractériser. Cette étape est délicate, car le nombre de positions à mesurer est important. Cela implique d'utiliser un dispositif adapté, qui introduit lui-même une diffraction spécifique. Les mesures sur des mannequins censés approximer un auditeur « moyen » sont un peu plus simples à mettre en œuvre, car il est plus facile de contrôler leur positionnement. Toutefois, la mesure sur ces mannequins pose déjà un problème de reproductibilité des mesures.

Des auteurs [Andreopoulou et al., 2015] se sont posés la question de la variabilité des mesures obtenues par différentes institutions et à différentes périodes à partir d'un même mannequin. Au total, douze jeux de HRTF mesurées par dix laboratoires différents sur le mannequin Neumann KU-100 ont été analysées. Les ITD obtenus par les différentes mesures ont été comparées, et des écarts allant jusqu'à 235 µs ont été observés pour certaines incidences. Selon le contenu du stimulus, la différence minimale perceptible (Just Noticeable Difference, JND) peut varier entre 2 µs et 60 µs [Blauert, 1997], alors que les variations introduites par la méthode de mesure sont très largement supérieures à ces JND. Des écarts sur les amplitudes spectrales ont également été mis en évidence, pouvant atteindre 6.7 dB pour des incidences arrières et 5 dB pour des incidences frontales, aux fréquences inférieures à 6 kHz. Au-dessus de 6 kHz, des écarts atteignant 22 dB ont été observés. Cette variabilité illustre la difficulté de mesurer des HRTF.

Une étude similaire a été proposée dans [Zhong et al., 2016], comparant cinq campagnes de mesures différentes sur le mannequin KEMAR. Une variation d'amplitude spectrale entre les différentes mesures est calculée pour chaque position spatiale. Les écarts atteignent jusqu'à 4 dB entre 5 kHz et 12 kHz. L'amplitude plus réduite de ces écarts que dans l'étude [Andreopoulou et al., 2015] se justifie peut-être par un nombre réduit de campagnes de mesures. Une évaluation perceptive de localisation complète ces indications objectives, et les résultats de cette expérience semblent révéler peu de différences perceptives. Les écarts entre mesures ne sont donc pas systématiquement reliés à la différence perçue.

I.4.B.b Mesures à proximité de la tête

Dans le cadre de ces travaux de thèse, les mesures à faible distance de la tête sont envisagées, ce qui les rend sensibles aux propriétés de la source. Les sources « usuelles » sont conçues pour être utilisées en champ lointain, où leur rayonnement est bien contrôlé. A moindre distance, leur rayonnement peut s'avérer plus complexe en raison de la déformée vibratoire de la membrane et de la diffraction introduite par la source, ce qui conduit à un champ variant selon la distance et la direction d'une manière dépendante de la source (ce qui n'est pas souhaitable pour mesurer les

HRTF). Il est donc impératif de placer la tête en champ lointain de la source, la distance minimale dépendant des longueurs d'ondes mises en jeu, mais également des propriétés de la source : elle est plus courte pour des sources de petite taille mais leur bande-passante est alors plus limitée en basses fréquences. Il est ainsi nécessaire de trouver un compromis entre une source suffisamment petite mais avec une bande passante adaptée aux mesures envisagées. Certains auteurs ont donc élaboré des sources spécifiques pour pouvoir réaliser des mesures très proches du mannequin [Hosoe et al., 2006], [Qu et al., 2009; Yu et al., 2010].

La plupart des campagnes de mesures présentes dans la littérature ont été effectuées en champ lointain [Gardner & Martin, 1994] [Algazi et al., 2001] [Warusfel, 2003] [Brinkmann et al., 2013], mais il existe quelques résultats de mesures à plus faible distance. Des auteurs [Brungart et al., 1999] ont comparé des HRTF mesurées sur un mannequin Kemar et sur une sphère aux distances 12.5 cm, 25 cm, 50 cm et 100 cm. D'autres auteurs [Hosoe et al., 2006] ont élaboré une source sonore adaptée à la mesure de HRTF en proximité et l'ont utilisée pour mesurer les HRTF d'un mannequin B&K 4128 à des distances allant de 20 cm à 1 m par pas de 10 cm. L'article ne fait pas état des résultats des mesures de HRTF, mais des données sont téléchargeables sur le site [Nagoya, 2016]. La technique de mesure par réciprocité utilisée dans [Zotkin et al., 2006] en 2006 permet une mesure de HRTF à 70 cm du centre de la tête. Par ailleurs, des auteurs [Qu et al., 2009] ont reporté des mesures effectuées sur un mannequin KEMAR à des distances de 20, 30, 40, 50, 75, 130 et 160 cm du centre de la tête (données téléchargeables sur le site [Qu et al., 2016]). D'autres auteurs [Yu et al., 2010] ont présenté des mesures également effectuées sur le mannequin KEMAR à des distances de 20 à 100 cm par pas de 10 cm, ainsi qu'à 25 cm. Ces données sont téléchargeables au format SOFA (Spatially Oriented Files for Acoustics [Majdak et al., 2013]) sur le site dédié [SOFA, 2016]. Les HRTF champ lointain issues de ces deux dernières bases ont d'ailleurs été comparées dans [Zhong et al., 2016]. Enfin, des auteurs [Wierstorf et al., 2011] ont présenté des mesures de HRTF du mannequin KEMAR dans le plan horizontal aux distances 3 m, 2 m, 1 m et 0.5 m (données téléchargeables [SOFA, 2016]).

I.4.B.c Individualisation de HRTF

Les mannequins dédiés constituent une approximation de la morphologie moyenne d'un auditeur. De nombreux travaux actuels concernent la prise en compte de la variabilité individuelle de morphologie [Katz, 2001], [Busson, 2006], [Iwaya, 2006], [Nicol et al., 2006], [Guillon et al., 2008], [Guillon & Nicol, 2008], [Guillon, 2009], [Rui et al., 2013], [Huttunen et al., 2014], [Wang & Chan, 2014], [Ghorbal et al., 2016], [Maazaoui & Warusfel, 2016]. La mesure individuelle pose de nombreux problèmes, et des approximations sont souvent préférées. Pour caractériser les particularités individuelles, certains travaux se basent directement sur quelques paramètres morphologiques [Guillon et al., 2008], d'autres sur une sélection de mesures acoustiques « représentatives » [Nicol et al., 2006], [Guillon & Nicol, 2008], [Wang & Chan, 2014], d'autres utilisent un maillage numérique de l'auditeur obtenu par scan optique ou photographie [Katz, 2001], [Rui et al., 2013], [Huttunen et al., 2014] et d'autres emploient des tests perceptifs [Iwaya, 2006]. A partir de l'identification de ces caractéristiques individuelles, les HRTF peuvent être sélectionnées parmi un ensemble de données mesurées ou calculées numériquement.

L'approche inverse est possible, utilisant la plasticité du cerveau de l'auditeur : la démarche vise alors à « adapter l'auditeur » à un jeu de HRTF données [Parseihian & Katz, 2012].

I.5 Technologies de son spatialisé

Un système de son spatialisé permet d'entendre des sources sonores situées en différentes positions de l'espace. Les principaux systèmes existants sont succinctement présentés ci-après, et les systèmes binauraux qui font l'objet de nos travaux sont alors décrits plus précisément. Deux principales familles de systèmes de diffusion peuvent se distinguer : ceux qui visent à créer une illusion de spatialisation en simulant des indices de localisation, et ceux qui sont basés sur une synthèse de champ physique visant à reproduire une pression cible.

I.5.A. La stéréophonie, systèmes surround et VBAP

La stéréophonie, les systèmes surround et le Vector Based Amplitude Panning (VBAP) sont des dispositifs visant à créer une « illusion spatiale ».

I.5.A.a Stéréophonie

La stéréophonie est probablement le dispositif de son spatialisé le plus répandu dans le grand public. Ce dispositif permet une spatialisation des sons entre deux haut-parleurs placés à ±30° face à l'auditeur. La sensation de spatialisation résulte des mécanismes de localisation d'ITD et d'ILD : par exemple, un son joué plus fort sur une enceinte est perçu comme provenant plutôt de cette enceinte.

L'acquisition de ce type de signaux se fait par des couples de microphones avec des positions et directivités particulières. Par exemple, le couple de type « XY » consiste à placer deux microphones cardioïdes orientés à 90°. En raison de la directivité des microphones, l'amplitude du signal de chaque microphone dépend de position de la source sonore. Il existe bien d'autres dispositifs d'acquisition, tenant aussi compte des différences de phase. La synthèse à partir d'une source virtuelle monophonique est possible en pondérant l'amplitude de cette source sur chacun des haut-parleurs.

Ce mécanisme est efficace, mais très limité en termes de performances : la localisation des sources est limitée aux positions entre les deux haut-parleurs. De plus la compatibilité d'une reproduction stéréophonique avec des contenus binauraux est discutable. Ce type de technologie ne semble donc pas directement adapté à notre problématique.

I.5.A.b Surround

Des technologies plus complexes reposant sur le paradigme précédent ont été développées par la suite. Ces technologies dites « *surround* » ont vu le jour pour des applications de type grand public, avec un certain nombre de formats particuliers. Le plus répandu est le 5.1, qui ajoute au système stéréophonique deux haut-parleurs à l'arrière (entre +/-100° et +/-120°), un haut-parleur central ainsi qu'un caisson de basse. Différents systèmes d'acquisition ont été spécialement développés pour cette restitution : croix IRT, decca tree, etc. ([Périaux et al., 2015] pages 181 à 237).

Des formats encore plus complexes consistent notamment à utiliser un plus grand nombre de haut-parleurs : 7.1, 22.2, etc. Une revue plus détaillée de ces formats, est proposée dans [Périaux et al., 2015] pages 20 à 46.

I.5.A.c Vector Based Amplitude Panning (VBAP)

L'effet de panoramique a également été généralisé à un nombre paramétrable de haut-parleurs placés assez arbitrairement : il s'agit du Vector Based Amplitude Panning (VBAP, [Pulkki, 1997]). Les performances de ce type de système sont directement liées au nombre de transducteurs, nécessairement élevé en 3D. Ce type d'approche ne semble donc pas pertinent dans le cadre d'un système léger. En outre, il n'existe pas de système de prise de son permettant de capter toute la complexité d'une scène et dédié à ce type de dispositif, et la transformation d'un contenu binaural pose de nombreux problèmes [Jakka, 2005]. La restitution VBAP est plus indiquée pour la synthèse de sources monophoniques.

I.5.B. Synthèse de champ de pression : WFS & HOA

Au contraire des techniques précédentes, la Wave Field Synthesis (WFS) et Higher Order Ambisonics (HOA) sont des dispositifs visant à reproduire physiquement un champ de pression dans une zone définie de l'espace.

I.5.B.a Wave Field Synthesis

La Wave Field Synthesis (WFS) est basée sur le principe de Huygens, qui permet de décomposer chaque front d'onde sonore comme la superposition de front d'ondes élémentaires. Un réseau de haut-parleurs où chaque haut-parleur génère un tel front d'onde élémentaire permet alors de diffuser des fronts d'onde complexes [Berkhout et al., 1993]. Ce principe est illustré par la Figure 4.

Cette méthode implique l'installation d'un réseau comportant de nombreux haut-parleurs, qui permettent de recréer le champ sonore physiquement par la somme de leurs contributions. Ainsi, le champ acoustique peut être recréé dans une zone étendue de l'espace, et donc la localisation des sources ne dépend pas de la position de l'auditeur. Une réalisation à grande échelle d'un tel dispositif existe à l'IRCAM [Noistering et al., 2012], basée sur un dispositif de 350 haut-parleurs. Une restriction couramment appliquée est la limitation de la restitution au plan horizontal, et un exemple de réalisation est proposé dans [Corteel, 2006], avec un réseau de 48 haut-parleurs formant un réseau linéaire de 8 m de long.

La distance entre deux haut-parleurs d_{HP} est conditionnée par le critère de Shannon spatial : il faut que $d_{HP} < \frac{\lambda}{2}$, λ étant la longueur d'onde la plus courte à reproduire. Si ce critère n'est pas respecté, les interférences entre ces contributions des différentes sources se traduisent par une reconstruction incorrecte du champ sonore. Ce critère est dimensionnant pour ce type de dispositif : par exemple, pour reproduire un champ sonore jusqu'à 10 kHz il faut que l'écartement entre sources soit inférieur à 1.7 cm. Pour un dispositif circulaire autour de l'auditeur à 1 m de distance, il faudrait alors compter 370 sources.

De plus, pour reproduire les basses fréquences il est nécessaire d'utiliser des transducteurs de grande taille, qui ne sont pas compatibles avec les écartements entre sources pour les hautes fréquences. Pour implémenter un dispositif large-bande il est alors nécessaire d'employer des réseaux multivoies, ce qui complexifie l'installation.

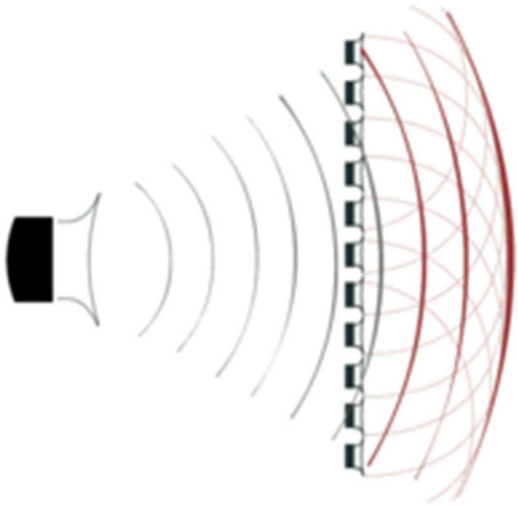


Figure 4 : illustration du principe de la WFS. La source virtuelle est représentée à gauche, et le champ de pression associé à cette source virtuelle est représenté en gris. La somme de chacune des contributions des haut-parleurs du réseau (en rouge) permet de générer le champ sonore de la source virtuelle.

I.5.B.b Ambisonics & Higher Order Ambisonics (HOA)

Les méthodes basées sur l'ambisonie ciblent une topologie différente : elles visent à reproduire le champ de pression acoustique autour d'un point de l'espace. Pour cela, une décomposition en série de Fourier-Bessel permet d'obtenir une représentation du champ sonore selon les harmoniques sphériques [Daniel, 2000]. Les harmoniques sphériques constituent une base orthogonale de l'espace L^2 sur la sphère unité paramétrée par (θ, ϕ). Avec cette représentation, le champ acoustique s'écrit en un point donné de coordonnées sphériques (r, θ, ϕ) :

$$p(kr, \theta, \phi) = \sum_{m=0}^{+\infty} i^m j_m(kr) \sum_{n=0}^{m} \sum_{\sigma=\pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \phi) \qquad (4)$$

Avec *k* le nombre d'onde, *r* le rayon d'observation, θ l'angle d'azimut et ϕ l'angle d'élévation dans le domaine fréquentiel, $j_m(kr)$ sont les fonctions de Bessel sphériques, $Y_{mn}^{\sigma}(\theta, \phi)$ les harmoniques sphériques et B_{mn}^{σ} les coefficients associés à ces harmoniques sphériques.

Pour définir parfaitement le champ acoustique dans le cas le plus général, cette somme doit être infinie. Dans la pratique, cela n'est pas réalisable, et cette décomposition doit être tronquée. L'ordre 0 définit uniquement une information omnidirectionnelle qui correspond à la pression acoustique au centre du repère, et les ordres supérieurs permettent de décrire les informations spatiales du champ sonore. Une décomposition jusqu'à l'ordre 1 permet de définir un système ambisonique, l'extension aux ordres supérieurs étant désignée par Higher Order Ambisonics (HOA) [Daniel, 2000].

L'un des avantages de cette technologie est qu'il existe un certain nombre de microphones permettant de capter des signaux adaptés (Soundfield, Eigenmike, etc.). Ces signaux permettent de définir les coefficients B_{mn}^{σ} en utilisant un encodage adapté aux caractéristiques du microphone. La restitution est possible sur de nombreux dispositifs de haut-parleurs, en utilisant un décodage adapté à la configuration. Le nombre minimal de haut-parleurs est fixé par l'ordre de décomposition. Les performances de ce type de système sont liées à l'ordre de restitution : d'après les travaux de Stéphanie Bertet [Bertet, 2009; Bertet et al., 2007], plus l'ordre est élevé et

plus la localisation est précise. Par exemple, pour le système le plus complexe testé (système d'ordre 4 avec 12 haut-parleurs dans le plan horizontal), la spatialisation est jugée « bonne » par des auditeurs en comparaison à un système de référence. Des comparaisons par paires ont également été effectuées dans ces travaux, montrant que les dimensions perceptives sont liées à l'ordre de décomposition et au nombre de haut-parleurs. En outre, la restitution de signaux binauraux est possible via une étape « d'up-mixing », en estimant la localisation d'une ou plusieurs sources [Jakka, 2005].

I.5.C. La technologie binaurale

Le principe de la technologie binaurale consiste à reproduire aux oreilles d'un auditeur la pression telle qu'elle aurait existé dans une situation d'écoute réelle. Les signaux compatibles avec ce format peuvent s'enregistrer avec des microphones au niveau des oreilles d'une tête artificielle ou humaine, et peuvent également être synthétisés à partir des HRTF. L'intérêt de ce format est de minimiser le volume d'information à enregistrer, ce qui explique sa popularité dans le monde industriel, et donc le volume de contenus disponibles. Par ailleurs, ce type de technologie tendrait à se développer pour le grand public [Nicol et al., 2014]. L'inconvénient d'un enregistrement binaural est qu'il contient une « signature spatiale » (HRTF) spécifique, ce qui limite la possibilité de le reproduire de manière générale.

Il a notamment été constaté que la localisation pouvait être approximative avec ce type de signaux, notamment pour les sources frontales et lorsque des HRTF non individuelles sont employées [Volk et al., 2008] : les sources ont tendance à être perçues à l'intérieur de la tête, (on parle de « perception intracrânienne »). Ce phénomène est illustré par la Figure 5 : la figure de gauche représente la trajectoire d'une source sonore, et les trois autres figures représentent des trajectoires perçues par les auditeurs. Cette figure illustre alors la variabilité de perception selon les auditeurs, et la difficulté à reproduire des sources frontales en reproduction binaurale.

Un autre inconvénient avec cette technique est que la scène sonore est « figée » : lorsque l'auditeur bouge la tête, la scène sonore suit ce mouvement contrairement à une écoute réelle. Ce phénomène peut être compensé par l'usage d'un head-tracking, permettant par ailleurs d'accroître les performances de localisation [Begault et al., 2001]. L'usage d'un head-tracking n'est toutefois compatible qu'avec la synthèse binaurale, un contenu pré-enregistré en binaural n'en bénéficie pas sans artifice (up-mixing).

Pour la diffusion, le moyen le plus simple consiste à utiliser un casque d'écoute, avec une égalisation du casque de restitution [Moller, 1992]. L'avantage d'une restitution au casque est qu'elle permet d'être indépendante du local d'écoute. En revanche, il existe un certain nombre d'inconvénients liés à cette diffusion. Tout d'abord, l'utilisation d'un casque d'écoute peut être source d'inconfort pour l'auditeur et il est assez intrusif. Une reproduction fidèle nécessite une égalisation précise de la réponse du casque, alors que son étalonnage pose encore problème. La manière de poser le casque d'écoute peut être à l'origine de variabilité dans la restitution. Le port de lunettes de vue par exemple peut favoriser la présence de fuites acoustiques. Des travaux ont montré que les différences liées à la variabilité de la pose du casque sont perceptibles [Paquier et al., 2011]. Enfin, l'utilisation du port du casque complique la comparaison directe du rendu avec d'autres systèmes de reproduction [A. H. Moore et al., 2007].

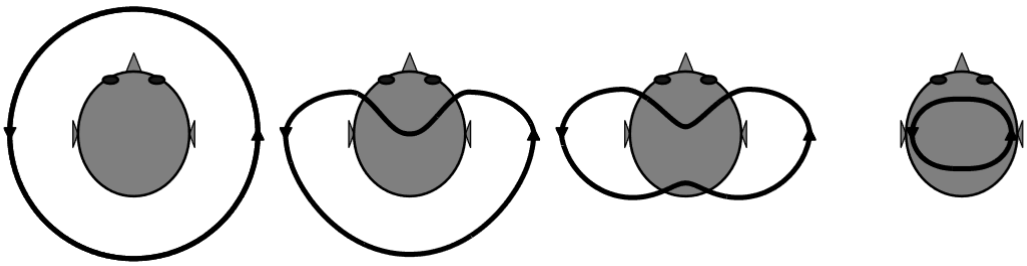


Figure 5 : exemples de trajectoires (au centre et à droite) perçues par des auditeurs à l'écoute d'un son spatialisé effectuant un cercle parfait autour de la tête (à gauche) synthétisé avec des HRTF non individuelles. Figure issue de [Parseihian, 2012], élaborée à partir de résultats de [Begault & Wenzel, 1993; Kim & Choi, 2005] et de discussions avec des utilisateurs de synthèse binaurale

I.5.C.a Diffusion transaurale

L'autre manière classique de diffuser un contenu binaural est l'utilisation de haut-parleurs, constituant alors un système transaural. Un traitement compensant les chemins croisés entre les haut-parleurs et les oreilles est alors nécessaire. Si le traitement permet effectivement d'annuler ces chemins croisés, la pression aux oreilles de l'auditeur correspond à celle qui aurait été diffusée par un casque d'écoute sans le traitement transaural. L'avantage d'un système transaural comparativement à une restitution au casque est qu'elle est moins intrusive, et élimine l'incertitude liée à la position du casque. Toutefois, comme pour tout dispositif de haut-parleur, la salle d'écoute peut influencer le rendu. Par ailleurs, le système vise à reproduire la pression au niveau des oreilles de l'auditeur : la zone d'écoute optimale est donc centrée sur un point précis. Enfin, les mouvements de tête ne conduisent plus à affiner la perception spatiale, et la restitution avec head-tracking est plus difficile que dans le cas d'un casque.

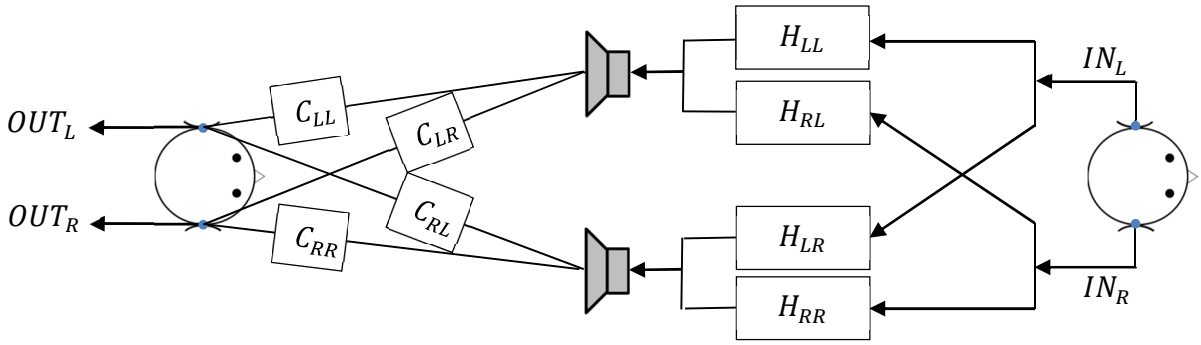


Figure 6 : schéma bloc d'une reproduction transaurale

Un système transaural à deux haut-parleurs est décrit dans le domaine fréquentiel de la manière suivante :

$$\begin{bmatrix} OUT_L \\ OUT_R \end{bmatrix} = \begin{bmatrix} C_{LL} & C_{RL} \\ C_{LR} & C_{RR} \end{bmatrix} \cdot \begin{bmatrix} H_{LL} & H_{RL} \\ H_{LR} & H_{RR} \end{bmatrix} \cdot \begin{bmatrix} IN_L \\ IN_R \end{bmatrix} \tag{5}$$

$$\boldsymbol{OUT} = C.H.\boldsymbol{IN} \tag{6}$$

L'équation (6) synthétise le processus, avec $[C]$ la matrice des filtres directs, $[H]$ la matrice des filtres transauraux et $\boldsymbol{IN}$ et $\boldsymbol{OUT}$ les entrées et sorties du système. La Figure 6 reprend l'équation

(5) de manière schématique. Idéalement $\boldsymbol{OUT} = \boldsymbol{IN}$, et pour satisfaire cette égalité, le calcul des filtres transauraux consiste à calculer une matrice $[H]$ inverse de $[C]$.

I.5.C.b Méthodes de calculs des filtres transauraux

L'estimation de la matrice inverse $[H]$ n'est pas un problème trivial, et plusieurs méthodes ont été proposées dans la littérature pour résoudre ce problème :

i Inversion du déterminant

En l'absence de singularités, l'inverse de la matrice C dans le domaine fréquentiel est donnée par la relation suivante :

$$[H] = \frac{1}{C_{LL}.C_{RR} - C_{LR}.C_{RL}}.\begin{bmatrix} C_{RR} & -C_{RL} \\ -C_{LR} & C_{LL} \end{bmatrix} \quad (7)$$

L'équation peut être reformulée à partir des fonctions de transfert interaurales (Interaural Transfer Functions) [Moller, 1992]:

$$[H] = \frac{1}{1 - ITF_L.ITF_R}.\begin{bmatrix} \frac{1}{C_{LL}} & 0 \\ 0 & \frac{1}{C_{RR}} \end{bmatrix}.\begin{bmatrix} 1 & -ITF_L \\ ITF_R & 1 \end{bmatrix} \quad (8)$$

Avec $ITF_L = \frac{C_{LR}}{C_{LL}}$ e $ITF_R = \frac{C_{RL}}{C_{RR}}$ les fonctions de transfert interaurales.

ii Minimisation au sens des moindres carrés

La matrice inverse peut aussi être estimée dans le domaine fréquentiel en minimisant un critère au sens des moindres carrés avec régularisation [Ole Kirkeby, Nelson, Hamada, et al., 1998]. La matrice $[H]$ des filtres inverses est donnée par la relation suivante :

$$[H] = ([C]^*.[C] + \beta[Id])^{-1}.[C]^*A \quad (9)$$

Avec β un terme de régularisation et A une réponse cible. Une alternative dans le domaine temporel est proposée dans [O. Kirkeby & Nelson, 1999].

iii Influence de la méthode de résolution

Dans tous les cas, plusieurs paramètres peuvent varier, concernant notamment le choix des filtres directs de la matrice $[C]$: les HRTF peuvent être mesurées sur des têtes artificielles ou réelles [A. H. Moore et al., 2010], en conditions anéchoïques ou conditions d'utilisation, et les HRTF peuvent être modélisées comme des filtres à phase minimale et un retard pur [Larcher, 2001].

Il semblerait que la méthode d'inversion à partir de la minimisation au sens des moindres carrés dans le domaine fréquentiel soit la plus usitée [Parodi, 2010]. Une étude a été menée en comparant trois méthodes [Parodi & Rubak, 2011], à partir d'indicateurs objectifs du rendu. Lorsque la longueur des filtres est courte, il semblerait que les méthodes calculées à partir des moindres carrés en temporel et fréquentiel soient équivalentes, et légèrement meilleures que la troisième méthode qui repose sur l'inversion du déterminant à partir de filtres à phase minimale.

Selon les auteurs, cette différence serait liée à l'approximation de phase minimale qui pourrait altérer le rendu.

I.5.C.c Placement des haut-parleurs

L'équation (5) a permis de présenter le principe général des systèmes transauraux à deux haut-parleurs qui en représentent le cas le plus simple. De nombreuses variantes en ont été proposées. Des travaux précurseurs sur le transaural [Schroeder, 1969] font état d'un positionnement de sources à +/- 22.5° face à l'auditeur. Peu de temps après, d'autres travaux [Damaske, 1971] ont proposé une restitution transaurale avec deux haut-parleurs situés à +/- 36°. L'influence de l'écartement angulaire des haut-parleurs a été étudiée par la suite : Kirkeby et al ont notamment proposé la solution dite du « Stéréo-Dipôle » [Ole Kirkeby & Nelson, 1998; Ole Kirkeby, Nelson, & Hamada, 1998]. Cette configuration consiste à placer deux sources avec un faible écart angulaire (typiquement +/- 5°). Des auteurs [Ole Kirkeby, Nelson, & Hamada, 1998] ont réalisé des simulations en champ libre (deux récepteurs à l'emplacement des oreilles, en l'absence d'auditeur) pour différents écartements angulaires (+/-5°, +/-10° et +/-30°). Le support temporel des filtres nécessaire pour effectuer l'annulation des chemins croisés est alors plus long dans le cas des grands écarts angulaires. Les auteurs montrent également que la zone d'annulation est plus grande dans le cas de petits écartements angulaires. Dans [Ole Kirkeby & Nelson, 1998], les auteurs incluent également des simulations à partir d'un modèle de tête sphérique et obtiennent des résultats similaires au cas en champ libre pour les petits écarts angulaires, avec toujours de meilleurs résultats qu'avec les plus grands écarts angulaires. Les propos sont nuancés pour la restitution de basses fréquences, pour lesquelles il est difficile d'obtenir une bonne annulation avec de petits écarts angulaires.

Par la suite, une étude [Bai & Lee, 2006] s'est également intéressée à l'effet de l'angle pour une restitution transaurale sur deux haut-parleurs. Contrairement aux travaux précédents, dans cette étude il est montré que les configurations aux grands écarts angulaires sont mieux conditionnées numériquement. La taille du sweet spot pour différentes configurations angulaires a été comparée, en définissant un sweet spot absolu correspondant à une zone où le contraste entre les deux oreilles atteint un certain seuil. A partir de cette définition, les écarts angulaires les plus importants (de +/- 60° à +/-75°) permettent d'avoir le sweet spot le plus grand. Enfin, un test perceptif de localisation a été mené sur une sélection de configurations, et les configurations les plus espacées donnent les meilleurs résultats.

Une étude plus récente s'est intéressée à l'évaluation systématique du rendu de systèmes transauraux sur deux et quatre haut-parleurs [Parodi & Rubak, 2011] en utilisant une base de HRTF de tête artificielle. Les paramètres variables sont l'écartement angulaire des haut-parleurs, la méthode de calcul et la longueur des filtres. Les auteurs n'ont pas mis en avant de claires différences en fonction de la position des haut-parleurs pour une restitution idéale. Les mêmes auteurs [Parodi & Rubak, 2010] se sont aussi intéressés à la taille du sweet spot dans différentes configurations. D'après ces travaux, la taille du sweet spot est plus réduite pour les configurations de haut-parleurs espacés, et les configurations avec des haut-parleurs élevés seraient les plus robustes aux déplacements de tête.

I.5.C.d Systèmes transauraux à plus de deux canaux

Il est possible d'approfondir le concept en utilisant plus de deux sources. Par exemple, [Bauck & Cooper, 1996] ont introduit la notion de « transaural généralisé », permettant d'implémenter des systèmes à plusieurs haut-parleurs. Le type d'implémentation proposé permet notamment une diffusion pour plusieurs auditeurs. Le principe a été repris par [Huang et al., 2007], où les auteurs comparent le rendu simulé de systèmes à deux et trois haut-parleurs dans plusieurs environnements acoustiques. D'après ces simulations, l'usage de trois haut-parleurs permet d'améliorer le rendu.

Une configuration à trois haut-parleurs a aussi été proposée par [Yang et al., 2003] afin d'améliorer la robustesse aux déplacements de la tête. Cette configuration a été élaborée à partir de simulations en champ libre, et de l'observation du conditionnement numérique du problème.

Des travaux ont mis en évidence l'effet de la fréquence sur la robustesse du rendu [Ward & Elko, 1999]. Ces travaux proposent un écartement angulaire adapté en fonction de la fréquence à restituer. Par la suite, Takeuchi et collaborateurs ont formalisé le principe de l'Optimal Source Distribution [Takeuchi & Nelson, 2002], qui consiste à réaliser un décodeur transaural par bandes de fréquences. Il y a autant de paires de haut-parleurs que de bandes de fréquences, et l'écartement angulaire associé est plus grand aux fréquences plus basses.

Un système composé de trois paires de stéréo-dipôles a été proposé [Parodi, 2010], où chacun des stéréo-dipôles décode les sources virtuelles dans son voisinage. Cela permet notamment de réduire les confusions avant/arrière par rapport à un système traditionnel, mais l'inconvénient de ce système est qu'il n'est compatible qu'avec la synthèse binaurale, et non avec les enregistrements binauraux.

Une proposition d'utilisation d'un réseau de haut-parleurs a été faite dans [Galvez & Fazi, 2015]. Comparativement à un système OSD, les auteurs montrent que l'énergie rayonnée est moindre au-dessus de 500 Hz, et permet d'obtenir un meilleur contraste lors de la simulation d'une salle d'écoute. Enfin, des configurations à plusieurs haut-parleurs ont été proposées pour assurer une restitution avec suivi des mouvements de tête de l'auditeur [Lentz, 2006] [Galvez et al., 2016].

I.6 Bilan

La position optimale de sources en reproduction transaurale ne fait pas encore l'objet d'un consensus clair, et l'influence de la distance n'a pas été abordée à notre connaissance. Les travaux de cette thèse visent à déterminer la configuration optimale selon plusieurs critères. La démarche est donc assez similaire aux travaux de Yesenia Lacouture Parodi [Parodi, 2010]. Toutefois, dans notre approche l'influence de la salle d'écoute est prise en compte et le rapprochement des sources à proximité de l'auditeur est le principal axe d'étude.

L'objet de cette thèse est d'optimiser un dispositif pour la reproduction sonore transaurale en préliminaire à une optimisation plus globale de « la » reproduction sonore par le dispositif proposé. Nous sommes en effet contraints *in fine* par l'usage de signaux binauraux existants, dont l'information spatiale est codée pour une morphologie particulière.

Chapitre II Influence de l'auditeur

Table des matières

II.1 Introduction

La diffusion transaurale consiste à reproduire une pression cible au niveau des oreilles de l'auditeur : pour son implémentation, il faut donc compenser la diffraction par l'auditeur. Les Head-Related Transfer Functions (HRTF) sont les fonctions de transfert caractérisant cet effet, leur étude est donc fondamentale dans l'optique d'optimiser le système de diffusion. Les HRTF varient avec la position de la source (azimut, élévation, distance) et avec la morphologie de l'auditeur.

La variabilité des HRTF concerne donc l'auditeur lui-même : idéalement, le système de reproduction doit être calibré pour lui, mais la caractérisation des HRTF pour chaque auditeur est délicate à mettre en œuvre. Des solutions existent toutefois pour approximer les HRTF. Notamment, des constructeurs proposent des têtes artificielles approximant une morphologie humaine moyenne. Ce type de dispositif est largement employé dans l'industrie et la recherche, car il permet notamment une acquisition reproductible de données au format binaural. Ces

mannequins sont tous censés approximer une morphologie moyenne, les mesures obtenues avec ces différents mannequins devraient donc être similaires. Les différences observables entre mannequins devraient pouvoir être considérées comme un minorant des différences entre individus.

La mesure de HRTF est une opération délicate, dont la durée dépend du nombre de positions de source à caractériser. Une robotisation du dispositif permet de réduire cette durée, mais n'est pas toujours possible. Un moyen simple de mesurer plusieurs HRTF d'azimuts différents consiste à placer le sujet sur une table tournante, et la mesure d'un nombre élevé d'azimuts n'est alors généralement pas problématique. La mesure de nombreuses élévations est plus difficile : une double rotation du mannequin nécessite un équipement spécifique, susceptible de perturber les mesures. Utiliser un réseau de haut-parleurs accroît leur diffraction du champ sonore mais l'utilisation d'une seule source sonore nécessite de la déplacer pour chaque élévation. Motoriser ce déplacement accroît aussi la diffraction par la structure, donc le déplacement manuel est finalement la solution la plus simple. Cette opération rend par contre le temps de mesure d'autant plus long que le nombre d'élévations à caractériser est important. Pour la même raison, les mesures à plusieurs distances sont longues : il faut systématiquement replacer la source avec précision pour chaque paire {distance, élévation}. Pour éviter de multiplier les mesures, une approche consiste à propager des données à partir de mesures de HRTF à une distance plus faible. La modélisation des HRTF à partir des harmoniques sphériques est proposée en première partie de ce chapitre permettant la propagation des données.

Les élévations très négatives sont difficilement caractérisables en pratique. En effet, la source doit alors être placée en dessous du support du mannequin et la mesure risque alors d'être fortement biaisée par son support. Une possibilité consiste à placer le mannequin à l'envers, comme il en est fait référence dans [Aussal et al., 2013] et [Parseihian, 2012]. Néanmoins, les mesures à élévations très négatives sur un mannequin ne correspondent pas à une situation réaliste : la partie inférieure d'un mannequin n'est pas représentative d'un corps humain.

Une manière alternative d'approximer les HRTF de l'auditeur consiste à estimer la diffraction par calcul numérique. Une méthode de type BEM peut être employée [Kreuzer et al., 2009], mais son principal inconvénient est son coût de calcul très important. Une autre solution est basée sur le calcul analytique de la pression diffractée par un objet simple. Une solution est connue pour un modèle de sphère, qui peut approximer raisonnablement la tête d'un auditeur [Busson, 2006]. L'avantage de ce modèle est qu'il permet de calculer des HRTF pour toute position de l'espace moyennant un coût de calcul très faible. Le principe du calcul analytique de diffraction par une sphère est présenté dans l'Annexe B. L'inconvénient de ce modèle est qu'il ne prend pas en compte les détails morphologiques fins, contribuant à la perception de l'espace. Dans l'optique d'une démarche d'optimisation préliminaire d'un système de diffusion, ce modèle peut néanmoins être suffisant : sa pertinence est vérifiée dans ce chapitre.

Des mesures de HRTF sur les mannequins Gras Kemar, Cortex MK2, Head HRS II.2, Bruël&Kjaer 4100-D et une sphère ont ainsi été réalisées aux distances de 40 cm et de 2 m. Ces mesures permettent d'étudier l'effet de la distance pour ces morphologies particulières, et de comparer

ces différents mannequins caractérisés par un dispositif commun. Ces mesures sont également utilisées pour évaluer les performances de l'algorithme de propagation proposé.

II.2 Modélisation des HRTF

Du fait de la difficulté à obtenir de manière fiable certains points de mesure, l'objet de cette partie est de déterminer des points non mesurés à partir d'une sélection de mesures, et de déduire les HRTF à une plus grande distance. Des travaux sur le sujet ont été réalisés par [Nguyen et al., 2010], où plusieurs méthodes de propagation ont été comparées à partir de 865 mesures sur un mannequin HEAD. Quatre méthodes de propagation sont exploitées :

- La première méthode principalement exploitée dans l'article repose sur une décomposition en harmoniques sphériques des HRTF et sera détaillée dans la suite du chapitre.
- La seconde consiste à compenser simplement les différentes mesures par un gain associé aux différences de distances (gain en 1/r avec r la distance)
- La troisième consiste à appliquer aux données un filtre de compensation de distance, calculé à partir d'un modèle de sphère.
- Enfin la dernière est une variante de la méthode précédente, considérant indépendamment les deux oreilles.

Pour l'estimation des données en champ lointain à partir de données à faible distance, les 3 premières méthodes donnent des résultats assez similaires en termes d'écarts objectifs, légèrement meilleurs avec la décomposition en harmoniques sphériques. C'est la méthode qui est étudiée dans la suite.

II.2.A. Modélisation à partir d'une décomposition en harmoniques sphériques

II.2.A.a Théorie

Dans la suite, $k = \frac{2\pi f}{c}$ est le nombre d'onde où c est la célérité du son et f la fréquence. Le principe de la méthode consiste à écrire l'équation d'ondes en coordonnées sphériques. Un champ de pression quelconque peut alors être exprimé via une combinaison linéaire de sess solutions à variables séparées (les harmoniques sphériques). La pression P au point de coordonnées (r, θ, ϕ) s'écrit ainsi de la manière suivante :

$$P(r, \theta, \phi, k) = \sum_{n=0}^{+\infty} \sum_{m=-n}^{n} a_{nm}(r, k) Y_n^m(\theta, \phi) \qquad (10)$$

$Y_n^m(\theta, \phi)$ est l'harmonique sphérique d'ordre n et de degré m définie de la manière suivante :

$$Y_n^m(\theta, \phi) = (-1)^m \sqrt{\frac{2n+1}{4\pi} \frac{(n-|m|)!}{(n+|m|)!}} P_n^{|m|}(cos\theta) e^{im\phi} \qquad (11)$$

Les termes $a_{nm}(r, k)$ correspondent à des ondes radiales décrites de la manière suivante :

$$a_{nm}(r, k) = b_{nm}(k) h_n(kr) \qquad (12)$$

où $P_n^{|m|}$ sont les polynômes de Legendre et h_n les fonctions de Hankel sphériques de première espèce. Ainsi, si $b_{nm}(k)$ est connu pour une distance quelconque, il est théoriquement possible de déduire $a_{nm}(r,k)$ puis $P(r,\theta,\phi,k)$ pour toute autre valeur de r.

Le principe de réciprocité permet de permuter les sources et les récepteurs, et appliqué au cas particulier de la diffraction par la tête d'un auditeur les HRTF peuvent être décomposées de la même manière. En pratique, l'équation (10) est tronquée à un ordre fini $n \leq N_{max}$ et peut s'écrire sous la forme matricielle suivante :

$$\boldsymbol{HRTF}(r,k) = [Y]\boldsymbol{a}(r,k) \tag{13}$$

Avec $\boldsymbol{HRTF}(r,k)$ un vecteur de N_{mes} rapports de pressions, $[Y]$ la matrice des harmoniques sphériques de taille N_{mes} x N_{harm} et $\boldsymbol{a}(r,k)$ le vecteur des N_{harm} coefficients associés. $N_{harm} = (N_{max}+1)^2$, où N_{max} est l'ordre de décomposition maximal des harmoniques sphériques.

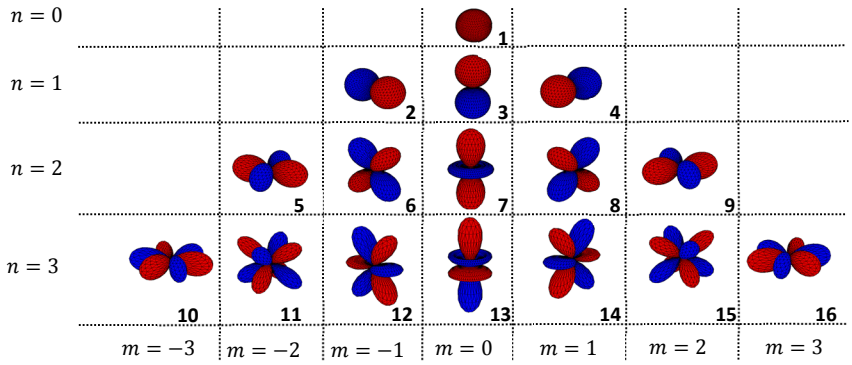


Figure 7 : représentation de la combinaison linéaire des parties réelle et imaginaire des harmoniques sphériques jusqu'à l'ordre 3. La couleur rouge correspond à des valeurs positives et la couleur bleue à des valeurs négatives. Les numéros en bas à droite de chaque harmonique correspondent à la convention d'indexation adoptée.

Des exemples d'harmoniques sphériques sont représentés sur la Figure 7. La convention d'indexation suivante est adoptée : l'harmonique Y_n^m est désignée par Y_i, avec $i = (n+1)^2 - n + m$. Elle se fait donc par ordre puis degré croissants comme l'illustre la Figure 7.

Dans l'équation (13), seuls les termes correspondant à des positions particulières de mesure sont présents. La matrice d'harmoniques sphériques correspond ainsi aux termes :

$$[Y] = \begin{bmatrix} Y_1(\theta_1,\phi_1) & Y_2(\theta_1,\phi_1) & \dots & Y_{N_{harm}}(\theta_1,\phi_1) \\ Y_1(\theta_2,\phi_2) & & & \vdots \\ \vdots & & & \vdots \\ Y_1(\theta_{N_{mes}},\phi_{N_{mes}}) & \cdots & \cdots & Y_{N_{harm}}(\theta_{N_{mes}},\phi_{N_{mes}}) \end{bmatrix} \tag{14}$$

La première étape de la propagation d'un jeu de mesures consiste à identifier les coefficients $a(r,k)$ associés aux harmoniques sphériques estimés à la distance de mesure r :

$$\boldsymbol{a}(r,k) = [Y]^{-1}\,\boldsymbol{HRTF}(r,k) \tag{15}$$

Conformément à l'équation (13), l'HRTF à la distance R peut aussi s'écrire de la manière suivante :

$$\boldsymbol{HRTF}(R,k) = [Y]\boldsymbol{a}(R,k) \tag{16}$$

Soit à partir de la relation (12) :

$$\boldsymbol{HRTF}(R,k) = [Y]\boldsymbol{a}(r,k)\frac{h_n(kR)}{h_n(kr)} \tag{17}$$

La matrice d'harmoniques sphériques et les fonctions de Hankel se déterminent analytiquement. La Figure 8 représente le logarithme du module des fonctions de Hankel ($20\log_{10}(|h_n(kr)|)$) pour une sélection d'ordres n et de distances r.

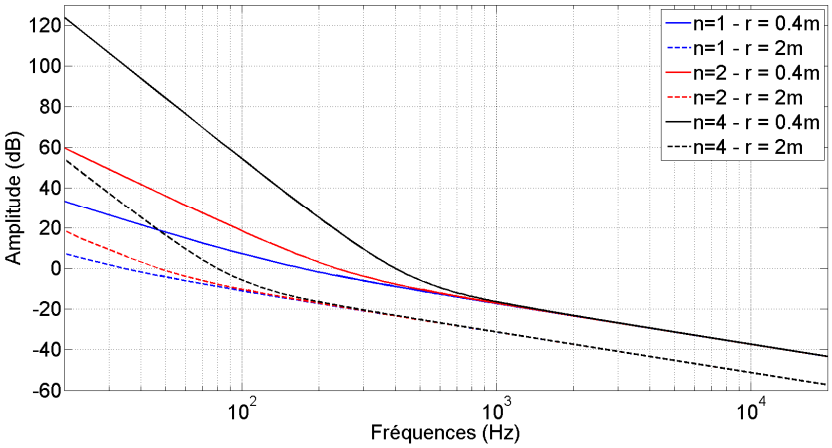


Figure 8 : module des fonctions de Hankel pour les ordres $n = 1$, 2 et 4 aux distances r = 0.4 m et r = 2 m

Aux basses fréquences, le module des fonctions de Hankel diminue lorsque la distance augmente : à 100 Hz par exemple le module vaut 55 dB à 40 cm pour l'ordre 4, alors qu'il vaut seulement -5 dB à 2 m. De ce fait, le rapport des fonctions de Hankel $\frac{h_n(kR)}{h_n(kr)}$ est toujours inférieur à 1 lorsque $R > r$, atténuant ainsi les coefficients des harmoniques sphériques. L'erreur liée à l'estimation des coefficients $\boldsymbol{a}(r,k)$ n'est donc pas amplifiée lors de leur propagation à une distance supérieure ou égale à celle des points de mesure.

En pratique les positions des points de mesure sont telles que la matrice $[Y]$ n'est pas toujours inversible. Un aspect délicat de la démarche consiste donc à estimer l'approximation de la matrice inverse, désignée $[Y]^{\dagger}$. Cette estimation est plus ou moins correcte selon le conditionnement du problème, qui dépend de la répartition des points de mesure et de l'ordre de troncature des harmoniques sphériques. Nous évoquons donc brièvement ci-après différentes méthodes permettant d'estimer $[Y]^{\dagger}$.

Comme cela a été évoqué en introduction de ce chapitre, la répartition des points de mesure s'écarte d'une distribution uniforme : il est plus simple de mesurer de nombreux azimuts et peu d'élévations pour des raisons de mise en œuvre expérimentale. De plus, les points de mesure ne maillent pas toute la sphère puisque les élévations très négatives sont difficiles à caractériser. Cette discrétisation fait qu'il n'est pas possible de définir un produit scalaire tel que les harmoniques sphériques soient une base des champs observables aux points de mesure. Le calcul des coefficients $b_{nm}(k)$ ne peut donc pas être effectué par projection.

L'approche proposée consiste à conserver une décomposition basée sur le formalisme des harmoniques sphériques, mais à en déterminer les coefficients comme solution d'un problème inverse. Dans cette approche, il est possible de sélectionner des termes si cela facilite l'inversion. L'approche que nous proposons consiste à utiliser des positions de mesure imposées, plutôt que de chercher à optimiser la répartition de ces positions comme cela a pu être fait dans [Bahu, 2013]. Plusieurs méthodes d'inversion adaptées au problème sont présentées dans les prochains paragraphes.

II.2.A.b Inversion par régularisation de Tikhonov

La méthode par régularisation de Tikhonov consiste à déterminer $[Y]^{\dagger}$ comme la solution minimisant $\|Ya - HRTF\|^2 + \|\beta a\|^2$, avec β un coefficient de régularisation. Ce type de résolution est employé dans plusieurs travaux concernant la modélisation d'HRTF [Duraiswami et al., 2004], [Nguyen et al., 2010], [Pollow et al., 2012], [Aussal et al., 2013] et la solution est la suivante :

$$[Y]^{\dagger} = ([Y]^{*}[W][Y] + \beta[D])^{-1}[W][Y]^{*} \tag{18}$$

$[W]$ est une matrice diagonale de taille N_{mes}xN_{mes} dont les termes diagonaux sont des poids associés aux positions de mesures, visant à se rapprocher d'une base orthonormée. Ces poids correspondent à la surface de l'angle solide, estimés à partir des surfaces de Voronoi. Les valeurs de ces poids sont reportées dans le Tableau 15 de l'Annexe C.

β est un coefficient de régularisation et dans notre cas il est fixé à partir d'une dynamique Dyn (exprimée en dB) par rapport à la norme de Frobenius N_{Fro} de la matrice $[Y]$:

$$\beta = \frac{N_{fro}^2}{10^{\frac{Dyn}{20}}} \tag{19}$$

$[D]$ est une matrice diagonale de taille N_{harm}xN_{harm} dont les éléments constituent des poids attribués aux termes de la décomposition. Ici, le $i^{ème}$ élément de la matrice $[D]$ est donné par la relation suivante avec n l'ordre correspondant à cet indice :

$$D_{ii} = 1 + n(n+1) \tag{20}$$

$[D]$ est ensuite normalisée par la somme de ses éléments diagonaux. Avec cette loi de pondération, le coefficient de régularisation associé aux ordres élevés d'harmoniques est alors plus élevé. Cette loi de pondération utilisée dans l'article [Nguyen et al., 2010] est la même que celle proposée dans une publication antérieure [Duraiswami et al., 2004] sans justification explicite. L'approche par inversion régularisée vise à se rapprocher autant que possible du résultat que donnerait une projection orthonormée, mais l'inconvénient de cette méthode est que le terme de régularisation introduit un biais systématique dans la solution.

II.2.A.c Inversion par SVD tronquée

Une autre approche d'inversion utilise la décomposition en valeurs singulières de la matrice $[Y]$:

$$\begin{cases} si & [Y] = [U][\Sigma][V]^* \quad , \quad [\Sigma] = Diag[\sigma_i] \\ alors & [Y]^\dagger = [V][\Sigma]^\dagger[U]^* \end{cases} \tag{21}$$

$$[\Sigma]^\dagger = Diag[\sigma_i^\dagger] \quad , \qquad avec \qquad \begin{matrix} \sigma_i^\dagger = \dfrac{1}{\sigma_i} & si\ i \leq N_{maxi} \\ \sigma_i^\dagger = 0 & sinon \end{matrix} \tag{22}$$

$[\Sigma]$ est la matrice diagonale des valeurs singulières de taille N_{harm}xN_{mes}, et $[U]$ et $[V]$ sont des matrices unitaires de dimensions N_{harm}xN_{harm} et N_{mes}xN_{mes} respectivement. Les valeurs singulières de la matrice $[\Sigma]$ désignées σ_i sont usuellement rangées par ordre décroissant, et la matrice $[\Sigma]^\dagger$ est la matrice dont les éléments diagonaux $\sigma_i^\dagger$ sont l'inverse des éléments diagonaux de Σ. Pour les faibles valeurs de σ_i, $\frac{1}{\sigma_i}$ serait très grand, impliquant une amplification excessive de composantes peu significatives, $\sigma_i^\dagger$ est alors fixé à 0. Ceci revient à tronquer la série des vecteurs propres de $[Y]$. N_{maxi} est l'indice maximal respectant la condition suivante :

$$20\log_{10}\left(\frac{\sigma_1}{\sigma_{N_{maxi}}}\right) \leq Dyn \tag{23}$$

Avec Dyn un paramètre à fixer. En d'autres termes, la troncature de la matrice $[\Sigma]$ limite à une valeur *a priori* le conditionnement $cond$ défini par :

$$cond = \frac{\sigma_1}{\sigma_{N_{harm}}} \tag{24}$$

Cette troncature permet d'augmenter la stabilité numérique du problème, en supposant que les faibles valeurs de σ_i correspondent à du bruit ou des composantes marginales de la matrice $[Y]$.

Cette suppression d'information introduit donc aussi un biais dans l'estimation, et n'est pas toujours exploitable aveuglément. Pour réaliser une sélection des termes de manière plus pertinente, une autre approche est proposée à partir d'un préconditionnement de la matrice $[Y]$.

II.2.A.d Inversion par SVD avec préconditionnement

La méthode suivante reprend le principe de la SVD tronquée, mais l'inversion n'est effectuée qu'après une sélection des termes adaptés à la résolution du problème, selon des considérations physiques. La répartition des points de mesure est telle que certaines composantes d'harmoniques sphériques sont très proches après discrétisation aux positions de mesure. Par exemple, la résolution en élévation est faible, seuls quelques termes sont donc probablement nécessaires pour coder l'élévation. Par contre l'utilisation de termes dont la trace aux positions de mesure est proche conduit à un mauvais conditionnement du problème. Une phase préliminaire de tri des termes de la décomposition est alors réalisée pour tenter d'éliminer les termes trop redondants.

Pour estimer la redondance entre les deux colonnes $\boldsymbol{Y_i}$ et $\boldsymbol{Y_j}$ de la matrice $[Y]$, un terme P_{ij} est calculé, analogue au produit scalaire discrétisé sur la sphère :

$$P_{ij} = \sum_{n=1}^{N_{mes}} Y_i(n) Y_j^*(n) \tag{25}$$

A partir de P_{ij}, une cohérence $Coh(i,j)$ est alors calculée :

$$Coh(i,j) = \left| \frac{P_{ij}}{\sqrt{P_{ii} . P_{jj}}} \right| \tag{26}$$

De cette manière, la cohérence d'une colonne avec elle-même vaut 1. La cohérence des termes de la matrice $[Y]$ pour un ordre de décomposition maximal de 4 correspondant à 25 harmoniques sphériques est représentée sous forme matricielle dans la Figure 9 pour le cas idéal et pour la discrétisation sur 433 points de mesure. Dans le cas idéal, seuls les termes de la diagonale sont unitaires, la cohérence valant 0 ailleurs. Avec la discrétisation sur 433 points de mesure, des motifs se répartissent en diagonales correspondant à des termes de degré identique et d'ordre consécutif. Par exemple, les harmoniques d'indices 2,3, 4 correspondent à l'ordre 1 et aux degrés -1,0,1 et les harmoniques d'indices 6,7,8 à l'ordre 2 et degrés -1,0,1. Les harmoniques de degrés identiques et d'ordres consécutifs ont des similitudes : pour un degré donné, les lobes se répartissent dans les mêmes directions. Les harmoniques à l'ordre supérieur comportent cependant plus de lobes, codant l'information avec plus de résolution spatiale.

Par exemple, les paires [2-6], [3-7] et [4-8] forment une succession de combinaisons pour lesquelles la redondance n'est pas nulle, particulièrement pour la paire [3-7] où elle vaut 0.5.

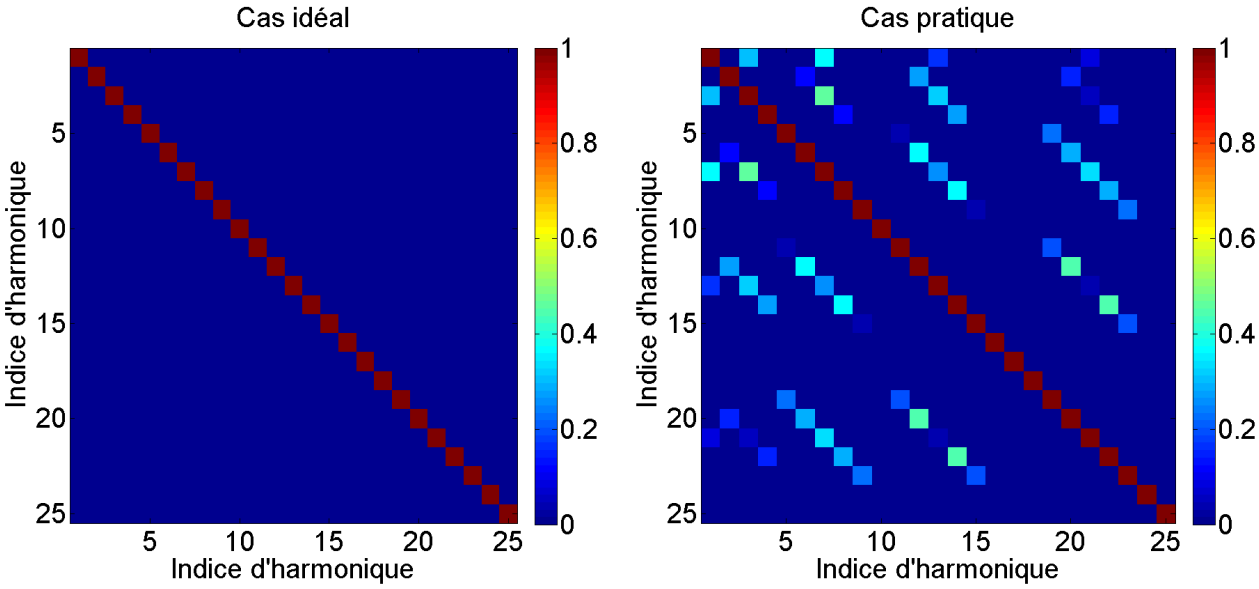


Figure 9 : matrice de cohérence des 25 premières harmoniques sphériques (ordre 4) dans le cas idéal (à gauche) et discrétisées pour 433 points de mesure (à droite).

Les harmoniques 3 et 7 codent principalement l'élévation : elles ont deux lobes principaux orientés vers les pôles. Il n'est pas surprenant de constater que ce sont des termes codant principalement l'élévation qui sont détectés comme redondants, puisque la répartition des points en élévation est plus faible. Notamment, l'absence de points à élévation très négative crée un

déséquilibre car l'information apportée par ces harmoniques est similaire sur les positions de mesures envisagées, leurs différences n'étant discernables qu'à des positions non mesurées.

Les positions de mesures accessibles expérimentalement ne permettant pas de séparer certains termes, une étape de tri préalable est alors proposée. Lorsque la cohérence entre deux termes dépasse un seuil pré-déterminé $Seuil_{scal}$, le terme d'indice le plus élevé (ordre et degré plus élevé) est éliminé. La taille de la matrice est d'autant plus réduite que $Seuil_{scal}$ est faible. Le conditionnement de la matrice $[\mathrm{Y}]$ ainsi modifié est calculé pour différentes valeurs de seuil et reporté dans la Figure 10 pour une décomposition à l'ordre 10. Une zone de transition brusque est visible autour du seuil 0.43 : jusqu'à 0.4 le conditionnement est inférieur à 20 dB, et au-delà de 0.45 il dépasse 300 dB, ce qui ne permet plus l'inversion. Dans ce cas, il est pertinent de choisir un seuil à 0.43. Avec un tel seuil, seul un petit nombre de paires d'harmoniques est détecté. Par exemple pour les indices [3-7] identifiés précédemment l'harmonique d'indice 7 est supprimée car redondante en termes d'observations aux points de mesure.

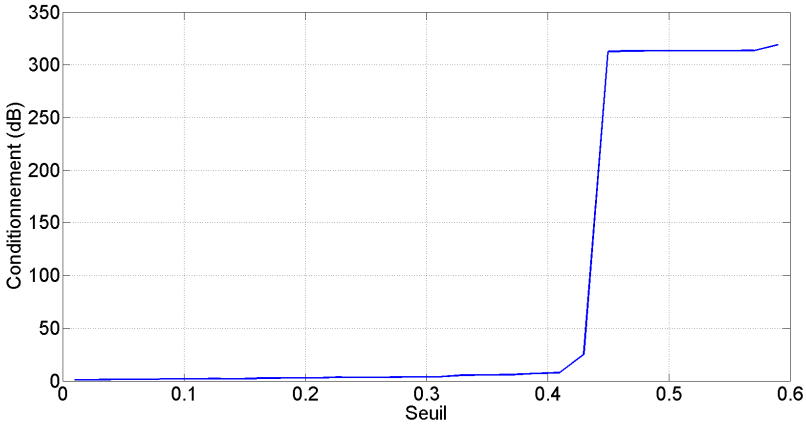


Figure 10 : conditionnement numérique en fonction du seuil d'élimination des termes pour une décomposition à l'ordre 10

La méthode de troncature proposée consiste ainsi à éliminer un petit nombre de termes, quitte à ne pas conserver une série complète pour chaque ordre. Il est possible de prolonger la démarche en appliquant une troncature plus fine : le seuil de troncature est alors déterminé spécifiquement pour chaque ordre d'harmonique sphérique. Pour cela, le seuil est déterminé de manière itérative, de telle sorte que chaque incrément de l'ordre ne fasse pas augmenter significativement le conditionnement de la matrice $[Y]$. Ceci garantit que les termes conservés jusqu'à l'ordre n sont identiques si l'ordre maximal est supérieur ou égal à n. Pour illustrer cette démarche, le seuil obtenu en fonction de l'ordre de décomposition pour obtenir un conditionnement de 40 dB est reporté dans la Figure 11.

La limite de cette approche est qu'elle ne permet pas de considérer *a priori* la redondance éventuelle entre combinaisons d'harmoniques sphériques. Le fait d'aboutir à une valeur de conditionnement maîtrisée permet néanmoins de tronquer la SVD sans introduire un biais significatif.

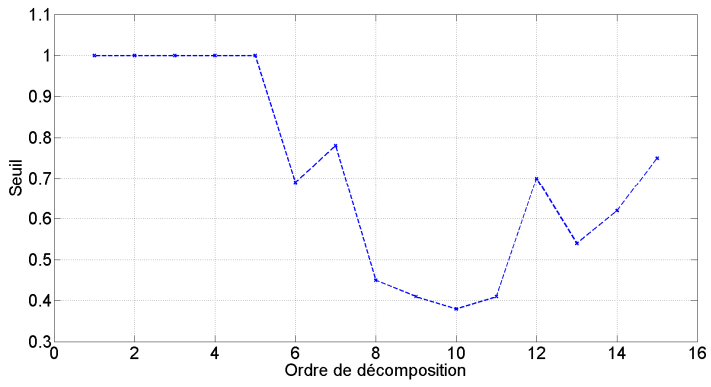


Figure 11 : Seuil de préconditionnement en fonction de l'ordre de décomposition pour que le conditionnement soit toujours inférieur à 40 dB

II.2.B. Validation de la propagation sur une sphère

Dans un premier temps, pour évaluer la faisabilité de la propagation, des données calculées analytiquement sur un modèle de sphère à la distance de 40 cm sont utilisées. Une répartition de points de mesure correspondant aux mesures effectuées sur les mannequins est employée, à partir de 72 azimuts espacés de 5° aux élévations -30°, -15°, 0°, 15°, 30°, 60° ainsi que la mesure à élévation 90°, soit un total de 433 points.

Pour calculer l'approximation de la matrice inverse $[Y]^{\dagger}$, deux méthodes sont exploitées : celle basée sur la régularisation de Tikhonov, et celle basée sur la SVD tronquée avec préconditionnement. Pour les deux méthodes, la dynamique de l'inversion Dyn est fixée à 40 dB. Dans un premier temps, l'influence de l'ordre de décomposition est étudiée, les algorithmes sont ensuite évalués à partir de données bruitées et enfin l'influence du nombre de points de mesure est abordée.

II.2.B.a Ordre maximal

Pour évaluer la performance de l'inversion, l'approximation de la matrice inverse $[Y]^{\dagger}$ est utilisée pour calculer N_{test} points répartis de trois manières différentes :

Cas A A 40 cm pour les azimuts entre les points du plan horizontal ($\theta = 2.5°, 7.5°, \dots, 357.5°$) $N_{test} = 72$ dans ce cas. Ce cas revient à interpoler entre les points qui ont servi à estimer les coefficients de décomposition.

Cas B A 40 cm pour $N_{test} = 677$ points répartis presque uniformément sur un anneau d'élévation comprise entre +30° et -30° à 40 cm de distance.

Cas C A 2 m dans le plan horizontal, pour les mêmes incidences que les points d'estimation ($N_{test} = 72$).

Les résultats sont comparés aux données calculées analytiquement à partir d'un calcul d'écarts spectraux moyens :

$$ES(f) = \frac{1}{N}\sum_{i=1}^{N_{test}} \left| 20\log_{10}\left(\frac{|HRTF_{calc}(i,f)|}{|HRTF_{extrap}(i,f)|}\right)\right| \qquad (27)$$

Pour une incidence d'indice i, $HRTF_{calc}$ est la HRTF calculée analytiquement, et $HRTF_{extrap}$ la HRTF estimée à partir de la décomposition en harmoniques sphériques. Cette métrique par ailleurs employée dans [Zhong et al., 2016] est limitée puisqu'elle ne prend pas en compte les erreurs liées à la phase des HRTF. Les écarts spectraux moyennés sur les différentes positions sont reportés dans la Figure 12, jusqu'à l'ordre 24. Une ligne blanche et une ligne noire sont également reportées sur cette figure, correspondant à la fréquence maximale pour laquelle l'écart est inférieur à 0.3 dB et 1 dB respectivement.

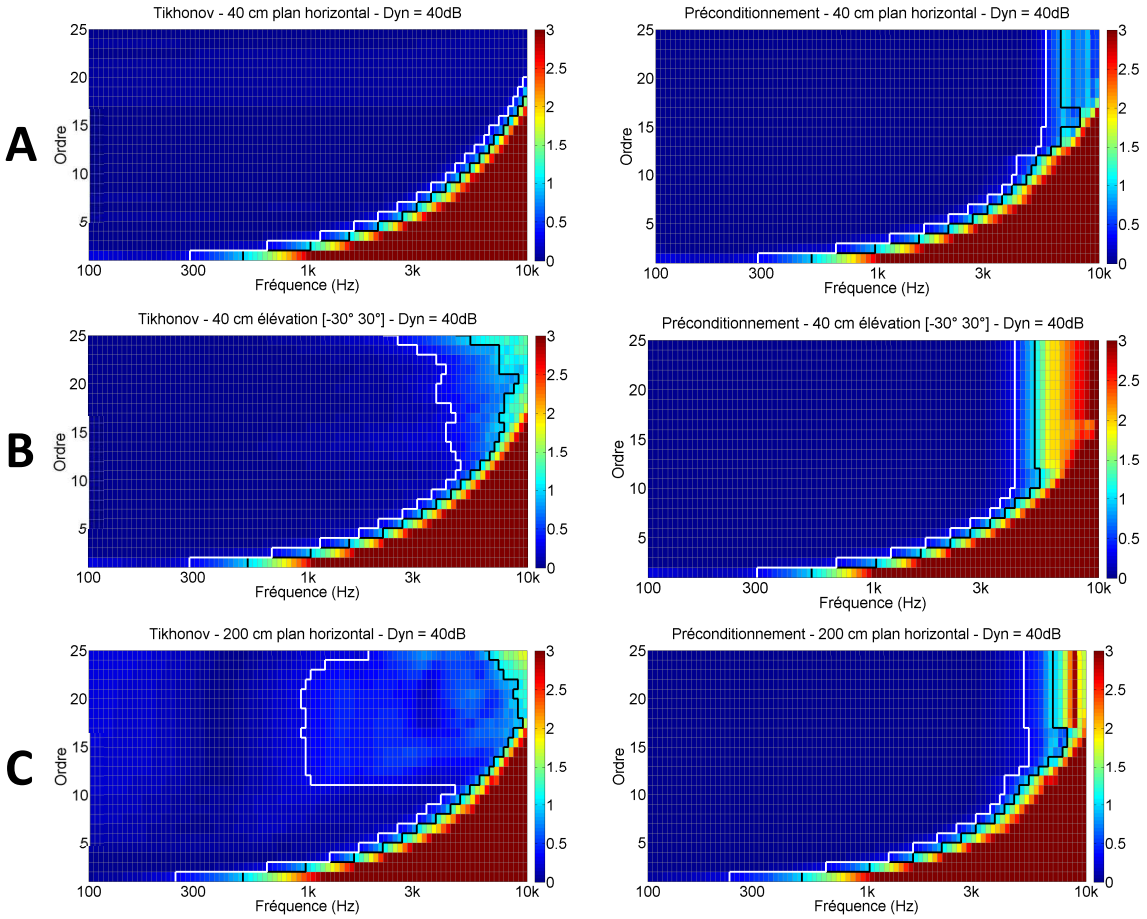


Figure 12 : Ecarts spectraux moyens pour deux méthodes d'inversions (Tikhonov à gauche et Préconditionnement à droite) et trois zones de reconstruction (figures du haut : interpolation à 40 cm entre les points de mesure dans le plan horizontal, figures milieu : interpolation à 40 cm entre les élévations +/-30° et figures du bas : propagation à 2 m dans le plan horizontal). Le seuil à 0.3 dB est représenté en blanc et le seuil à 1 dB est représenté en noir. Dyn = 40 dB

Un ordre faible suffit à reconstruire les basses fréquences, à l'ordre 1 les écarts spectraux sont inférieurs à 1 dB jusqu'à 500 Hz pour la reconstruction à 40 cm pour les deux méthodes. L'ordre de décomposition d'harmoniques sphériques doit nécessairement être élevé pour pouvoir appliquer la méthode aux hautes fréquences, les écarts spectraux moyens sont systématiquement supérieurs à 3 dB en hautes fréquences et à ordre faible. Une relation entre l'ordre minimal de décomposition N_{min} et la fréquence peut être définie [Aussal et al., 2013] :

$$N_{min} \geq \left\lfloor \frac{2\pi f}{c} r_s \right\rfloor + 1 \qquad (28)$$

Où ⌊ ⌋ désigne la partie entière. Cette loi se vérifie avec les résultats obtenus ici : par exemple à 3 kHz l'ordre minimal N_{mini} obtenu est de 5, et correspond approximativement au seuil à 1 dB pour les deux méthodes.

Jusqu'à l'ordre 10, il n'y a pas de différence significative entre les deux méthodes : les limites à 0.3 dB et 1 dB sont atteintes pour des fréquences comparables, à 4.5 kHz et 5.5 kHz respectivement à l'ordre 10. Néanmoins, une légère différence est visible dans le cas C (propagation à 2 m) : pour les fréquences inférieures au seuil de 0.3 dB, les écarts sont de 0 dB avec la méthode de préconditionnement, alors qu'ils ne sont pas nuls avec la méthode par régularisation de Tikhonov. Ce léger écart est lié au terme de régularisation qui introduit un biais systématique.

Pour les ordres supérieurs, les performances de la propagation dépendent de la méthode utilisée. Pour la méthode avec régularisation de Tikhonov, les seuils à 0.3 dB et 1 dB augmentent toujours lorsque l'ordre de décomposition augmente dans le cas A. En revanche, au-delà d'un certain ordre les seuils diminuent dans les autres cas, traduisant une moins bonne propagation aux ordres élevés. En particulier pour le cas C, le seuil à 0.3 dB est atteint à 4.7 kHz à l'ordre 10 et à 1 kHz à l'ordre 11. Cette augmentation de l'erreur en hautes fréquences aux ordres élevés pourrait être liée à un phénomène de sous-échantillonnage spatial. Lorsque les points de mesure sont trop espacés par rapport à la longueur d'onde, une ambiguïté apparait se traduisant par un repliement aux ordres inférieurs. Le choix de l'ordre est donc primordial, un ordre trop élevé pouvant conduire à une augmentation des erreurs.

Pour la méthode avec préconditionnement, les performances sont plus stables pour les trois cas, et ne dépendent pas de l'ordre de décomposition pour les ordres élevés. Le seuil à 0.3 dB est atteint à 5.7 kHz à partir de l'ordre 15 pour le cas A, et à 5.5 kHz à partir de l'ordre 13 pour le cas C. Toutefois, les écarts ont tendance à augmenter rapidement avec la fréquence au-delà de ce seuil. Cet effet est particulièrement visible pour le cas B : le seuil à 0.3 dB est atteint à 4.2 kHz, et au-delà de 6 kHz les écarts sont supérieurs à 2 dB quel que soit l'ordre supérieur à 13. Pour les ordres élevés, la méthode avec préconditionnemment a donc un comportement assez binaire : en dessous d'une fréquence limite, l'erreur est non-significative et au-delà l'erreur augmente rapidement avec la fréquence.

Pour les deux méthodes, la modélisation est la moins précise pour calculer des points à élévation non nulle. Cette limite est liée à la répartition des points de mesure qui n'est pas très dense en élévation. Par ailleurs, l'étape de propagation n'introduit pas d'artefact significatif supplémentaire : l'essentiel du problème consiste à estimer la matrice inverse $[Y]^{\dagger}$, et l'erreur de reconstruction ne semble pas dépendante du terme propagatif et donc de la distance de reconstruction. Plutôt que d'effectuer des mesures à plusieurs distances, il serait donc avantageux de caractériser des élévations en nombre suffisant à faible distance, et déduire les données aux autres distances par propagation.

La méthode avec pré-conditionnement a un comportement plus stable pour les ordres élevés, qui dépend moins de la zone de reconstruction qu'avec la méthode de Tikhonov. Cette stabilité est

préférable pour estimer des données inconnues, pour lesquelles on ne peut pas vérifier quel est l'ordre optimal. Elle semble donc *a priori* préférable à ce stade. En pratique, les données mesurées sont cependant inévitablement bruitées et la propagation peut y être sensible. L'influence du bruit de mesure est donc évaluée dans la prochaine section.

II.2.B.b Influence du bruit de mesure

Dans ce paragraphe, les calculs précédents ont été repris en ajoutant du bruit pour se rapprocher des conditions réelles de mesure. Pour cela, deux types de bruit sont introduits sur les données : le premier est un bruit spatial uniforme entre +/- 1 cm perturbant la position des sources servant à calculer analytiquement les HRTF. Le second est un bruit additif uniforme entre +/- 0.3 dB perturbant les amplitudes des HRTF calculées. En présence de bruit, le paramètre de la dynamique d'inversion Dyn doit être choisi judicieusement. Les écarts spectraux moyens associés à deux paramétrages ($Dyn = 40\ dB$ et $Dyn = 15\ dB$) sont représentés dans la Figure 13 pour la propagation à 2 m dans le plan horizontal (cas C).

L'effet du paramètre Dyn est marqué pour la méthode avec préconditionnement pour les ordres supérieurs à 5 : le seuil à 1 dB est à 200 Hz pour une dynamique de 40 dB, alors qu'il est de l'ordre de 4 kHz avec une dynamique de 15 dB. Pour cette méthode, le paramètre Dyn correspond au conditionnement du problème : une valeur de 40 dB correspond donc à un cas mal conditionné et sensible aux erreurs de mesure.

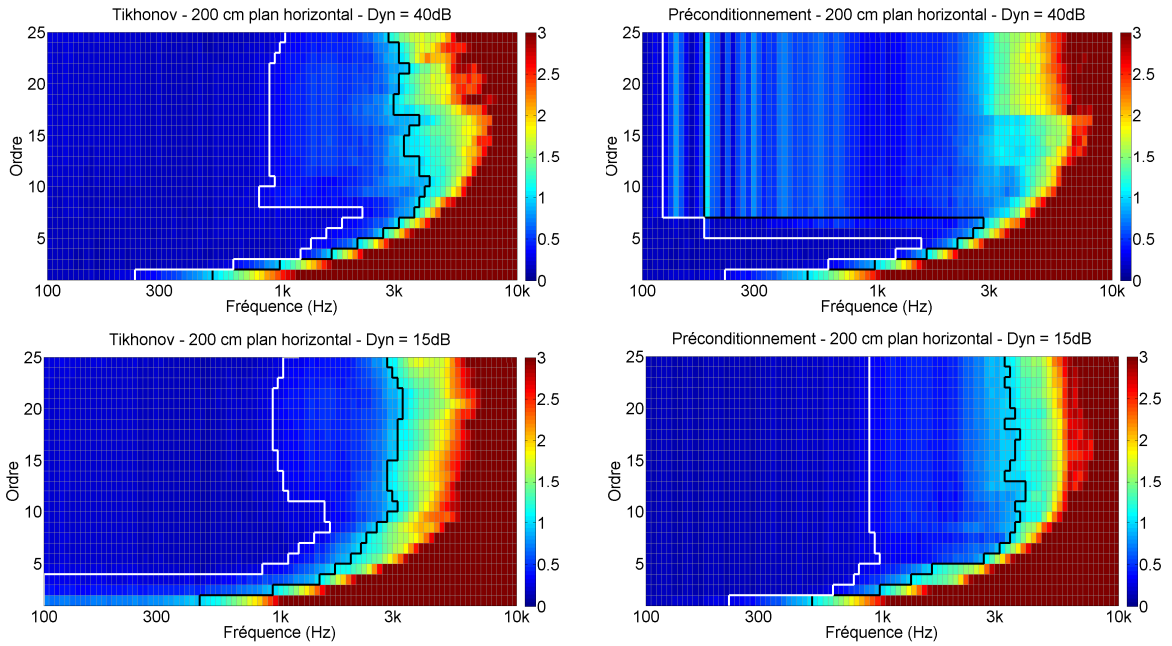


Figure 13 : Erreur de reconstruction dans le plan horizontal à partir de données bruitées. Le seuil à 0.3 dB est représenté en blanc et le seuil à 1 dB est représenté en noir. Figure de gauche : résolution avec régularisation de Tikhonov, figure de droite : résolution avec préconditionnement

Une dynamique de 15 dB permet d'obtenir des résultats similaires avec les deux méthodes pour les ordres supérieurs à 12 : le seuil à 0.3 dB est atteint autour de 1 kHz et le seuil à 1 dB est atteint autour de 3 kHz. En revanche pour les ordres inférieurs à 4, les écarts spectraux sont toujours supérieurs à 0.3 dB avec la méthode de Tikhonov, alors que le seuil à 0.3 dB est atteint entre 200 Hz et 800 Hz avec la méthode de préconditionnement. Pour les ordres compris entre 4 et 10, le

seuil à 0.3 dB est atteint plus haut en fréquence qu'avec la méthode de préconditionnement : en particulier à l'ordre 8 il est atteint à 1.6 kHz avec la méthode de Tikhonov et à 0.9 kHz avec la méthode de préconditionnement. Au contraire, le seuil à 1 dB est atteint plus haut en fréquence avec la méthode de préconditionnement : à 3.7 kHz à l'ordre 8, alors qu'il n'est jamais supérieur à 3 kHz avec la méthode de Tikhonov. En conditions bruitées, il semblerait que la méthode la plus précise soit celle avec préconditionnement si le seuil de 1 dB est visé, et la méthode de Tikhonov si le seuil à 0.3 dB est visé.

Le paramétrage de Dyn est déterminant pour les deux méthodes en conditions bruitées. Ce paramétrage semble un peu moins sensible pour la méthode de Tikhonov, cependant pour cette méthode le paramétrage de l'ordre de troncature est tout aussi important. Pour la méthode avec préconditionnement, la tendance est inversée : le paramètre Dyn a une grande importance, et l'ordre de troncature n'a pas une influence significative. Pour le traitement des données mesurées (potentiellement bruitées), une dynamique de 15 dB semble préférable. Alors, la méthode avec préconditionnement semble la plus robuste (peu sensible au choix de l'ordre), même si elle n'est pas forcément précise.

II.2.B.c Effet du nombre de points

Une autre simulation est réalisée, en considérant cette fois un plus grand nombre de points, mais sans considérer un maillage extrêmement dense (qui tendrait vers la solution idéale). Seules deux élévations sont rajoutées de manière à compléter judicieusement les points de mesure : l'élévation -90° (1 point) et l'élévation -60° (72 points). Cet ajout de points permet d'effectuer le calcul avec un maillage symétrique en élévation, et en limitant les zones sans points. Les écarts spectraux obtenus lors de la propagation dans le plan horizontal sont représentés dans la Figure 14. L'ajout de ces points supplémentaires améliore sensiblement les résultats de propagation : avec la méthode de préconditionnement les écarts sont inférieurs à 1 dB jusqu'à 6.4 kHz à partir de l'ordre 14, alors qu'ils ne l'étaient que jusqu'à 5 kHz à partir de l'ordre 8. Les points rajoutés sont toutefois impossibles à mesurer comme cela a été évoqué en introduction car ils correspondent à des élévations très négatives. Par ailleurs, la partie basse d'un mannequin ne ressemble plus à un corps, et l'usage de ces HRTF n'est pas envisageable pour un système de diffusion. Ce résultat illustre bien le fait que l'impossibilité concrète de mesurer ces points limite les possibilités de propagation.

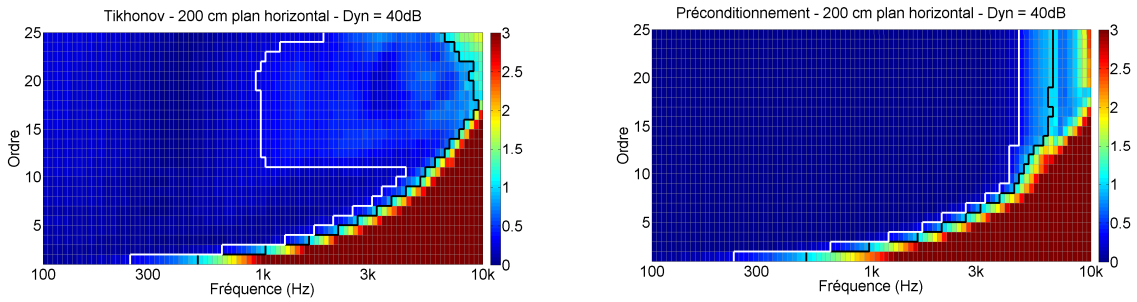


Figure 14 : Ecarts spectraux moyens en fonction de l'ordre de troncature à 2 m. Le seuil à 0.3 dB est représenté en blanc et le seuil à 1 dB est représenté en noir. Figure de gauche : inversion avec régularisation de Tikhonov, figure de droite : inversion avec préconditionnement

II.3 Mesures de HRTF à deux distances

La propagation des HRTF par décomposition en harmoniques sphériques semble prometteuse, à condition de limiter la gamme fréquentielle en tenant compte du nombre de points de mesure. Une campagne de mesure de HRTF sur différents mannequins a donc été réalisée et l'algorithme de propagation est testé sur ces mesures.

Une première campagne de mesure a été réalisée dans la grande chambre anéchoïque du Laboratoire de Mécanique et d'Acoustique (site de Joseph Aiguier). Pour ces essais, les mannequins étaient disposés sur un pied de microphone, lui-même placé sur une table tournante permettant de caractériser plusieurs azimuts. La source était manuellement positionnée pour chaque élévation et distance souhaitée. Seuls les azimuts allant de 0° à 180° ont été caractérisés pour plusieurs élévations, l'idée étant de reconstruire le cercle complet par symétrie entre les deux oreilles. Malheureusement, des écarts importants ont été observés entre les deux oreilles pour les incidences 0° et 180°, de l'ordre de 1 dB jusqu'à 1 kHz et de plusieurs dB aux fréquences supérieures. Ces écarts nous ont paru trop importants pour provenir de dissymétries des mannequins, et nous avons estimé que le dispositif mis en place ne permettait pas un positionnement assez précis pour notre objectif.

Une seconde campagne de mesure a donc été réalisée, en s'appuyant sur le dispositif mis en place dans le cadre du projet Loudnat (ANR-11-BS09-016-01) destiné à mesurer des HRTF sur des auditeurs à la distance de 2 m. Le dispositif a été adapté pour pouvoir mesurer des HRTF de mannequins, et pouvoir réaliser des mesures à plus faible distance. Les mesures que nous avons réalisées et celles du projet Loudnat ont été effectuées sur une période commune, ne facilitant pas la mise en œuvre. Suite au déménagement du laboratoire, il n'a en particulier pas été possible de modifier le dispositif existant pour ajouter des positions de mesure.

Toutes les mesures ont été réalisées dans la chambre anéchoïque de Psychoacoustique du Laboratoire de Mécanique et d'Acoustique sur le site de Joseph Aiguier. Ses dimensions sont de 5.3 x 6.0 x 5.0 m, et les parois sont recouvertes de dièdres de 80 cm d'épaisseur. La fréquence de coupure est estimée à 150 Hz et le bruit de fond est de 25 dBA.

Ces HRTF ont été estimées dans la bande de fréquences comprises entre 200 Hz et 15 kHz, qui contient la majeure partie de l'information de localisation. L'estimation en dessous de cette bande de fréquences semble peu pertinente car les variations entre les mannequins sont négligeables : les longueurs d'onde sont alors très supérieures aux dimensions de l'objet diffractant. L'estimation au-delà de cette bande de fréquences n'est pas réalisée, car elle nécessite une extrême précision de positionnement difficilement réalisable en pratique alors qu'un auditeur est réputé être moins sensible à ces fréquences.

II.3.A. Mannequins caractérisés

Quatre mannequins différents et une sphère ont été mesurés :

- Gras KEMAR 45 BB-3 (Sound Quality Recording), avec les grandes oreilles (VA-style, modèle KB0090 et KB0091). Les microphones sont de type GRAS 40AG (1/2'') et

préamplificateurs 26AS, conditionnés à l'aide d'un Nexus 2690. La tête sera désignée « Kemar » dans la suite du chapitre.

- Bruël & Kjaer 4100-D. Les microphones sont de type B&K 4189 et préamplificateurs B&K 2671 conditionnés à l'aide du boitier fourni avec la tête. La tête sera désignée « B&K » dans la suite du chapitre.
- Cortex MK2. Les capsules microphoniques sont de type MK 231 E (type ICP), alimentés par une alimentation fantôme fournie par une carte audionumérique Focusrite Saffire pro et un adaptateur isemCOM SA-P48/CCP-C. La tête sera désignée « Cortex » dans la suite du chapitre.
- Head Acoustics HRS-II.2. Nous n'avons pas d'information concernant le type de microphones, seulement qu'ils nécessitent une alimentation fantôme. Elle est fournie par une carte audiomérique Focusrite Saffire Pro. Cette tête dispose d'égalisations (« Free Field » et « Independant of Direction »), aucune égalisation n'a été sélectionnée pour les mesures. La tête sera désignée « Head » dans la suite du chapitre.
- Une sphère a été fabriquée en ABS par impression 3D. Elle est constituée de deux demi-sphères creuses dont les parois ont une épaisseur de 4 mm, rigidifiées par des nervures et un enduit d'étanchéité. Le rayon extérieur est de 8.75 cm et un emplacement pour microphone a été prévu de chaque côté de sorte qu'ils soient affleurants à la surface avec un angle d'incidence de 100° par rapport au plan médian. Des microphones GRAS 40PR (1/4") ont été employés pour les mesures. Pour la sphère, seule une sélection de mesure a été réalisée pour comparer une partie des résultats.

Le centre de la tête de chacun de ces mannequins est considéré comme étant le point situé au milieu des deux microphones. Ce choix de convention a l'avantage d'être le plus facilement identifiable, et correspond avec la définition proposée par la recommandation ITU [ITU-T P.58, 2013] (le repère gradué sur le mannequin B&K visible en Figure 15 ne correspond pas à la convention adoptée). Pour la sphère, le centre est par contre pris en son centre, qui n'est pas exactement le point entre les deux microphones (situés légèrement en arrière du diamètre). Le positionnement des mannequins est ajusté à l'aide d'un niveau laser en croix. La précision d'un tel dispositif est de l'ordre de 5 mm.

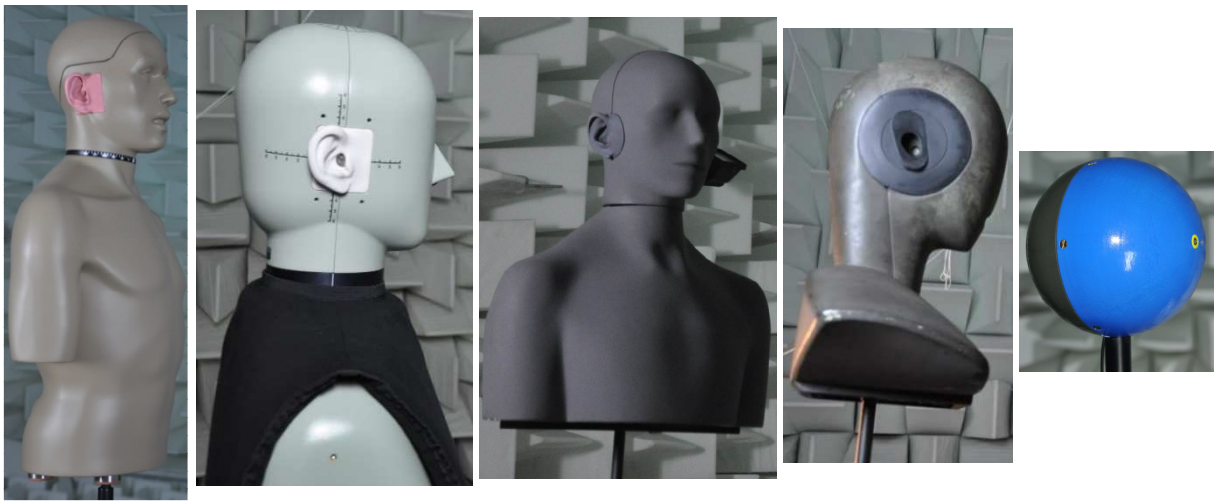


Figure 15 : Photos des mannequins caractérisés. De gauche à droite : KEMAR, B&K 4100D, Cortex MK2, HRS-II.2 et sphère

Quand cela était possible, les microphones des mannequins ont été calibrés à l'aide d'un pistonphone générant un niveau de 94 dBSPL à 1 kHz. Le mannequin Head est particulier puisque ses microphones ne sont pas des microphones de mesure, et il n'est pas possible de les calibrer. Une estimation du facteur de calibration a été réalisée à partir d'une mesure à l'incidence frontale à 200 Hz : nous estimons qu'à cette fréquence le niveau sonore est équivalent quel que soit le mannequin.

II.3.B. Dispositif pour la mesure à 2 m

Les mesures en champ lointain ont été effectuées à une distance de 2 m, en utilisant un réseau de haut-parleurs disposés en deux arcs-de-cercle autour du mannequin. Les haut-parleurs sont des Meyersound MM-4XP (enceintes actives), et seuls les haut-parleurs disposés dans le plan médian sont utilisés. Ils sont placés aux élévations 0°, 15°, 30°, 45°, 60° et 90° comme l'illustre la Figure 16.

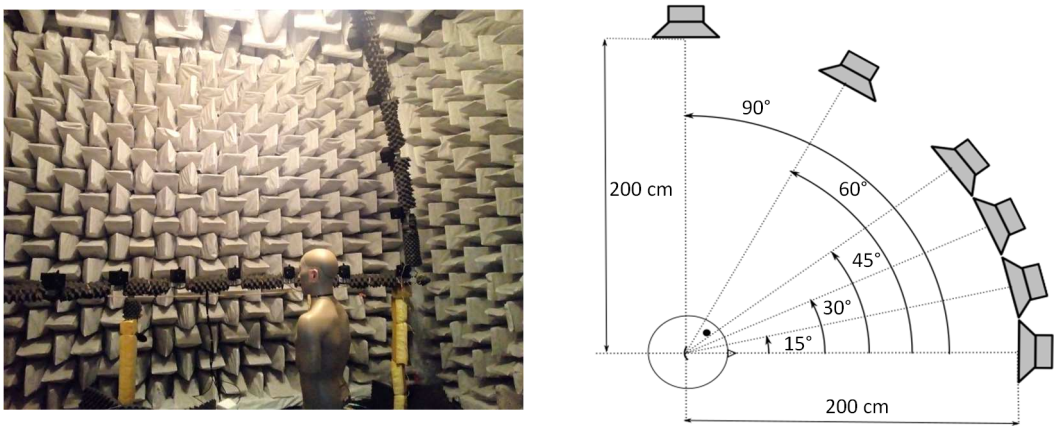


Figure 16 : illustration du dispositif de mesure en champ lointain. Figure de gauche : photo avec le mannequin KEMAR, figure de droite : schéma du dispositif

II.3.C. Dispositif pour la mesure à 40 cm

Les mesures à 40 cm ont été réalisées en positionnant un haut-parleur à l'élévation voulue. Une enceinte a été spécialement conçue pour s'approcher d'une source ponctuelle avec une réponse en basse fréquence suffisante pour effectuer la mesure désirée. Les détails de la conception de cette source sont présentés dans l'Annexe A. Le signal d'entrée de la source est préalablement amplifié par un amplificateur Trends TA 10.1.

La bande passante de cette source est de [200 Hz – 18 kHz] à +/- 10 dB, ce qui permet l'acquisition de HRTF dans la bande-passante désirée ([200 Hz – 15 kHz]). L'atténuation liée à la directivité est inférieure à 1 dB pour les bandes d'octaves inférieures ou égales à 8 kHz et pour les angles d'incidence inférieure à 15°, et inférieure à 4 dB pour les angles d'incidence inférieure à 30°. La directivité est considérée comme négligeable pour les angles inférieurs à 15°, donc la diffraction par la tête ne sera pas influencée significativement par la directivité de la source. La directivité pour les angles entre 15° et 30° n'est pas négligeable (la diffraction par les épaules pourra être influencée par la directivité) mais est relativement faible. Le dispositif permettant les mesures en champ lointain était toujours présent lors des mesures à proximité du mannequin.

Les mesures ont été réalisées pour 7 élévations : -30°, -15°, 0°, 15°, 30°, 60° et 90°. Pour chaque élévation, la source est placée à la position associée et dirigée vers le centre du dispositif.

II.3.D. Matériel utilisé

Les haut-parleurs et l'acquisition étaient pilotés par une carte son MOTU 24I/O. Le mannequin était placé sur une table tournante LT360 EX capable d'une résolution de 1°. Une interface programmée en Max/MSP permettait de piloter l'ensemble des opérations (rotation de la table, lecture et enregistrement des signaux). Le signal utilisé pour la mesure était un sweep exponentiel entre 50 Hz et 20 kHz, de 6.1 s de durée et calculé à partir de la méthode présentée par Antonin Novak [Novak et al., 2009]. Les signaux bruts ont été enregistrés en wav 24 bits à la fréquence d'échantillonnage de 48 kHz, puis post-traités pour obtenir des réponses impulsionnelles.

II.3.E. Mesure du champ libre (sans mannequin)

Une partie de l'estimation des HRTF correspond à la mesure de la pression au centre du dispositif sans le mannequin, soit le terme $P_{champ\ libre}$ de l'équation (3). Pour cela, les fonctions de transfert de tous les haut-parleurs ont été caractérisées. Un microphone champ libre de type B&K 4190 dirigé vers la source a été utilisé pour ces mesures.

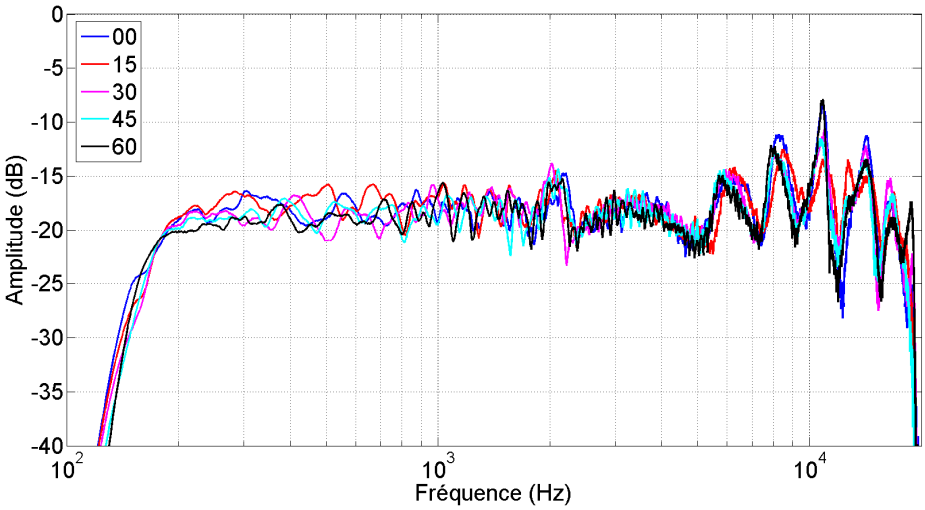


Figure 17 : Fonctions de transfert des différents haut-parleurs utilisés pour les mesures en champ lointain

Les fonctions de réponse en fréquences des différents haut-parleurs sont reportées dans la Figure 17 pour les haut-parleurs en champ lointain et dans la Figure 68 (Annexe A) pour le haut-parleur à faible distance. Les différents haut-parleurs pour le champ lointain ont des réponses en fréquences similaires, avec des écarts dans toute la bande passante, atteignant 5 dB à 500 Hz et jusqu'à 8 dB à 2,2 kHz. Cette variabilité n'est pas problématique pour l'estimation de HRTF puisqu'elle est compensée selon la définition donnée par l'équation (3). La bande passante s'étend de 150 Hz à 18 kHz à +/- 10 dB et est donc satisfaisante pour réaliser les mesures envisagées. La directivité n'a pas été caractérisée, car considérée comme négligeable en champ lointain.

II.3.F. Mise en forme des HRTF

Plusieurs étapes de traitement sont réalisées pour obtenir des HRTF correspondant à la définition donnée dans l'équation (3) et minimiser les biais liés au protocole de mesures. Cinq opérations sont réalisées, elles sont détaillées dans l'Annexe C.

La première opération consiste à fenêtrer les réponses impulsionnelles, permettant de limiter leur durée et la présence d'écho liée à la présence d'appareils de mesure. Un fenêtrage en deux bandes de fréquences est réalisé, limitant davantage la longueur des RI aux hautes fréquences. La longueur finale des RI est de 42.7 ms (2048 échantillons à 48 kHz).

La deuxième opération consiste à normaliser les fonctions de transfert mesurées avec les mannequins par la fonction de transfert mesurée au centre de la tête (sans le mannequin). En pratique, cette normalisation est réalisée par la division des spectres complexes des fonctions de transfert. Les HRTF ainsi obtenues ne sont donc pas systématiquement causales.

En raison des limites des sources électro-acoustiques, les mesures en basses et hautes fréquences sont bruitées. Pour les fréquences inférieures à 230 Hz, les HRTF sont calculées à partir d'un modèle de sphère de rayon 8.75 cm dont les récepteurs sont sur le diamètre. Les hautes fréquences sont éliminées à partir d'un filtre de Linkwitz-Riley d'ordre 12 à la fréquence de coupure de 19 kHz.

Enfin, la dernière étape consiste à égaliser les HRTF. Elle permet notamment de compenser l'influence des microphones et de leur position dans le conduit auditif. Une égalisation de type champ diffus est appliquée, où la HRTF en champ diffus est estimée à partir d'une moyenne des mesures, pondérées en fonction de l'angle solide représenté.

II.4 Comparaison de quatre mannequins et une sphère à deux distances

Une analyse préliminaire est menée, concernant la validité du dispositif de mesure : des mesures sur une sphère rigide ont été réalisées et comparées au calcul analytique. De plus, la symétrie entre les mesures de chaque oreille a été vérifiée. Ces résultats sont présentés dans l'Annexe D. Il ressort de cette analyse que les mesures sur la sphère sont similaires aux calculs analytiques, avec quelques variations de 1 à 2 dB jusqu'à 8 kHz. Des différences plus importantes ont été observées pour les mesures à 40 cm pour les fréquences supérieures à 11 kHz, pouvant provenir de la directivité de la source. Des dissymétries entre les oreilles des mannequins ont été relevées, et sont d'un ordre de grandeur similaire aux observations de [Andreopoulou et al., 2015].

L'analyse suivante se focalise sur les mesures effectuées dans le plan horizontal qui permettent une analyse déjà très riche, car nous estimons que l'effet de l'azimut est prépondérant sur celui de l'élévation et mérite plus d'attention. En effet, la perception de l'azimut d'une source est plus fine que la perception de l'élévation [Blauert, 1997]. L'analyse porte sur l'amplitude des HRTF et sur l'estimation de l'ITD par une méthode de seuil.

II.4.A. HRTF dans le plan horizontal

Les HRTF dans le plan horizontal sont représentées par la Figure 18 sous forme de cartographie pour différents azimuts, où l'amplitude est codée par une échelle de couleurs.

II.4.A.a Remarques générales

Toutes les HRTF représentées se ressemblent, avec deux principales zones selon l'azimut : une zone d'amplification pour les azimuts inférieurs à 180°, et une zone d'atténuation pour les autres incidences. Ces zones sont particulièrement marquées aux fréquences élevées, supérieures à 1 kHz : la dynamique est alors de plus de 15 dB. Au niveau de la zone d'atténuation autour des incidences 270°, une zone est moins atténuée voire amplifiée pour les fréquences comprises entre 500 Hz et 3 kHz. Il s'agit du « bright spot » [Duda & Martens, 1998], correspondant à une arrivée en phase de plusieurs trajets acoustiques. Ce bright spot est identifiable pour tous les mannequins et à toutes les distances.

II.4.A.b Effet de la distance

L'effet de la distance est visible sur les différentes HRTF. Premièrement, les HRTF à faible distance sont davantage contrastées entre les incidences ipsilatérales et contralatérales. En basses fréquences (f<800 Hz), l'amplitude associée aux incidences ipsilatérales est de l'ordre de 5 dB à 40 cm, alors qu'elles sont proches de 0 dB à 2 m. De manière générale, des différences de l'ordre de 5 dB entre les extrema des deux distances sont observables pour tous les mannequins, y compris la sphère : l'amplitude maximale à 40 cm est de 10 dB à 10 kHz pour l'incidence 90°, alors qu'elle est seulement de 5 dB à 2 m. La zone d'atténuation est plus prononcée à faible distance, et le bright spot est également plus atténué.

Une conséquence de ce contraste plus élevé à faible distance est que les HRTF varient plus rapidement d'une incidence à l'autre. Visuellement, cela se traduit par un étalement plus important des zones d'amplification et d'atténuation selon l'axe des azimuts en champ lointain. Par exemple, la zone où le gain est supérieur à 5 dB de la tête Head entre 500 Hz et 1 kHz s'étend entre les azimuts 20° et 160°, alors qu'elle s'étend entre les azimuts 10° et 190° à 2 m.

Pour expliquer ce contraste plus élevé, il faut rappeler qu'une HRTF correspond au rapport entre la pression au niveau de l'oreille et la pression au centre. En approximant le rayon de la tête à 10 cm, à 40 cm le rapport entre les distances source-micro et source-centre est de 40/30, et correspond à un facteur 1.33. A l'inverse, à 2 m de distance le facteur est seulement de 1.05 : l'amplification en champ lointain est alors principalement liée aux effets de diffraction, alors qu'à proximité du mannequin la différence de distance intervient davantage. Pour les positions contralatérales, la tête fait aussi davantage écran à l'onde incidente sphérique et atténue davantage l'énergie à faible distance, y compris pour le bright spot.

Une métrique d'écarts spectraux $ES(f)$ permettant de caractériser des différences entre l'amplitude spectrale de deux fonctions de transfert H_X et H_Y est définie :

$$ES(f) = \left| 20 \log_{10} \left(\frac{|H_X(f)|}{|H_Y(f)|} \right) \right| \qquad (29)$$

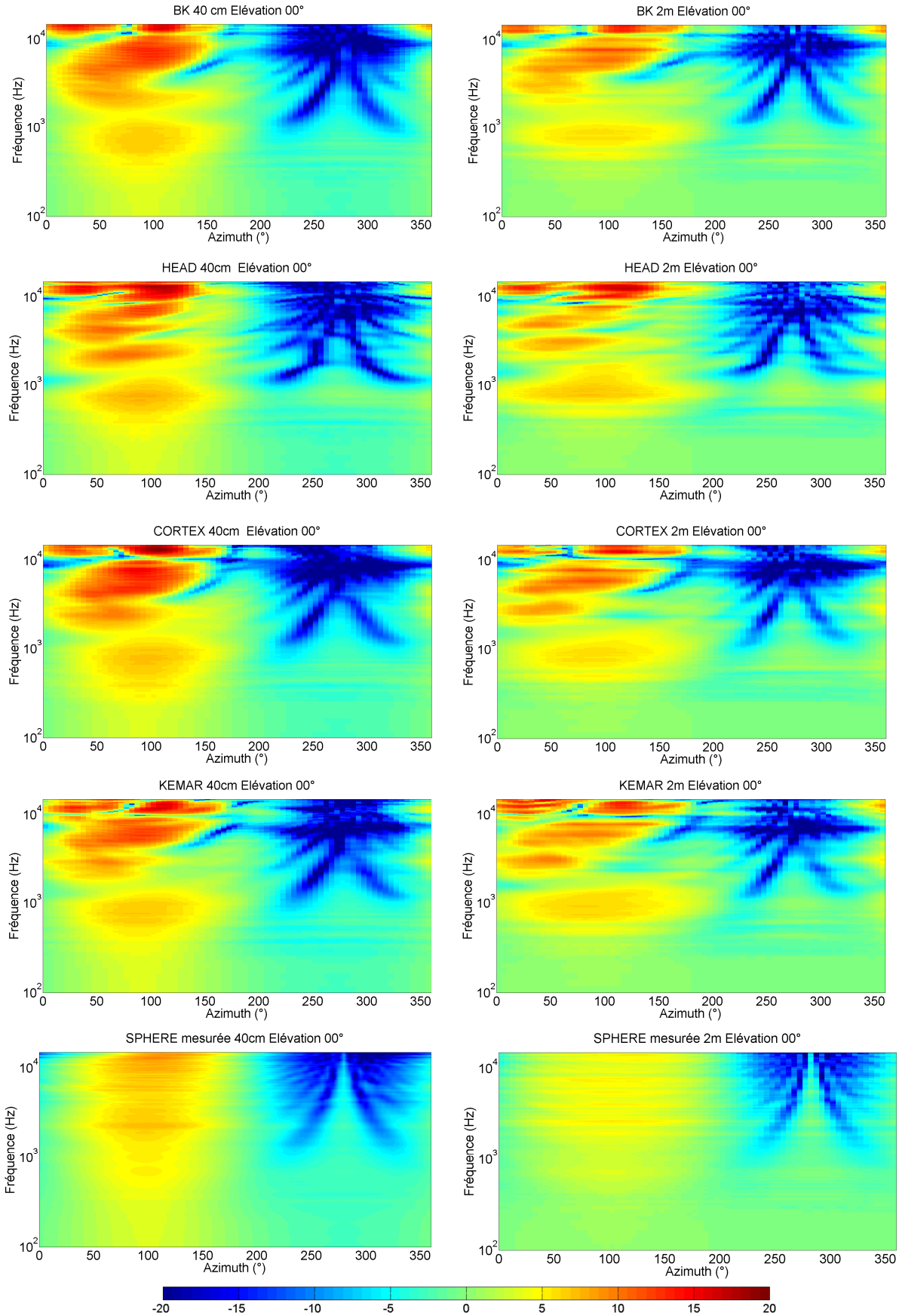


Figure 18 : HRTF mesurées dans le plan horizontal pour les quatre mannequins et la sphère à 40 cm (figures de gauche) et à 2 m (figures de droite). Les HRTF ont été égalisées par la réponse en champ diffus. De haut en bas : B&K, HEAD, CORTEX, KEMAR et SPHERE. L'amplitude des HRTF est codée par une échelle de couleurs de -20 dB à +20 dB

Les écarts spectraux moyens sont calculés entre les deux distances à partir de l'équation (29), et moyennés pour l'ensemble des azimuts. Ils sont reportés dans la Figure 19. Ils sont de l'ordre de 2 à 3 dB dans l'essentiel de la bande passante, et augmentent avec la fréquence (dépassant ponctuellement 5 dB au-delà de 10 kHz). L'effet de la distance sur les mesures est donc important. Pour la sphère, cet effet se rapproche de celui des autres mannequins, bien que minorant plutôt des écarts par rapport aux autres mannequins. La comparaison entre mesure et calcul de la sphère permet de quantifier l'incertitude de mesure. Ainsi, autour de 400 Hz les écarts sont de l'ordre de 2 dB pour tous les mannequins (y compris la sphère mesurée), alors qu'ils sont seulement de l'ordre de 1.5 dB pour la sphère calculée. De manière analogue, des pics autour de 2 kHz, 2.5 kHz, 7 kHz et 10 kHz sont identifiables pour tous les mannequins et la sphère, alors que le modèle de sphère prédit une variation régulière. Il est donc probable que ces accidents soient en partie liés à l'environnement de mesure, la présence de nombreuses structures à l'intérieur de la salle anéchoïque pouvant en particulier expliquer certains accidents.

La comparaison des écarts entre mesures à deux distances pour la sphère d'une part, et pour les mannequins d'autre part, fait apparaître assez peu de différences notamment avec le mannequin B&K. L'irrégularité constatée pour le mannequin Head peut résulter de l'utilisation de microphones « audio » de réponse moins contrôlée que des microphones de mesure.

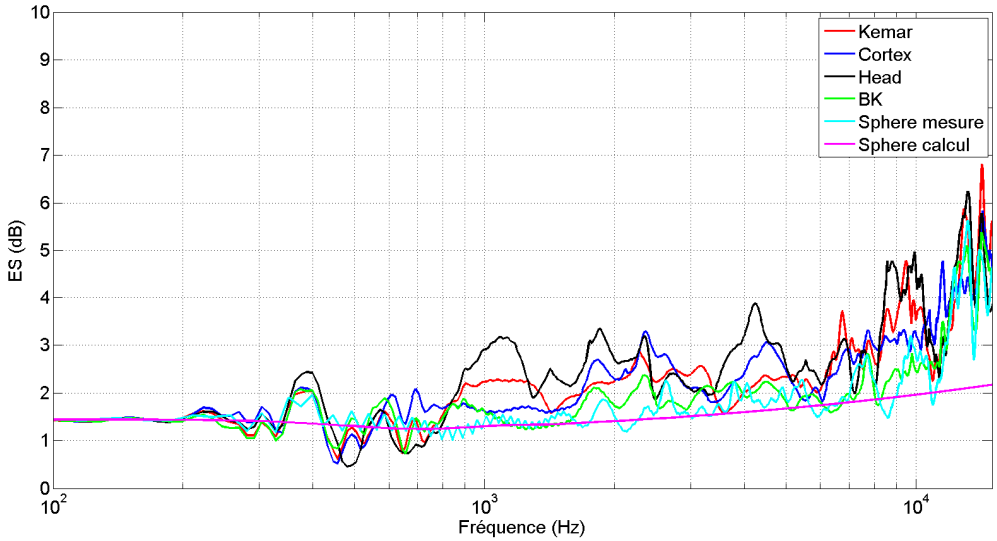


Figure 19 : Ecarts spectraux moyens dans le plan horizontal entre les HRTF à 2 m et 40 cm pour les 4 mannequins et la sphère

II.4.A.c Effet du mannequin

Pour une distance donnée, les HRTF des mannequins se ressemblent donc, mais il est clairement possible de les identifier. Premièrement, l'allure du bright spot est variable selon le mannequin considéré. La sphère est un cas particulier, puisque son bright spot s'étend jusqu'à 15 kHz, alors qu'il n'est plus identifiable au-delà de 3 kHz pour les autres mannequins. Cela peut s'expliquer par la plus grande complexité des morphologies des mannequins : pour une sphère il existe de nombreuses symétries, permettant l'arrivée en phase de plusieurs trajets acoustiques. Pour les mannequins au contraire, la présence de détails morphologiques (pavillons d'oreille, nez, etc.) limite le nombre de symétries. En basses fréquences, les longueurs d'ondes sont supérieures aux

dimensions des détails morphologiques et les différents trajets acoustiques peuvent arriver en phase au niveau de l'oreille.

La zone d'ombre correspondant à un gain inférieur à -10 dB est similaire pour les mannequins Cortex, B&K et Kemar. Elle s'étend entre les azimuts 150° et 330° pour les fréquences autour de 8 kHz. Pour le mannequin Head, elle est moins étendue (entre 200° et 330°) et se rapproche de celle de la sphère. Cette particularité est probablement liée à la forme des pavillons d'oreille, qui sont assez simplifiés dans le cas de la tête Head. A 8 kHz, les longueurs d'onde sont de l'ordre de 4 cm, correspondant à des dimensions caractéristiques des pavillons.

La zone d'amplification correspondant à un gain supérieur à 5 dB pour les incidences ipsilatérales est similaire selon les mannequins, à l'exception de la sphère. Pour la sphère, cette zone est uniforme en azimut et en fréquence, contrairement aux mannequins où plus de variabilité est visible selon la fréquence et l'azimut. Ceci est notamment lié aux pavillons d'oreilles, créant des amplifications selon l'incidence de la source et la forme du pavillon. En l'absence de pavillons comme c'est le cas pour la sphère, ces amplifications sont limitées.

Les écarts spectraux moyens entre les différents mannequins estimés dans le plan horizontal sont reportés dans la Figure 20 pour la distance de 40 cm. Les écarts à la distance de 2 m sont équivalents, et ne sont donc pas représentés. Au-delà de 7 kHz, les écarts entre mannequins sont importants, atteignant localement 8 dB. Notamment entre 600 Hz et 1500 Hz les écarts les plus importants sont systématiquement obtenus pour la comparaison de la tête Head et des autres mannequins, y compris la sphère. Ces écarts sont de l'ordre de 2 dB et traduisent une particularité liée à ce mannequin peut-être encore liée à ses microphones « audio ».

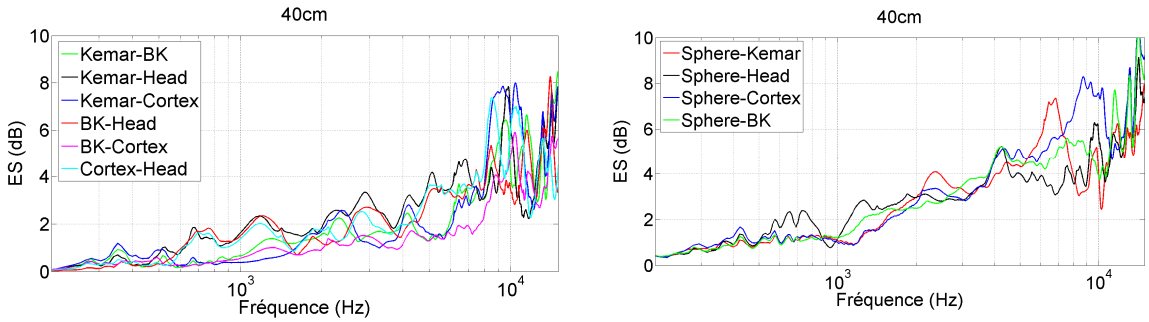


Figure 20 : Ecarts spectraux moyens dans le plan horizontal entre mannequins à 40 cm. Figure de gauche : comparaison entre mannequins, figure de droite : comparaison avec la sphère

Les écarts entre la sphère et les autres mannequins sont similaires quel que soit le mannequin, croissant régulièrement avec la fréquence jusqu'à des valeurs de l'ordre de 4 dB vers 5 kHz. Cet ordre de grandeur est assez faible, proche de celui existant entre les différentes paires de mannequins.

Ainsi, l'effet de distance semble similaire pour tous les mannequins y compris pour la sphère. Cet effet se traduit par un contraste plus élevé à faible distance et des extrema de plus grande amplitude. L'effet du mannequin se traduit par une répartition fréquentielle et azimutale différente des zones d'amplification et d'atténuation suivant le modèle. Les différences moyennes entre la sphère et chaque mannequin semblent être du même ordre de grandeur que les différences entre les mannequins, les détails des HRTF étant moins fins pour la sphère, mais spécifiques à chaque mannequin.

II.4.B. Calcul d'ITD

Une caractéristique temporelle des HRTF est estimée par l'ITD, qui correspond à la différence des temps de l'arrivée de l'énergie entre les deux oreilles. En synthèse binaurale les HRTF sont très communément modélisées sous une forme compacte, en une composante à phase minimale et un retard pur qui correspond à l'ITD [Busson et al., 2005; Larcher, 2001]. Cette approximation consiste à considérer que la composante d'excès de phase des HRTF est linéaire, et en pratique cela n'est pas systématiquement le cas : il n'existe pas toujours un retard commun à toutes les fréquences associées aux HRTF. L'estimation de ce retard n'est alors pas triviale, et plusieurs méthodes d'estimation de l'ITD des HRTF ont été proposées, avec de nombreuses variantes selon les auteurs qui les utilisent.

Une présentation de différentes méthodes a été proposée dans [Katz & Noistering, 2014]. Au total, 32 manières de calculer l'ITD ont été comparées sur une sphère rigide et sur le mannequin Kemar. De larges variations ont été observées selon les méthodes, dépassant 200 µs pour le mannequin Kemar. Selon le contenu du stimulus, la différence minimale perceptible (Just Noticeable Difference, JND) peut varier entre 2 µs et 60 µs [Blauert, 1997], les variations introduites par la méthode d'estimation sont largement supérieures à ces JND. Les auteurs ont mis en évidence des différences selon les méthodes, sans en préconiser une. Le choix de la méthode dépend de l'usage que l'on souhaite faire de l'ITD.

Les méthodes d'estimation de l'ITD à partir de HRTF peuvent se classer en trois familles, pour lesquelles il existe plusieurs variations dans l'implémentation.

II.4.B.a Méthodes d'intercorrélation

Une famille de méthodes consiste à calculer l'intercorrélation entre les réponses impulsionnelles droite et gauche, et l'ITD est déduite de la position du maximum de la fonction d'intercorrélation. Une alternative consiste à calculer cette fonction à partir de l'enveloppe des réponses impulsionnelles, et il est également envisageable de filtrer au préalable les réponses impulsionnelles. L'inconvénient de cette méthode est la présence potentielle de plusieurs pics dans la fonction d'intercorrélation. Dans ce cas, une ambiguïté apparait, biaisant l'estimation. Cela apparait dans le cas de plusieurs arrivées successives d'énergie, ce qui est typiquement le cas en présence de réflexions. L'effet de diffraction par la tête se traduit par une propagation de l'onde sonore autour de la tête, et l'énergie peut aussi arriver en différents instants. Le cas particulier du « bright spot » correspond à une arrivée en phase de plusieurs trajets, mais pour de nombreuses incidences les arrivées sont légèrement décalées, créant une ambiguïté sur le temps d'arrivée.

II.4.B.b Méthodes du retard de groupe

D'autres méthodes sont basées sur l'estimation du retard de groupe des réponses en fréquences. Une manière communément répandue [Larcher, 2001] consiste à réaliser pour chacune des oreilles une régression linéaire sur la phase déroulée dans une bande de fréquences particulière. L'ITD se déduit de la différence de pente entre les régressions issues des deux oreilles. Le choix de la bande de fréquences est déterminant dans cette approche, et en fait l'avantage de cette méthode : il est facilement possible de sélectionner la bande de fréquences d'intérêt. L'approximation large-bande est possible, mais est biaisée si la phase n'est pas linéaire.

II.4.B.c Méthodes de seuils

Enfin, pour la troisième famille de méthode une estimation du temps d'arrivée sur chaque oreille est donnée par l'instant où l'amplitude de la réponse impulsionnelle dépasse un seuil. L'ITD correspond à la différence entre ces deux temps d'arrivée. Le choix du seuil est un paramétrage important de cette méthode : il est possible que plusieurs pics soient présents autour du pic principal, perturbant alors la détection si leur amplitude est proche du pic principal. Il est également possible de filtrer préalablement les réponses impulsionnelles pour estimer l'ITD dans une bande de fréquences particulière. Dans ce cas, plus la largeur de la bande de fréquences est petite, et plus la résolution temporelle est faible, réduisant la précision de l'estimation.

Seule la méthode de seuil va être employée ici, car elle semble plus robuste pour des mesures en salle d'écoute, qui correspond à l'application de nos travaux. Pour effectuer le calcul, le premier échantillon de la HRIR dépassant le seuil fixé est préalablement identifié. Cette identification n'est pas suffisante, car la résolution temporelle à la fréquence d'échantillonnage de 48 kHz est de 21 µs, ce qui est supérieur aux JND. Pour affiner l'estimation, les HRIR sont sur-échantillonnées d'un facteur 100 autour du premier échantillon dépassant le seuil. Pour effectuer le sur-échantillonnage, une régression polynomiale d'ordre 2 est effectuée en utilisant le premier échantillon dépassant le seuil, le précédent et le suivant. Les coefficients polynomiaux sont ceux qui minimisent l'écart au sens des moindres carrés sur les trois échantillons utilisés. Le sur-échantillonnage permet alors d'avoir une résolution temporelle de 0.21 µs et il est possible d'estimer l'instant de dépassement du seuil avec plus de précision.

Les ITD sont estimées pour les quatre mannequins et la sphère en utilisant un seuil de -6 dB et un seuil de -12 dB. Les résultats dans le plan horizontal sont représentés dans la Figure 21. Cette figure permet de mettre en évidence l'importance du choix du seuil : lorsqu'il est trop élevé (-6 dB), les variations d'ITD d'une incidence à l'autre peuvent être élevée. Notamment, pour les azimuts proches de 90° (autour de l'axe interaural), des « sauts » d'ITD de l'ordre de 200 µs sont visibles pour les mannequins Kemar, Cortex et Head. Ces variations n'apparaissent pas avec le seuil à -12 dB, et les résultats semblent alors plus cohérents. Avec le seuil à -12 dB, les différences entre mannequins sont faibles, les principales différences apparaissent autour de l'axe interaural : la différence atteint jusqu'à 60 µs entre le mannequin Kemar et le mannequin Head. L'ITD obtenue avec la sphère est légèrement supérieure à celui des autres mannequins, la différence étant maximale avec le mannequin Kemar : environ 100 µs à 90°, notamment du fait de la dissymétrie relevée sur ce mannequin. Globalement, en dehors du mannequin Kemar les ITD des autres

mannequins sont très similaires et leurs écarts dépassent légèrement les JND. L'effet de la distance n'est quant à lui pas significatif.

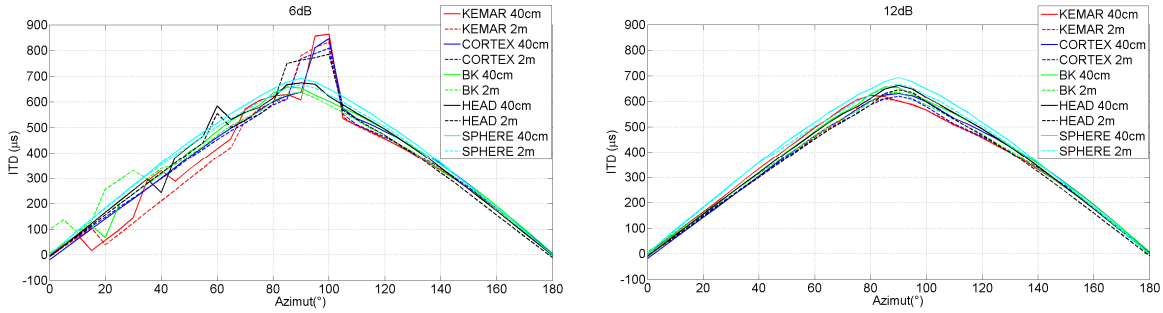


Figure 21 : ITD des quatre mannequins et la sphère estimées aux deux distances de mesure (40 cm et 2 m) dans le plan horizontal. Deux seuils différents sont employés : 6 dB (figure de gauche) et 12 dB (figure de droite)

II.4.C. Analyse des différences observées

L'effet de la distance a été identifié, les HRTF à faible distance sont davantage contrastées que les HRTF en champ lointain. Ce contraste plus élevé implique que les variations d'une incidence à l'autre sont plus importantes, et cela confirme l'intérêt de mesurer les HRTF à plusieurs distances. Des différences ont été constatées entre la sphère et les mannequins, mais les valeurs moyennes sont similaires à celles existant entre les autres mannequins. Le modèle de sphère est donc approprié pour approximer l'effet de distance sur les HRTF.

Les HRTF mesurées sur les différents mannequins se ressemblent : la répartition de l'énergie en fonction de l'azimut et de la fréquence est similaire dans le plan horizontal, et les ITD estimées sont très proches pour tous les mannequins. Des différences entre mannequins ont été observées, et sont du même ordre de grandeur que les différences entre la sphère et les mannequins. Concernant la sphère, la répartition de l'énergie dans le plan horizontal en fonction de la fréquence est plus lisse que pour les mannequins, on observe moins de variations selon l'incidence. Une légère surestimation de l'ITD a été observée dans le cas de la sphère pour les incidences proches de l'axe interaural, les différences avec les mannequins étant de l'ordre des JND. La sphère n'est donc pas significativement différente des autres mannequins censés approximer un auditeur moyen, sa simplicité conduisant par contre à des HRTF plus lisses.

Jusqu'à 1 kHz, les différences observées entre deux distances pour un même mannequin semblent plus importantes que les différences entre deux mannequins. Entre 1 kHz et 8 kHz, les différences sont d'un ordre de grandeur équivalent, et au-delà les différences entre mannequins sont les plus importantes. Pour implémenter un système de diffusion, il est plus important de prendre en compte les effets de distance que les effets d'individualisation jusqu'à 1 kHz. Au-delà, les particularités individuelles semblent aussi importantes que les particularités liées à la distance, mais les écarts moyens entre les mannequins sont du même ordre que ceux avec la sphère. Ces différences sont toutefois un minorant de différences individuelles, car tous les mannequins sont censés approximer une morphologie moyenne. Par ailleurs, la mesure de HRTF sur des auditeurs réels pose le problème de leurs éventuels mouvements pendant la mesure, introduisant encore des artefacts en plus de ceux suspectés à une distance de 2 m.

II.5 Propagation des HRTF : application aux mannequins

Une base de données de 433 points de mesure pour chaque mannequin à 40 cm a été constituée et la méthode de propagation présentée en début de chapitre est maintenant utilisée avec ces données expérimentales. Dans le cas de la sphère, seules les mesures dans le plan horizontal ont été réalisées : la propagation est alors réalisée à partir de données calculées, et les données mesurées sont utilisées pour évaluer le résultat de la propagation.

II.5.A. Propagation en hautes fréquences

La méthode de décomposition en harmoniques sphériques donne des résultats encourageants en basses fréquences, mais plus limités en hautes fréquences. Une méthode plus robuste est donc proposée pour les fréquences supérieures à 3 kHz. Cette méthode se base sur le comportement favorable d'une sphère en termes de prise en compte de la distance des sources. Elle consiste à calculer une fonction de compensation de la distance à partir des HRTF d'un modèle sphérique, définie de la manière suivante :

$$H_{r\to R}(\theta,\phi,f) = \frac{\left|HRTF_{sphere}(R,\theta,\phi,f)\right|}{\left|HRTF_{sphere}(r,\theta,\phi,f)\right|} \tag{30}$$

Avec $HRTF_{sphere}(R,\theta,\phi,f)$ les HRTF d'une sphère de 8.75 cm de rayon et oreilles sur le diamètre pour une onde incidente au point (R,θ,ϕ). Pour assurer la continuité des solutions obtenues en basses fréquences par décomposition en harmoniques sphériques et celles en hautes fréquences obtenues à partir du modèle, les HRTF sont propagées sur toutes la bande fréquentielle pour les deux méthodes. Ces HRTF sont ensuite converties en HRIR par transformée de Fourier inverse et un filtre de cross-over de Linkwitz-Riley d'ordre 8 est appliqué à la fréquence de coupure de 3 kHz.

II.5.B. Application aux mesures sur les mannequins

Pour les fréquences inférieures à 3 kHz, les HRTF des mannequins sont propagées avec la méthode avec préconditionnement avec un ordre 7 de décomposition en harmoniques sphériques et une dynamique $Dyn = 15\ dB$. Les écarts spectraux entre les HRTF propagées et mesurées dans le plan horizontal sont représentés dans la Figure 22 pour deux cas de figure. Le premier correspond à la propagation en considérant simplement les données mesurées. Le second cas correspond à une symétrisation artificielle des données : des données manquantes aux élévations -60° et -90° ont été introduites dans l'algorithme, en utilisant les données aux élévations 60° et 90°. Cette manière de procéder permet de légèrement diminuer les écarts spectraux, notamment pour le mannequin Cortex autour de 300 Hz : avec les données brutes les écarts sont de l'ordre de 1 dB, alors qu'ils sont de l'ordre de 0.5 dB avec la symétrisation. Cette opération semble donc permettre une meilleure reconstruction des données. Avec la symétrisation, les écarts spectraux sont de l'ordre de 0.5 dB à 1 dB pour tous les mannequins jusqu'à 2 kHz, puis augmentent avec la fréquence et dépassent 2 dB pour les mannequins Head, Kemar et B&K à 2.5 kHz. Ces écarts sont du même ordre de grandeur que ceux estimés dans le cas de la sphère jusqu'à 3 kHz, et plus importants que ceux estimés dans le cas de la sphère sans bruit (paragraphe II.2.B.a). Le bruit de mesure semble donc plus important que l'erreur liée à la propagation et le résultat de la propagation des mesures à 40 cm est donc probablement meilleur que ces écarts.

Les HRTF propagées dans le plan horizontal avec symétrisation des données sont représentées dans la Figure 23, ainsi que les HRTF mesurées à 2 m pour comparaison. Une différence apparait systématiquement pour les fréquences autour de 2 kHz : le gain est plus faible pour les HRTF mesurées que pour les HRTF propagées. Cette différence fait penser à un artefact de mesure, car elle rend discontinue la répartition énergétique des HRTF mesurées, alors qu'elle est plus lisse pour les HRTF propagées. La discontinuité sur les mesures est particulièrement visible pour le mannequin Head.

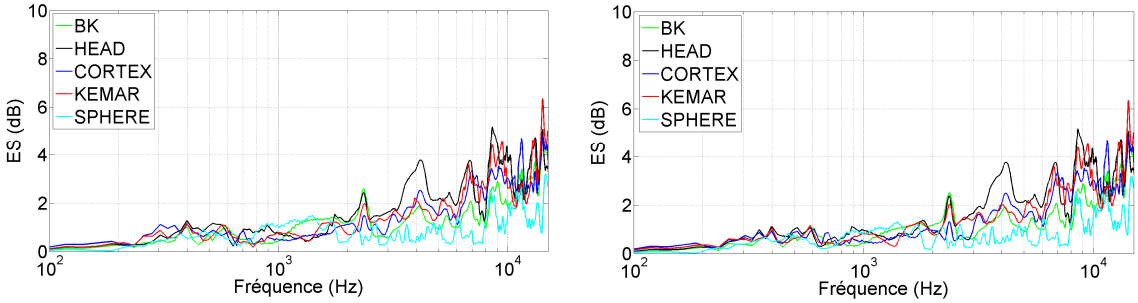


Figure 22 : Ecarts spectraux moyens entre les mesures et les données propagées à 2 m dans le plan horizontal. Figure de gauche : sans ajout de données, Figure de droite : ajout de données aux élévations -60° et -90°

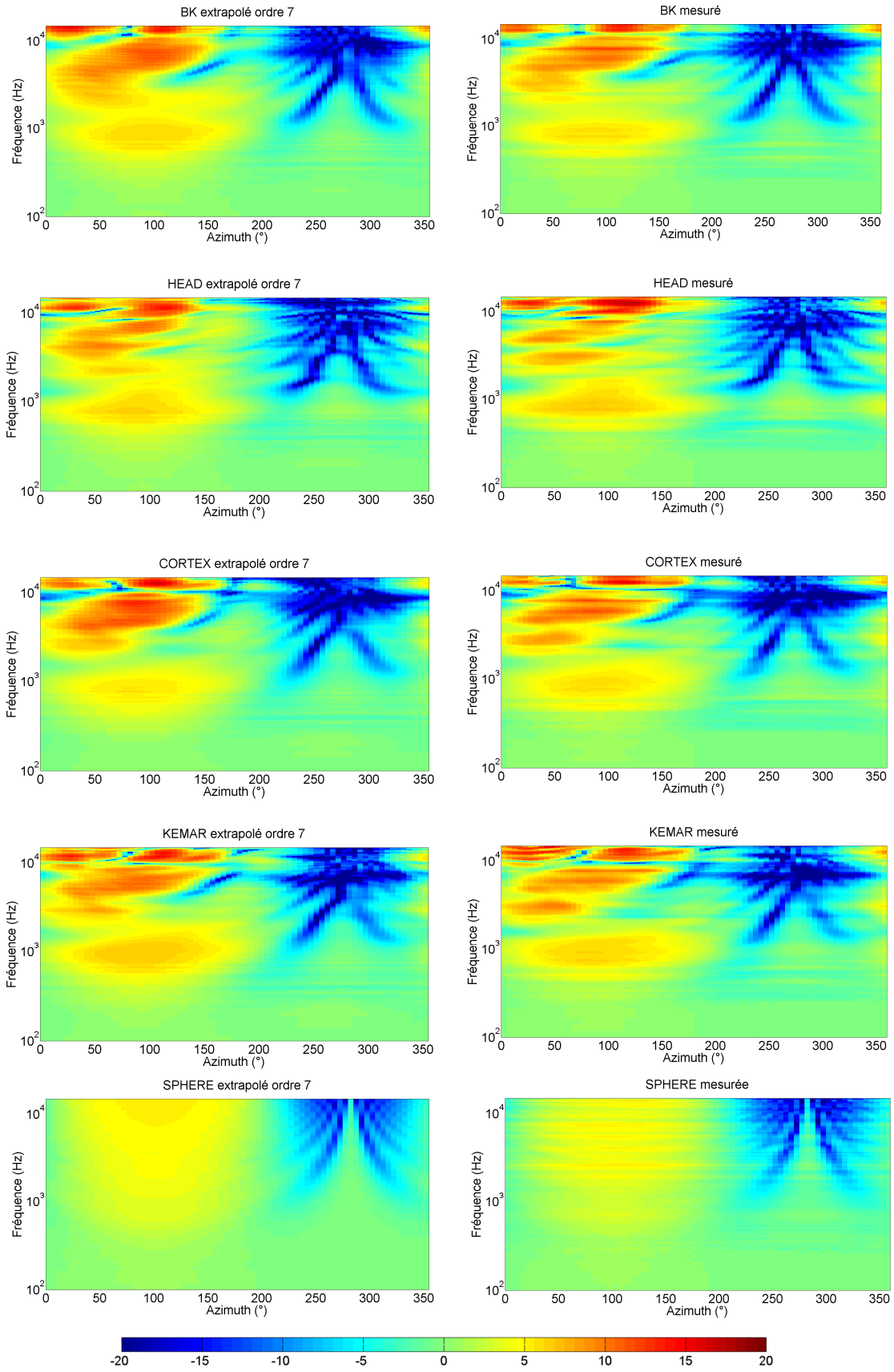


Figure 23 : HRTF propagées à 2 m (gauche) à partir de données symétrisées à 40 cm et mesurées à 2 m (droite) dans le plan horizontal. Pour la propagation, les données à 40 cm sont mesurées pour les quatre mannequins, et calculées pour la sphère.

II.6 Conclusion du chapitre

Une méthode de propagation de HRTF basée sur une décomposition en harmoniques sphériques a été proposée. La démarche consiste à sélectionner les termes séparables aux positions de mesure. Les résultats obtenus avec cette méthode sont similaires en basses fréquences à la méthode par régularisation de Tikhonov, mais elle est plus facile à paramétrer. Il est alors nécessaire de fixer une dynamique d'inversion, l'ordre de décomposition ayant peu d'influence s'il est suffisant. La limite majeure de la décomposition en harmoniques sphériques est qu'il est nécessaire de caractériser un grand nombre de points, y compris aux élévations inférieures à -30°, qui sont malheureusement inaccessibles expérimentalement.

Quatre mannequins et une sphère ont été caractérisés avec un même dispositif aux distances de 40 cm et 2 m pour quelques élévations. Des comparaisons ont été réalisées entre les différentes mesures et sont résumées dans le Tableau 1. Pour les fréquences inférieures à 1 kHz, les ordres de grandeur des écarts entre les deux distances sont les plus importants : cela confirme la nécessité de caractériser les mannequins à plusieurs distances. Notamment, les HRTF à faible distance sont plus contrastées : cela traduit une plus grande variabilité d'une incidence à l'autre. Les différences observées sont toutefois influencées par le bruit de mesure, qui a été quantifié dans l'Annexe D. Il est possible que ce bruit de mesure soit encore plus important pour des auditeurs réels : leurs mouvements possibles peuvent perturber la mesure.

Ecarts moyens	100 Hz → 1 kHz	1 kHz → 10 kHz
Entre 40 cm et 2 m	De 1 à 3 dB	De 1 à 5 dB
Entre mannequins	De 0 à 2 dB	De 1 à 8 dB
Entre sphère et mannequins	De 0 à 2 dB	De 1 à 7 dB
Entre propagation et mesure	1 dB	De 1 à 5 dB

Tableau 1 : ordre de grandeur d'écarts moyens observés dans le plan horizontal pour différents effets

Il semblerait que les mesures à proximité soient moins bruitées que les mesures à plus grande distance. Il serait donc préférable de mesurer à faible distance les HRTF, et utiliser une technique de propagation pour estimer les données à plus grande distance.

Un résultat important pour la suite de notre travail est que les écarts entre la sphère et les mannequins sont du même ordre que les écarts entre mannequins. Les mannequins diffèrent donc à peu près autant entre eux qu'avec la sphère, tous étant censés approximer un auditeur « moyen ». Pour concevoir un système de diffusion sonore, l'utilisation d'un modèle de sphère semble donc aussi justifiée que l'utilisation de HRTF issues des mannequins qui ne correspondent de toute façon pas à l'auditeur. L'utilisation d'un modèle de sphère est par contre beaucoup plus flexible que des HRTF mesurées directement sur des auditeurs : le calcul analytique permet d'obtenir des HRTF en tout point de l'espace, pour un coût de calcul très faible. L'absence de détails morphologiques conduit à une moindre description des détails, mais dans le cadre d'une approche préliminaire d'optimisation d'un système de diffusion sonore ce modèle semble suffisant. Il va donc être employé dans la suite des travaux, permettant de simuler la diffusion sonore pour de nombreuses configurations.

Chapitre III Efficacité des sources

Table des matières

III.1 Introduction

La diffusion transaurale consiste pour beaucoup à compenser une différence de diffraction entre les sources sonores « physiques » du système de reproduction et les sources « virtuelles » à reproduire. Cette compensation se fait par un double filtrage : pour un système transaural à deux canaux, deux filtres servent à égaliser les sources, et deux filtres cherchent à annuler les trajets croisés entre les deux sources et les deux oreilles.

Ce filtrage « croisé » conduit à annuler partiellement le champ produit par les sources, ce qui implique une plus grande sollicitation alors qu'elles sont limitées en performances. Ce compromis prend une importance particulière aux basses fréquences où les limites des petites sources sont particulièrement gênantes (voir Annexe A).

Ce chapitre cherche à évaluer l'influence de la géométrie d'un système de reproduction transaural sur les sources électro-acoustiques. Il s'agit ici d'un problème de dimensionnement électro-acoustique : la capacité des sources doit être adaptée à la dynamique nécessaire. Dans le cas où les sources sont mal dimensionnées, de la distorsion non-linéaire peut apparaître modifiant alors le rendu. L'aspect perceptif n'a pas été abordé dans ce mémoire, mais une thèse a été réalisée sur le sujet au sein du laboratoire [Michaud, 2012].

III.2 Calcul de filtres transauraux

Différentes méthodes de calcul de filtres transauraux ont été présentées dans l'état de l'art et seule la méthode avec régularisation de Tikhonov est employée par la suite.

Le système transaural est constitué de deux haut-parleurs placés aux coordonnées $(r, \pm\theta, \phi)$. Pour une configuration de système transaural à deux haut-parleurs, il existe quatre transferts source-récepteur : pour chaque haut-parleur il y a deux transferts correspondant aux deux oreilles. Les quatre fonctions de transfert sont assemblées sous forme de matrice de taille 2x2, désignée par $[C]$:

$$[C(f)] = \begin{bmatrix} C_{LL}(f) & C_{RL}(f) \\ C_{LR}(f) & C_{RR}(f) \end{bmatrix} \tag{31}$$

La dépendance fréquentielle sera omise dans la suite, les calculs sont effectués indépendamment pour chaque fréquence. Les signaux des sources sont préalablement filtrés par une matrice $[H]$ qui cherche à annuler les termes non diagonaux de $[C]$. La relation entre les signaux en entrée et la pression en sortie du système transaural est donnée par la relation suivante :

$$\boldsymbol{OUT} = [C].[H].\boldsymbol{IN} \tag{32}$$

où $[H]$ est la matrice de filtres transauraux à déterminer et $\boldsymbol{IN}$ et $\boldsymbol{OUT}$ les entrées et sorties du système. $\boldsymbol{IN}$ correspond donc au signal à reproduire, et $\boldsymbol{OUT}$ au signal reproduit aux oreilles de l'auditeur.

$[H]$ est calculée ici d'après la pseudo-inverse de Moore-Penrose de $[C]$ régularisée par une matrice de Tikhonov [Ole Kirkeby, Nelson, Hamada, et al., 1998] :

$$[H] = ([C]^*.[C] + \beta[Id])^{-1}.[C]^* A \tag{33}$$

β est un terme de régularisation, choisi d'après une dynamique Dyn : $\beta = \max(C^2).10^{\frac{-Dyn}{20}}$. La valeur de ce paramètre n'a pas une influence significative sur les données non bruitées. Pour les simulations ci-après, une dynamique $Dyn = 80\ dB$ est employée, ce qui revient à régulariser marginalement la solution. $[Id]$ est la matrice identité et A est une réponse cible correspondant à un filtre passe-bande. En pratique, un filtre FIR de 85 ms de durée (4096 échantillons à la fréquence d'échantillonnage de 48 kHz) est jugé suffisant. Un retard d'une demi-longueur de filtre est introduit dans A par permutation circulaire des échantillons.

Les signaux peuvent ensuite être reconstruits aux oreilles de l'auditeur dans la configuration simulée. La reconstruction des signaux est réalisée dans le domaine temporel. Pour cela, les réponses impulsionnelles associées à toutes les réponses en fréquence sont obtenues par transformée de Fourier inverse. Les variables dans le domaine temporel sont désignées par les mêmes lettres que l'équivalent en fréquentiel mais en minuscules. Le son hp_L généré sur le haut-parleur gauche est ainsi défini de la manière suivante :

$$hp_L(t) = in_L * h_{LL} + in_R * h_{RL} \tag{34}$$

Le signal joué sur le haut-parleur droit se définit de manière analogue, en utilisant les signaux dont les indices L et R sont permutés.

III.3 Critère d'évaluation de configurations

La démarche employée ici sera utilisée pour évaluer l'influence de différents aspects de ce travail, et elle est représentée par la Figure 24. La tête de l'auditeur y est représentée par une sphère, les points rouges correspondant à ses oreilles et le point vert correspondant à son nez. Pour chaque configuration, l'évaluation se fait en trois temps : le rendu est simulé aux oreilles de l'auditeur à partir du modèle de diffraction de sphère validé au Chapitre II, un indicateur est calculé à partir de ces signaux, et cet indicateur est normalisé sur une échelle bornée. Cette opération est réalisée pour de nombreuses configurations, et pour les représenter les notes attribuées sont codées par une échelle de couleurs à l'emplacement correspondant aux sources physiques. Cette représentation est limitée à un demi-cercle à gauche de l'auditeur, car un système transaural est constitué de deux sources placées de manière symétrique par rapport à l'axe X : la connaissance de la position d'une des deux sources permet de déduire la position de la source complémentaire.

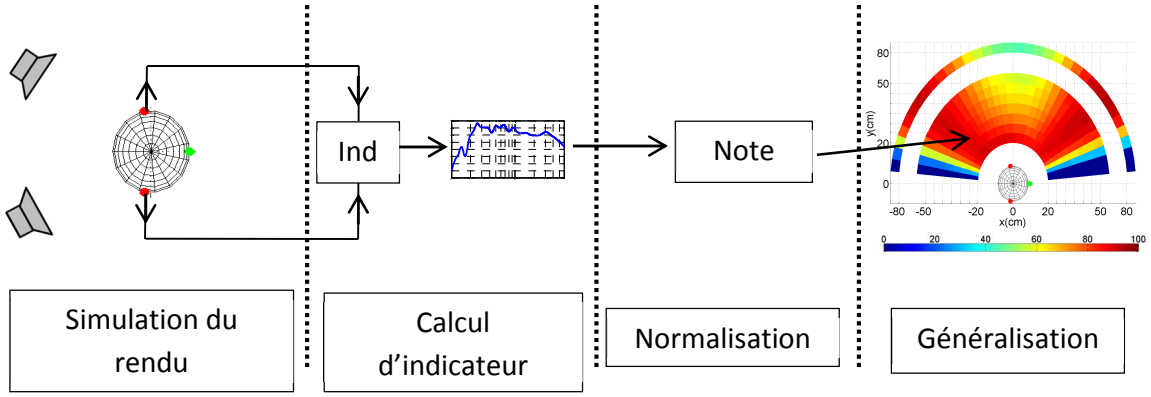


Figure 24 : illustration de la démarche proposée pour évaluer chaque aspect influençant la diffusion sonore

III.4 Influence des filtres transauraux

III.4.A.Notion de « contraste »

Pour comprendre l'influence des filtres transauraux sur les sources électro-acoustiques, il est pratique d'introduire la notion de contraste. Il correspond à la différence du niveau sonore généré par une source entre l'oreille gauche et l'oreille droite : il dépend de la position de la source et de la fréquence. Par exemple, les fonctions de réponse en fréquence des oreilles gauche et droite sont représentées par la Figure 25 pour des sources d'incidence 5° et 90° à 40 cm de distance. Le contraste n'est jamais supérieur à 3 dB pour l'incidence de 5°, alors qu'il est toujours supérieur à 5 dB pour la source à 90°. Par ailleurs, le contraste dépend de la fréquence : pour la source à 90°, le contraste est de l'ordre de 5 dB pour les fréquences inférieures à 400 Hz, et dépasse 25 dB à 8 kHz.

Pour certaines configurations de paires de haut-parleurs, le contraste est naturellement élevé (par exemple pour l'azimut +/-90°), alors que pour d'autres il est très limité (par exemple pour l'azimut +/-5°). Le fait que la position des sources du système conduise à un contraste élevé est *a priori* un atout pour l'implémentation de filtres transauraux : l'effort de filtrage est alors réduit. Pour le vérifier, un cas particulier de diffusion transaurale est calculé dans la prochaine section, et un indicateur associé au coût lié à l'obtention du contraste est proposé.

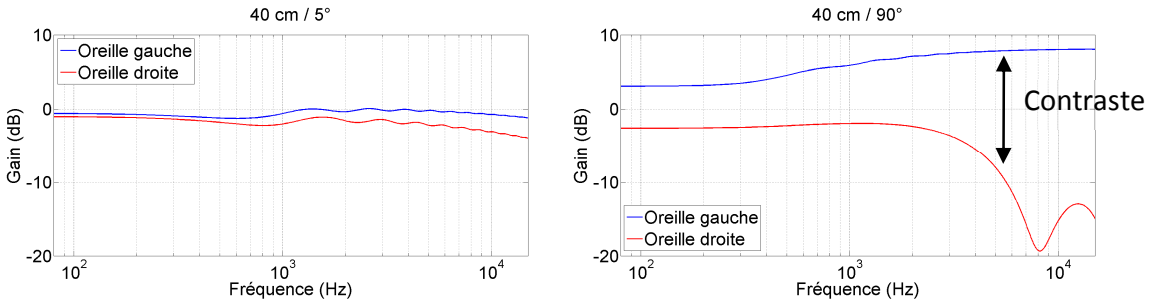


Figure 25 : FRF calculées avec un modèle de sphère pour deux incidences de source : θ = 5° (figure de gauche) et θ = 90° (figure de droite)

III.4.B.Coût lié à l'obtention du contraste

Pour évaluer le coût lié à l'obtention du contraste, le signal employé correspond à un cas extrême de contraste entre les deux oreilles : $in_L(t)$ est un Dirac et $in_R(t)$ est un vecteur de zéros. L'évaluation symétrique consistant à utiliser un Dirac sur la voie droite n'est pas présentée car son résultat serait identique. Pour évaluer le coût d'obtention du contraste transaural, le débit des sources en reproduction transaurale est comparé au débit des sources sans utiliser de filtres pour obtenir la même pression acoustique au niveau d'une oreille. Pour cela, le rapport $R_{débit}$ est défini de la manière suivante :

$$R_{débit} = \frac{Q_{trans}}{Q_{monopole}} \tag{35}$$

Avec Q_{trans} la somme des débits des sources lorsque les filtres transauraux sont employés, et $Q_{monopole}$ la somme des débits des sources seules sans interférences. Les sources sont considérées comme des monopoles, et le débit Q d'une source en un point de l'espace à une distance r de la source en champ libre est lié à la pression P de la manière suivante :

$$Q(f) = \frac{2rP}{\rho f} \tag{36}$$

Pour obtenir la même pression OUT_L sur la position de l'oreille sans utiliser de filtre transauraux, les contributions des deux sources sont additionnées en module. La transformée de Fourier du signal joué sur chacune des sources est alors :

$$HP_m = \frac{OUT_L}{|C_{LL}| + |C_{LR}|} \tag{37}$$

Le rapport de la somme des débits peut alors s'écrire :

$$R_{débit} = \frac{|HP_L| + |HP_R|}{2|HP_m|} \tag{38}$$

Le rapport de débit ainsi obtenu dans trois situations est présenté dans la Figure 26. Il s'agit du rapport de débit pour trois configurations à 40 cm de distance, pour un positionnement angulaire des sources à +/-5°, +/-10° et +/-20°. Dans le cas de la solution classique du stéréo-dipôle (+/-5°), le rapport de débit est élevé en basses fréquences : il est supérieur à 10 pour les fréquences inférieures à 200 Hz. Le rapport de débit décroît avec la fréquence jusqu'à 6 kHz, avec un pic vers

10.5 kHz atteignant la valeur de 4. Le coût en débit est donc maximal aux basses fréquences, alors que les sources électro-acoustiques y sont les plus limitées (voir Annexe A). Cette augmentation du débit en basses fréquences s'explique par le fait que le contraste naturel est particulièrement faible (voir Figure 25) : pour créer le contraste, le gain des filtres transauraux annulent une grande part du rayonnement, ce qui se traduit par une augmentation de débit. Le pic de rapport de débit observable autour de 10.5 kHz s'explique plutôt par la différence de distance de propagation entre une source et chaque oreille, qui est un multiple d'une demi-longueur d'onde (les deux contributions sont en opposition de phase).

Plus l'écartement angulaire entre les sources est faible, plus les fonctions de transfert de chacune des sources vers une oreille tendent à être identiques. Le cas extrême est le cas où les deux sources sont coïncidentes, correspondant alors à deux transferts identiques. L'augmentation de l'écart angulaire permet de réduire le rapport de débit aux basses fréquences : pour un écartement angulaire de +/-20° il n'est jamais supérieur à 5. Pour les plus grands écarts angulaires, le contraste naturel est plus élevé, simplifiant le filtrage transaural : les solutions à plus grand écart angulaire permettent donc de limiter le coût lié à l'obtention du contraste en basses fréquences, là où les sources électro-acoustiques sont les plus limitées.

Par ailleurs, aux hautes fréquences un plus grand nombre de pics de rapport de débit est visible lorsque l'écartement angulaire augmente. L'amplitude des pics décroît avec l'augmentation de l'écart angulaire, pour la même raison qu'en basses fréquences : le contraste naturel simplifie le filtrage.

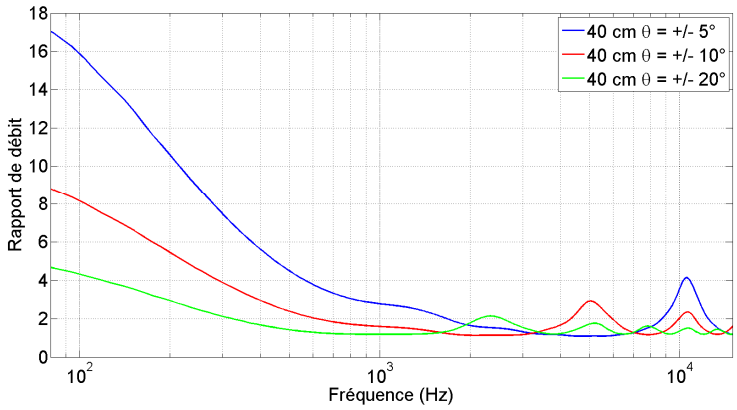


Figure 26 : Rapport de débit pour trois configurations à 40 cm : écart angulaire de 5°, 10° et 20°

Les augmentations de débit aux hautes fréquences peuvent être prédites à partir de la géométrie du problème, illustrée par la Figure 27. Sur cette figure, les sources sont localisées aux points S_1 et S_2 et les récepteurs aux points R_1 et R_2. La modélisation des trajets acoustiques est approximée, en considérant qu'il s'agit d'une combinaison d'un segment entre la source et la sphère, et d'un arc de cercle de cette intersection jusqu'à l'oreille. Un pic de débit a lieu pour une différence de marche voisine d'un multiple d'une demi-longueur d'onde, soit :

$$|x_1 - x_2| = \frac{n\lambda}{2} \tag{39}$$

Et cela permet d'identifier les fréquences centrales des pics f_{pics} :

$$f_{pic} = \frac{nc}{2|x_1 - x_2|} \tag{40}$$

A partir de cette relation, les quatre premières fréquences centrales des pics sont calculées pour les configurations à 40 cm d'écartement angulaire +/-5°, +/-10° et +/-20° et reportées dans le Tableau 2. Les ordres de grandeur des fréquences centrales des pics correspondent à ceux observés sur la Figure 26.

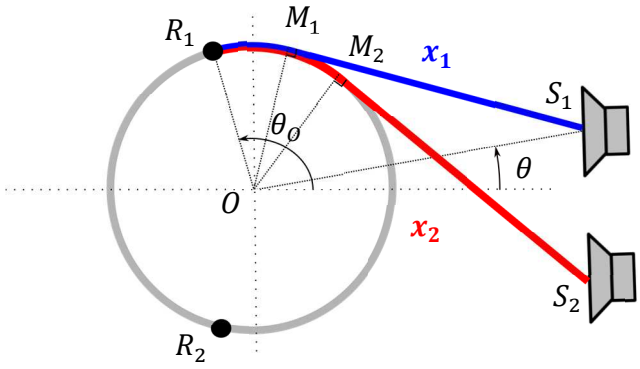


Figure 27 : représentation schématique de la tête de l'auditeur. Les points noirs correspondent aux oreilles, le trait bleu au trajet de la source 1 vers le récepteur 1 et le trait rouge au trajet entre la source 2 et le récepteur 1.

n	1	2	3	4
$\theta = +/-5°$	11.3 kHz	22.5 kHz	33.8 kHz	45.1 kHz
$\theta = +/-10°$	5.6 kHz	11.3 kHz	16.9 kHz	22.5 kHz
$\theta = +/-20°$	2.8 kHz	5.6 kHz	8.5 kHz	11.3 kHz

Tableau 2 : fréquences centrales des pics estimées

III.4.C. Conversion en note

Pour convertir le rapport de débit en une évaluation entre 0 et 100, une loi arbitraire a été définie sur la base de deux bornes. Un doublement de débit est considéré comme la limite de l'acceptable, et un rapport unitaire correspond au cas idéal. La loi choisie doit donc passer par le cas idéal (100) et tendre asymptotiquement vers 0 quand le débit augmente. Une loi basée sur une fonction gaussienne est ainsi utilisée par la suite :

$$Note_{dynamique} = 100e^{\frac{-(R_{débit}-1)^2}{2\sigma^2}} \tag{41}$$

Cette loi est illustrée à la Figure 28, le choix de $\sigma = 0.5$ permet de fixer la « limite acceptable ». Cette « note » respecte qualitativement le comportement attendu, mais est choisie de manière relativement arbitraire, les notes n'ayant pas de lien avec la perception. Cette normalisation permettra cependant de la combiner avec les autres aspects évalués dans la thèse (effet de salle, déplacement) pour lesquels les notations ont un lien avec la perception.

Comme le rapport de débit dépend de la fréquence considérée, il est calculé par bandes d'octave, pour des fréquences centrales allant de 125 Hz à 8 kHz. Le dimensionnement de sources électro-acoustiques est le plus problématique aux basses fréquences : la surface de la membrane doit être inversement proportionnelle au carré de la fréquence à reproduire (voir Annexe A). Un rapport de débit « global » est aussi évalué comme la moyenne des débits par bandes d'octave, pondérées par le carré de la fréquence centrale. Il y a donc 8 $Note_{dynamique}$ calculées pour chaque configuration.

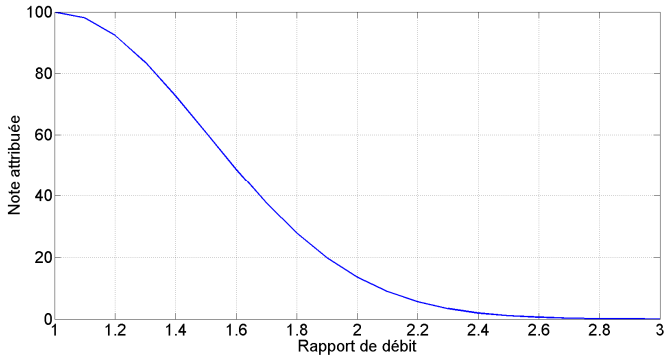


Figure 28 : loi de conversion du rapport de débit en « note » comprise entre 0 et 100 (100 étant la meilleure note)

III.5 Evaluation de configurations de sources

Le critère proposé est désormais employé pour évaluer un grand nombre de configurations de sources dans le plan horizontal, et en élévation.

III.5.A.Sources dans le plan horizontal

La simulation est premièrement réalisée pour des sources situées dans le plan horizontal, pour les positions des sources physiques allant de +/-5° à +/- 175° par pas de 5°, et pour les distances allant de 20 cm à 50 cm par pas de 5 cm, et à 80 cm. Les résultats pour les bandes d'octave de 125 Hz à 8 kHz, ainsi que le cas global sont représentés dans la Figure 29.

Ces figures permettent d'illustrer que plusieurs facteurs influent l'augmentation de débit, de manière analogue aux observations précédentes. Premièrement, le coût lié à l'obtention du contraste est particulièrement élevé aux basses fréquences. La majorité des configurations ont une note inférieure à 50 pour la bande d'octave de 125 Hz, alors que la majorité des configurations ont une note supérieure à 50 pour la bande d'octave de 4 kHz par exemple. Ces observations confirment qu'il est difficile d'obtenir du contraste aux basses fréquence, mais certaines configurations sont toutefois plus favorables : en particulier les configurations proches de l'oreille sont celles qui obtiennent les meilleures notes. Les résultats obtenus pour le rapport de débit global sont très similaires à la bande d'octave de 125 Hz, car la pondération appliquée donne plus de poids aux basses fréquences : pour un système large-bande ce sont les performances aux basses fréquences qui sont dimensionnantes. Pour l'ensemble des bandes de fréquences considérées, les notes sont toujours supérieures à 70 pour les configurations à 20 cm et d'écartement angulaire compris entre +/-70° et +/-110°. Pour un système de reproduction large-bande, il est donc pertinent de placer les sources à cet emplacement du point de vue de l'efficacité des sources. Un système multivoies par bande de fréquences peut également être envisagé, comme cela a été proposé [Takeuchi & Nelson, 2002].

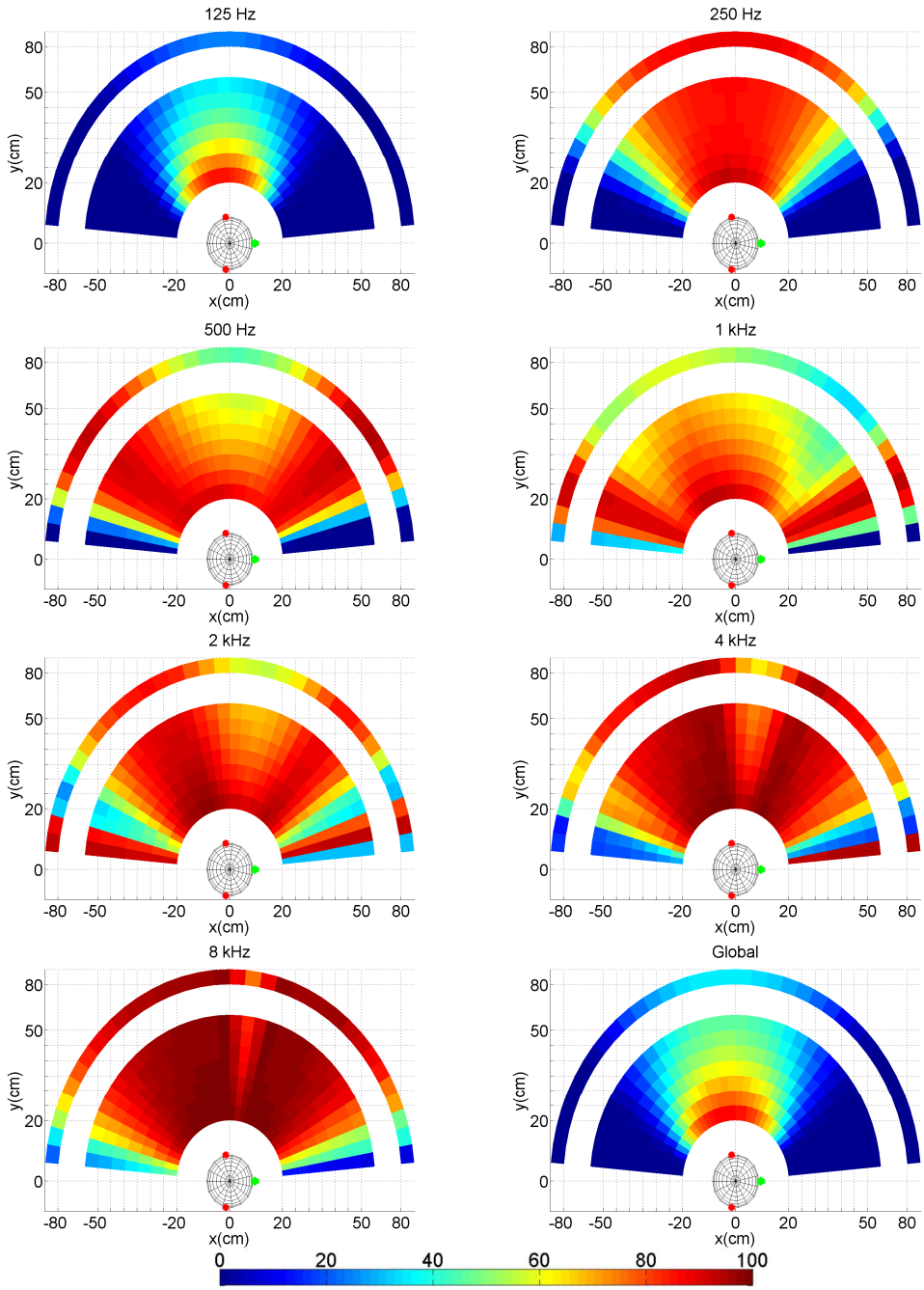


Figure 29 : notes attribuées par l'indicateur de dynamique dans le plan horizontal pour des bandes d'octave de fréquence centrale allant de 125 Hz à 8 kHz, ainsi que pour le cas « global » (moyenne pondérée)

III.5.B.Sources positionnées en élévation

Pour évaluer l'effet de l'élévation des sources physiques, seules les configurations à la distance de 40 cm sont évaluées. Ce choix est arbitraire, mais il a été montré que l'effet de distance est limité et nous avons par ailleurs constaté un effet similaire de l'élévation aux autres distances. Les résultats pour les bandes d'octave centrées sur 125 Hz et 4 kHz et le cas « global » représentatifs de l'ensemble des bandes de fréquences sont reportés dans la Figure 30.

Pour le cas global et la bande d'octave centrée sur 125 Hz, la note attribuée est systématiquement réduite pour les configurations avec élévation. Pour le cas global, les meilleures notes sont obtenues pour les configurations à élévation nulle et pour un écartement angulaire voisin de +/-

90°. A 4 kHz, l'effet de l'élévation est plus variable selon l'azimut : autour de 80° une faible élévation permet d'obtenir de meilleures notes, passant de 80 à élévation nulle à 95 pour une élévation de 15°. Pour les faibles écarts angulaires, les minima et maxima de note sont décalés en fonction de l'élévation : par exemple à élévation nulle le maxima est atteint pour l'azimut 5°, alors qu'il est obtenu pour l'azimut 10° à élévation 45°. Le placement des sources en élévation n'a donc pas une influence bénéfique sur le contraste naturel pour la majorité des cas.

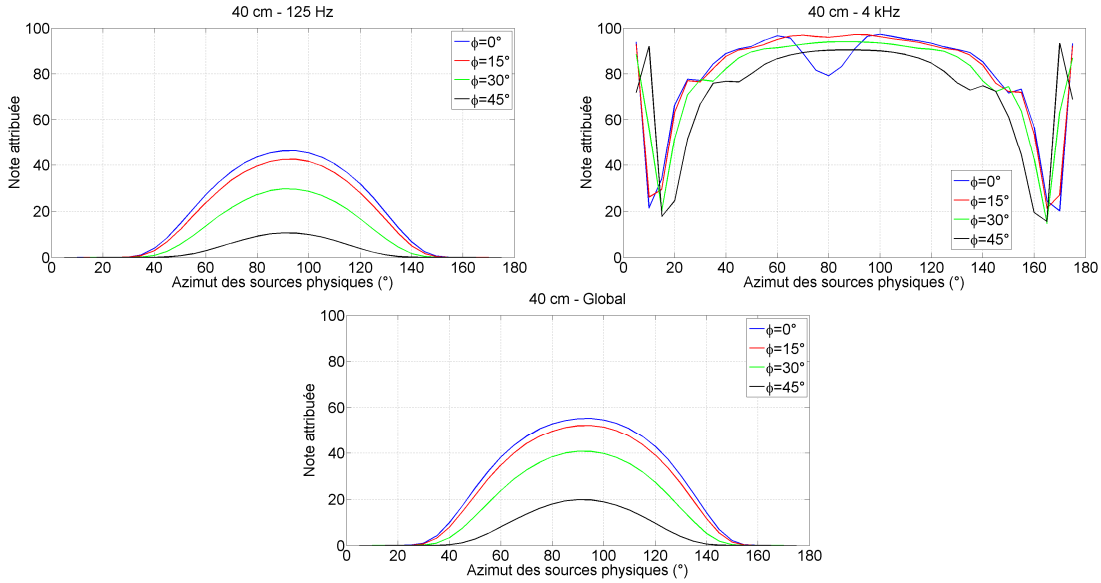


Figure 30 : notes attribuées pour plusieurs élévations en fonction de l'écartement azimutal des sources pour les bandes de fréquences 125 Hz, 4 kHz et cas global (moyenne pondérée)

III.6 Conclusion

Ce chapitre évalue l'impact des filtres transauraux sur le cahier des charges des sources électro-acoustiques. Cet impact est défini comme un surcoût lié à l'obtention du contraste, qui se traduit par une augmentation du débit des sources. Il est le plus élevé aux basses fréquences alors que c'est à ces fréquences que les sources électro-acoustiques sont les plus limitées : le coût lié à l'obtention du contraste est donc fondamental pour dimensionner un système de diffusion. Pour limiter ce coût, le placement des sources à proximité de l'oreille est la meilleure solution : le contraste naturel élevé y limite alors l'effort de filtrage alors que le niveau à reproduire est déjà plus faible.

Par ailleurs, le coût lié à l'obtention du contraste reflète des pics de rapport de débit aux hautes fréquences. Ces pics sont toutefois d'amplitude plus réduite que celle observée aux basses fréquences et sont liés à la géométrie particulière du système. Comme pour les basses fréquences, l'amplitude de ces pics diminue quand le contraste naturel augmente.

La démarche proposée pour évaluer l'influence de ce paramètre est la suivante : le rendu de système transauraux est simulé à partir d'un modèle de sphère, et une « note » de performance est définie. Ceci permet de représenter ce critère particulier sur une échelle de 0 à 100. D'autres aspects vont être étudiés dans les prochains chapitres à partir de cette démarche, dans le but d'en permettre la hiérarchisation.

Chapitre IV Influence de l'environnement acoustique

Table des matières

IV.1 Introduction

Pour une restitution sonore par haut-parleurs, la salle d'écoute joue un rôle important. Son influence se traduit par une coloration du rendu, et n'est donc pas souhaitée dans le cas d'une restitution de référence. L'égalisation d'un système audio dans une salle d'écoute est un problème complexe, qui est généralement un compromis entre la finesse de l'égalisation et l'étendue de la zone de validité de l'égalisation. L'égalisation de système audio est un domaine de recherche depuis plusieurs années, et plusieurs méthodes ont été proposées [Fielder, 2003; O. Kirkeby & Nelson, 1999; Mertins et al., 2010; Radlovik & Kennedy, 2000].

Dans ce chapitre, l'influence de la salle et les moyens de la limiter sont étudiés indépendamment des autres aspects de la diffusion sonore : la reproduction sonore est monophonique, et la diffraction par le corps de l'auditeur n'est pas prise en compte. Une partie des résultats de ce test ont par ailleurs été présentés en conférence [Vidal et al., 2016].

Pour limiter l'influence de la salle, deux approches sont explorées : la première consiste à appliquer une égalisation du système de diffusion. La seconde approche consiste à rapprocher les sources du point d'écoute, maximisant ainsi le rapport du champ direct sur le champ réverbéré. L'objectif de ce chapitre est de comparer ces deux approches en termes de performances et de complexité pour plusieurs salles de qualités différentes (taille et traitement).

Les performances sont évaluées de manière perceptive : un test d'écoute a été mis en place, où il est demandé à l'auditeur d'évaluer la similarité entre un son anéchoïque et un son reproduit en salle d'écoute. Les sons préalablement enregistrés ont été diffusés au casque d'écoute. Cela permet d'évaluer plusieurs paramètres sans en informer l'auditeur, minimisant les biais sur son jugement. Au total, cinq salles et deux distances d'écoute différentes ont été employées et quatre types d'égalisation ont été appliqués. La restitution des signaux avec égalisation a été réalisée au point d'écoute, ainsi qu'en un point légèrement décalé : ceci permet d'évaluer la robustesse des égalisations lorsque la position d'écoute n'est pas la même que pour l'égalisation.

Le principal objectif de ce chapitre est de bâtir un indicateur lié à l'effet de salle et son égalisation, permettant d'anticiper la perception du rendu sonore d'un système d'écoute dans une salle usuelle. L'idée est de pouvoir généraliser les résultats du test perceptif à d'autres configurations : par exemple d'autres distances ou d'autres salles d'écoute. Dans la suite du document, nous appellerons « configuration » la combinaison d'une salle d'écoute et d'une distance source-point de mesure.

IV.2 Protocole de test MUSHRA

Un son associé à une « configuration » correspond à son dégradé par un canal de transmission : la salle d'écoute. Cette dégradation est plus ou moins importante selon les caractéristiques de la salle et la distance source-micro. Cette dégradation est éventuellement limitée par l'utilisation d'un filtre d'égalisation.

Une configuration peut donc être assimilée à un filtre, qui se caractérise simplement par sa réponse impulsionnelle. Ce problème est similaire à la dégradation du signal par un CODEC audio. Pour ce type de problème, il existe un test spécifique permettant d'évaluer la « qualité » d'un tel système : il s'agit du test MUSHRA (MUltiple Stimuli with Hidden Reference and Anchor) [ITU-R BS.1534-1, 2001], [ITU-R BS.1534-3, 2015]. Dans la version initiale de ce test, il est demandé à l'auditeur d'évaluer la qualité de plusieurs sons présentés simultanément par rapport à un son de référence sur une échelle allant de « Excellente » à « Mauvaise », la meilleure qualité étant *a priori* le signal de référence (qui fait partie du panel de stimuli).

Comparativement aux tests de similarités, ce type de test présente l'avantage de pouvoir évaluer un grand nombre de sons en peu de temps. La principale contrainte associée est de ne pouvoir comparer simultanément que quelques stimuli. Un panel de grande taille doit donc être divisé en

« séries », et tous les sons ne sont alors pas présentés simultanément. Si les dégradations varient d'une série à l'autre, la notation des auditeurs peut évoluer, et la comparaison des résultats des différentes séries n'est pas immédiate. Pour assurer une continuité dans l'échelle de notation par l'auditeur entre les séries, des signaux communs à toutes les séries sont ajoutés, désignés par le terme « ancre ». D'après [ITU-R BS.1534-1, 2001], il est recommandé d'employer au moins deux ancres : une ancre haute et une ancre basse permettant de borner l'échelle des dégradations. Il est par ailleurs optionnellement proposé d'inclure des ancres intermédiaires supplémentaires, et d'après [ITU-R BS.1534-3, 2015] l'utilisation d'une ancre de qualité intermédiaire est clairement recommandée.

Dans notre implémentation, l'échelle de notation a été modifiée, et il est demandé à l'auditeur d'évaluer la « proximité » des sons à la référence. Le maximum de l'échelle (100) correspond au son « le plus proche » et le minimum de l'échelle (0) correspond au son « le plus différent ». Les auditeurs disposent d'un curseur associé à chaque son à évaluer, à positionner entre ces deux extrema. Il est demandé à l'auditeur de juger au moins un son à 0 et un autre à 100 et il lui est possible d'attribuer des notes identiques à plusieurs sons. Initialement, tous les curseurs sont positionnés au milieu de l'échelle, à la note 50. En préalable à l'expérience, une feuille de consignes présentant le test a été donnée à l'auditeur. Cette feuille de consigne est reportée en Annexe E. Suite à sa lecture, une discussion avec l'expérimentateur permettait éventuellement de clarifier certains points. L'interface est implémentée en Matlab, basée sur un programme disponible au téléchargement [Vincent, 2005]. Une interface de test MUSHRA est présentée en Annexe F.

L'élaboration de ce test a été guidée par la recommandation [ITU-R BS.1534-1, 2001], mais à ce jour une révision de cette recommandation a été publiée [ITU-R BS.1534-3, 2015] et certains points sont différents concernant notamment le nombre maximal de signaux et l'ancre intermédiaire.

Notre objectif pour ce test d'écoute est d'estimer l'importance de l'égalisation de la salle pour un auditeur moyen, sans *a priori* sur le processus de reproduction. Nous avons donc fait appel à des auditeurs non entraînés spécifiquement, sans critère de sélection particulier.

IV.3 Mesures en salles d'écoute

La première étape de la démarche consiste à rassembler des signaux à évaluer. Pour cela plusieurs mesures ont été réalisées et présentées ci-après.

IV.3.A. Source

Les enregistrements ont été réalisés avec une enceinte Tannoy System 600. Il s'agit d'une enceinte de monitoring coaxiale bass reflex, pour laquelle nous avons bouché les évents. En les bouchant, la réponse en basses fréquences est alors sensiblement modifiée, sans être aberrante. La même procédure a été employée dans des travaux antérieurs [Michaud, 2012]. Cette enceinte a été préalablement caractérisée dans la chambre anéchoïque de psychoacoustique du Laboratoire de Mécanique et d'Acoustique sur le site de Château-Gombert. Un sweep exponentiel entre 20 Hz et 20 kHz à la fréquence d'échantillonnage de 48 kHz a été utilisé, et la réponse impulsionnelle associée à cette mesure a été calculée par la méthode présentée dans [Novak et al., 2009].

Cette enceinte est égalisée à partir d'un filtre FIR à phase minimale calculé à partir de ces mesures, permettant alors d'obtenir une bande passante de [80 Hz – 13 kHz] à +/- 1 dB. Toutes les configurations sont établies à partir de la même source égalisée. Comme il s'agit d'écoutes comparatives, la réponse de la source n'a pas une influence significative sur le test d'écoute mais permet de maîtriser le contenu spectral des signaux testés. Par rapport au réglage d'un vrai système de diffusion, une partie importante du travail d'égalisation est donc déjà effectuée : l'objet de ce chapitre porte exclusivement sur l'effet de la salle. La caractérisation et l'égalisation de cette source sont détaillées dans l'Annexe G.

IV.3.B. Salles évaluées

Les caractéristiques des cinq salles d'écoute sont présentées dans le Tableau 3. Ce tableau mentionne également les acronymes pour chacune des salles, qui seront utilisés dans la suite du chapitre. Ces cinq salles sont de petites dimensions : la surface n'excède pas 20 m² ce qui est inférieur aux préconisation de l'ITU [ITU-R BS.1116-3, 2015] pour l'évaluation des systèmes de reproduction stéréophonique (qui préconise une surface comprise entre 20 m² et 60 m²). L'AES recommande également d'utiliser une salle d'écoute d'au moins 20 m² [AES, 1996]. Cependant, ces salles sont chères et non représentatives de locaux usuels : elles ne correspondent donc pas à notre base de travail.

Type	Salle de réunion	Studio d'écoute	Bureau moyen	Petit bureau	Cabine audiométrique
Nom	Reu	Stu	BuM	BuP	Cab
Surface au sol (m²)	19	18	16	7.5	4
Volume (m³)	47.5	45	40	18.8	10

Tableau 3 : propriétés des salles d'écoute

Les salles utilisées n'ont pas de traitement acoustique spécifique, à l'exception du studio d'écoute dont les parois sont recouvertes de mousse alvéolée de 5 cm, et la cabine audiométrique qui est également spécifiquement traitée. Les salles « Reu », « BuP » et « Cab » sont situées dans le même bâtiment, et les salles « Stu » et « BuM » sont situées dans un autre bâtiment. Dans ces deux dernières salles, le faux plafond est très absorbant. Une particularité de ces salles est que le plenum entre le faux plafond et le plafond communique avec d'autres locaux, ce qui contribue à la dispersion d'énergie aux basses fréquences et donc à en réduire le temps de réverbération. Ces salles ne sont donc pas « génériques » mais correspondent à une réalisation usuelle, et permettent de tester des configurations d'écoute variées.

Pour chacune des salles, le temps de réverbération estimé pour les bandes d'octaves allant de 125 Hz à 8 kHz est présenté à la Figure 31, qui comporte également le gabarit recommandé par l'ITU pour une salle de 20 m² et de hauteur sous plafond de 2.5 m.

Les salles d'écoutes ont des comportements acoustiques différents. Le studio et la cabine sont très mats : leur temps de réverbération est inférieur à 300 ms pour toutes les bandes de fréquences représentées. Il est de plus toujours inclus dans l'intervalle préconisé par l'ITU pour

une salle de 20 m², voire inférieur. Ces salles ne sont pas usuelles, mais leurs caractéristiques se rapprochent de la recommandation ITU : elles peuvent être considérées comme des salles de bonne qualité. La salle de réunion et le petit bureau sont les plus réverbérants, notamment en basses fréquences où le temps de réverbération est de l'ordre de la seconde. Ces salles correspondent à des locaux usuels dans les constructions modernes. Le bureau moyen a un gabarit de temps de réverbération assez atypique comparativement aux autres, il est plus élevé pour les hautes fréquences que les basses fréquences. Pour ces trois dernières salles, le temps de réverbération est toujours plus élevé que celui préconisé par l'ITU. Les salles évaluées sont donc assez hétérogènes : certaines sont très réverbérantes, tandis que d'autres ne le sont que légèrement.

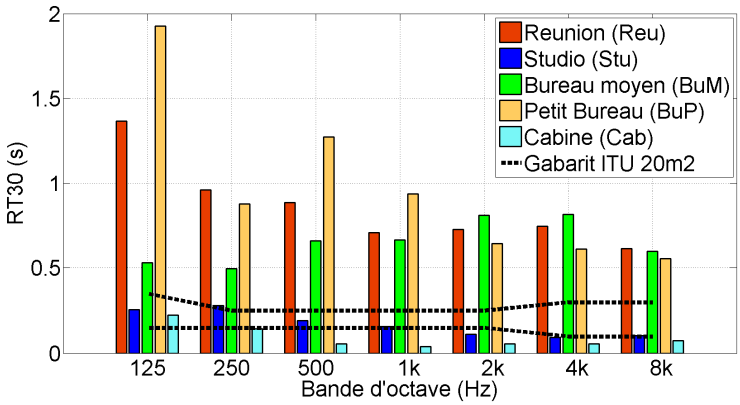


Figure 31 : Temps de réverbération des différentes salles d'écoutes

IV.3.C. Positions mesurées

Pour toutes les salles, la source et le microphone ont été positionnés sur la plus grande diagonale de la salle d'écoute, à une hauteur de 120 cm du sol. La source a été positionnée au tiers de cette diagonale, et le microphone à 40 cm et 80 cm de distance, sur cette même diagonale en direction de l'angle le plus éloigné. Ce placement correspond à un placement réaliste d'un système de diffusion : les sources et récepteurs sont éloignés des parois sans être au centre de la pièce où la signature modale peut être particulière. Pour la cabine, la mesure à 80 cm n'a pas été réalisée car ses dimensions ne s'y prêtent pas. Dans la suite du chapitre, la notation d'une configuration associant une salle d'écoute et une distance sera condensée par le nom de la salle (voir Tableau 3) et la distance. Par exemple, la configuration à 40 cm dans le studio est désignée par Stu40.

Pour toutes les configurations, une mesure a été systématiquement effectuée en décalant la position du microphone de 10 cm par rapport à sa position initiale, à distance à la source constante. L'orientation de la source n'est pas modifiée entre la mesure dans l'axe et décalée : la mesure décalée est donc influencée par la directivité de la source. Cette mesure permet d'évaluer l'effet d'un déplacement du point de mesure similaire au déplacement d'un auditeur.

IV.4 Méthodes d'égalisation

Quatre types d'égalisation ont été évalués. D'autres méthodes ont été implémentées, mais elles ont été éliminées lors d'écoutes informelles car elles introduisaient des artefacts excessifs.

IV.4.A. Egalisation IIR

Une égalisation de type IIR a été employée, et repose sur la publication de Ramos et Lopez [Ramos & Lopez, 2006]. Leur approche est de quantifier un défaut à partir d'une réponse en fréquence, défini par une aire « gain x largeur fréquentielle ». La représentation en Figure 32 d'une FRF à égaliser permet d'identifier ces défauts représentés en gris.

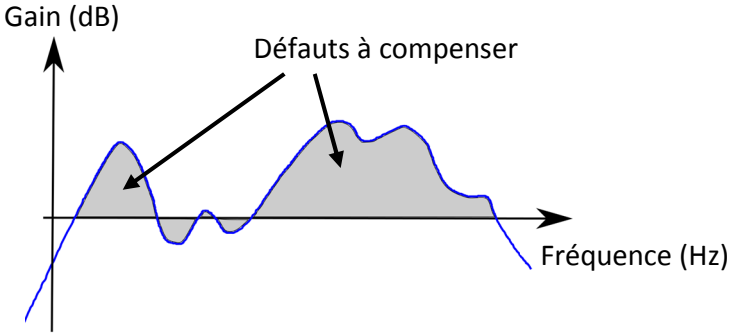


Figure 32 : représentation d'une FRF à égaliser et des aires « gain x largeur fréquentielles » grisées

Le principe consiste à mettre en cascade plusieurs filtres IIR d'ordre 2, chacun d'entre eux visant à corriger un défaut particulier de la réponse fréquentielle à égaliser. Ces filtres IIR sont des filtres biquadratiques (« biquads » en anglais) de type « peak » ou « notch ».

La procédure est itérative, en corrigeant les défauts par ordre d'importance : les plus grandes aires sont corrigées en premier. L'identification des défauts est réalisée sur une échelle logarithmique de fréquence par bandes de 1/48ème d'octave, la réponse fréquentielle à égaliser étant préalablement lissée par bandes de 1/12ème d'octave. Le filtre d'égalisation est composé d'au plus 31 biquads dans notre implémentation. Pour chaque défaut à corriger, les paramètres des biquads (fréquence centrale, amplitude et facteur de qualité) sont déterminés par un algorithme de Monte-Carlo. Dans notre implémentation, lorsque l'aire du plus grand défaut observé est inférieure à un certain seuil, l'algorithme est arrêté. Deux alternatives sont comparées ici, correspondant à deux valeurs du seuil d'arrêt : 0.1 et 0.5 dBoctave (1 dBoctave correspondant à un défaut de 1 dB sur une bande d'octave).

Le nombre de filtres permettant d'atteindre cet objectif est un indicateur de la quantité de correction appliquée : plus le nombre de biquads est élevé plus il corrige de petits défauts, mais plus le filtre d'égalisation est complexe. Le nombre de biquads utilisés en fonction de la configuration est représenté dans le Tableau 4. Une différence significative du nombre de biquads utilisés est visible selon les deux critères d'arrêt : pour le critère à 0.5 dBoctave le nombre de biquads ne dépasse pas 17, alors qu'il atteint le nombre maximal dans la majeure partie des cas à 0.1 dBoctave. Pour le critère à 0.5 dBoctave, le nombre de biquads est très faible dans certains cas : seulement 4 biquads sont utilisés pour les configurations Cab40 et Stu40. Les corrections apportées sont alors peu nombreuses, traduisant le fait que sans égalisation ces configurations ne modifient que légèrement le spectre d'amplitude.

La réalisation de ce filtrage pour la configuration BuM40 est présentée en Figure 33 avec un seuil fixé à 0.1 dBoctave et avec un seuil fixé à 0.5 dBoctave. La différence de complexité des filtres

s'observe sur cette figure : la FRF inverse est plus lisse dans le second cas, corrigeant moins la FRF initiale. Dans le cas présenté, la reconstruction hors axe est très similaire à la reconstruction dans l'axe jusqu'à 1kHz. Un écart entre les deux de l'ordre de 2 dB apparait autour de 1 kHz, et de l'ordre de 3 dB entre 7 kHz et 9 kHz. L'égalisation est donc ici plutôt robuste au déplacement.

	Reu40	Reu80	Stu40	Stu80	BuM40	BuM80	BuP40	BuP80	Cab40
0.5 dBoctave	7	9	4	6	5	11	11	17	4
0.1 dBoctave	20	31	25	31	31	31	31	24	29

Tableau 4 : nombre de biquads utilisés

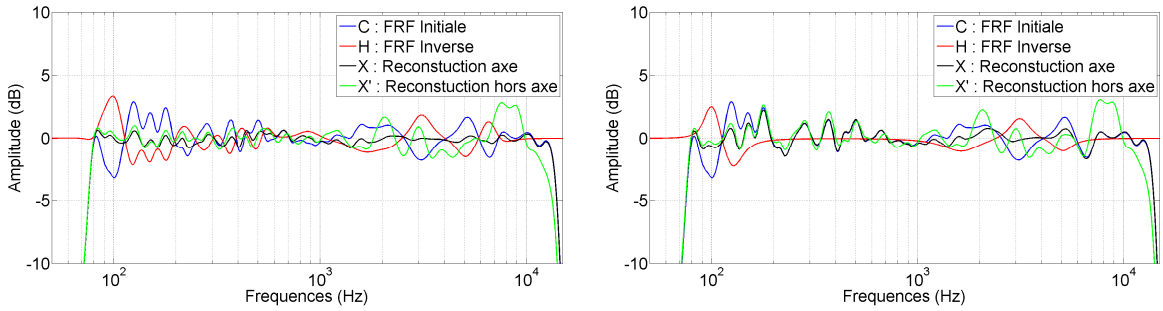


Figure 33 : filtre d'égalisation pour le bureau moyen à 40 cm avec la méthode IIR. Figure de gauche : seuil à 0.1 dBoctave, Figure de droite : seuil à 0.5 dBoctave

IV.4.B. Egalisation FIR à phase minimale

Une autre méthode est basée sur un filtrage FIR à phase minimale. Pour cette méthode, les gains en bandes fines de la cible et de la FRF à égaliser sont préalablement lissés en 1/6ème d'octave. Un gabarit fréquentiel de filtre inverse est déterminé à partir du rapport entre les gains lissés de la cible et de la réponse à égaliser. On en déduit un filtre FIR inverse à phase linéaire en appliquant la méthode de la fenêtre [Mitra, 1998]. L'excès de phase est compensé, de telle sorte que le filtre soit à phase minimale ([Oppenheim & Schafer, 1975], fonction *rceps* de Matlab). Le filtre FIR ainsi obtenu est d'ordre 2048 (soit un support temporel de 43 ms pour un échantillonnage à 48 kHz).

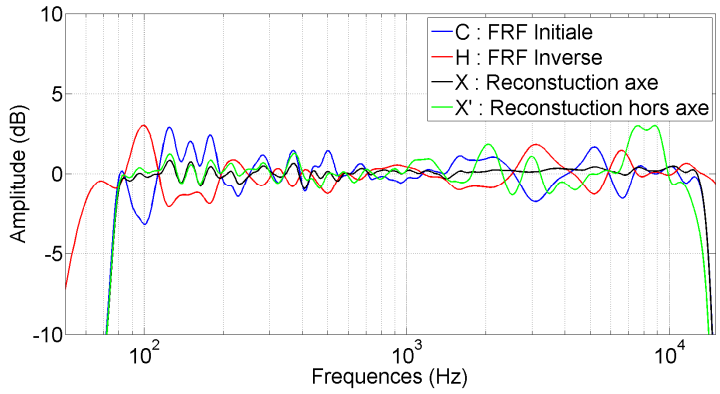


Figure 34 : filtre d'égalisation pour le bureau moyen à 40 cm avec la méthode FIR à phase minimale.

La réalisation de ce filtrage, toujours pour la configuration BuM40, est présentée en Figure 34. Les allures des FRF obtenues sont très similaires à celles obtenues avec la méthode IIR 0.1 dBoctave

(Figure 33). La méthode pour obtenir ce résultat est différente, mais les performances semblent similaires.

IV.4.C. Egalisation FIR gain & phase

Enfin, la dernière méthode repose sur le calcul d'un filtre inverse avec régularisation de Tikhonov, comme explicité dans [O. Kirkeby & Nelson, 1999] et implémenté dans le domaine temporel. Le filtre obtenu prend alors en compte les fluctuations de phase. Le filtre inverse h est déterminé de la manière suivante :

$$h = [C^T C + \beta B^T B]^{-1}.C^T a_m \qquad (42)$$

Avec C la matrice de covariance de forme Toeplitz déterminée à partir de la réponse impulsionnelle à égaliser, préalablement tronquée à la demi-longueur du filtre et apodisée sur la seconde moitié en utilisant une demi-fenêtre de Hann. De plus, elle est complétée avec des échantillons nuls de telle sorte que sa durée soit celle du filtre. a_m est un vecteur correspondant à un retard pur de la demi-longueur du filtre. La longueur du filtre est différente selon les cas, et déterminée en lien avec β, le paramètre de régularisation. Il s'agit d'un scalaire dont la valeur est définie par approches successives dans le cadre d'écoutes informelles, en tentant de maximiser l'égalisation tout en limitant les artéfacts liés à l'égalisation. Ces artefacts sont principalement de type « pré-échos », correspondant à l'arrivée d'un signal audible précédant le champ direct. La longueur du filtre est également adaptée par écoutes successives. Les valeurs de ces deux paramètres pour chacune des configurations sont données dans le Tableau 5.

	Reu40	Reu80	Stu40	Stu80	BuM40	BuM80	BuP40	BuP80	Cab40
β	10^{-5}	10^{-2}	10^{-5}	10^{-3}	10^{-5}	3.10^{-2}	10^{-5}	5.10^{-2}	10^{-5}
Longueur de filtre (ms)	200	133	200	200	200	200	200	133	200

Tableau 5 : paramètres des filtres FIR gain & phase

Compte tenu de la taille maximale des matrices pouvant être inversées sur notre plateforme (Matlab 32 bits et 2 Go de RAM allouée), l'ordre maximum des filtres qui peuvent être calculés par cette méthode est limité à 6000. Pour maximiser la durée de réponse impulsionnelle prise en compte dans l'égalisation, celle-ci a été sous-échantillonnée à 30 kHz avant le calcul du filtre. La longueur maximale du filtre est ainsi de 200 ms, qui est une durée encore trop faible pour pouvoir traiter la totalité de la réverbération (le temps de réverbération pouvant être de l'ordre de la seconde dans certaines salles, notamment en basses fréquences : voir Figure 31). Une piste d'amélioration de cet algorithme serait de calculer un filtre spécifique à chaque bande d'octave : dans la majeure partie des cas, les bandes de fréquences basses ont les temps de réverbération les plus longs et il serait possible de les traiter avec une fréquence d'échantillonnage plus basse. Cette piste n'a pas été développée, car l'objectif de ces travaux de thèse ne concerne pas spécifiquement l'égalisation d'un système SISO, mais l'optimisation d'un système MIMO selon plusieurs critères.

Les FRF obtenues pour la même configuration que pour les cas précédents (BuM40) sont présentées dans la Figure 35. L'allure de la FRF inverse est un peu plus complexe, et la FRF reconstruite dans l'axe est très régulière. Seule une variation résiduelle de l'ordre de 1 dB est visible autour de 230 Hz. La reconstruction hors axe suit la même tendance que pour les autres égalisations : jusqu'à 1 kHz elle est très proche de la reconstruction dans l'axe, puis des écarts significatifs sont visibles au-delà.

Cette méthode d'égalisation est la seule qui tente de corriger la phase de la réponse impulsionnelle. C'est donc la seule méthode étudiée ici qui pourrait compenser finement la réponse de la salle, elle se distingue donc particulièrement des autres en termes d'objectifs d'égalisation.

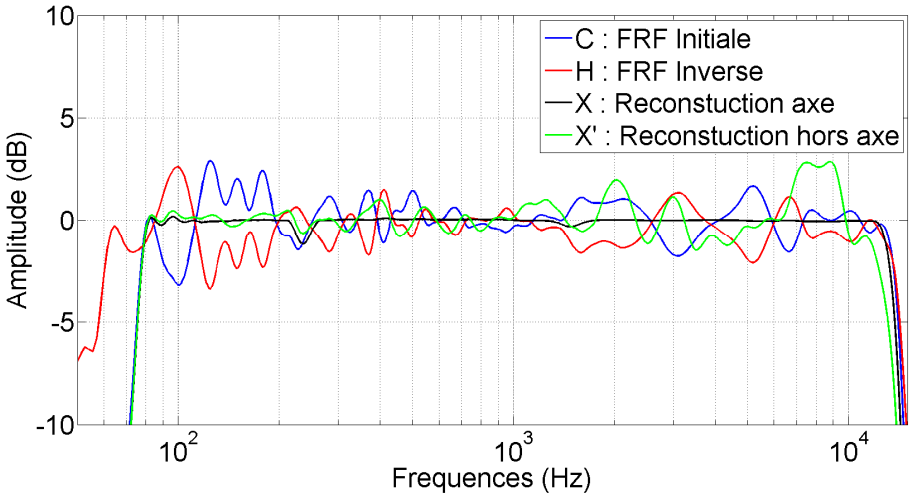


Figure 35 : filtre d'égalisation pour le bureau moyen à 40 cm avec la méthode FIR gain & phase.

IV.5 Evaluation perceptive

Neuf configurations et deux points d'écoute ont été envisagés. Pour ces dix-huit cas, il y a cinq réalisations possibles (quatre égalisations et l'absence d'égalisation). En incluant le signal de référence 91 sons doivent donc être évalués.

IV.5.A. Organisation du test

Le protocole de type MUSHRA consiste à comparer simultanément une série de sons à une référence. Ici, la référence est le signal anéchoïque mesuré à 40 cm dans l'axe. Chaque série correspond à une « configuration » : l'auditeur n'évalue qu'une seule salle et distance à la fois. Il y a neuf configurations différentes, évaluées via neuf tests. Pour chacun de ces neuf tests, il y a alors treize stimuli à évaluer :

- Trois ancres
    - Une ancre haute, qui correspond à la référence cachée
    - Une ancre basse, qui dans le cadre d'évaluation de CODEC correspond à la référence filtrée passe-bas. Dans le cadre de nos travaux où il s'agit d'évaluer une dégradation liée à un effet de salle, le filtrage passe-bas n'est pas représentatif des autres dégradations : un signal avec effet de salle particulièrement important est

préféré. Il s'agit du signal non égalisé mesuré dans le petit bureau à 80 cm avec le récepteur décalé de 20 cm par rapport à la configuration dans l'axe.
    - Une ancre intermédiaire qui est choisie comme le signal non égalisé dans le bureau moyen, à 80 cm de distance.
- Huit stimuli correspondant aux quatre types d'égalisation, pour laquelle la reconstruction a été effectuée au point initial, ainsi qu'au point décalé.
- Deux stimuli non égalisés : dans l'axe et décalé

D'après [ITU-R BS.1534-1, 2001], il est recommandé de ne pas évaluer plus de quinze signaux à la fois, notre expérience satisfait ce critère. En revanche, dans la version révisée de 2015 [ITU-R BS.1534-3, 2015] il est recommandé de ne pas évaluer plus de douze signaux. A l'issue du test, certains auditeurs ont effectivement signalé avoir éprouvé de la difficulté à évaluer un panel aussi important.

Pour choisir un stimulus, des signaux artificiels et des extraits d'enregistrement musicaux et vocaux ont été comparés dans le cadre d'écoutes informelles. Pour différencier les méthodes d'égalisation, la plus discriminante s'est avérée être une salve de quatre bouffées de bruit rose, chaque bouffée durant 200 ms avec 100 ms de silence. Ce signal a de nombreux avantages :

- Il est large-bande et varie au cours du temps, ce qui lui permet d'être discriminant.
- Ce stimulus est de courte durée, de l'ordre de 1 seconde lorsqu'il n'y a pas de réverbération. Cela permet à l'auditeur de comparer rapidement différentes configurations et limiter ainsi l'effort de mémorisation.
- Les bouffées de bruit large-bande ont par ailleurs été identifiées comme discriminantes pour évaluer perceptivement la localisation de sources [Santala et al., 2014].

Idéalement, il faudrait réaliser le test avec plusieurs types de stimuli, afin de quantifier l'effet du contenu sonore. L'évaluation a plutôt été focalisée sur un grand nombre de paramètres de systèmes, en supposant que les tendances soient similaires pour d'autres signaux. L'approche générale consiste à hiérarchiser des systèmes de manière préliminaire, dans l'optique d'en sélectionner quelques-uns et de les évaluer de manière plus approfondie.

La Figure 36 représente un schéma résumant la réalisation des différents stimuli. Tous les stimuli sont égalisés en niveau en s'appuyant sur estimation de la sonie définie dans la norme ISO 532B [ISO 532B, 1975].

Les signaux sonores sont diffusés à l'aide d'un casque Beyerdynamic DT990 Pro et d'une carte audionumérique ESI U24 XL à un niveau de 30 sones. Une phase d'entrainement a précédé le test afin de familiariser l'auditeur avec l'interface et les sons à évaluer. Cet entrainement se présente de la même manière que pour les 9 configurations à évaluer. Les sons associés correspondent aux ancres haute et basse, aux 9 configurations non égalisées dans l'axe, ainsi que les configurations Reu80 et BuM80 égalisées avec le FIR gain et phase. Les résultats obtenus lors de cet entrainement ne sont pas analysés.

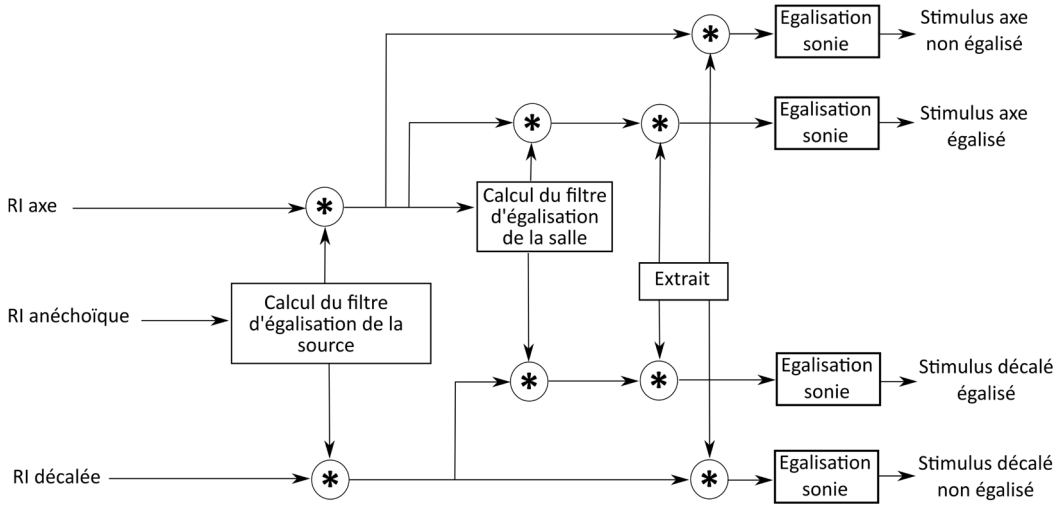


Figure 36 : Schéma de la confection des stimuli. * représente l'opération de convolution

IV.5.B. Analyse préliminaire des résultats

Vingt-deux auditeurs ont participé au test d'écoute : leur audition n'a pas été vérifiée mais aucun n'a reporté de problèmes d'écoute. La fiabilité de leur réponse a pu être contrôlée indirectement via la discrimination de la référence cachée, conformément à la recommandations [9] : il est recommandé d'écarter les réponses d'un auditeur qui aurait attribué à la référence cachée une note inférieure à 90 pour plus de 15 % des tests. Pour cette expérience, cela correspond à ne pas identifier la référence cachée dans au moins deux conditions, et cette procédure a été appliquée. Un seul auditeur se trouvait dans cette situation, ses réponses ne sont donc pas prises en compte dans la suite de l'analyse.

La variabilité des réponses entre les auditeurs est étudiée en calculant le coefficient de corrélation de Bravais-Pearson C_{XY} des réponses de chaque paire d'auditeur : $C_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$. X et Y correspondent aux vecteurs de réponses de deux auditeurs, σ_X et σ_Y les écarts-type associés et Cov désigne l'opération de covariance Il ressort de cette analyse que les réponses de tous les auditeurs sont plutôt corrélées ($C_{XY} > 0.6$), : cette analyse ne permet donc pas d'identifier un auditeur dont les réponses seraient significativement différentes des autres.

IV.5.C. Analyse statistique

Pour ce type d'expérience, une analyse de la variance (ANOVA) est généralement envisagée. Pour un tel test, plusieurs hypothèses concernant les données doivent être respectées. Notamment, leurs distributions doivent être normales et les variances doivent être homogènes. Pour le quantifier, un test de Jarque-Bera est employé. Ce test compare les moments d'ordre 2 et 3 (skewness et kurtosis) de la distribution testée à ceux d'une distribution normale. Au total, seulement 17 % des notes ont une distribution proche d'une distribution normale.

Bien que la normalité des distributions ne soit clairement pas respectée pour l'ensemble des signaux, une ANOVA est appliquée sur les données à titre d'information, en excluant les ancres. Sont considérés les facteurs « configuration », « égalisation » et « décalage ». Le facteur « auditeur » est considéré comme une variable aléatoire. L'ANOVA est menée avec la fonction *anovan* de la Statistics Toolbox de Matlab, et les résultats sont reportés dans le Tableau 6.

Facteur	SCE	D.L.	F	p>F
Auditeur	41646.1	20	1.00	0.46
Configuration	1620538.5	8	107.37	<0.01
Egalisation	29221.9	4	39.75	<0.01
Décalage	14634.2	1	55.09	<0.01
Auditeur*Configuration	301854.2	160	14.60	<0.01
Auditeur*Egalisation	14704.7	80	1.42	<0.01
Auditeur*Décalage	5313.3	20	2.06	<0.01
Configuration*Egalisation	21197.2	32	5.13	<0.01
Configuration* Décalage	19850.1	8	19.2	<0.01
Egalisation*Décalage	2341.2	4	4.53	<0.01

Tableau 6 : Résultats de l'ANOVA. SCE signifie "Somme des Carrés des Ecarts", DL signifie Degrés de Liberté, F est la statistique du test de Fisher et p la probabilité associée

Cette ANOVA révèlerait de la significativité pour tous les facteurs (à l'exception du facteur auditeur) et de toutes les interactions au seuil de signification de 0.01. Cela signifie que tous les facteurs seraient évalués différemment, et leurs interactions deux par deux auraient une influence significative. Ceci n'est pas un résultat effectif puisque l'hypothèse de normalité n'est pas validée. Aucun test post-hoc n'est donc mené, les données sont simplement représentées graphiquement dans les prochaines sections.

IV.5.D.Regroupement des configurations

Les signaux ont été évalués configuration par configuration. Une tentative de regrouper les notes ainsi attribuées est réalisée. Pour cela, les notes médianes attribuées aux ancres et au signal non égalisé dans l'axe pour chacune des configurations sont préalablement représentées à la Figure 37. Les ancres haute et basse ont été bien identifiées car les notes médianes sont respectivement de 100 et 0 pour la majorité des configurations. Un cas particulier est visible pour la configuration BuP80 : les notes attribuées à l'ancre basse sont très similaires à celles attribuées au signal non égalisé. Ces deux signaux ont été élaborés à partir de réponses impulsionnelles issues de la même salle d'écoute à la même distance, et seul le désaxage diffère dans les deux cas. L'influence du désaxage n'est donc pas significative dans ce cas, et les deux signaux peuvent être pris comme « pire cas ».

L'évaluation de l'ancre intermédiaire évolue assez fortement selon la configuration, alors qu'il s'agit toujours du même son. Cela est visible par les variations de médiane (variant entre 15 et 50) et des écarts interquartiles très différents selon les configurations (allant de 10 à 50). En particulier, la note médiane attribuée au signal non égalisé pour la configuration Reu80 est plus élevée que la note attribuée à l'ancre intermédiaire, alors que c'est l'inverse dans le cas de la

configuration Reu40. Or, la note médiane attribuée au signal non égalisé de la configuration Reu80 est légèrement plus élevée que celle de la configuration Reu40 : cette note ne traduit donc pas à elle seule la comparaison avec l'ancre intermédiaire. L'évaluation des signaux semble ainsi dépendre du « contexte » de présentation.

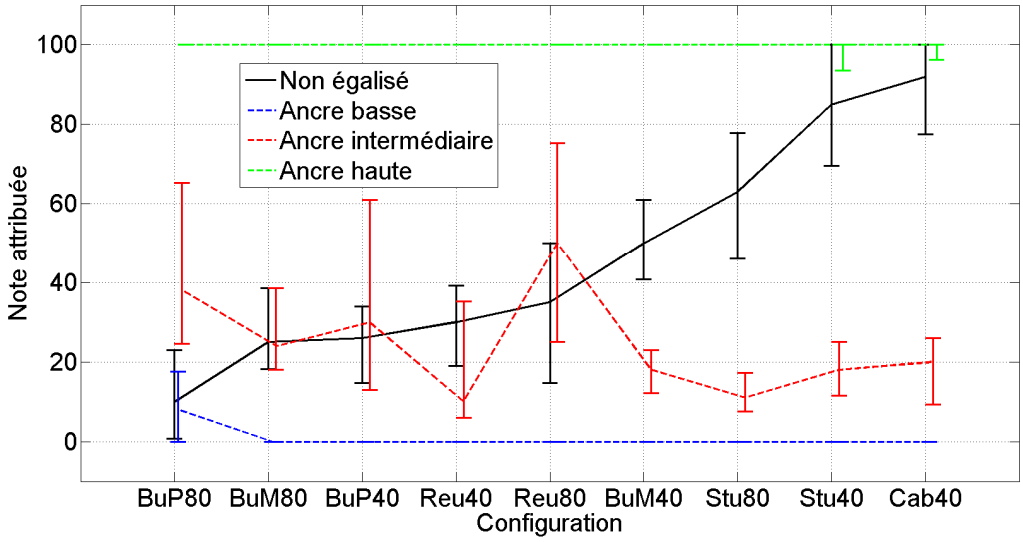


Figure 37 : notes médianes attribuées aux ancres et au signal non égalisé dans l'axe. Les barres d'erreurs correspondent aux écarts interquartiles

Il est toutefois souhaitable de comparer toutes les configurations entre elles. Pour cela un post-traitement a été appliqué à l'ensemble des résultats en utilisant les notes attribuées aux trois ancres. Le post-traitement correspond à la procédure suivante :

1) Détermination de la note de l'ancre intermédiaire en prenant la moyenne de la note attribuée à ce signal pour toutes les configurations et tous les auditeurs. Cette note vaut alors 27.
2) Pour chaque auditeur et chaque configuration, une normalisation a été appliquée de sorte que les notes attribuées aux ancres haute, intermédiaire et basse soient respectivement 100, 27 et 0. La normalisation est effectuée en deux segments : le premier concerne les notes comprises entre l'ancre intermédiaire et l'ancre basse et le second les notes comprises entre l'ancre intermédiaire et l'ancre haute. Par ce processus, il est possible que des notes soient négatives (si l'ancre basse n'a pas été évaluée à 0) ou supérieures à 100 (si l'ancre haute n'a pas été évaluée à 100).

Dans un petit nombre de tests, l'ancre basse a obtenu une note plus élevée que l'ancre intermédiaire. Les deux notes ont alors été mises à 0, et la normalisation est effectuée en un seul segment. Parmi toutes les réalisations possibles (21 auditeurs x 9 configurations = 189 réalisations), seulement 6 étaient dans ce cas (soit 3.2 % des résultats).

Les notes post-traitées sont représentées en Figure 38. Le classement des configurations pour le signal non égalisé y est conservé, à l'exception des configurations Reu40 et Reu80 qui se trouvent ainsi permutées.

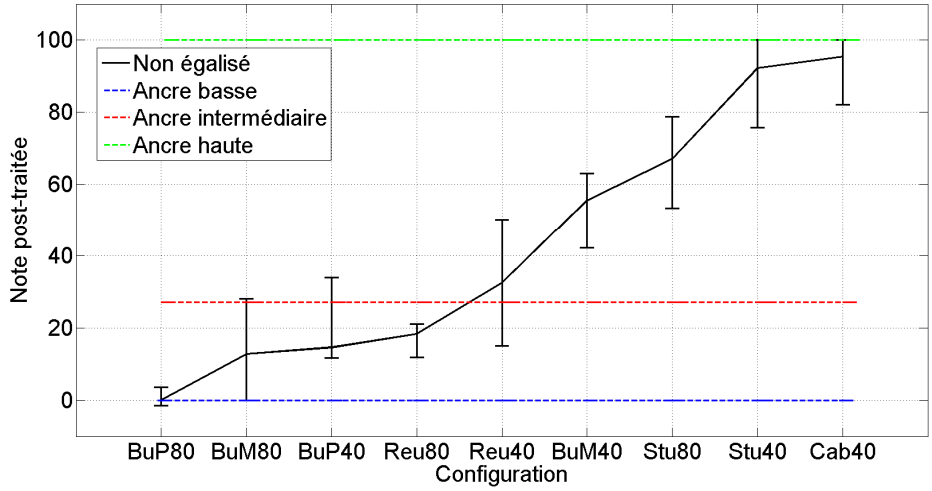


Figure 38 : notes médianes post-traitées des ancres et du signal non égalisé dans l'axe. Les barres d'erreurs correspondent aux écarts interquartiles

IV.6 Analyse descriptive

L'analyse proposée ci-après est basée sur la représentation graphique des notes post-traitées. Premièrement, l'effet de la configuration (distance et salle) est étudié, puis l'effet de l'égalisation et enfin l'effet du désaxage.

IV.6.A.a Effet de la configuration

La Figure 38 regroupe l'évaluation des signaux non égalisés pour les différentes configurations. Les sons associés à différentes configurations n'ont pas été évalués simultanément, mais les résultats sont toutefois regroupés à partir de l'évaluation des ancres. La portée de cette analyse est donc limitée, mais l'effet de la configuration semble cependant significatif. Les salles les plus traitées acoustiquement (Studio et Cabine) obtiendraient ainsi les meilleures notes. L'effet de la distance est également important, il est le plus visible pour la salle Bureau Moyen : la configuration BuM40 obtient une note médiane de 55 tandis que la configuration BuM80 obtient une note médiane de 15. Dans le premier cas la configuration fait partie des meilleures et dans le second cas elle fait partie des moins bonnes

IV.6.A.b Effet de l'égalisation

L'effet de l'égalisation peut s'observer sur la Figure 39, où les notes obtenues pour les différentes égalisations sont représentées en fonction de la configuration. Visiblement, l'effet de l'égalisation est moindre : pour la majorité des configurations les notes sont similaires quelle que soit la méthode d'égalisation. Dans certain cas, les notes obtenues avec l'égalisation FIR gain & phase sont plus basses que les notes obtenues sans égalisation : ces configurations sont Bup80, Reu80, BuM40 et Stu80. Une telle égalisation peut donc avoir un effet contre-productif, probablement à cause d'artefacts audibles. Il est cependant impossible de généraliser ce résultat car le filtrage réalisé ne prend pas en compte l'intégralité du support temporel de la réponse impulsionnelle.

Pour les configurations qui sont déjà très bonnes sans égalisation (Stu40 et Cab40), les notes après égalisations sont excellentes pour toutes les égalisations. L'égalisation semble donc avoir un intérêt lorsque les modifications à apporter sont mineures. La configuration BuM80 est toutefois

un cas particulier de configuration évaluée dans le bas de l'échelle pour laquelle les égalisations à phase minimale permettent néanmoins d'augmenter légèrement la note.

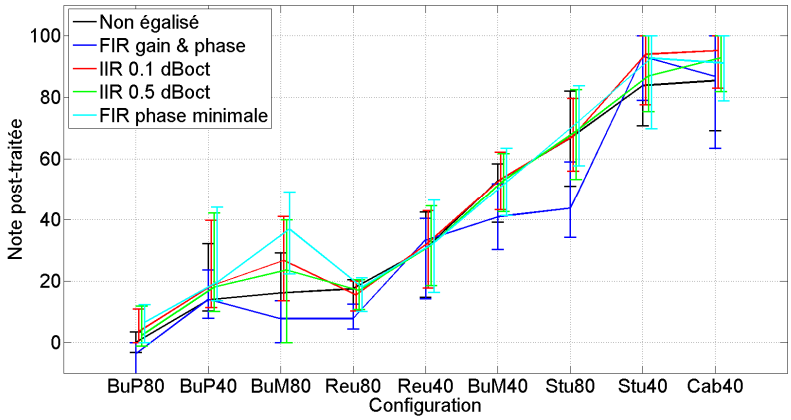


Figure 39 : notes médianes post-traitées du signal non égalisé et des égalisations dans l'axe. Les barres d'erreurs correspondent aux écarts interquartiles

Pour toutes les configurations, les égalisations à phase minimale sont très similaires. Cette observation est intéressante puisque les coûts de calcul ne sont pas les mêmes dans les deux cas : les filtres IIR ont été implémentés avec un nombre bien moindre de coefficients. Pour un résultat équivalent, il semble donc plus avantageux ici d'employer une égalisation IIR.

IV.6.A.c Effet du désaxage

Dans les principales configurations pour lesquelles l'égalisation semble fonctionner (BuM80, Stu40 et Cab40), l'effet du déplacement du microphone entre la phase d'égalisation et celle de restitution est abordé. Les notes attribuées pour ces configurations dans l'axe et décalées sont représentées dans la Figure 40 pour trois situations : sans égalisation, avec l'égalisation FIR gain & phase et avec l'égalisation FIR à phase minimale.

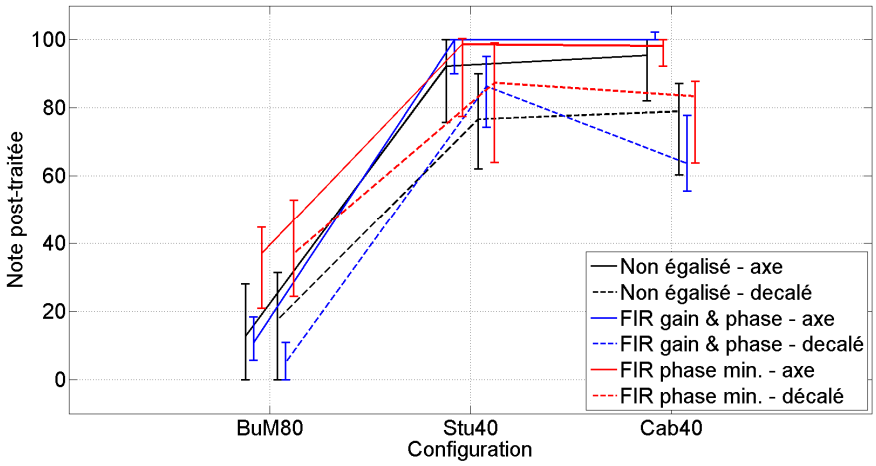


Figure 40 : notes médianes post-traitées du signal non égalisé et des égalisations dans l'axe et hors axe. Les barres d'erreurs correspondent aux écarts interquartiles

Sans égalisation, les notes des configurations Stu40 et Cab40 sont légèrement plus basses lorsque le microphone est décalé.

Avec une égalisation FIR à phase minimale, une fois désaxé le rendu est au moins aussi bon que sans égalisation, et dans certains cas il peut être meilleur. Cela traduit une certaine robustesse de ce filtrage.

Avec une égalisation FIR gain & phase, une fois désaxé le rendu est au mieux aussi bien noté que sans égalisation, et dans certains cas il peut être pire. Cela traduit un manque de robustesse de ce filtrage.

IV.7 Objectivation des résultats

Cette partie a pour but de représenter les résultats du test perceptif sous forme d'indicateurs objectifs. D'après les résultats obtenus, l'effet de salle est clairement dominant, alors que l'effet des égalisations testées semble mineur. Pour l'objectivation, nous nous focalisons donc sur l'effet de salle, et écartons l'égalisation. Seules les notes obtenues pour les signaux non égalisés et dans l'axe (post-traitées) sont donc utilisées.

IV.7.A. Calculs d'indicateurs objectifs

Beaucoup d'indicateurs ont été proposés pour évaluer des salles, et ils sont généralement basés sur l'analyse d'une réponse impulsionnelle. Une partie d'entre eux sont présentés, puis des indicateurs basés sur les paramètres de la configuration (salle et distance) sont proposés.

IV.7.A.a Indicateurs obtenus à partir de la RI

Les indicateurs obtenus à partir de la RI visent à estimer la « quantité de réverbération » à partir de la réponse impulsionnelle $ri(t)$. La clarté C_X est définie de la manière suivante :

$$C_X = 10\log_{10}\left(\frac{\int_{t=0}^{X} ri^2(t)dt}{\int_{t=X}^{N} ri^2(t)dt}\right) \tag{43}$$

N est la longueur de la réponse impulsionnelle, et X le paramètre de durée d'intégration de clarté pour lequel les valeurs classiques de 50 ms et 80 m sont testées. De plus, une durée d'intégration plus courte de 20 ms est également testée, une analyse préliminaire ayant montré sa pertinence.

La définition D_X est calculée de la manière suivante :

$$D_X = 100.\frac{\int_{t=0}^{X} ri^2(t)dt}{\int_{t=0}^{N} ri^2(t)dt} \tag{44}$$

De la même manière que la clarté, X est le paramètre de durée d'intégration. Une durée d'intégration de 50 ms est classiquement utilisée, et la longueur de 20 ms est également testée par analogie avec la C_{20}.

Le temps central T_c est un indicateur qui correspond au barycentre temporel de l'énergie de la réponse impulsionnelle :

$$T_c = \frac{\int_{t=0}^{N} t.ri^2(t)dt}{\int_{t=0}^{N} ri^2(t)dt} \tag{45}$$

IV.7.A.b Indicateurs obtenus à partir du TR

Une alternative à l'utilisation d'une réponse impulsionnelle est de modéliser l'effet de salle par un champ diffus, caractérisé par une pression uniforme à phase aléatoire en tout point de la salle. En l'absence de données précises, cette pression P_{diffus} peut être estimée à partir d'une hypothèse de rayonnement monopolaire de la source et d'une absorption de l'énergie aux parois [Gade, 2007] :

$$P_{diffus} = \sqrt{\frac{16\pi r^2 P_{anech}^2}{A}} \tag{46}$$

r est à la distance de mesure (40 cm ou 80 cm selon les cas). A correspond à l'aire équivalente d'absorption estimée en supposant que l'absorption aux parois peut se modéliser par une surface complètement absorbante [Gade, 2007], ce qui conduit à la relation :

$$A = \frac{kV}{TR_{60}} \tag{47}$$

Avec $k = 0.16$, V le volume de la salle d'écoute et TR_{60} le temps de réverbération. Les données de temps de réverbération par bandes d'octave présentées dans la Figure 31 sont utilisées.

Une grandeur analogue à une clarté $Rapport_{direct/diffus}$ peut ainsi être estimée :

$$Rapport_{direct/diffus}\,(f) = \left|20\log\left(\frac{|P_{anech}(f)|}{|P_{diffus}(f)|}\right)\right| \tag{48}$$

L'indicateur associé $R_{D/D}$ correspond alors à la moyenne des N_{tot} bandes d'octave centrées de 125 Hz à 8 kHz (indicées par N_f) :

$$R_{D/D} = \frac{1}{N_{tot}}\sum_{N_f=1}^{N_{tot}} Rapport_{direct/diffus}\,(N_f) \tag{49}$$

Une alternative à l'indicateur précédent est proposée, en utilisant un modèle de réverbération, pour se ramener à un calcul similaire à celui d'une clarté (équation (43)). Schématiquement, la modélisation de l'énergie au cours du temps est supposée évoluer de la manière suivante :

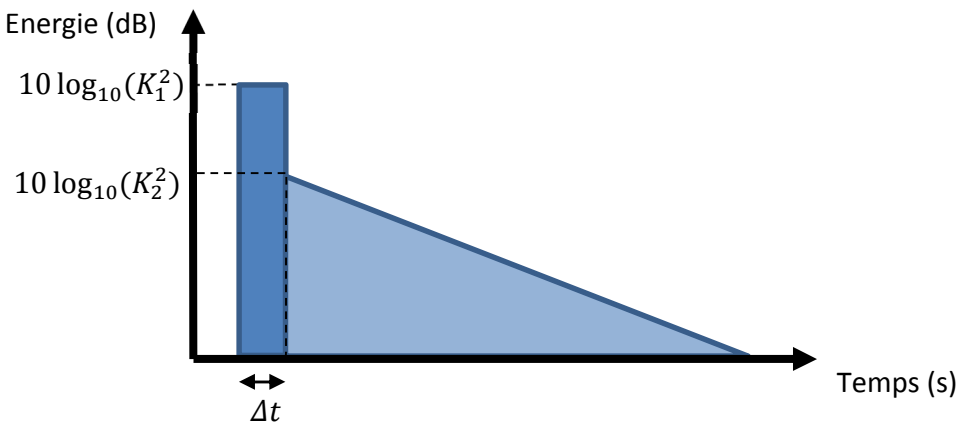


Figure 41 : modélisation de réponse impulsionnelle en salle

Le temps de réverbération est le temps pour lequel l'énergie a diminué de 60 dB, ce qui permet de déterminer la pente de la partie réverbérée. La réponse impulsionnelle est donc modélisée de la manière suivante :

$$\begin{cases} ri^2(t) = \dfrac{E_{direct}}{\Delta T} & pour\ t \leq \Delta T \\ ri^2(t) = K_2 e^{\frac{-13.8}{TR_{60}}t} & pour\ t > \Delta T \end{cases} \tag{50}$$

Δt est le support temporel du champ direct, et l'énergie de la contribution directe E_{direct} est définie de la manière suivante :

$$E_{direct} = K_1 \Delta t = 1 \tag{51}$$

Le paramètre ΔT est fixé de sorte qu'il soit très inférieur à 20 ms : sa valeur n'a pas une influence significative sur l'indicateur estimé. Une valeur de 0.4 ms est choisie arbitrairement, et permet alors de déduire K_1. L'énergie de champ diffus $E_{diffus} = P_{dif}^2$ est déterminée à partir de l'équation (46). Il vient alors :

$$K_2 = \frac{E_{diffus}}{\sum_{t=0}^{T_{max}} \left(\exp\left(\frac{-13.8}{TR_{60}}\right) t\right)} \tag{52}$$

Avec T_{max} la durée de la réponse impulsionnelle modélisée, fixé à $1.5TR_{60}$. Une clarté est alors calculée avec une durée d'intégration de 20 ms de la manière suivante :

$$C_{modele} = 10 \log_{10} \frac{\sum_{t=0}^{20} ri^2(t)}{\sum_{t=20}^{T} ri^2(t)} \tag{53}$$

Pour toutes les configurations, la modélisation est réalisée par bandes d'octaves à partir des mesures de temps de réverbération effectuées dans chacune des salles. A partir de ces modélisations, l'indicateur associé C_{TR} correspond alors à la moyenne de C_{modele} calculée sur les N_{tot} bandes d'octave centrées de 125 Hz à 8 kHz (indicées par N_f) :

$$C_{TR} = \frac{1}{N_{tot}} \sum_{N_f=1}^{N_{tot}} C_{modele}\ (N_f) \tag{54}$$

La clarté calculée à partir de la RI est reportée en fonction de la clarté C_{TR} dans la Figure 42. Cette figure permet d'illustrer que la valeur de ΔT n'est pas significative pour l'estimation de la clarté. A l'exception des salles à faible clarté, la clarté C_{TR} semble bien corrélée avec la clarté C20 calculée à partir de la RI. Une sous-estimation d'environ 5 dB de la clarté C_{TR} par rapport à la clarté calculée est toutefois visible. Cette sous-estimation n'est pas très gênante pour notre application puisque nous cherchons un modèle décrivant les résultats du test perceptif, et non pas un modèle qui décrive au mieux un indicateur d'acoustique des salles.

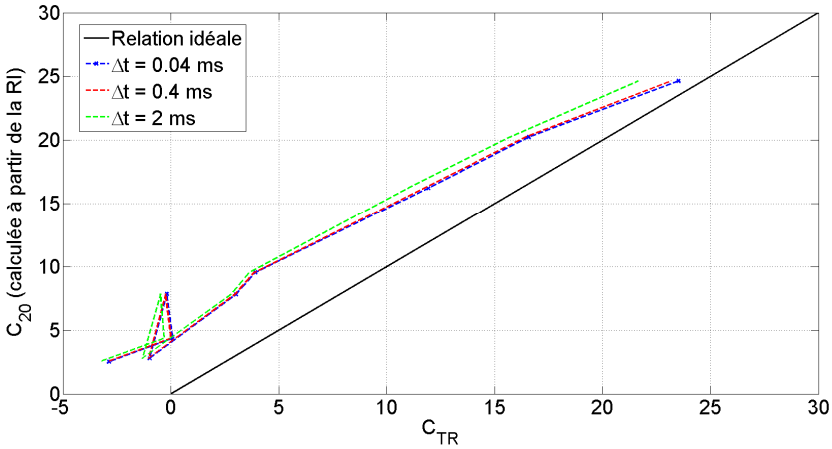


Figure 42 : clarté C_{20} calculée à partir de la RI en fonction de la clarté C_{TR} estimée à partir du TR_{60}

IV.7.B. Evaluation des indicateurs

La relation entre les valeurs des indicateurs et les notes attribuées par les auditeurs est supposée linéaire *a priori*. Pour prédire les notes attribuées par les auditeurs, une régression linéaire est donc estimée sous forme $Z = aX + b$, avec a et b les coefficients de la régression linéaire et X les données issues d'un indicateur. Pour évaluer la pertinence de l'indicateur, les coefficients suivant sont calculés :

- Le coefficient de corrélation de Bravais-Pearson :

$$R = \frac{Cov(X,Y)}{STD_X STD_Y} \tag{55}$$

  Avec X et Y les données associées aux notes du test et à l'indicateur évalué et STD_X et STD_Y les écarts-type de ces données. Ce coefficient compris entre -1 et 1 traduit la linéarité de la relation entre deux vecteurs : une valeur proche de 0 traduit une absence de corrélation entre les grandeurs.
- Le coefficient de corrélation de rang de Spearman :

$$\rho_{Spear} = 1 - \frac{6\sum_{i=1}^{N}[rg(X_i) - rg(Y_i)]^2}{N^3 - N} \tag{56}$$

  Avec $rg(X_i)$ et $rg(Y_i)$ les rangs du $i^{ème}$ élément de X et Y respectivement et N le nombre d'éléments. Ce coefficient également compris entre -1 et 1, permet d'évaluer le lien d'ordonnancement de deux vecteurs : une valeur proche de 0 traduit une absence de relation d'ordre. Le principal avantage de coefficient est qu'il ne présuppose pas de relation linéaire entre les grandeurs.
- La différence entre le modèle et les données issue du test (« Root Mean Square Error ») :

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(X_i - Z_i)^2}{N}} \tag{57}$$

  Avec Z_i le $i^{ième}$ élément de la régression linéaire.

IV.7.B.a Corrélation entre les indicateurs

Tous les indicateurs décrivent le même effet : la quantité de réverbération associée à chaque configuration. Ces indicateurs sont donc censés se ressembler, et seuls des particularités d'implémentation les différencient. Pour quantifier cette ressemblance, le coefficient de corrélation défini par l'équation (55) est calculé entre chaque paire d'indicateurs et la valeur absolue est représentée à la Figure 43.

On peut constater que tous les indicateurs sont très corrélés entre eux, avec une corrélation supérieure à 0.75 dans tous les cas. La clarté estimée à partir du TR est particulièrement corrélée avec les clartés calculées à partir de la RI : le coefficient de corrélation est toujours supérieur à 0.9. Cela semble montrer que la méthode d'estimation permet d'obtenir des résultats proches de ceux obtenus à partir de la RI, et est donc pertinente, tout en utilisant des données accessibles dès la phase de conception d'une salle.

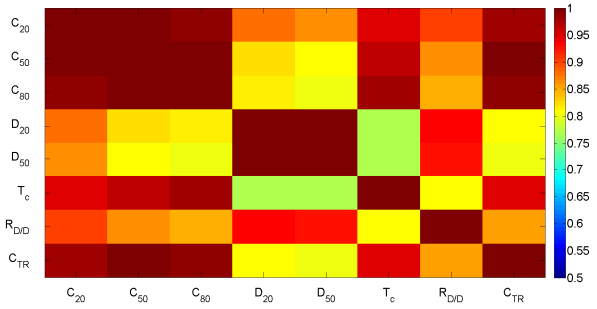


Figure 43 : Matrice de corrélation entre les différents indicateurs

IV.7.B.b Lien avec les données du test

Les trois coefficients (R, ρ_{Spear} et $RMSE$) présentés en début de section sont calculés entre chaque indicateur et les notes attribuées lors du test. Ces différents coefficients sont représentés dans le Tableau 7. Les coefficients de corrélation R sont toujours supérieurs à 0.88. Les indicateurs qui décrivent au mieux les résultats du test sont les indicateurs de clarté. Parmi ces indicateurs, C_{20} est celui qui décrit le mieux les données : son coefficient de corrélation R vaut 0.97, le coefficient de corrélation de rang ρ_{Spear} vaut 0.95 et l'erreur moyenne $RMSE$ vaut 8.63. La durée d'intégration de la clarté a donc une influence, et une valeur inférieure aux durées « usuelles » permet de mieux décrire les données du test. Le signal employé lors du test correspond en effet à des bouffées de bruit, qui a des transitoires très brefs, alors que les indicateurs « usuels » ont été établis pour décrire des signaux musicaux ou l'intelligibilité de la parole.

Les indicateurs obtenus à partir du TR ($R_{D/D}$ et C_{TR}) sont également très corrélés aux résultats du test : une modélisation minimaliste de l'acoustique des salles permet donc de retrouver les notes du test avec une pertinence voisine de celle obtenues avec les indicateurs calculés sur les réponses impulsionnelles. En particulier, la prise en compte de la décroissance temporelle de l'énergie permet d'expliquer quasiment aussi bien les notes du test que le meilleur indicateur calculé à partir de la RI : l'indicateur C_{TR} est donc retenu *in fine*.

Indicateur	C_{20}	C_{50}	C_{80}	D_{20}	D_{50}	T_c	$R_{D/D}$	C_{TR}
R	0.97	0.96	0.95	0.88	0.87	-0.93	0.94	0.96
ρ_{Spear}	0.95	0.95	0.95	0.95	0.95	-0.85	0.93	0.93
$RMSE$	8.63	9.94	10.90	18.58	19.22	13.75	12.31	10.17

Tableau 7 : coefficient de corrélation, de corrélation de rang et erreur moyenne obtenus pour les différents indicateurs

Des améliorations ont été tentées en calculant les indicateurs par bandes d'octave et en analysant les minima, maxima et moyennes des bandes d'octave, avec ou sans pondération A. Aucune des variantes envisagées n'a permis d'expliquer les résultats du test avec une corrélation supérieure à celle de l'indicateur C_{20}. Une autre manière de calculer la quantité d'énergie est proposée dans l'Annexe H : cette méthode consiste à éliminer la contribution du champ direct à partir de la réponse impulsionnelle anéchoïque. Cette méthode ne permet pas de prédire les résultats du test avec une meilleure corrélation qu'avec la C_{20}.

Les paramètres de la régression linéaire permettant de prédire les notes à partir de ce modèle sont alors définis par la relation suivante :

$$Note_{prédite} = 3.78C_{TR} + 20.5 \qquad (58)$$

Les notes attribuées lors du test en fonction des notes prédites à partir de la relation (58) sont présentées en Figure 44. Cette figure permet d'illustrer que les notes prédites à partir de C_{TR} sont proches des notes attribuées lors du test. L'ordre du classement des configurations qui obtiennent des notes très basses, entre 10 et 30, est toutefois inversé. Cela représente la principale limite de l'indicateur, mais elle n'est pas très gênante pour notre utilisation : ces configurations sont très proches de l'ancre basse, et ne correspondent pas à des situations de reproduction souhaitables .

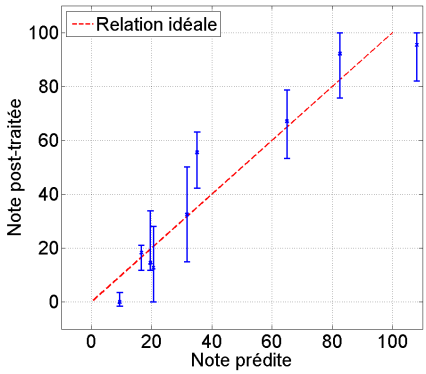


Figure 44 : notes médianes attribuées lors du test en fonction des notes prédites par l'indicateur C_{TR}

IV.8 Conclusion

L'évaluation de l'effet de salle a été abordée dans le cas particulier d'un système SISO, simplement caractérisé par sa réponse impulsionnelle. Pour cela, une évaluation perceptive a été menée, permettant de hiérarchiser qualitativement quatre paramètres : la proximité source-récepteur, les caractéristiques de la salle, un filtre d'égalisation, et la robustesse de ce filtre au changement

de position. Les résultats du test perceptif montrent que le rapprochement des sources semble être une approche très intéressante : pour une salle aux performances acoustiques « moyenne » le rapprochement des sources permet d'obtenir une évaluation par les auditeurs proche d'une salle bien traitée acoustiquement.

D'un point de vue pratique, il semble beaucoup plus facile de rapprocher les sources que de traiter la salle ou d'égaliser le système de diffusion. Une salle traitée acoustiquement permet d'obtenir de bons résultats, y compris à 80 cm, mais ce traitement est alors très coûteux. Une salle qui n'est pas assez bien traitée ne peut pas être compensée par l'un des filtrages testés ici, et de manière générale les filtres testés ont un intérêt marginal : ils n'améliorent que les salles déjà bonnes et peuvent dégrader le rendu dans certaines situations. Pour une égalisation SISO, une approche IIR est simple et efficace lorsque l'égalisation est nécessaire.

Un indicateur a été bâti à partir des résultats du test. L'indicateur est proche de la clarté C_{20} mais calculé à partir de paramètres simples de la salle (dimensions et temps de réverbération) et permet de prédire les résultats du test de manière quasiment équivalente aux indicateurs calculés à partir de réponses impulsionnelles. Le coefficient de corrélation entre les notes obtenues avec l'indicateur proposé et les résultats du test est en effet de 0.96.

L'effet d'un déplacement par rapport à la position d'égalisation a été abordé, mais son influence n'est pas très significative ici : l'effet de salle semble dominer et masquer les autres effets. Ce résultat mériterait cependant d'être vérifiée dans une application plus complexe, plus proche de notre sujet d'étude : un système MIMO (deux oreilles et système de diffusion à plusieurs canaux).

Chapitre V Interaction de l'auditeur avec son environnement

Table des matières

V.1 Introduction

L'effet de la salle d'écoute sur le rendu sonore a été abordé précédemment, concluant à l'intérêt de placer les sources à proximité de l'auditeur. Cependant, le chapitre II a montré que dans ce cas le contraste est plus élevé à proximité de l'auditeur, et que les HRTF varient donc plus rapidement d'une position à l'autre. Ceci se traduit par une plus grande variation potentielle du rendu lorsque l'auditeur n'est pas à la position d'écoute prévue. Il importe donc de savoir dans quelle mesure ces variations sont perceptibles, afin de déterminer la distance minimale compatible avec un déplacement raisonnable de l'auditeur au sein d'un système transaural.

Pour y répondre, un test perceptif a été mis en place, en simulant des situations d'écoute de différents systèmes transauraux pour lesquels l'auditeur ne serait pas à la position optimale. La restitution de ces situations est effectuée au casque d'écoute afin de reproduire de manière contrôlée le champ à des positions spécifiques. Des configurations de systèmes transauraux dont les sources sont aux distances de 20 cm à 80 cm par rapport au centre de la tête sont évaluées, la distance de 20 cm étant choisie dans le but de révéler des dégradations liées à une proximité *a priori* exagérée.

Une sélection de déplacements de l'auditeur a été opérée, car le nombre de déplacements envisageables est conséquent : a minima, six degrés de libertés peuvent être combinés pour générer un déplacement de la tête de l'auditeur. Pour être exhaustif il faudrait évaluer tous les déplacements en situation d'écoute, impliquant l'évaluation d'un grand nombre de stimuli. Cela n'est pas compatible avec un test perceptif d'une durée raisonnable pour les auditeurs. Pour cette raison, seulement deux déplacements ont été sélectionnés, avant tout pour bâtir un indicateur à partir des résultats du test. Cet indicateur devrait suffire à éliminer des configurations inacceptables et peut-être permettre de classer les plus appropriées.

V.2 Evaluation perceptive

L'objectif de ce test est d'évaluer l'effet d'un déplacement de l'auditeur, et de mettre en relation les résultats obtenus avec ceux du test précédent concernant l'effet de salle. Pour assurer la continuité avec les précédents résultats, de nombreux points sont communs entre les deux protocoles. Le test est donc de type MUSHRA, et certains signaux du premier test sont inclus dans le second.

Il est ici demandé à l'auditeur d'évaluer la « proximité » entre un son binaural de référence, et un son issu d'un système transaural tel que la position d'écoute n'y est pas toujours optimale. La maîtrise des conditions d'écoute est obtenue par simulation : les fonctions de transfert entre les haut-parleurs et les oreilles sont calculées pour une position donnée de la sphère utilisée pour la simulation, cette position respectant le décalage à évaluer. La convolution des signaux à tester par ces fonctions de transfert permet la restitution au casque d'écoute de la situation simulée. La première section de cette partie est consacrée à la génération des stimuli. L'interface du test et la feuille de consigne donnée aux auditeurs sont similaires à l'expérience précédente (voir Annexe E et Annexe F), en adaptant le nombre de stimuli.

V.2.A. Configurations évaluées

Les tests consistent à reproduire une situation virtuelle, via différents systèmes physiques. Ces situations virtuelles et systèmes physiques sont présentés dans les prochains paragraphes.

V.2.A.a Configurations des sources physiques

L'effet de la distance des sources physiques est l'un des principaux paramètres à évaluer ici, et les distances de 80 cm et 40 cm correspondant au test précédent sont reprises. La distance de 80 cm semblait correspondre à un maximum au-delà de laquelle l'effet de la salle devient majeur. Nous ajoutons dans ce nouveau test la distance de 20 cm, qui semble être la distance minimale à laquelle il est possible d'envisager de placer un auditeur dans le système. En effet, le référentiel est centré sur la tête dont le rayon est de l'ordre 10 cm, la configuration à 20 cm revient donc à placer les sources à une distance de l'ordre de 10 cm de la surface de la tête.

L'effet de l'écartement angulaire des haut-parleurs joue également un rôle important, et quatre écarts angulaires différents sont évalués : +/-5°, +/- 30°, +/- 60° et +/-90°. Pour toutes les configurations, les haut-parleurs sont placés dans le plan horizontal. Il a été montré que le placement des sources en élévation pouvait être un avantage [Parodi, 2010], toutefois, cet aspect semble être de second ordre après l'écartement angulaire, et n'est pas étudié dans ce test

perceptif. Son influence sera évaluée ultérieurement (au chapitre VI) par simulations. La Figure 45 représente les différents systèmes évalués.

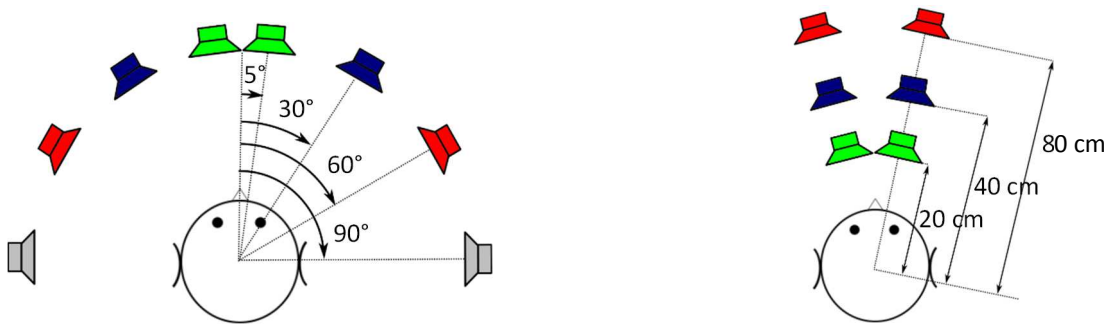


Figure 45 : Représentation des différents systèmes évalués. A gauche, les quatre écartements angulaires possibles et à droite les trois distances possibles. Toutes les combinaisons sont évaluées, soit douze systèmes.

V.2.A.b Sources virtuelles

Les sources virtuelles évaluées ici sont monophoniques et statiques. Des sources virtuelles simples ont été préférées pour notre approche préliminaire d'optimisation. L'utilisation de sources virtuelles plus complexes, avec une largeur apparente ou une dynamique dans la localisation par exemple, pourra être évaluée ultérieurement pour affiner les résultats.

L'effet de la position de la source virtuelle est évalué. Elle est localisée en trois incidences différentes : face à l'auditeur (0°), légèrement sur le côté (45°) et fortement sur le côté (90°). Le côté gauche a été choisi arbitrairement, en supposant que les résultats seraient équivalents du côté droit. Une source virtuelle à 90° correspond donc à l'emplacement d'une des sources physiques d'un système transaural. Ce cas de reproduction sonore triviale est inclus car il peut correspondre à une situation d'utilisation d'un système de spatialisation. Par la suite, l'acronyme « SRV » sera utilisé pour définir une source virtuelle.

V.2.A.c Stimulus

Le stimulus choisi est le même que pour le test précédent, il s'agit de bouffées de bruit rose préalablement convoluées avec la réponse impulsionnelle de l'enceinte Tannoy System 600 égalisée avec le filtre calculé dans le chapitre concernant le test sur la restitution en salle d'écoute. Le spectre reproduit couvre alors l'intervalle de [80 Hz-13 kHz] à +/-1 dB.

V.2.A.d Ancres

De la même manière que pour le test précédent, des signaux communs (les « ancres ») sont inclus dans les stimuli:

- **« REF »** désigne la référence cachée (signal binaural)
- **« ANCH »** désigne une ancre haute : il s'agit du signal anéchoïque monophonique en restitution diotique. Il s'agit de la référence du test précédent.
- **« ANCI »** désigne une ancre intermédiaire. Ce signal correspond à un son évalué dans le haut de l'échelle du test précédent : Studio 80 cm (note médiane de 67).
- **« ANCB »** désigne une ancre basse. Ce signal correspond à un son évalué dans le bas de l'échelle dans le test précédent : Bureau moyen 80 cm (note médiane de 15).

Ces trois ancres (haute, intermédiaire et moyenne) sont des signaux monophoniques, alors que les signaux évalués sont des sons spatialisés. Cette combinaison *a priori* surprenante, a deux objectifs : d'une part elle permet de faire le lien avec le test précédent et de présenter les notes attribuées par les auditeurs sur une échelle commune aux deux tests. D'autre part, le fait de combiner des signaux plus ou moins dégradés sur le plan spectral avec des signaux spatialisés avec plus ou moins de qualité peut permettre de comparer l'importance relative accordée par les auditeurs à ces deux aspects.

V.2.A.e Déplacements de l'auditeur

Nous ne souhaitons pas présenter plus de dix sons simultanément, en accord avec la recommandation [ITU-R BS.1534-3, 2015]. En plus des quatre ancres, cela permet d'évaluer six sons associés à un déplacement. Par ailleurs, l'effet de distance est celui que nous cherchons à évaluer prioritairement : il est donc souhaitable d'évaluer les signaux issus de différentes distances au sein d'un même test. Trois distances ont été sélectionnées, et il n'est alors possible d'évaluer que deux déplacements. C'est une autre limite de ce test : les échantillons sélectionnés ne permettent pas de représenter exhaustivement les situations d'écoute. Cependant, ils sont supposés être suffisamment représentatifs de l'ensemble des situations d'écoute pour servir à bâtir un indicateur objectif, appliquable à d'autres déplacements.

Il existe au moins six degrés de liberté de déplacements de la tête de l'auditeur : par exemple trois translations selon les axes cartésiens du repère et trois rotations autour de ces axes. Pour chacun, deux directions sont possibles. Pour chaque amplitude de déplacements, il existe donc douze déplacements potentiels. Pour effectuer un premier tri sur les déplacements à évaluer, un certain nombre d'hypothèses sont émises sur la plausibilité des déplacements de l'auditeur :

- L'incertitude de placement en hauteur de l'auditeur est facilement contrôlable par réglage de la hauteur du siège et l'amplitude de déplacement possible pour l'auditeur est alors réduite.
- La rotation autour de l'axe X (tête penchée à gauche/droite) peut être limitée en donnant à l'auditeur un repère visuel.
- De même, la rotation autour de l'axe Y (tête vers le haut/bas) n'est pas considérée, car un repère visuel permet d'orienter le regard de l'auditeur. De plus, les variations selon cette rotation n'ont que peu d'effets sur un modèle sphérique, pour lequel les microphones sont très proches de l'axe de rotation.

Seule la rotation autour de l'axe Z (rotation gauche/droite) et les translations selon l'axe X (avant/arrière) et Y (gauche/droite) sont donc considérées. L'amplitude de ces déplacements a été fixée à 5 cm pour les translations et 10° pour les rotations, ce qui semble compatible avec l'incertitude de placement d'un auditeur. L'effet du signe du déplacement (la direction) ne semble pas avoir d'influence significative sur toutes les configurations testées. De plus, des écoutes informelles ont révélé qu'un déplacement en rotation de 10° est peu discernable de la référence. Les déplacements en translation de 5 cm selon l'axe X (+5X) et Y (+5Y) sont donc finalement ceux qui semblent les plus significatifs pour comparer les systèmes de restitution. Les configurations « sans déplacement » ne sont pas présentées lors du test car le rendu était jugé indiscernable de la référence lors d'écoutes informelles.

V.2.A.f Egalisation en sonie des signaux

Les signaux correspondant à un déplacement de la tête ne sont pas égalisés en sonie : le changement de sonie lié à un déplacement fait partie de la dégradation correspondante et doit être pris en compte dans le test.

Seules les ancres sont égalisées en sonie à partir d'une méthode d'estimation normalisée (ISO 532B [ISO 532B, 1975]). La sonie cible pour ces signaux monophoniques est la sonie estimée pour le son correspondant à la source virtuelle à 0°, qui est diotique. La procédure utilisée consiste à appliquer un gain sur les signaux monophoniques pour que l'estimation de sonie soit identique à celle de la sonie cible.

V.2.B. Mise en œuvre du test

V.2.B.a Organisation du test

Chaque test MUSHRA correspond à une position de source virtuelle, et un écartement angulaire du système transaural. Ainsi, au sein d'un même test sont comparés les sons correspondant aux deux déplacements de l'auditeur et trois distances de sources. Cela représente six stimuli, auxquels s'ajoutent les quatre ancres : la référence cachée, le signal monophonique anéchoïque et les deux signaux monophoniques en salle d'écoute. Le signal de référence est spécifique à chaque test : il s'agit du signal binaural correspondant à la source virtuelle objet du test. L'ensemble des signaux à tester est ainsi réparti entre douze tests, correspondant chacun à une série de dix signaux.

De manière analogue au test précédent, une phase d'entraînement est proposée avant le début du test, pour familiariser l'auditeur avec l'interface et les sons à évaluer. Cet entraînement se présente de la même manière que pour les douze configurations à évaluer, à la différence qu'il ne contient qu'une sélection de huit signaux. La référence dans ce cas est le signal binaural correspondant à la source virtuelle à 45°. Les notes attribuées lors de cette phase d'entraînement ne sont pas exploitées.

V.2.B.b Déroulement du test

Dix-huit auditeurs (quatre femmes et quatorze hommes) ont participé au test d'écoute. Ils étaient âgés de 21 à 40 ans (moyenne de 30 ans). Douze de ces auditeurs avaient participé au test précédent. Leur audition n'a pas été vérifiée, mais la validité de leur réponse a été vérifiée à travers l'évaluation de la référence cachée comme le suggère la recommandation [ITU-R BS.1534-3, 2015]. La durée moyenne du test (hors explication des consignes et phase d'entraînement) était de 40 minutes.

Les sons ont été diffusés avec le même matériel que pour le test précédent, soit un casque Beyerdynamic DT990 Pro et une carte audionumérique ESI U24 XL. Le niveau sonore du signal monophonique était identique à celui du test précédent (sonie monaurale de 30 Sones).

V.3 Analyse des résultats

A partir des réponses des auditeurs, plusieurs étapes d'analyse sont effectuées : premièrement la validité de leurs réponses est étudiée, puis une analyse statistique est menée et enfin les principaux résultats sont issus des représentations graphiques des notes attribuées.

V.3.A. Validité et corrélation des réponses des auditeurs

L'analyse préliminaire concerne la validité des données collectées. Nous vérifions que les auditeurs ont su identifier la référence cachée, pour valider leurs réponses. L'homogénéité des réponses entre les différents auditeurs est également vérifiée, permettant de supposer qu'ils appartiennent à une même population.

V.3.A.a Référence cachée non identifiée

D'après [ITU-R BS.1534-3, 2015], il est recommandé d'écarter les réponses d'un auditeur qui aurait attribué à la référence cachée une note inférieure à 90 pour plus de 15 % des tests. Dans cette expérience, cela correspond à ne pas identifier la référence cachée dans au moins deux conditions. Un auditeur qui n'avait pas participé au test précédent se trouve dans cette situation, et aucune de ses réponses n'est prise en compte dans la suite de l'analyse.

Au total, les réponses de dix-sept auditeurs sont donc exploitées. Parmi ces dix-sept auditeurs, sept n'ont pas identifié la référence cachée pour une des configurations, en lui attribuant une note inférieure à 90 %. Cela est relativement fréquent, puisque cela signifie que près de la moitié des auditeurs se sont trompés une fois. D'un point de vue statistique, cette fréquence n'est pas si importante : étant donné que chaque auditeur a évalué douze configurations, cela représente 3.4 % des cas.

V.3.A.b Corrélation des réponses entre auditeurs

Le test proposé demande de comparer des sons sur un seul critère (la similarité avec une référence) et ce jugement peut mettre en jeu plusieurs dimensions perceptives, que nous chercherons à hiérarchiser par la suite. Il est possible que les auditeurs aient adopté des stratégies de réponse différentes, en accordant plus au moins d'importance aux différentes dimensions perceptives. Pour mettre en évidence les différences inter-individuelles entre les auditeurs, la matrice de corrélation entre les réponses des différents auditeurs est calculée. Pour chaque auditeur, l'ensemble des réponses est réuni en un vecteur. Un coefficient de corrélation est calculé par paires d'auditeurs de la manière suivante :

$$C_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \tag{59}$$

Où X et Y correspondent aux vecteurs de réponses de deux auditeurs, σ_X et σ_Y les écarts-type associés et Cov désigne la covariance. Les notes attribuées à la référence cachée ont déjà été exploitées pour effectuer un premier tri, et ne sont plus utilisées ici. La matrice obtenue pour l'ensemble des auditeurs est reportée à la Figure 46. Les notes attribuées par les différents auditeurs sont bien corrélées ($C_{XY} > 0.7$) et ne permettent pas d'identifier un groupe d'auditeurs pour lequel les réponses seraient différentes.

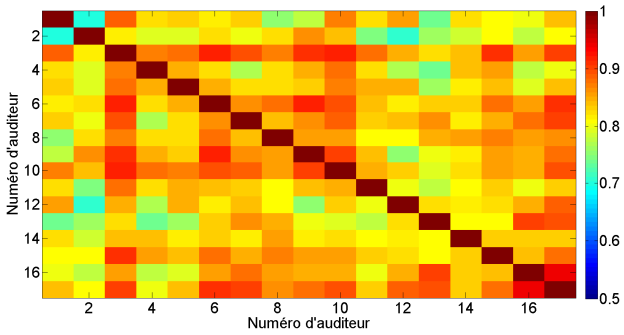


Figure 46 : Matrice de corrélation entre les réponses des auditeurs

V.3.B. Analyse statistique

V.3.B.a Distribution des réponses par son

L'une des hypothèses des tests paramétriques tels que l'ANOVA concerne la normalité des distributions. Pour tester la normalité des distributions, le test de Jarque-Bera a été appliqué pour chaque son. Il compare les moments d'ordre 2 et 3 (skewness et kurtosis) de la distribution testée à ceux d'une distribution normale. Le nombre de sons dont la distribution est normale est reporté dans le Tableau 8, pour chaque type de son. Chaque case correspond donc à douze réalisations différentes (quatre écartements angulaires et trois sources virtuelles).

	REF	ANCH	ANCB	ANCI	20_{+5X}	40_{+5X}	80_{+5X}	20_{+5Y}	40_{+5Y}	80_{+5Y}
Nombre de distributions jugées « normales »	0	7	0	9	10	7	2	12	11	11
Pourcentage (sur douze réalisations)	0 %	58 %	0 %	75 %	83 %	58 %	17 %	100 %	92 %	92 %

Tableau 8 : nombre de sons dont la distribution des notes attribuées par les auditeurs est normale selon le test de Jarque-Bera. Pour chaque case, le pourcentage est calculé pour douze réalisations différentes (quatre écarts angulaires de système transaural et trois sources virtuelles)

La référence cachée et l'ancre basse n'ont jamais une distribution normale car ces signaux ont été spécialement choisis pour qu'ils soient évalués aux extrêmes de l'échelle. Ceci implique une distribution asymétrique.

Au total, 74 % des signaux avec déplacements ont une distribution normale : ceci est nettement supérieur au test sur l'effet de salle (où seulement 27% des sons avaient une distribution normale des notes). Une ANOVA est donc plus légitime pour ce deuxième test.

V.3.B.b ANOVA

Une Analyse de la Variance (ANOVA) est appliquée sur les données, en excluant les ancres. Sont considérés les facteurs « distance », « écartement angulaire », « source virtuelle » et « déplacement ». Le facteur « auditeur » est considéré comme un effet aléatoire. L'ANOVA est menée avec la fonction *anovan* de la Statistics Toolbox de Matlab et les résultats sont reportés dans le Tableau 9.

Facteur	SCE	D.L.	F	p>F
Auditeur	74227.3	16	3.75	<0.01
Distance	16429.4	2	44.39	<0.01
Ecartement angulaire	61889.9	3	47.57	<0.01
Source virtuelle	33735.3	2	61.66	<0.01
Déplacement	218694.7	1	201.65	<0.01
Auditeur*Distance	5921.2	32	0.75	0.84
Auditeur*Ecartement angulaire	20817.4	48	1.76	<0.01
Auditeur*Source virtuelle	8754.0	32	1.11	0.31
Auditeur*Déplacement	17352.1	16	4.40	<0.01
Distance*Ecartement angulaire	9827.7	6	6.65	<0.01
Distance*Source virtuelle	4001.4	4	4.06	<0.01
Distance*Déplacement	51256.9	2	104.06	<0.01
Ecartement angulaire*Source virtuelle	20616.9	6	13.95	<0.01
Ecartement angulaire*Déplacement	50112.0	3	67.82	<0.01
Source virtuelle*Déplacement	18365.6	2	37.29	<0.01

Tableau 9 : Résultats de l'ANOVA. SCE signifie "Somme des Carrés des Ecarts", DL signifie Degrés de Liberté, F est la statistique du test de Fisher et p la probabilité associée

L'ANOVA révèle de la significativité au seuil de 0.01 pour tous les facteurs et toutes les interactions n'impliquant pas le facteur auditeur. De la même manière qu'au chapitre IV, aucun test post-hoc n'est mené et l'analyse porte sur la représentation graphique des notes.

V.3.C. Analyse descriptive

Les notes médianes attribuées à chacune des configurations sont regroupées en fonction de leur écartement angulaire dans la Figure 47. Chaque courbe représentée correspond à une série évaluée par les auditeurs alors que des courbes différentes ne correspondent pas à des sons évalués simultanément par l'auditeur.

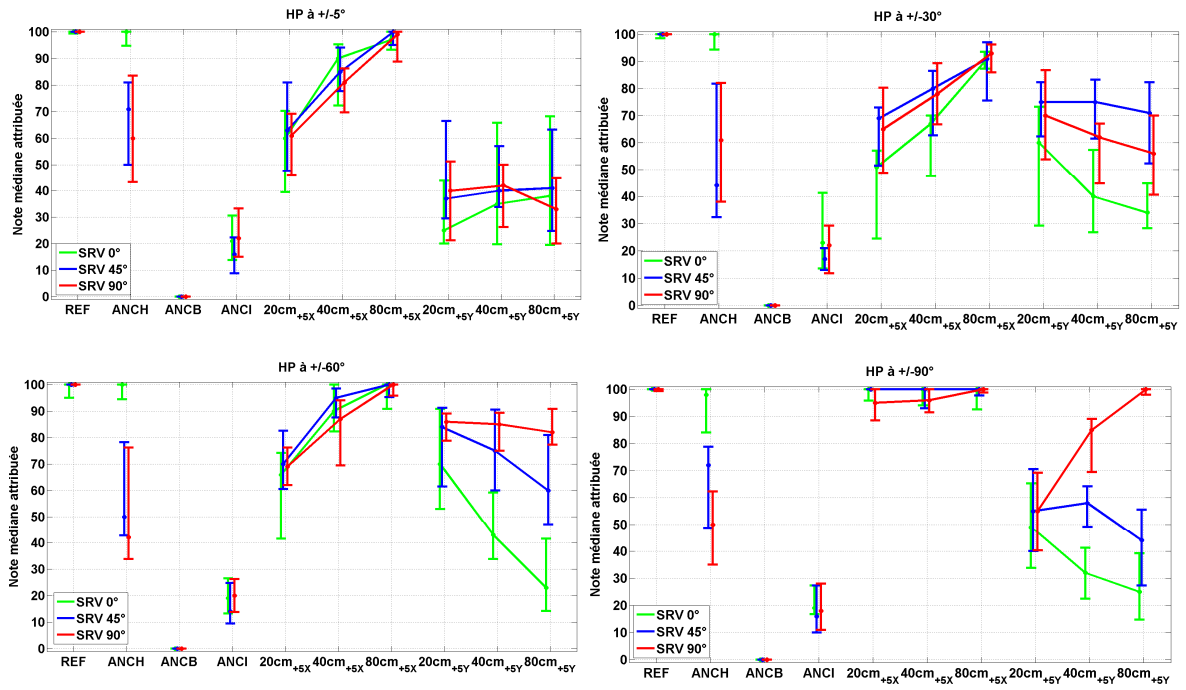


Figure 47 : Notes médianes attribuées aux différents sons pour différents systèmes. Chaque figure correspond à un écart angulaire de sources physiques : En haut à gauche : +/-5°, en haut à droite : +/-30°, en bas à gauche : +/-60°, en bas à droite : +/-90°. Les barres d'erreurs correspondent aux écarts interquartiles. Chaque courbe correspond à une série évaluée par les auditeurs au cours d'un même test MUSHRA

V.3.C.a Ancres

La référence cachée et l'ancre basse obtiennent des notes médianes de 100 et 0 respectivement pour toutes les configurations. L'ancre basse a bien joué son rôle, puisque systématiquement évaluée au bas de l'échelle. Il ne semble pas y avoir d'ambiguïté sur cette ancre basse puisque les écarts interquartiles sont nuls dans chaque cas.

Pour la source virtuelle à 0°, la note médiane du signal mono (ANCH) est systématiquement évaluée proche de 100. Les auditeurs ne parviennent pas à faire la différence entre une source monophonique et une source censée être spatialisée devant l'auditeur. Cela traduit probablement le fait que la perception d'une source frontale est souvent intracrânienne en restitution binaurale, notamment avec l'usage de HRTF non individualisées. De plus, cela permet de faire le lien avec le test précédent, où la référence était le signal monophonique : pour une source virtuelle frontale, le signal de référence est sensiblement équivalent dans les deux tests.

Pour les sources virtuelles à 45° et 90°, les notes médianes obtenues pour le signal monophonique (ANCH) sont similaires et significativement inférieures à la référence : l'écart entre les deux notes médianes est au plus de 22 pour la configuration à +/- 90°. S'il y a une nette différence avec l'absence de latéralisation (SRV0°), le degré de latéralisation (SRV45° ou SRV90°) semble avoir peu d'influence. Pour les deux sources virtuelles latérales, il apparaît une variabilité dans l'évaluation du signal monophonique (ANCH) entre les séries : pour la source virtuelle à 45°, la note médiane est de 71 pour la série associée aux écartements de +/- 5°, alors qu'elle vaut seulement 44 pour la série à +/- 30°. L'écart interquartile est étendu pour cette configuration, le quartile supérieur atteint 82 (dépassant ainsi la valeur médiane pour la série à +/-5°).

La note de l'ancre intermédiaire est plutôt basse et proche de 20 dans tous les cas, avec un écart interquartile assez réduit (toujours inférieur à 33). L'effet de localisation ne semble pas intervenir pour l'évaluation de ce son, puisque la note est du même ordre quelle que soit l'incidence de la source virtuelle. La présence d'autres signaux n'influe pas non plus le jugement de l'auditeur, les notes étant similaires quelle que soit la présentation. Lors du test du chapitre précédent, les notes attribuées aux signaux correspondant aux ancres basse et intermédiaire étaient respectivement de 15 et 67 : l'ancre intermédiaire faisait partie des signaux les plus similaires à la référence, alors qu'ici c'est le contraire. La majorité des signaux évalués ici aurait probablement eu une note supérieure à 67 dans le test précédent, donc les différences perçues ici sont plus faibles que lors du précédent test.

Ainsi, la comparaison des deux tests semble montrer que l'effet de la salle détériore davantage le rendu que le déplacement de l'auditeur, et confirme l'intérêt de traiter cet aspect prioritairement. La présence d'autres signaux (effet de contexte) influe le jugement de l'auditeur, mais de manière limitée : seule de la variabilité est observée pour le signal monophonique en comparaison aux sources virtuelles à 45° et 90°, mais les écarts inter quartiles se recoupent. Une comparaison des différentes séries semble donc légitime au moins sur un plan qualitatif.

V.3.C.b Effet de la distance

L'effet de la distance est particulièrement visible pour le déplacement de +5X (translation de 5 cm vers l'avant) : pour tous les écarts angulaires les notes augmentent avec la distance quelle que soit la source virtuelle. La note obtenue est alors quasiment maximale à 80 cm pour toutes les configurations. Un cas particulier concerne l'écartement +/- 90° pour lequel les notes sont toujours maximales : cet écart angulaire est particulièrement robuste au déplacement vers l'avant.

Pour le déplacement +5Y (translation de 5 cm vers la gauche), l'effet est inverse : les configurations en champ proche sont généralement plus robustes. Cet effet est le plus visible pour l'écartement angulaire de +/-60°, où la note médiane est de 23 à 80 cm, et de 70 à 20 cm. Cette robustesse en proximité de l'auditeur est surprenante car elle est contraire à notre intuition initiale. Il a été montré dans le chapitre III que le contraste naturel est élevé pour les configurations proches de l'oreille, et que le rôle des filtres transauraux est alors réduit. La simplicité des filtres pourrait ainsi expliquer cette robustesse au déplacement de l'auditeur.

L'évaluation de la source virtuelle pour l'écart angulaire à +/- 90° est particulière car les notes obtenues augmentent avec la distance pour le déplacement +5Y. Dans ce cas l'incidence de la source virtuelle correspond à une source physique, et correspond donc à un cas trivial de reproduction sonore. A faible distance, ce type de reproduction n'est pas robuste à un déplacement en direction de la source. A 20 cm, l'oreille du côté de la source virtuelle est initialement à 11 cm de la source physique. Lors d'une translation de 5 cm en direction de cette source physique, l'oreille est seulement à 6 cm de la source physique ce qui est presque deux fois plus proche. Une différence de sonie significative peut alors apparaître dans ce cas, impliquant une forte sensibilité aux déplacements en direction de la source physique pour les grands écartements angulaires.

Un cas particulier où la distance n'a pas d'effet significatif concerne l'écart angulaire de +/- 5° et le déplacement de +5Y : les notes sont toujours inférieures à 40. Ce système est particulièrement fragile aux déplacements latéraux.

Une tendance générale est visible : les configurations en champ lointain sont plus sensibles aux déplacements latéraux et les configurations en champ proche sont plus sensibles aux déplacements avant/arrière.

V.3.C.c Effet de l'écart angulaire entre sources

L'effet de l'écartement angulaire entre sources physiques est très lié à celui de la distance. Par exemple, la configuration à 20 cm d'écartement angulaire +/-60° est la plus robuste : les notes sont toujours supérieures à 60. En revanche, à 80 cm la configuration d'écartement angulaire +/-60° est peu robuste au déplacement +5Y où la note médiane est de 20 pour une source virtuelle frontale. La configuration la moins robuste du test correspond est celle à 20 cm d'écartement +/-5°, dont les notes sont toujours inférieures à 60.

Les configurations à faible écart angulaire semblent plus robustes aux déplacements vers l'avant et les configurations avec un grand écart angulaire semblent les plus robustes aux mouvements latéraux, mais avec une influence croisée de la distance. La configuration à 20 cm et +/-60° est la plus robuste du test, et la configuration à 20 cm et +/- 5° en est la moins robuste.

V.3.C.d Importance du déplacement de l'auditeur

La robustesse au déplacement est étroitement liée à la distance et à l'écart angulaire comme évoqué précédemment. De manière générale, les configurations évaluées semblent plus robustes au déplacement vers l'avant +5X. En effet, pour ce déplacement toutes les notes sont supérieures à 50, alors que pour le déplacement +5Y seulement 18 configurations (soit la moitié) ont des notes supérieures à 50. En pratique, les déplacements +5Y sont mieux contrôlables : avec un repère visuel face à l'auditeur (un écran), il peut s'aligner instinctivement. En revanche le repérage avant/arrière est plus délicat sans contrôle extérieur, et par fatigue l'auditeur peut avoir tendance à pencher peu à peu la tête vers l'avant.

V.3.D. Synthèse des résultats du test

Les dégradations évaluées dans ce test sont plus subtiles que celles évaluées dans le test précédent. Un signal commun aux deux tests a été évalué à 67 dans le test précédent et à 20 dans ce test. La majorité des signaux de ce test a ici des notes supérieures à 20 : ils auraient donc *a priori* été évalués à plus de 67 dans le premier test. Ceci tend à indiquer que les dégradations liées aux déplacements évalués ici sont moins importantes que les dégradations liées à l'effet de salle.

Les configurations à faible écart angulaire sont peu robustes aux déplacements latéraux, quelle que soit la distance. Même si les systèmes proches sont moins robustes aux déplacements latéraux, la solution du « stéréo-dipôle » ne s'avère donc pas optimale selon ce critère.

L'effet de la distance sur la robustesse a été mis en évidence, et n'est pas toujours celui attendu : dans certain cas les configurations en champ proche sont les plus robustes. En particulier, la configuration d'écartement angulaire +/-60° à 20 cm est la plus robuste aux deux déplacements

testés pour les trois sources virtuelles. Cependant ce résultat n'est pas définitif car tous les déplacements envisageables n'ont pas été testés. Pour ce type de configuration, le contraste naturel est élevé (voir chapitre III), ce qui semble accroître la robustesse aux déplacements. Cet effet va dans le même sens que le test précédent : placer les sources à proximité de l'auditeur peut permettre de préserver le timbre d'origine.

Les notes sont très proches de 100 avec la source virtuelle frontale et traduisent une difficulté à discerner une source spatialisée devant l'auditeur d'une source monophonique : ceci indique probablement une perception intracrânienne de ces stimuli. Cet effet résulte probablement des HRTF très génériques employées ici (modèle sphérique) et les résultats auraient sans doutes été différents avec l'usage de HRTF individualisées.

Parmi les 72 signaux avec effet de déplacement, 22 ont obtenu des notes médianes supérieures ou égales à 90. Le test mis en place ne permet pas de savoir si ces modifications sont imperceptibles ou s'il s'agit d'un effet d'échelle : les autres modifications pourraient être trop importantes, incitant les auditeurs à attribuer une note élevée aux modifications plus tolérables. Le fait que de nombreux auditeurs se soient trompés une fois pour identifier la référence cachée laisse toutefois penser que les différences bien notées sont peu perceptibles. Pour s'en assurer, une perspective consisterait à réaliser un test complémentaire, de type ABX, capable de détecter des modifications très subtiles.

Le test a mis en évidence des différences importantes selon les configurations testées. Il serait souhaitable d'étendre les résultats à d'autres configurations, et de tester d'autres déplacements. Pour évaluer toutes les combinaisons possibles, il faudrait alors multiplier exagérément les tests perceptifs, ce qui n'est pas envisageable concrètement. La suite du travail présenté ici consiste donc plutôt à déterminer un indicateur objectif, issu de l'analyse des signaux et permettant de généraliser les notes obtenues lors du test.

V.4 Détermination d'un indicateur objectif

Seul un indicateur permettant de décrire les effets de déplacement est étudié ici, sans chercher à intégrer l'effet de salle. En effet, l'objectif de cet indicateur n'est pas de prédire les effets cumulés de la salle et d'un déplacement : aucun signal dans le test ne correspond à cette configuration, il serait risqué de vouloir l'estimer sans avoir d'éléments perceptifs de référence. De plus, il y a un fort déséquilibre dans ce test entre le nombre de signaux avec effet de salle et celui des signaux avec effet de déplacement : seulement deux stimuli monophoniques comportant un effet de salle sont présents, alors qu'il y a 72 signaux générés à partir d'un déplacement. Enfin, d'après les résultats du chapitre IV il semblerait que la répartition temporelle de l'énergie permette déjà d'estimer la note prédite pour les signaux comportant un effet de salle.

Le choix des indicateurs testés est issu de remarques formulées de manière informelle par les auditeurs à l'issue du test. Les auditeurs ont en effet évoqué des différences de timbre, de niveau, de localisation et de réverbération. Ces remarques sont étayées par l'analyse de travaux antérieurs.

Dans la littérature, la sonie est considérée de manière générale comme la dimension perceptive ayant le plus d'influence dans les jugements perceptifs [Lavandier, 2005]. Il est rappelé que les signaux avec déplacement n'ont pas été égalisés en sonie : des différences de sonie pouvaient donc apparaître. Par ailleurs, pour l'évaluation d'enceintes acoustiques, il a été montré que la balance spectrale des signaux est une dimension perceptive dominante, une fois la sonie égalisée [Lavandier, 2005]. Dans ce contexte, une dimension supplémentaire a été identifiée, *a priori* liée à la notion d'espace. Ces trois dimensions (sonie, timbre et espace) sont donc envisagées dans la suite.

Par ailleurs, l'usage de modèles auditifs a permis de décrire efficacement les résultats d'évaluation perceptive [Lavandier, 2005]. Pour ces raisons, des indicateurs basés sur des modèles auditifs décrivant les dimensions de sonie, de timbre et de localisation sont présentés, en tentant de les relier aux résultats de notre test.

V.4.A. Calcul d'indicateurs objectifs

Lors du test, il était demandé aux auditeurs d'évaluer une « proximité », et les différents indicateurs présentés estiment au contraire une « dissemblance » entre deux sons car elle est plus simple à objectiver (notion de « distance »). Les comparaisons des auditeurs sont ainsi considérées comme commutatives : la comparaison du son A avec B est supposée équivalente à la comparaison du son B avec le son A. Ceci est intégré dans les indicateurs envisagés ici.

V.4.A.a Dissemblance de sonie

Dans notre cas, les signaux évalués sont binauraux et instationnaires (bouffées de bruit) or il n'existe pas de consensus pour estimer la sonie dans cette situation. Une approximation est réalisée, en utilisant un modèle de sonie stationnaire adapté au signaux binauraux proposé par Sivonen et Ellermeier (voir Annexe I et [Sivonen & Ellermeier, 2008]). Ce modèle considère la sonie binaurale comme étant toujours inférieure à la somme des contributions gauche et droite, avec un paramètre qui est le gain de sommation binaurale g. Une valeur de $g = 3\ dB$ est préconisé par les auteurs, mais d'autres travaux ont permis de montrer que ce paramètre peut varier selon le type de son [Vannier, 2016].

La sonie monaurale est calculée à l'aide de la Loudness Toolbox [Loudness Toolbox, 2016] selon le modèle de la norme ISO 532B [ISO 532B, 1975] en considérant une durée de 1.8 secondes, soit 500 ms de plus que la durée du signal anéchoïque, pour prendre en compte l'effet de réverbération lorsqu'elle est présente. Pour calculer les différents indicateurs de sonie et de timbre, un gain binaural $g = 3.5\ dB$ a été préalablement identifié comme pertinent, son influence sera étudiée dans un second temps.

Un son avec une sonie de N sones est un son N fois plus fort qu'un son d'un sone : l'écart de sonie se traduit par le rapport de deux sonies. L'indicateur associé à la dissemblance de sonie est le rapport entre les sonies de deux signaux, en mettant au dénominateur la plus petite valeur. L'indicateur lié à la sonie est ainsi D_{sonie} :

$$D_{sonie} = \frac{\max\{N_{ref}, N_{sig}\}}{\min\{N_{ref}, N_{sig}\}} - 1 \qquad (60)$$

Où N_{ref} est la sonie du son de référence et N_{sig} la sonie du son à comparer, exprimée en sone. $D_{sonie} \in [0; +\infty[$ avec $D_{sonie} = 0$ lorsqu'il n'y a pas de différence de sonie. La dissemblance de sonie est calculée entre le son de référence et tous les signaux avec déplacement, et tracée en fonction des notes attribuées lors du test en Figure 48. La dissemblance de sonie n'est pas équivalente pour tous ces sons, traduisant que cet indicateur décrit effectivement des variations entre les signaux. En particulier, les dissemblances de sonie sont les plus élevées pour les notes comprises entre 40 et 70. Toutefois, l'indicateur ne suffit pas à décrire les notes du test : il n'y a pas une relation monotone entre D_{sonie} et les notes du test.

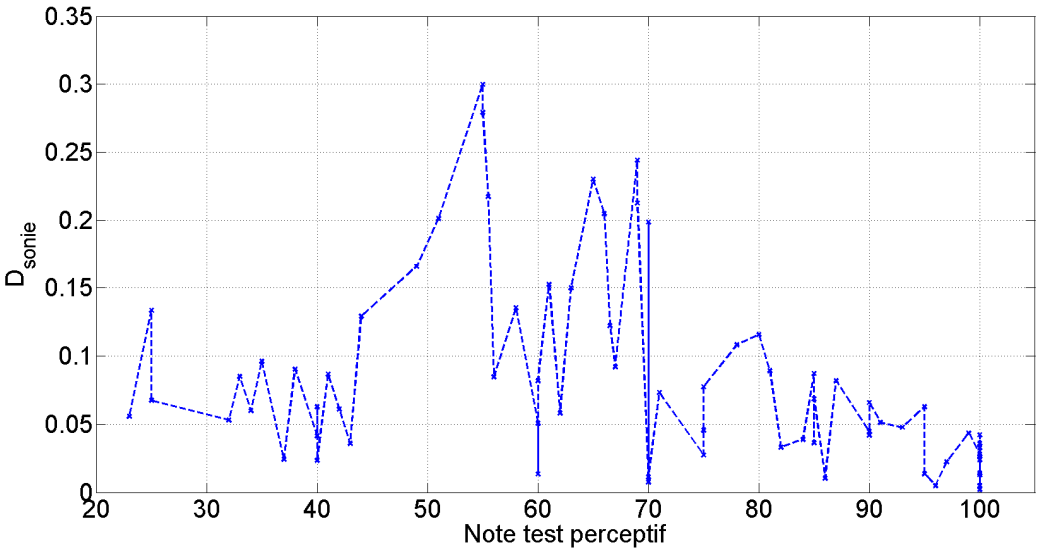


Figure 48 : Dissemblance de sonie estimée à partir des signaux en fonction des notes attribuées par les auditeurs lors du test d'écoute.

V.4.A.b Dissemblance de timbre

Le timbre est considéré comme une dimension perceptive dominante dans les jugements de similarité [Lavandier, 2005]. Il existe plusieurs approches pour décrire le timbre perçu à partir de signaux [Lavandier, 2005; Peeters, 2004], et pour cette raison, plusieurs implémentations sont testées pour déterminer celle qui correspond le mieux aux données du test.

Pour estimer les dissemblances de timbres, l'approche est similaire aux travaux réalisés dans [Lavandier, 2005; Michaud et al., 2015], qui se basaient sur la sonie spécifique estimée à partir d'un modèle monaural. La sonie spécifique $S(E)$ correspond au niveau sonore perçu pour une bande de fréquences E, et peut s'estimer à partir d'un modèle de sonie. Elle est ici estimée à partir du même modèle que précédemment (modèle de sonie binaurale de Sivonen et Ellermeier [Sivonen & Ellermeier, 2008] avec $g = 3.5\ dB$), l'échelle des fréquences étant alors l'échelle des Barks.

Dans la suite du calcul, la sonie spécifique est normalisée par la sonie binaurale totale. Cette opération permet alors de dissocier l'influence de la sonie de celle du timbre.

i Dissemblance globale

Pour une bande de fréquences donnée E, la dissemblance de timbre est donnée par la relation :

$$D(E) = \begin{cases} \dfrac{S_{ref}(E)}{S_{sig}(E)} - 1 & si\ S_{ref} > S_{sig} \\ -\left[\dfrac{S_{sig}(E)}{S_{ref}(E)} - 1\right] & si\ S_{sig} > S_{ref} \end{cases} \tag{61}$$

Avec S_{ref} la sonie spécifique du signal de référence et S_{sig} celle du signal comparé. La Figure 49 illustre le calcul de dissemblance entre deux signaux.

Deux indicateurs de timbre sont calculés à partir de cette dissemblance. Le premier indicateur est analogue à celui employé par Mathieu Lavandier dans sa thèse [Lavandier, 2005], et désigné ici par dissemblance globale ou D_{glob}. Il s'agit des dissemblances timbrales totales, qui sont la somme des dissemblances de chaque bande de fréquences :

$$D_{glob} = \sum_{E=2}^{E_{max}} |D(E)| \tag{62}$$

Où E désigne l'indice de bande Bark, et E_{max} l'indice maximal fixé à 23, correspondant à une fréquence maximale de 12 kHz. L'indice minimal est fixé à 2, correspondant à une fréquence minimale de 100 Hz. Cet indicateur correspond à l'aire sous la courbe de la Figure 49 (en gris). $D_{glob} \in [0; +\infty[$ avec $D_{glob} = 0$ en l'absence de différences.

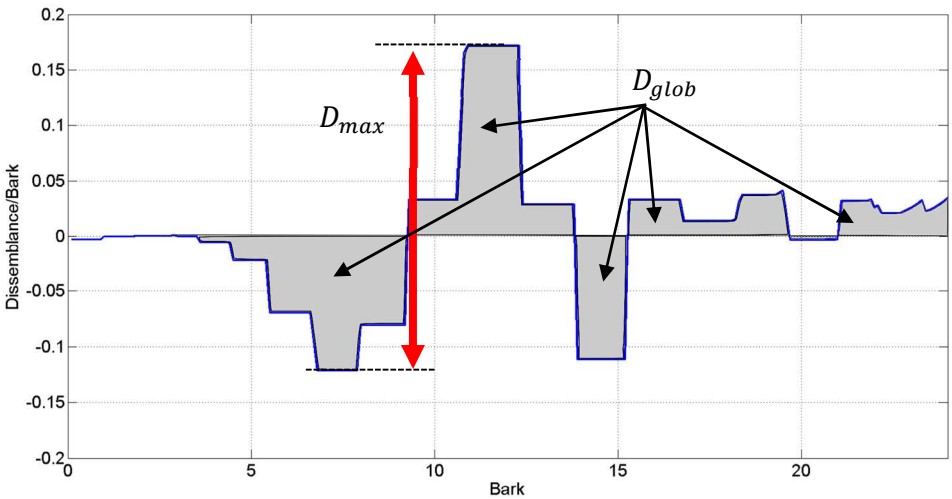


Figure 49 : dissemblance de timbre entre le signal de référence et le signal comparé

La dissemblance de timbre global est calculée pour tous les signaux du test avec déplacement, et reportée en fonction des notes attribuées lors du test en haut à gauche de la Figure 50. La dissemblance de timbre décroit assez systématiquement lorsque les notes du test augmentent. Cet indicateur semble donc mieux adapté que la sonie pour décrire les notes obtenues. Des fluctuations sont toutefois visibles, signifiant que cet indicateur seul n'est pas suffisant : il peut être nécessaire de l'enrichir.

ii Dissemblance d'écart maximal

Un autre indicateur de timbre est calculé, basé sur l'écart maximal entre bandes de fréquences. Cet indicateur est censé révéler l'émergence de certaines bandes de fréquences par rapport à d'autres. Cet indicateur dit de dissemblance d'écart maximal D_{max} est calculé de la manière suivante :

$$D_{max} = \max_{E}[D] - \min_{E}[D] \tag{63}$$

La dissemblance d'écart maximal est illustrée dans la Figure 49 par le trait rouge. $D_{max} \in [0; +\infty[$ avec $D_{max} = 0$ en l'absence de différences.

La dissemblance de timbre d'écart maximal est calculée pour tous les signaux avec déplacement du test, et tracée en fonction des notes attribuées lors du test, en haut à droite de la Figure 50. Là aussi, on peut constater que la dissemblance a tendance à décroître lorsque les notes augmentent. Les fluctuations sont toutefois plus importantes que dans le cas de D_{glob}, donc cet indicateur serait plutôt moins adapté à décrire les résultats du test.

iii Centre de gravité spectral

Une autre manière d'identifier des variations de timbre est de calculer le centre de gravité spectral (CGS). Il se définit comme le barycentre fréquentiel de l'énergie en Hz [Peeters, 2004]. Il est ici calculé d'après la sonie spécifique définie précédemment :

$$CGS = \frac{\sum_{i=1}^{E_{max}} S(E_i)E_i}{\sum_{i=1}^{E_{max}} S(E_i)} \tag{64}$$

Avec $S(E_i)$ la sonie spécifique binaurale de la bande de Bark E_i. La dissemblance de CGS est définie par la valeur absolue de la différence entre le CGS du signal de référence et celui du signal à comparer :

$$D_{CGS} = |CGS_{ref} - CGS_{sig}| \tag{65}$$

Avec CGS_{ref} et CGS_{sig} les centres de gravité spectraux du signal de référence et du signal à comparer. Cet indicateur est à valeurs dans $[0; +\infty[$ où une valeur nulle traduit l'absence de différence.

La dissemblance de CGS est calculée pour tous les signaux avec déplacement du test, et tracée en fonction des notes attribuées lors du test en bas à gauche de la Figure 50. Par rapport aux deux précédents indicateurs de timbre, le lien avec les résultats du test est moins marqué, même si une tendance générale semble montrer que les plus grandes dissemblances sont attribuées aux signaux dont la note est faible.

iv Dissemblance d'étalement spectral

L'étalement spectral est défini de la manière suivante [Peeters, 2004] :

$$ET = \frac{\sum_{i=1}^{E_{max}} (E_i - CGS)^2 S(E_i)}{\sum_{i=1}^{E_{max}} S(E_i)} \tag{66}$$

La dissemblance d'étalement de spectre est calculée de manière analogue au CGS :

$$D_{ET} = |ET_{ref} - ET_{sig}| \tag{67}$$

Cet indicateur est à valeurs dans $[0; +\infty[$ où une valeur de 0 traduit une absence de différences. La dissemblance D_{ET} est calculée pour tous les signaux avec déplacement du test, et tracée en fonction des notes attribuées lors du test, en bas à droite de la Figure 50. L'allure de cette représentation est similaire à l'indicateur CGS : le lien inversement proportionnel entre dissemblance et note est visible, mais d'importantes fluctuations apparaissent entre des signaux ayant des notes similaires.

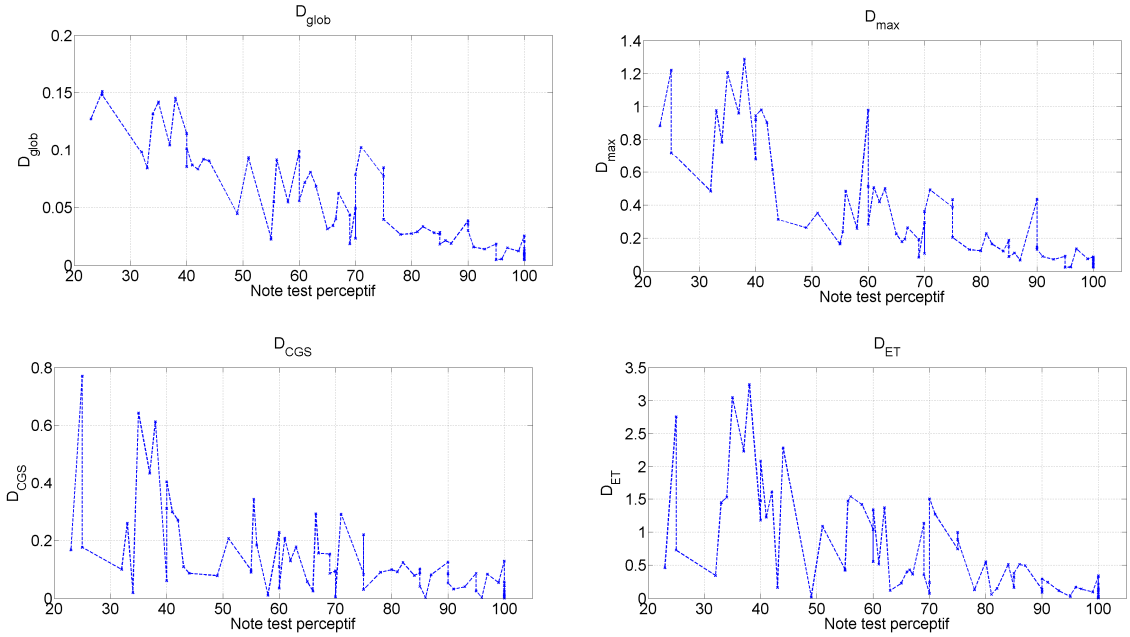


Figure 50 : Dissemblances de timbre calculées à partir des signaux en fonction des notes attribuées par les auditeurs lors du test d'écoute.

V.4.A.c Dissemblance de localisation

Enfin, le dernier type d'indicateur calculé concerne la perception de la localisation à partir de l'ITD et l'ILD. Même si les auditeurs sont censés être moins sensibles à cet aspect, il pourrait être un complément d'information utile.

L'ITD est déterminé à partir d'une détection de seuil à -12 dB du maximum de l'énergie des réponses impulsionnelles, comme au chapitre II. L'indicateur associé correspond à l'écart par rapport à l'ITD du signal de référence :

$$D_{ITD} = \left|ITD_{ref} - ITD_{sig}\right| \tag{68}$$

L'ILD est défini comme étant le rapport de sonie calculée entre les deux oreilles. Les sonies N_L et N_R des oreilles gauche et droite respectivement sont calculées indépendamment pour chaque oreille avec la définition de la norme ISO 532B. L'ILD est défini de la manière suivante :

$$ILD = \begin{cases} \frac{N_L}{N_R} - 1 & si\ N_L > N_R \\ -\left[\frac{N_R}{N_L} - 1\right] & si\ N_R > N_L \end{cases} \quad (69)$$

L'indicateur associé à l'ILD est défini de manière analogue à D_{ITD} :

$$D_{ILD} = \left|ILD_{ref} - ILD_{sig}\right| \quad (70)$$

Ces deux indicateurs sont à valeurs dans $[0; +\infty[$ où une valeur nulle traduit une absence de différences. Les dissemblances de localisation sont calculées pour tous les signaux du test « avec déplacement », et tracées en Figure 51 en fonction des notes attribuées lors du test. Aucune tendance claire n'est visible entre la dissemblance d'ITD et les notes du test. Toutefois, l'indicateur fluctue selon les signaux : il pourrait éventuellement apporter une information complémentaire à d'autres indicateurs. Cependant, des valeurs de dissemblance élevée peuvent paradoxalement être observées pour des signaux qui ont des notes élevées, dont les différences sont quasiment imperceptibles.

Pour l'indicateur d'ILD, la tendance générale n'est pas non plus évidente, mais des valeurs de dissemblances élevées apparaissent pour des signaux notés au milieu de l'échelle (entre 40 et 70). Ce comportement particulier a également été relevé pour l'indicateur de sonie : ces deux indicateurs décrivent peut-être un aspect commun.

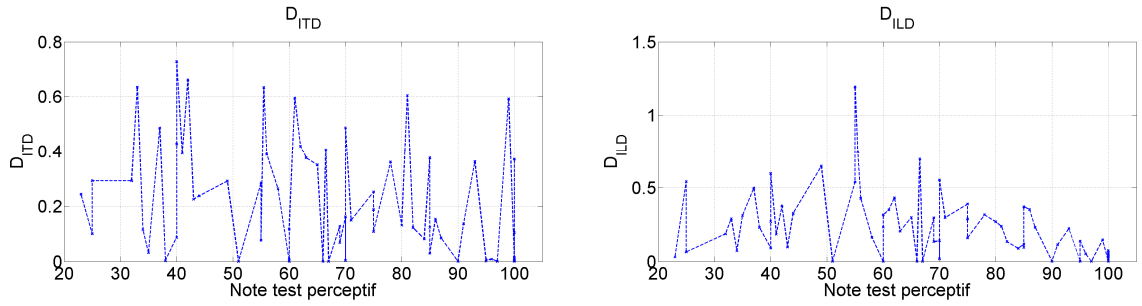


Figure 51 : dissemblances de localisation calculées à partir des signaux en fonction des notes attribuées par les auditeurs lors du test perceptif

V.4.A.d Corrélation entre indicateurs

Plusieurs indicateurs réagissent qualitativement de la même manière aux signaux du test. La démarche est maintenant de combiner plusieurs indicateurs, or il n'est pas nécessaire de le faire pour des indicateurs qui portent la même information. Pour quantifier le lien entre chaque indicateur, la matrice de corrélation entre les différents indicateurs (calculée à partir de l'équation (59)) est reportée en Figure 52.

Le groupe des indicateurs de timbre émerge significativement des autres : le coefficient de corrélation est supérieur à 0.65 entre ces indicateurs, et inférieur à 0.4 avec les autres indicateurs. Il ne semble donc pas utile de combiner deux indicateurs de timbre.

L'indicateur de sonie est peu corrélé avec tous les autres indicateurs, à l'exception de l'indicateur d'ILD (coefficient de corrélation égal à 0.46). Ceci résulte probablement de la normalisation de la sonie spécifique pour les calculs d'indicateurs de timbre. Le modèle de sonie binaurale est basé sur une combinaison des contributions des deux oreilles, alors que l'ILD correspond à la différence des deux contributions. Une modification significative du signal sur l'une des deux oreilles pourrait ainsi modifier l'indicateur de sonie et l'indicateur d'ILD de manière similaire.

Les indicateurs de localisation (ITD et ILD) sont moyennement corrélés : le coefficient de corrélation vaut 0.5. Ces deux indicateurs portent donc eux aussi une part d'information commune.

La combinaison de trois indicateurs (timbre, sonie et ITD) semble *a priori* pertinente car elle correspond à des composantes peu corrélées : l'information serait ainsi codée de manière « parcimonieuse ». Toutefois, la matrice de corrélation des indicateurs ne permet pas de conclure quant à la meilleure manière de grouper les indicateurs et il est nécessaire d'en tester plusieurs combinaisons.

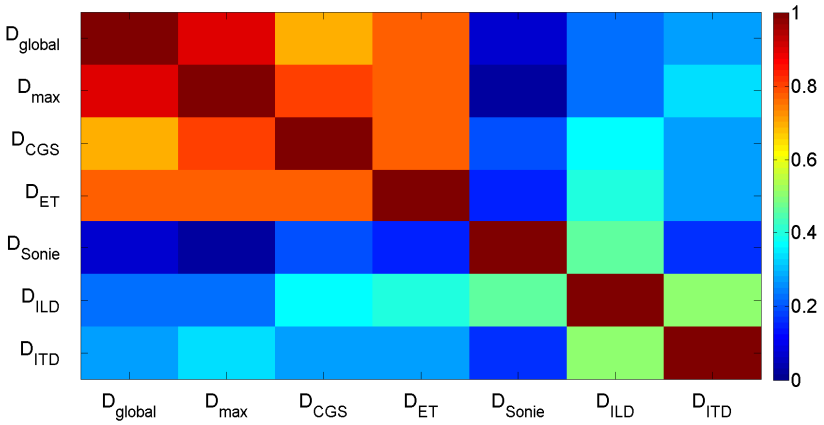


Figure 52 : valeur absolue du coefficient de corrélation entre les différents indicateurs testés

V.4.B. Combinaison d'indicateurs

Plusieurs indicateurs ont été proposés, certains étant très corrélés entre eux. La démarche suivante vise à déterminer la meilleure combinaison de ces indicateurs.

V.4.B.a Effet du nombre d'indicateur

Une note $Note_{prédite}$ est calculée à partir d'un modèle de régression multiple combinant plusieurs des indicateurs précédents :

$$Note_{prédite} = \sum_{i=1}^{I} A_i I_i + 100 \qquad (71)$$

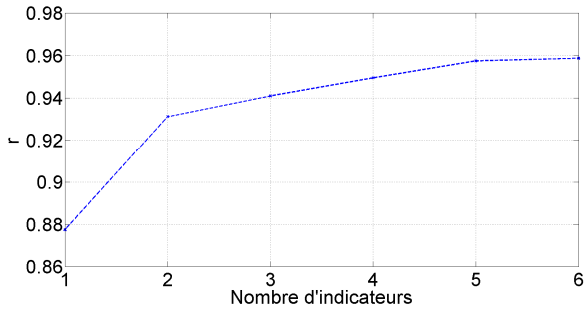
Avec I_i la valeur de l'indicateur d'indice i et A_i les paramètres de la régression minimisant l'erreur de prédiction au sens des moindres carrés. Pour tous les indicateurs évalués, une absence de différence se traduit par une valeur nulle. Dans ce cas, la note attribuée est fixée à 100, le

maximum de l'échelle. L'estimation des paramètres A_i est réalisée avec la fonction *regress* de la Statistics Toolbox de Matlab. I est le nombre d'indicateurs pris en compte dans le modèle.

Pour déterminer le nombre minimal d'indicateurs permettant de décrire au mieux les données, toutes les combinaisons d'indicateurs possibles sont testées en utilisant un à six indicateurs. Pour évaluer la pertinence de l'indicateur, le coefficient de corrélation est calculé entre les notes prédites et les réponses des auditeurs. Le coefficient maximal obtenu pour un nombre d'indicateurs donné est reporté dans la Figure 53, sa valeur correspondant à la meilleure combinaison possible. Les indicateurs correspondant à la meilleure combinaison sont reportés sur le tableau à droite de cette figure.

Avec un seul indicateur, le coefficient de corrélation vaut 0.88 et l'indicateur associé est la dissemblance de timbre global. Un seul indicateur semble suffisant pour décrire assez bien les données. Toutefois, l'estimation peut être améliorée : avec deux indicateurs le coefficient de corrélation vaut 0.93, et l'indicateur supplémentaire est la sonie. Pour trois indicateurs, incluant en plus l'ITD, le coefficient de corrélation atteint 0.94. La différence de corrélation obtenue entre un et deux indicateurs est significative, en revanche la différence avec un nombre supérieur d'indicateurs est ténue. Les indicateurs D_{global} et D_{sonie}, qui sont très peu corrélés entre eux (coefficient de 0.08 en Figure 52) semblent donc les plus pertinents pour décrire les résultats du test. L'ajout d'un troisième indicateur pour représenter les données est discutable : il pourrait être l'ITD, qui est moyennement corrélé avec les deux premiers. Par ailleurs, l'ITD est le seul à décrire des variations temporelles et peut s'obtenir simplement à partir de la RI. Toutefois, l'apport de cet indicateur est limité : la corrélation avec les résultats du test n'augmente pas significativement.

L'augmentation du nombre d'indicateurs tend mécaniquement à augmenter ses performances pour décrire les données d'apprentissage. En revanche, un modèle plus complexe peut s'écarter de la solution optimale en dehors des points d'estimation, le rendant plus « fragile ». Pour cette raison, il est préférable d'utiliser un modèle simple. Le modèle à deux indicateurs est donc retenu, en tant que meilleur compromis entre simplicité et efficacité.



	D_{sonie}	D_{glob}	D_{max}	D_{CGS}	D_{ET}	D_{ITD}	D_{ILD}
1		■					
2	■	■					
3	■	■				■	
4	■	■	■	■			
5	■	■	■	■			■
6	■	■	■	■		■	■

Figure 53 : Figure de gauche : coefficient de corrélation maximal en fonction du nombre d'indicateurs. Tableau de droite : indicateurs sélectionnés en fonction du nombre d'indicateurs. Une case grisée correspond à un indicateur sélectionné

V.4.B.b Adaptation du gain de sommation binaurale

Le modèle de sonie binaurale se paramètre avec g le gain de sommation binaurale, et des valeurs de gain différentes peuvent être appliquées selon les cas [Vannier, 2016]. Différentes valeurs de gain sont testées et le coefficient de corrélation est calculé avec les trois indicateurs D_{glob}, D_{sonie}

et D_{ITD}. Dans chaque cas étudié, le paramètre g est identique pour le calcul de sonie et le calcul de timbre. L'évolution de ce coefficient de corrélation en fonction du paramètre g est reportée dans la Figure 54. Ce paramètre a un effet sur le coefficient de corrélation, variant de 0.90 à plus de 0.93 pour les valeurs testées. La valeur maximale du coefficient de corrélation 0.93 est obtenue pour g=3.5 dB et cette valeur est donc utilisée dans la suite. Elle correspond à une valeur proche de celle qui est proposée par les auteurs [Sivonen & Ellermeier, 2008] et correspond à la valeur proposée par [Vannier, 2016] dans le cas de bruit à bande étroite (alors que dans notre cas, le bruit est large bande mais non-stationnaire). La valeur de g n'est toutefois pas critique dans le cadre de ce test : le coefficient de corrélation est toujours supérieur à 0.91 pour g compris entre 2 et 5.5 dB.

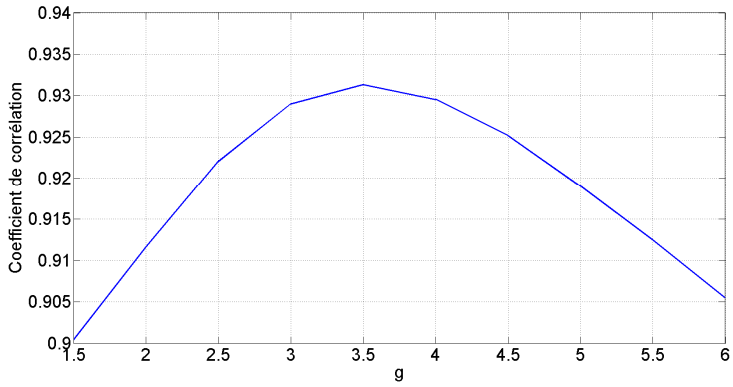


Figure 54 : Coefficient de corrélation obtenu avec les indicateurs D_{glob}, D_{sonie} et D_{ITD} en fonction du paramètre g

V.4.B.c Modèle non-linéaire et modèle avec interactions

Le modèle à deux indicateurs semble satisfaisant, mais des possibilités d'améliorations sont tout de même explorées. Le modèle présenté dans l'équation (71) considère que chaque indicateur est lié de manière linéaire avec les données. En utilisant les mêmes conventions de notation que pour l'équation (71) un modèle non-linéaire peut s'écrire de la manière suivante :

$$Note_{prédite} = \sum_{i=1}^{I} A_i . I_i^{K_i} + 100 \qquad (72)$$

Avec K_i un paramètre associé à chaque indicateur. Les paramètres correspondant aux données sont estimés à partir d'un algorithme de Levenberg-Marquardt [Seber & Wild, 2003] implémenté dans la fonction *nlinfit* de la Statistics Toolbox de Matlab. L'algorithme est initialisé avec les paramètres A_i minimisant l'erreur au sens des moindres carrés du modèle linéaire ($K_i = 1$).

En utilisant les deux mêmes indicateurs que précédemment (D_{sonie} et D_{glob}) l'algorithme converge vers une solution où les K_i s'écartent faiblement de 1. La corrélation avec les notes du test alors obtenue est très similaire au cas linéaire et le modèle non-linéaire ne sera donc pas retenu.

Les indicateurs sélectionnés sont faiblement corrélés entre eux, mais leur corrélation n'est pas nulle et il est possible que les indicateurs interagissent entre eux : un déplacement peut entraîner une modification du timbre et de la sonie sans que les effets se cumulent d'un point de vue

perceptif. Un modèle prenant en compte les interactions est alors envisagé, elles y sont modélisées à partir du produit des différents indicateurs. Le modèle est alors le suivant :

$$Note_{prédite} = \sum_{i=1}^{I=2} A_i I_i + \sum_{i=1}^{I=2} \sum_{j=1}^{I=2} C_{ij} I_i I_j + 100 \qquad (73)$$

Ce type de modèle permet d'obtenir un coefficient de corrélation de 0.94 ce qui est quasiment équivalent au cas sans interactions. La prise en compte des interactions ne semble donc pas avoir un intérêt majeur, alors qu'elle complexifie davantage le modèle augmentant le risque de diverger en dehors du domaine d'identification. L'objectif étant de généraliser les résultats, cette complexification n'est donc pas souhaitable, et les interactions sont négligées par la suite.

V.4.B.d Objectivation à partir de la RI

L'indicateur proposé est relativement complexe, puisqu'il est basé sur des calculs de sonie pour chaque signal écouté. La généralisation des résultats à partir de cet indicateur nécessite donc de simuler la reproduction d'un signal temporel pour en estimer la sonie.

Dans le but de mettre en place un outil prédictif de rendu de systèmes spatialisés, un modèle plus simple est préférable. Dans l'idéal, ce modèle se calculerait directement à partir des réponses impulsionnelles : elles sont indépendantes du signal, peuvent se mesurer sur une installation, et peuvent éventuellement être modélisées.

Pour se rapprocher des conditions de test, les réponses impulsionnelles $out_L(t)$ et $out_R(t)$ sont préalablement filtrées par un filtre à -3 dB par octave, pour « rosir » les réponses impulsionnelles car le signal écouté pendant le test est une succession de bouffée de bruit rose.

Le niveau de reproduction était 30 sones, soit 89 phones. Pour approximer la sensibilité de l'oreille en fonction de la fréquence, une pondération B est appliquée par filtrage. Un indicateur est alors calculé selon le schéma précédent, mais à partir de ces RI filtrées (donc sans utiliser de modèle de sonie). Pour cela, une dissemblance de niveau D_{niv} est calculée par analogie à la dissemblance de sonie D_{sonie}. Le niveau monaural Niv_L de l'oreille gauche est défini de la manière suivante :

$$Niv_L = 10 \log_{10}[\langle out_L^2(t) \rangle] \qquad (74)$$

Où $\langle . \rangle$ est la moyenne arithmétique et out_L la réponse impulsionnelle à l'oreille gauche. Le niveau monaural droit Niv_R se définit de la même manière à partir de out_R. Le niveau binaural Niv est calculé de manière analogue à la sonie binaurale avec $g = 3.5\ dB$:

$$Niv = g log_2(2^{\frac{Niv_L}{g}} + 2^{\frac{Niv_R}{g}}) \qquad (75)$$

A partir du niveau binaural du signal de référence Niv_{ref} et du signal évalué Niv_{sig}, une dissemblance de niveau est calculée :

$$D_{niv} = |Niv_{ref} - Niv_{sig}| \qquad (76)$$

Une dissemblance de timbre D_{timbre} est calculé de manière analogue à D_{global}. Pour cela, les niveaux par bandes de tiers d'octaves sont calculés à partir de l'équation (75) où Niv_L et Niv_R

correspondent aux niveaux en tiers d'octave des réponses impulsionnelles filtrée rose, avec une pondération B, et normalisée. La dissemblance de timbre est alors :

$$D_{timbre} = \sum_{i=1}^{i_{max}} \left| Niv_{ref}(i) - Niv_{sig}(i) \right| \tag{77}$$

Avec i l'indice de bande de tiers d'octave et i_{max} l'indice maximal. Le calcul est effectué pour les bandes de tiers d'octave dont la fréquence centrale varie de 100 Hz à 12.5 kHz. La note prédite est estimée à partir d'une combinaison de ces deux indicateurs, minimisant l'erreur de prédiction au sens des moindres carrés (cf V.4.C). Le coefficient de corrélation entre les notes du test et les notes prédites est de 0.94, ce qui est équivalent au modèle basé sur la sonie. Un modèle non-linéaire ne permet pas d'augmenter la corrélation, et le paramètre g optimal reste de 3.5 dB. La pondération C permet d'obtenir le même résultat, alors que l'utilisation d'une pondération A est légèrement moins précise, le coefficient de corrélation atteignant dans ce cas 0.92.

V.4.B.e Robustesse de l'indicateur

Ce paragraphe vise à évaluer la robustesse du modèle simplifié. Pour cela, les paramètres du modèle sont estimés à partir des deux tiers des données, et les données associées au tiers restant sont prédites avec l'indicateur et comparées aux résultats du test. Les données permettant d'estimer les paramètres du modèle sont désignées par « données d'apprentissage » et les autres sont désignées par « données d'évaluation ». Le coefficient de corrélation est calculé pour les données d'évaluation pour 2000 tirages de données d'apprentissage. Le coefficient de corrélation moyen obtenu pour ces 2000 tirages est 0.94, ce qui correspond au coefficient de corrélation lorsque toutes les données sont prises en compte. Le coefficient de corrélation minimal pour ces 2000 tirages est de 0.86, ce qui est significativement inférieur au cas où toutes les données sont prises en compte. Cela traduit une certaine fragilité de l'indicateur à prédire les points en dehors des données d'apprentissage, mais ce cas de figure n'est observé que pour le « pire » cas, et qui reste bien corrélé.

V.4.C. Modèle retenu

V.4.C.a Paramètres du modèle

Le modèle retenu pour prédire la note se définit de la manière suivante :

$$Note_{prédite} = A_1 D_{niv} + A_2 D_{timbre} + 100 \tag{78}$$

Les coefficients de ce modèle sont reportés dans le Tableau 10. Les valeurs des coefficients A_1 et A_2 sont toutes négatives, traduisant que l'augmentation d'une des valeurs de dissemblance conduit à une diminution de la note prédite, ce qui est cohérent.

Pour évaluer le poids relatif de chacun des deux indicateurs, les variables D_{niv} et D_{timbre} sont préalablement centrées et réduites. Un autre jeu de paramètres « normalisés » A_1 et A_2 est alors calculé pour comparaison, et reporté dans le Tableau 10. Le coefficient normalisé associé à la dissemblance de timbre est plus élevé en valeur absolue que celui associé au niveau. Cela est cohérent avec le fait que les dissemblances de timbre permettent de mieux décrire les données

du test. La normalisation n'a pas d'influence sur les performances de l'indicateur, et n'est donc plus utilisée dans la suite.

	A_1	A_2
Coefficient	-5.78	-1.24
Coefficient normalisé	-6.84	-19.29

Tableau 10 : valeurs des paramètres du modèle retenu

Le modèle retenu s'avère être celui qui permet d'obtenir la prédiction la plus corrélée aux résultats du test (r=0.94) avec deux indicateurs, et il semble être suffisament robuste lorsque les données d'évaluation ne sont pas les mêmes que les données d'apprentissage.

V.4.C.b Comparaison du modèle aux résultats du test

Les notes attribuées par les auditeurs sont représentées par la Figure 55 en fonction des notes estimées avec le modèle retenu. Ces dernières sont également représentées par la Figure 56, de la même manière que pour les notes obtenues lors du test (Figure 47). Les ancres basse et intermédiaire ne sont pas représentées, car l'indicateur proposé ne permet pas d'évaluer l'effet de salle. La différence entre la note prédite et la note obtenue lors du test est représentée par la Figure 57.

Concernant l'ancre haute (signal monophonique), l'erreur de prédiction dépend de l'écart angulaire des sources physiques : pour la SRV 45° l'erreur est de 26 pour l'écartement +/-30° et de -2 pour l'écartement +/-90°. Il s'agit pourtant du même signal dans les deux cas, la note prédite est donc identique (Figure 56). Le critère objectif proposé ici est « absolu » : il dépend uniquement du signal, alors que le jugement des auditeurs peut dépendre de la présence d'autres signaux. Les variations entre résultats du test et indicateur objectif sont toutefois inclues dans les écarts interquartiles : pour la configuration à +/-30°, l'écart interquartile est supérieur à 50.

Pour l'écartement angulaire +/-5°, l'erreur de prédiction est particulièrement élevée pour les SRV 45° et 90° pour le déplacement +5Y. Il apparait alors que les notes obtenues par la modélisation sont significativement plus basses avec la SRV 0°, alors que les résultats du test ne permettaient pas d'identifier de différence significative entre les SRV. Cependant, les notes obtenues avec le test pour les déplacements +5Y pour l'écartement angulaire +/-5° sont toutes inférieures à 45 : il s'agit donc de configurations peu robustes. L'erreur de prédiction tend donc à dévaluer ces configurations, déjà parmi les moins bien notées. Par ailleurs, pour l'écartement angulaire de +/-90° et le déplacement de +5Y, l'erreur de prédiction pour la SRV 0° est de 24. La modélisation de la note pour cette configuration n'est alors pas significativement différente de la SRV 45°, alors que les auditeurs ont identifié des différences.

De plus, pour les systèmes à +/-30° l'erreur de prédiction est négative pour le déplacement +5Y pour la SRV à 45°, et augmente en valeur absolue avec la distance atteignant -16 à 80 cm. L'indicateur sous-estime donc légèrement les configurations en champ lointain dans ce cas. La décroissance des notes avec la distance est visible sur les notes obtenues lors du test, mais l'erreur de prédiction tend à amplifier légèrement cette tendance.

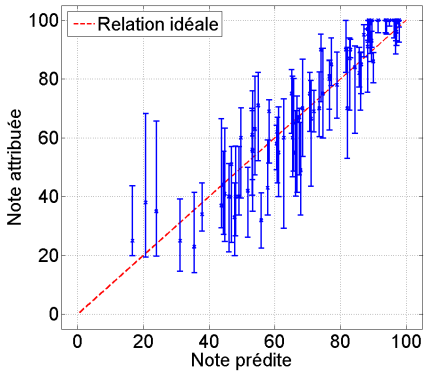


Figure 55 : notes médianes attribuées lors du test en fonction des notes prédites par l'indicateur (sauf les ancres intermédiaire et basse)

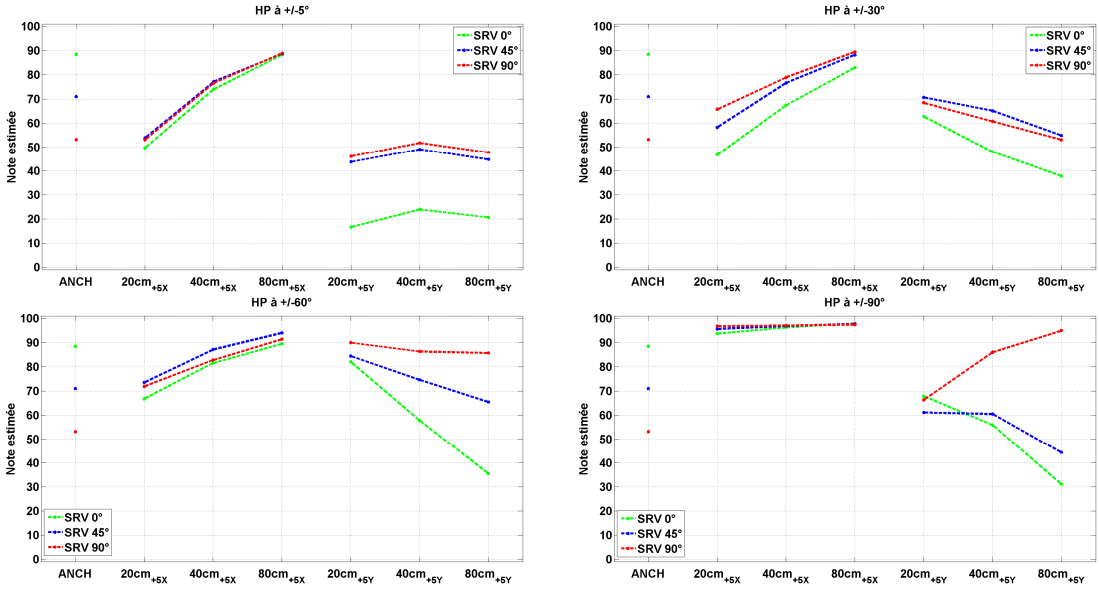


Figure 56 : Notes modélisées par l'indicateur retenu. Chaque figure correspond à un écart angulaire de sources physiques : En haut à gauche : +/-5°, en haut à droite : +/-30°, en bas à gauche : +/-60°, en bas à droite : +/-90°.

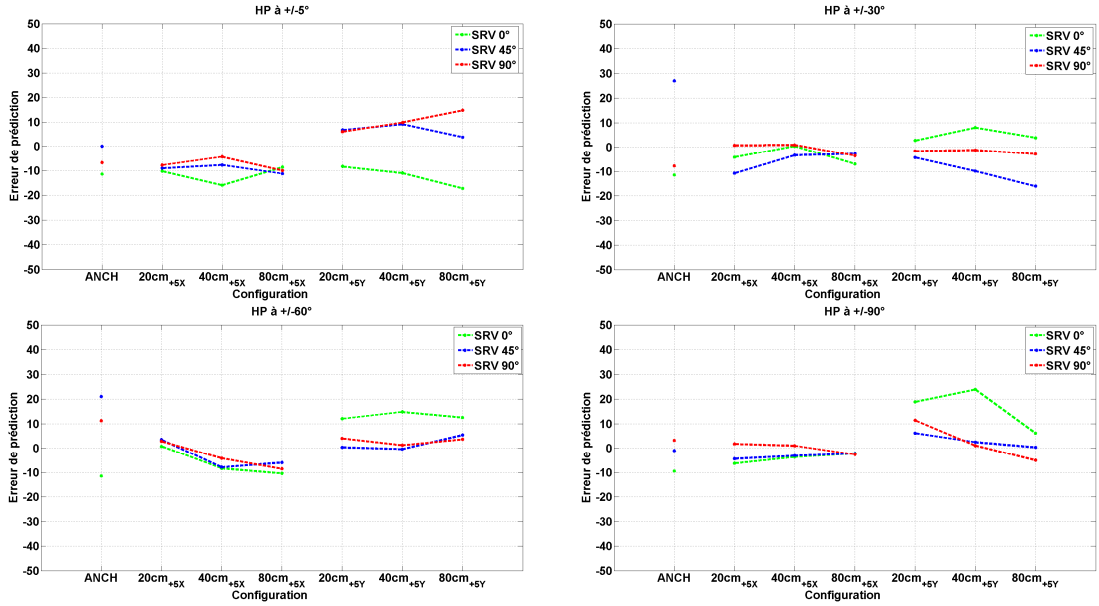


Figure 57 : erreur de prédiction pour les différents écartements angulaire. En haut à gauche : +/-5°, en haut à droite : +/-30°, en bas à gauche : +/-60°, en bas à droite : +/-90°.

Un certain nombre de différences sont observées entre les notes prédites et les notes attribuées par les auditeurs : l'indicateur objectif reste, assez logiquement, moins révélateur qu'un test perceptif. L'utilisation de l'ITD pour complexifier le modèle a été testée, mais ne permet pas d'améliorer significativement l'erreur de prédiction. Globalement, les erreurs de prédictions sont cependant toujours inférieures à 25, et dans la majeure partie des cas inférieures à 15. Ces erreurs sont tolérables, car les écarts interquartiles sont de cet ordre de grandeur : les erreurs de prédiction sont du même ordre de grandeur que l'incertitude de mesure liée à la taille restreinte de l'échantillon des auditeurs sondés.

V.5 Conclusion du chapitre

Un test perceptif a été mis en place, permettant d'évaluer le rendu sonore de systèmes transauraux lorsque l'auditeur s'est déplacé par rapport à la position optimale. Pour cela, le rendu sonore a été simulé à partir d'un modèle de sphère et le rendu diffusé au casque d'écoute.

Seuls deux types de déplacements ont été testés, pour de nombreuses configurations de systèmes transauraux : quatre écartements angulaires (de +/-5° à +/-90°), trois distances (20 cm, 40 cm et 80 cm) et pour trois incidences de sources virtuelles (0°, 45°, 90°). Les deux déplacements testés, considérés comme les déplacements les plus critiques, correspondent à deux translations de 5 cm : l'une vers le côté gauche et une l'autre vers l'avant.

Les signaux ont été élaborés à partir de fonctions de transfert calculées sur un modèle de sphère. Les HRTF employées sont donc assez sommaires car elles ne prennent pas en compte les pavillons d'oreille, les épaules et le torse. La perception spatiale peut alors n'être qu'approximative, et certaines différences perceptives n'ont peut-être pas été mises en évidence. Ce choix a été effectué par simplicité, dans le cadre d'une approche préliminaire pour sélectionner des configurations dont le fonctionnement semble optimal. Ces résultats mériteront d'être approfondis en prenant en compte la complexité des HRTF individuelles.

Les réponses des dix-sept auditeurs ont permis de dégager plusieurs éléments. Les résultats de ce test ont pu être comparés à ceux du test précédent qui consistait à comparer l'effet de salle en restitution monophonique. Cette comparaison a permis de montrer que les différences évaluées ici sont plus fines : un son évalué à 67 dans le premier test est évalué à 20 dans le second. De manière générale, les sons issus d'un déplacement semblent moins dissemblants que les sons avec un effet de salle ce qui confirmerait que l'effet de la salle d'écoute soit celui qui modifie le plus le rendu sonore. Ce point peut cependant résulter en partie de notre protocole (choix des signaux, écoute diotique pour l'effet de salle).

De manière générale, les configurations en proximité sont plus robustes aux déplacements vers l'avant et les configurations plus éloignées sont plus robustes aux déplacements latéraux. De plus, quelle que soit la distance les configurations à faible écart angulaire sont plus robustes aux déplacements vers l'avant et les configurations avec un grand écart angulaire sont les plus robustes aux mouvements latéraux. Ainsi la configuration à 20 cm et +/-60° est la plus robuste du test avec des notes toujours supérieures à 65, et la configuration à 20 cm et +/- 5° est la moins robuste du test avec des notes toujours inférieures à 65.

Le fait que certaines configurations en proximité de l'auditeur soient robustes aux déplacements contredit notre intuition initiale. Cette contradiction est significative, notamment dans le cas des déplacements vers l'avant pour lesquelles les configurations en champ proche sont les plus robustes. Il existe une correspondance entre les configurations robustes et les configurations pour lesquelles le filtrage est simplifié par le contraste naturel évoqué au chapitre III. Cet effet est très intéressant puisqu'il va dans le même sens que l'effet de salle : plus les sources sont proches de l'auditeur et plus le timbre de la restitution est préservé. Cette remarque doit cependant être nuancée puisqu'il faut rappeler que l'approche n'est pas exhaustive et tous les déplacements envisageables n'ont pas été testés.

Pour évaluer d'autres déplacements, un indicateur objectif permettant de prédire les notes du test a été implémenté. Plusieurs indicateurs associés à des perceptions identifiées ont été calculés : la sonie, le timbre et la localisation. Une combinaison de la sonie binaurale et du timbre global permet de prédire les notes du test avec une corrélation de 0.94. Une version simplifiée de l'indicateur permet d'obtenir un résultat très proche en se basant sur le calcul du niveau du signal et de niveaux par bandes de tiers d'octave, à partir des réponses impulsionnelles. Cet indicateur est préféré car il est indépendant du signal de test, et plus simple à implémenter. En excluant les ancres, l'erreur de prédiction avec cet indicateur est toujours inférieure à 25, et généralement inférieure à 15. Ces écarts sont considérés comme modérés car d'un ordre de grandeur voisin des plus petits écarts interquartiles.

Chapitre VI Optimisation globale du système

Table des matières

VI.1 Introduction

Différents facteurs influençants la diffusion sonore ont été abordés de manière séparée dans les précédents chapitres, en cherchant à privilégier des critères pratiques, quitte à limiter leur portée. L'objectif de ce chapitre est de proposer des configurations « optimales » selon ces différents facteurs. Pour cela, les critères obtenus dans les chapitres précédents sont utilisés pour tester des configurations plus exhaustives que celles envisagées jusqu'ici.

Les différents facteurs étudiés sont la diffraction par l'auditeur (Chapitre II), les limites des sources sonores (Chapitre III), l'influence de l'environnement d'écoute (Chapitre IV) et l'interaction de l'auditeur avec le système de restitution (Chapitre V). Des tests perceptifs ont en particulier été mis en place dans les chapitres IV et V pour évaluer les effets de la salle d'écoute et du déplacement de l'auditeur, et en déduire deux indicateurs objectifs pour la suite du travail.

Dans ce travail, la variabilité des HRTF n'a pas été évaluée de manière perceptive : nous avons considéré dans les chapitres précédents que la sphère était une approximation acceptable de l'auditeur. En pratique, les systèmes de restitution sont généralement réglés à l'aide d'une tête artificielle, qui est de toute façon de morphologie différente de celle d'un auditeur donné. Pour estimer l'importance relative de la morphologie individuelle de l'auditeur, nous pouvons considérer la différence entre les HRTF utilisées pour le calcul des filtres transauraux et celles qui seraient mesurées sur l'auditeur selon une démarche similaire à un déplacement de l'auditeur : l'indicateur implémenté au chapitre V est donc à nouveau utilisé. Pour cela, la simulation consiste à calculer un système transaural pour les HRTF résultantes d'une sphère, puis simuler la reconstruction des signaux aux oreilles de l'auditeur en utilisant des HRTF mesurées au Chapitre II sur des mannequins.

VI.2 Généralisation des résultats perceptifs

Les indicateurs définis aux chapitre IV et V sont exploités pour estimer les résultats pour un grand nombre de configurations de sources, à des distances allant de 20 cm à 50 cm par pas de 5 cm et à 80 cm, et pour des positions angulaires variant de +/-5° à +/-175° par pas de 5° en azimut, et de 0° à 45° par pas de 15° en élévation. Ces indicateurs attribuent une note de similarité comprise entre 0 et 100, où 100 correspond à une similarité maximale.

VI.2.A. Effet de salle

L'indicateur mis au point dans le chapitre IV est basé sur un système SISO (Single Input Single Output), or les configurations transaurales sont de type MIMO (Multiple Input Multiple Output). Pour utiliser l'indicateur du chapitre IV, il est nécessaire de se rapprocher d'une configuration de type SISO. Les résultats du test mis en place au chapitre V ont mis en évidence que la perception d'une source virtuelle frontale et d'un son monophonique sont similaires. Pour évaluer l'effet de l'environnement d'écoute, la diffusion d'une source virtuelle frontale est simulée : la pression est censée être équivalente au niveau des deux oreilles et s'apparente ainsi à un système SISO.

VI.2.A.a Paramètres de la simulation

La simulation consiste à calculer des filtres transauraux d'une manière similaire aux Chapitres III et IV, et à en simuler le rendu dans une salle d'écoute. Pour cette simulation, le signal correspondant à une source virtuelle frontale à 1 m de distance est calculé à partir d'un modèle de sphère.

L'effet de salle est modélisé de manière sommaire : il correspond à un champ diffus homogène dans la salle d'écoute, et décroît exponentiellement au cours du temps. Cette approche très simple s'est révélée pertinente pour modéliser les résultats du test du chapitre IV : elle permet en effet d'estimer un indicateur de clarté qui est bien corrélé aux résultats du test perceptif.

Pour estimer le champ diffus, la puissance acoustique W rayonnée par le système transaural est calculée comme celle rayonnée en dehors de la zone d'écoute en champ libre. Cette pression rayonnée P_{ext} est calculée sur une sphère de rayon 1.5 m centrée sur la tête de l'auditeur. Cette distance est supérieure à toutes celles des configurations testées, et reste cohérente avec les distances aux parois dans une salle d'écoute. La sphère est discrétisée en $N = 1000$ points répartis de manière quasiment uniforme, et la pression moyenne sur la sphère est la suivante :

$$P_{ext} = \sqrt{\frac{1}{N}\sum_{\theta,\Phi}^{N}[P_1(r,\theta,\phi) + P_2(r,\theta,\phi)]^2} \qquad (79)$$

P_1 et P_2 sont respectivement les pressions émises par les haut-parleurs ipsilatéral et contralatéral au point de la sphère (r,θ,ϕ) :

$$P_i(r,\theta,\phi) = \frac{hp(i)}{d_i(r,\theta,\phi)} e^{\frac{-2j\pi f d_i(r,\theta,\phi)}{c}} \qquad (80)$$

Avec $d_i(r,\theta,\phi)$ la distance entre le haut-parleur i et le point de coordonnées (r,θ,ϕ). $hp(i)$ est le signal d'entrée du haut-parleur i. La puissance acoustique W équivalente au système transaural est alors définie de la manière suivante :

$$W = \frac{4\pi r^2 P_{ext}^2}{\rho c} \tag{81}$$

Cette puissance acoustique permet d'estimer l'énergie E_{diffus} associée au champ diffus en tout point de l'espace [Gade, 2007] :

$$E_{diffus} = \frac{4\rho c W}{A} \tag{82}$$

Avec A l'aire équivalente d'absorption définie de la manière suivante :

$$A = \frac{0.16V}{TR_{60}} \tag{83}$$

TR_{60} et V sont respectivement le temps de réverbération et le volume d'une salle d'écoute.

A partir de l'énergie du champ diffus E_{diffus}, une réponse impulsionnelle $h(t)$ d'une durée $T = 1.5TR_{60}$ est modélisée de la manière suivante :

$$\begin{cases} h^2(t) = \dfrac{E_{direct}}{\Delta T} & pour\ t \leq \Delta T \\ h^2(t) = Ke^{\frac{-13.8}{TR_{60}}t} & pour\ t > \Delta T \end{cases} \tag{84}$$

Pour le modèle choisi, $\Delta T = 0.4\ ms$ mais sa valeur n'a pas d'influence significative sur les résultats si elle est inférieure à 20 ms. E_{direct} correspond à l'énergie à l'oreille de l'auditeur et le coefficient K est défini par la relation suivante :

$$K = \frac{E_{diffus}}{\sum_{t=0}^{T}\left(e^{\frac{-13.8}{TR_{60}}t}\right)} \tag{85}$$

A partir de cette modélisation de réponse impulsionnelle, le calcul de clarté est effectué de la manière suivante :

$$Clarté = 10\log_{10}\frac{\sum_{t=0}^{20} h^2(t)}{\sum_{t=20}^{T} h^2(t)} \tag{86}$$

Enfin, une note est estimée à partir des résultats du test du chapitre IV :

$$Note = 3.78\ Clarté + 20.5 \tag{87}$$

La simulation est réalisée en adaptant la valeur de A pour quatre salles différentes : le studio, le bureau moyen et la cabine qui ont été évaluées lors du test perceptif. Une salle de 20 m² de surface et 50 m³ de volume, avec un TR de 0.2 s pour toutes les bandes de fréquences est aussi simulée. Cette salle satisfait les critères de l'ITU [ITU-R BS.1116-3, 2015]. Les valeurs des aires équivalentes

d'absorption A_{Stu}, A_{BuM}, A_{cab} et A_{ITU} respectivement associées sont reportées dans le Tableau 11.

Bande d'octave (Hz)	125	250	500	1000	2000	4000	8000
A_{Stu} (m^2)	28.35	25.90	37.49	46.45	65.45	78.26	70.59
$A_{BuM}(m^2)$	12.08	12.96	9.73	9.64	7.94	7.89	10.74
$A_{Cab}(m^2)$	7.14	11.27	28.57	40.00	29.63	29.63	21.62
$A_{ITU}(m^2)$	40.00	40.00	40.00	40.00	40.00	40.00	40.00

Tableau 11 : Aires équivalentes d'absorption par bandes d'octave pour les quatre salles considérées

VI.2.A.b Sources physiques dans le plan horizontal

Dans un premier temps, la simulation est réalisée pour des configurations dans le plan horizontal. Les notes obtenues pour les quatre salles sont représentées par la Figure 58, de la même manière qu'au chapitre III : seuls des demi-cercles sont représentés car les configurations sont symétriques.

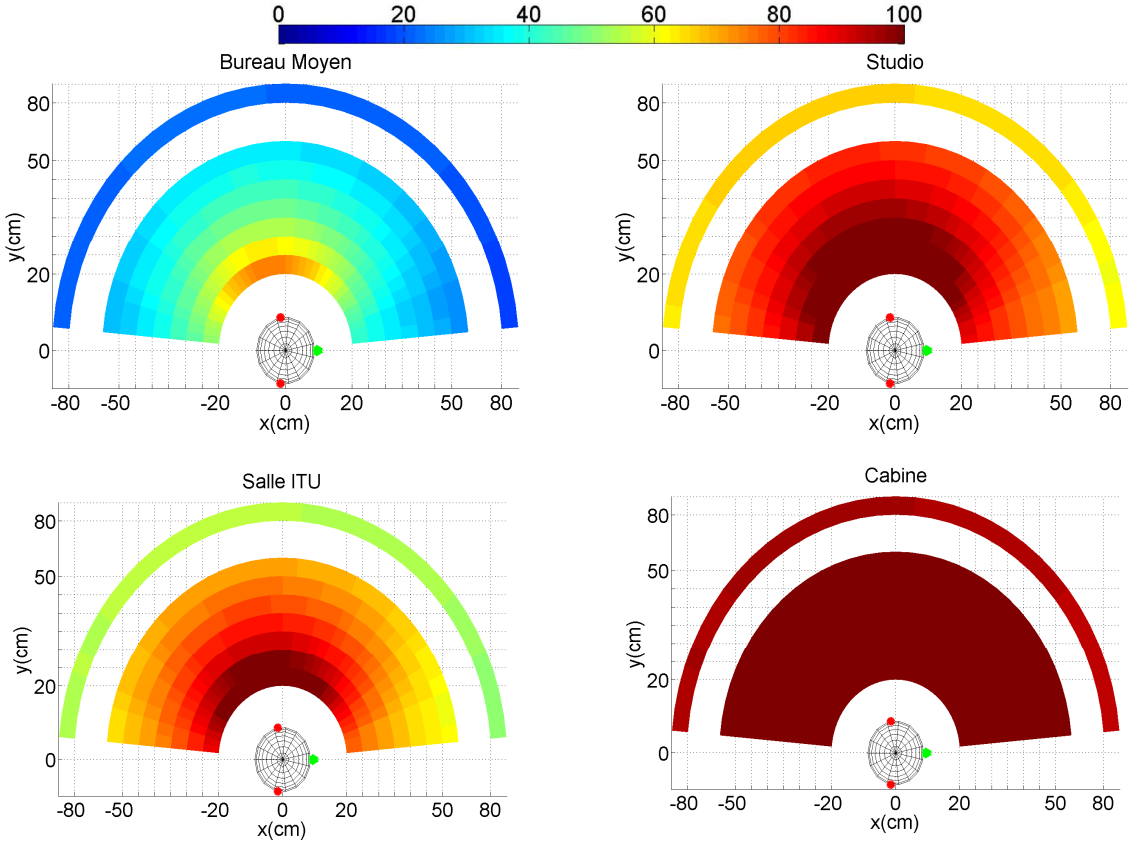


Figure 58: Notes estimées à partir de l'indicateur objectif d'effet de salle dans le plan horizontal pour quatre salles différentes. La sphère est représentée au centre du dispositif à l'échelle, les points rouges correspondent aux oreilles et le point vert au nez. Les configurations à 80 cm ne sont pas représentées à l'échelle.

De manière générale, les notes obtenues pour les salles traitées acoustiquement sont les plus élevées : pour le studio, les notes sont toujours supérieures à 60 pour les distances inférieures à 50 cm, alors que pour le bureau moyen les notes sont presque systématiquement inférieures à 60

pour toutes les configurations. Pour la cabine qui est particulièrement traitée en absorption, les notes sont excellentes quelle que soit la configuration des sources. L'exception pour le bureau moyen concerne les configurations en extrême proximité de l'oreille : les configurations à 20 cm autour de 110° obtiennent une note de l'ordre de 70. Cette tendance s'observe également pour le studio et la salle ITU : les configurations proches de l'oreille obtiennent les meilleures notes. Les configurations légèrement en arrière obtiennent ainsi de meilleures notes car les oreilles du modèle sont ici placées légèrement en arrière. Le type de salle influe donc fortement le rendu aux oreilles de l'auditeur, mais le placement des sources en proximité de l'oreille permet d'en limiter l'effet.

VI.2.A.c Sources physiques en élévation

Pour évaluer l'effet de l'élévation des sources, les azimuts sont les mêmes que pour le plan horizontal, aux élévations 15°, 30° et 45°. Les notes attribuées à ces élévations pour le bureau moyen sont représentées dans la Figure 59 en fonction de l'azimut, aux distances de 20 cm et 80 cm. L'effet de l'élévation est léger pour les configurations à 20 cm, et inexistant à 80 cm. A 20 cm pour les azimuts compris entre 20° et 165° les configurations en élévation obtiennent de moins bonnes notes. Les écarts sont les plus prononcés autour de 100° où l'écart entre les notes du plan horizontal et celles de l'élévation 45° est de l'ordre de 10.

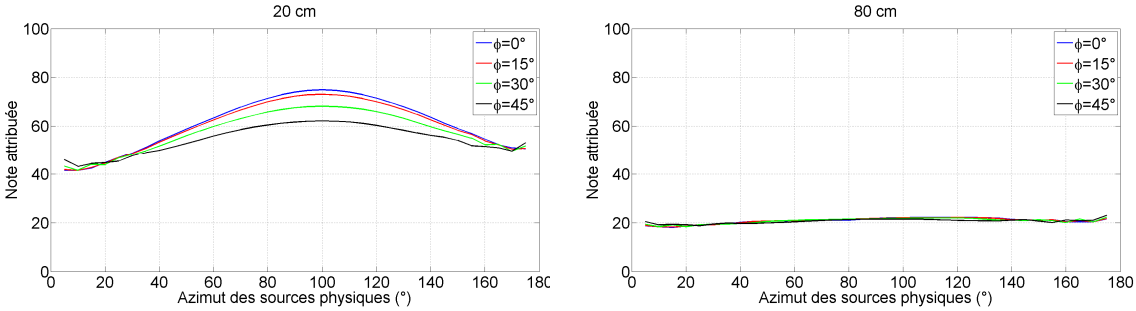


Figure 59 : Notes attribuées en fonction de l'azimut pour différentes élévations pour le bureau moyen. La figure de gauche correspond à une distance de 20 cm et la figure de droite à une distance de 80 cm

En conclusion, l'effet de la salle a une forte influence sur le critère retenu, mais la distance et l'écartement angulaire des sources sonores permettent de compenser en grande partie cet effet. Le placement des sources à proximité des oreilles permet d'obtenir les meilleures notes, quelle que soit la salle d'écoute. L'élévation a moins d'influence, et a tendance à faire baisser les notes en champ proche pour les grands écarts angulaires. Selon notre critère d'effet de salle, un système bien positionné dans une salle d'écoute aux propriétés acoustiques médiocres a de meilleures performances qu'un système mal positionné dans une bonne salle, ce qui reflète bien les résultats du Chapitre IV.

VI.2.B. Déplacement de l'auditeur

L'effet du déplacement de l'auditeur est évalué à partir de l'indicateur proposé dans le chapitre V et issu de résultats perceptifs, pour les mêmes positions des sources physiques que précédemment. Pour l'évaluer, le rendu de systèmes transauraux est simulé lorsque l'auditeur n'est pas à la position optimale. Trois sources virtuelles et douze déplacements de l'auditeur sont

testés. L'indicateur déterminé au chapitre V est alors calculé, et la situation retenue est la « pire », c'est-à-dire celle qui conduit à obtenir la note la plus basse.

VI.2.B.a Paramètres de la simulation

La simulation consiste à générer des signaux aux oreilles de l'auditeur, comme au chapitre V. L'influence de l'incidence virtuelle sur le rendu a été montrée, et trois simulations sont réalisées pour des incidences de sources virtuelles différentes : 0°, 45° et 90° dans le plan horizontal. Nous considérons que ces trois sources virtuelles sont représentatives d'un champ sonore 3D. Aucune source virtuelle arrière n'est évaluée, car les fonctions de transfert sont ici calculées analytiquement à partir d'un modèle de sphère et il y a donc peu de différences entre la diffraction par une onde incidente frontale et arrière.

La simulation consiste à convoluer les réponses impulsionnelles simulées avec des filtres transauraux calculés à partir de HRIR issues du même modèle de sphère. Le résultat est alors à nouveau convolué avec les HRIR entre les sources et les oreilles calculées lorsque l'auditeur n'est pas à la position prévue. Les réponses impulsionnelles résultantes aux oreilles de l'auditeur sont filtrées par un filtre à -3 dB/octave, permettant d'approcher le spectre d'un bruit rose. Pour calculer l'indicateur objectif, ces réponses impulsionnelles sont de plus filtrées avec une pondération B. Les notes associées à ces simulations sont calculées à partir des résultats du chapitre V :

$$Note = -5.78D_{Niv} - 1.24D_{Timbre} + 100 \qquad (88)$$

Avec D_{Niv} la dissemblance de niveau et D_{Timbre} la dissemblance de timbre calculées à partir des réponses impulsionnelles. Ce calcul est répété pour douze déplacements de l'auditeur :

- Deux déplacements de rotation autour de l'axe Z (+10°, -10°)
- Six déplacements de translation de 5 cm sur les axes X, Y et Z (+5X, +5Y, +5Z, -5X, -5Y, -5Z)
- Quatre déplacements de translation de 5 cm sur les diagonales des axes X et Y (+X+Y, -X+Y, -X-Y, +X-Y)

VI.2.B.b Sources physiques dans le plan horizontal

Pour chaque configuration, la note associée à la « pire » situation (déplacement et SRV) est retenue et représentée par la Figure 60.

L'effet de l'écartement angulaire est clairement identifiable : les configurations faiblement écartées ont les notes les plus faibles. La distance a un effet significatif uniquement pour les configurations d'écartement angulaire compris entre +/-35° et +/- 140°, les configurations en proximité obtenant les meilleures notes. Les écarts angulaires compris entre +/-110° et +/-130° à 30 cm de distance semblent être les plus robustes : les notes sont supérieures à 70. Les configurations les plus proches (20 cm) semblent légèrement moins robustes que les configurations à 30 ou 40 cm pour les grands écartements angulaires. Toutefois, les notes obtenues pour ces configurations restent significativement supérieures aux notes obtenues pour les faibles écarts angulaires.

Chacune de ces notes correspondent à la « pire » situation. Le déplacement et la source virtuelle correspondant à cette « pire » situation sont représentés par la Figure 61. Les déplacements en translation sont ceux qui minimisent systématiquement la note de déplacement : les systèmes sont robustes aux rotations. Notamment l'écart en translation selon Y est celui qui est majoritairement présent, la translation +5 cm en Y étant celle évaluée dans le test perceptif du chapitre V. Ce déplacement est donc assez représentatif des plus grandes dégradations que l'on peut obtenir, et il était judicieux de le sélectionner. Par ailleurs, les déplacements sur l'axe Z ne minimisent jamais la note. Pour les configurations où un autre déplacement que Y minimisent la note, les notes attribuées pour ces configurations sont quasiment toujours supérieures à 50, traduisant des dégradations modérées. Les déplacements les plus problématiques sont donc les déplacements latéraux, qui modifient le plus le contraste entre les deux oreilles.

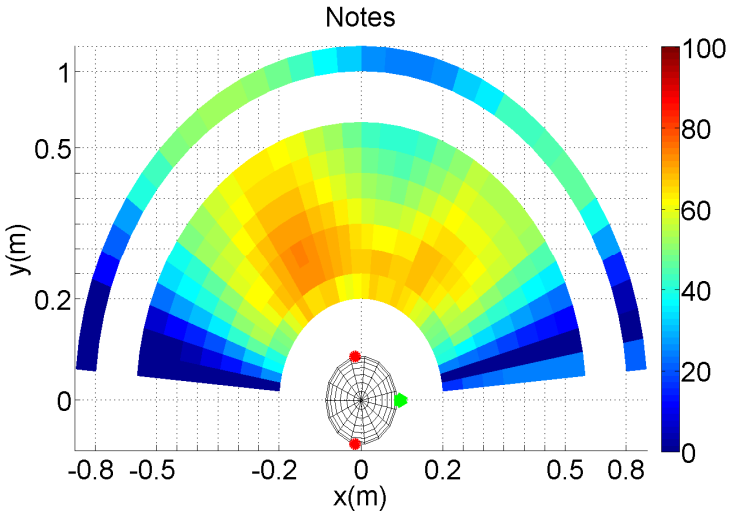


Figure 60 : notes estimées à partir de l'indicateur objectif de déplacement dans le plan horizontal pour trois incidences de sources virtuelles et douze déplacements. Seule la note la plus basse parmi ces trois incidences et douze déplacements est représentée. La sphère est représentée au centre du dispositif à l'échelle, les points rouges correspondent aux oreilles et le point vert au nez. Les configurations à 80 cm ne sont pas représentées à l'échelle.

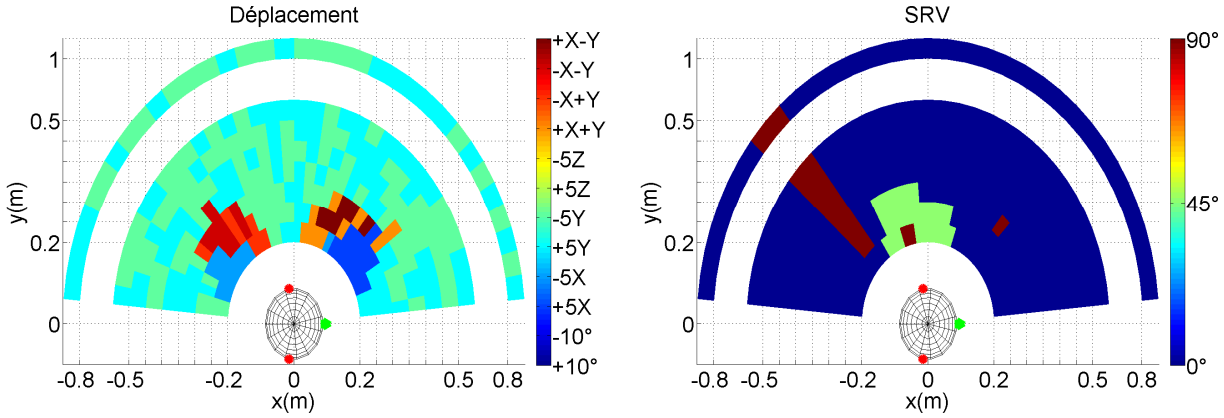


Figure 61 : « pires » déplacement (figure de gauche) et source virtuelle (figure de droite). La sphère est représentée au centre du dispositif à l'échelle, les points rouges correspondent aux oreilles et le point vert au nez. Les configurations à 80 cm ne sont pas représentées à l'échelle.

La source virtuelle à 0° est celle qui minimise quasiment systématiquement les notes : l'indicateur retrouve donc le fait que les sources frontales sont les plus délicates à reproduire.

VI.2.B.c Sources physiques en élévation

De manière analogue aux sources physiques dans le plan horizontal, seule la « pire » situation est étudiée pour les configurations avec élévation. Les notes obtenues pour différentes élévations aux distances 30 cm et 80 cm sont représentées par la Figure 62. L'effet de l'élévation des sources physiques est le plus marqué pour la configuration d'écartement angulaire +/-5° à 80 cm, pour laquelle la note passe de 20 à élévation nulle à 77 pour l'élévation à 45°, devenant la configuration la plus robuste en champ lointain. L'effet de l'élévation est également visible pour les grands écarts angulaires. En particulier, à 30 cm pour la configuration à +/-80° la note vaut 62 à élévation nulle et 73 à 30° d'élévation.

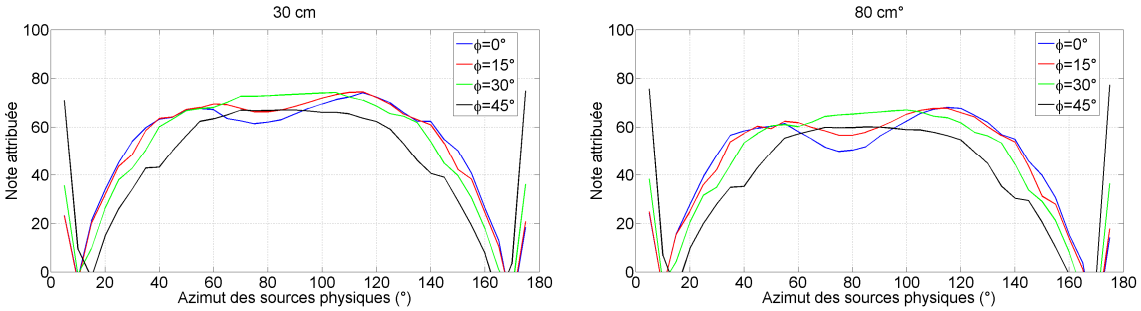


Figure 62 : Notes attribuées en fonction de l'azimut pour différentes élévations pour la source virtuelle 0°. Figure de gauche : distance de 30 cm, figure de droite : distance de 80 cm.

Pour conclure en ce qui concerne l'évaluation de l'influence du déplacement de l'auditeur, les configurations à faibles écarts angulaires ne sont jamais robustes quelle que soit la distance des sources physiques. Inversement, les configurations correspondant à la zone située à proximité des oreilles de l'auditeur, légèrement en arrière, obtiennent toujours des notes élevées. Les positions faiblement en élévation sont par ailleurs légèrement plus robustes. Les déplacements latéraux sont les plus critiques : ce sont eux qui obtiennent les notes les plus basses pour la majeure partie des configurations. Ces déplacements sont probablement ceux qui modifient le plus le contraste aux oreilles de l'auditeur : la préservation du contraste serait l'élément clef d'une restitution fidèle. Il semblerait que les sources virtuelles frontales soient les plus difficiles à reproduire, y compris lorsque les sources physiques sont placées devant l'auditeur.

VI.2.C. Système calibré pour une tête différente

Le paramètre étudié dans cette section est l'effet d'un changement de forme de la tête entre la phase de calibration du système et au moment de l'écoute. L'objectif est de quantifier la sensibilité à la non-individualisation du système. Pour cela, les HRTF mesurées sur des mannequins dans le chapitre II sont exploitées, préalablement égalisées en champ diffus. Cette simulation ne correspond pas à une évaluation perceptive, l'indicateur du chapitre V a donc été réutilisé pour mettre en évidence ces différences. La simulation réalisée ici est donc une extrapolation qui mériterait d'être approfondie à partir de tests perceptifs, tâche considérable que nous ne pouvons pas envisager à ce stade du travail.

Le principe de la simulation est donc exactement le même que dans le cas de l'étude de l'effet de déplacement, à l'exception que la matrice des filtres de reconstruction correspond aux HRTF mesurées sur les différents mannequins. Le calcul de dissemblance est effectué en utilisant l'indicateur de déplacement, et est calculé pour la source virtuelle à 0° et pour des configurations à 40 cm de distance. Les notes obtenues sont reportées dans la Figure 63 pour les élévations 0° et 30°.

Une tendance commune est visible pour tous les mannequins : les configurations arrières sont celles qui obtiennent les meilleures notes. Par exemple à élévation nulle, les notes sont toujours supérieures à 40 pour les écartements angulaires supérieurs à +/-110°, et toujours inférieures à 35 pour les écartements angulaires inférieurs à +/-30°. Par ailleurs, les configurations en élévation obtiennent de meilleures notes, particulièrement pour les configurations arrières. Pour les azimuts entre 120° et 140° les notes sont toujours supérieures à 70.

Les notes attribuées dépendent du mannequin considéré : par exemple pour les configurations à 0° d'élévation et d'écart angulaire inférieur à +/-100°, les notes pour le mannequin Cortex sont inférieures à celles du mannequin Kemar de l'ordre de 20 points. A l'élévation 30°, il n'y a pas d'influence significative du mannequin, à l'exception du mannequin Kemar pour lequel les notes sont plus basses que les autres aux faibles écarts angulaires. Cela confirmerait que les configurations avec une légère élévation sont plus robustes aux variations de morphologie. Par ailleurs, cela illustre à nouveau le fait que les mannequins sont significativement différents alors qu'ils sont tous censés approximer un auditeur moyen. L'influence de la morphologie réelle d'un auditeur est donc très probablement sous-évaluée par les simulations basées sur des mannequins.

De manière générale, les notes attribuées ici sont d'un ordre de grandeur proche de celui des notes liées au déplacement de l'auditeur. Cela signifie que la variation de morphologie a un effet significatif et que son influence nécessiterait d'être vérifiée par des tests perceptifs. Néanmoins, une tendance est visible : les configurations derrière la tête et avec une légère élévation semblent être les plus robustes aux variations de morphologie testées. Les mannequins approximant tous un auditeur moyen, les variations inter-individuelles sont probablement plus marquées.

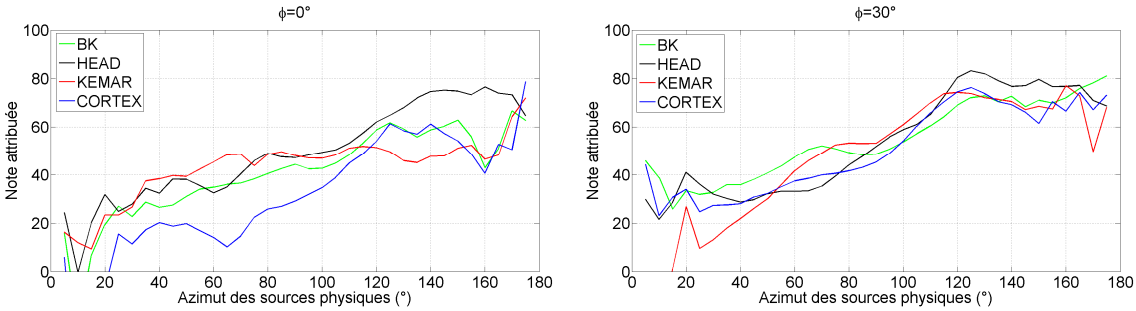


Figure 63 : notes attribuées en fonction de l'azimut lorsque le système a été calibré sur un modèle de sphère, et diffusé vers différents mannequins à 40 cm de distance pour la source virtuelle 0°. La figure de gauche correspond à une élévation de 0° et la figure de droite à une élévation de 30°

VI.3 Discussion

Des tests perceptifs ont permis d'évaluer séparément les effets de la salle et du déplacement de l'auditeur, mais la correspondance des résultats mérite d'être discutée afin d'établir une hiérarchisation des effets. L'effet de la morphologie de l'auditeur a été étudié à travers des mesures de HRTF sur différents mannequins, mais n'a pas été évalué perceptivement. Par ailleurs, l'influence des filtres transauraux a été quantifiée à partir de l'augmentation de débit des sources électro-acoustiques.

VI.3.A. Hiérarchisation des effets

Deux tests perceptifs ont été réalisés de manière indépendante, visant à évaluer deux types de dégradations sur le rendu : l'effet de salle et l'effet de déplacement de l'auditeur. Des signaux communs aux deux tests permettent d'envisager une hiérarchie entre ces deux effets : un signal correspondant à une situation d'écoute « classique » est présent dans les deux tests. Ce signal a été enregistré à une distance de 80 cm dans une salle d'écoute dont le temps de réverbération n'est jamais supérieur à ce que préconise l'ITU.

Dans le cadre du test sur l'effet de salle, ce signal a obtenu une note de similarité de 67, alors que dans le cadre du test évaluant le déplacement de l'auditeur sa note est de 20. Tous les déplacements évalués dans les tests ont par ailleurs obtenu des notes supérieures à 20 : les différences introduites par la salle sont donc jugées plus importantes. L'amplitude des déplacements envisagés est réduite (5 cm par rapport à la position nominale), mais semble réaliste avec le positionnement d'un auditeur en conditions contrôlées. Par ailleurs les sons avec un effet de salle ont été présentés de manière diotique : ce type de situation n'est pas la plus représentative d'une situation d'écoute « réelle » où les composantes réfléchies sont séparées dans l'espace. Notre objectif était d'optimiser un système pour le plus grand nombre, et utiliser des HRTF individuelles pour le test perceptif aurait nettement complexifié sa mise en place. L'importance de l'effet de salle a néanmoins probablement été sur-évaluée dans notre démarche.

L'effet de la différence de morphologie a été abordé dans le deuxième chapitre à partir de mesures de fonctions de transfert sur différents mannequins. Aucun test perceptif n'a été conduit pour comparer les différents mannequins, mais le critère obtenu pour les déplacements a été utilisé pour estimer son effet. Les différences perçues lorsque le système a été calibré à partir d'une tête différente seraient du même ordre de grandeur que les différences perçues lorsque l'auditeur n'est pas à l'emplacement prévu. Ces différences dépendraient également de la position des sources physiques. L'importance de la morphologie individuelle a probablement été sous-estimée par l'utilisation exclusive de mannequins.

L'effet des filtres sur les sources électro-acoustiques a été étudié dans le chapitre III. Cet effet se traduit par un coût lié à l'obtention du contraste, et les configurations pour lesquelles le contraste naturel est élevé sont les plus performantes. Il est délicat de comparer cet effet aux autres, car il n'a pas d'équivalent perceptif. Toutefois, ce critère peut devenir majeur si le niveau sonore à reproduire est important (automobile, aéronautique).

Globalement, l'effet de salle est celui qui semble prépondérant parmi tous les effets étudiés ici. L'effet de déplacement et l'effet de morphologie semblent modifier le rendu perçu d'un ordre de grandeur similaire, mais sensiblement inférieur à celui lié à l'effet de la salle.

L'influence des différents effets étudiés est reportée dans le Tableau 12 en fonction du placement des sources physiques en distance, azimut et élévation. La significativité du placement est spécifiée par une échelle allant de « non significatif ou non mis en évidence » (-) à « très significatif » (+++). L'indication spécifiée dans chaque case correspond alors au meilleur placement.

Effet / Position des sources physiques	Coût lié aux filtres	Effet de salle	Morphologie	Déplacement auditeur
Distance	Proximité **(+)**	Proximité **(+++)**	**(-)**	Proximité ssi grands écarts **(+)**
Azimut	Grands écarts **(+++)**	Grands écarts **(+++)**	Positions arrières **(+)**	Grand écarts **(+++)**
Elévation	**(-)**	**(-)**	Elévation 30° **(+)**	Elévation ssi grande distance et faible écart angulaire **(+)**

Tableau 12 : position optimale résumée en fonction des différents paramètres. La mention « + » correspond à un effet légèrement significatif, « +++ » correspond à un effet très significatif et « - » à un effet non significatif ou non mis en évidence

Cette classification des effets conduit à proposer des configurations pour lesquels l'effet de salle est réduit, c'est-à-dire placer des sources à proximité des oreilles de l'auditeur. Il s'avère que ces positions sont également robustes au déplacement de l'auditeur pour les trois sources virtuelles testées. Dans le plan horizontal, une distance de 30 cm semble la plus adaptée. Le placement des sources en élévation semble améliorer légèrement la robustesse aux variations de morphologie et aux déplacements de l'auditeur.

VI.3.B. Propositions de configurations optimales

D'après la hiérarchisation proposée, l'effet de la salle est le plus important. Il serait donc intéressant d'évaluer une configuration qui soit la plus performante selon ce critère, et correspond à des sources placées à 20 cm et d'écartement angulaire de sources de +/-100°.

La zone qui est assez robuste au déplacement correspond à peu près à la zone optimale pour l'effet de salle, avec les sources légèrement en arrière. Toutefois, les configurations « trop proches » sont un peu moins robustes que les configurations à 30 cm. En particulier, la

configuration d'écartement +/-115° et à 30 cm permet d'obtenir les meilleures notes. Afin de faire cohabiter cette solution avec la première, un écartement de +/-120° est préféré.

Les simulations de robustesse au changement de morphologie semblent montrer une sensibilité du même ordre de grandeur que pour le déplacement. Les configurations permettant d'obtenir les meilleures notes pour tous les mannequins sont les configurations avec une élévation de 30° et légèrement en arrières, notamment entre +/- 110° et +/-140°. Cela coïncide avec les autres zones optimales pour les autres paramètres. La configuration la plus robuste est dans ce cas à +/-120° d'écartement angulaire. Il est alors possible que les sources virtuelles soient perçues à l'arrière alors qu'elles sont censées être perçues devant. Par ailleurs, il semblerait que les sources frontales soient les plus délicates à reproduire. Il est possible que le placement des sources physiques dans l'hémisphère frontal y soit plus favorable, et une configuration alternative est proposée, en plaçant les sources avec un écartement angulaire de +/-60° à l'élévation nulle. Les paramètres de placement de trois configurations sont alors présentés dans le Tableau 13. Les points forts et faibles sont précisés pour chaque configuration. Pour ces trois configurations, les haut-parleurs sont placés dans le plan horizontal, car il semblerait que le placement en élévation n'ait pas beaucoup d'influence. Il serait judicieux de tester des configurations avec élévation, néanmoins le dispositif actuel ne permet pas d'implémenter simplement ces configurations.

	1	2	3
r	30 cm	30 cm	30 cm
θ	+/-100°	+/-120°	+/- 60°
ϕ	0°	0°	0°
Points forts	La plus robuste à l'effet de salle	La plus robuste au déplacement et assez robuste à la variation de morphologie	Localisation plus précise des sources frontales ?
Points faibles	Localisation des sources frontales ? Un peu sensible aux déplacements	Localisation des sources frontales ?	Plus sensible à la salle et au déplacement

Tableau 13 : paramètres de placement de trois configurations proposées, pertinentes selon les indicateurs proposés.

Pour évaluer ces configurations, un dispositif a été spécialement conçu, permettant de fixer plusieurs haut-parleurs à une distance contrôlée de l'auditeur. Il est représenté dans la Figure 64. Les haut-parleurs peuvent être placés aux distances de 40 cm et 50 cm, ou rapprochés à l'aide de barrettes filetées. Le dispositif proposé n'est pas très encombrant, ce qui permet de l'installer facilement dans n'importe quelle salle. Ceci ouvre la possibilité de tester la robustesse du système de diffusion à l'environnement d'écoute. La petite taille des sources mises au point dans ce travail

(voir Annexe A) permet de juxtaposer ces trois configurations, permettant alors de les évaluer simultanément dans un même test perceptif. De plus, ces configurations peuvent coexister avec d'autres sources physiques, ce qui rend possible la comparaison de sources « virtuelles » avec des sources « réelles ». Ce dispositif permet également d'envisager de comparer plusieurs systèmes de restitution : en particulier une configuration en cercle permettrait d'implémenter des solutions de type HOA, WFS ou VBAP.

Il est enfin possible de superposer les différentes configurations, en espérant profiter des avantages de chacune. Ces solutions à plus grand nombre de canaux devraient aussi pouvoir être optimisées à l'aide de l'outil prédictif proposé. L'augmentation du nombre de canaux permettrait par exemple de concentrer encore le champ rayonné, réduisant encore l'influence de la salle d'écoute [Warusfel et al., 1997; Warusfel & Misdaris, 2004].

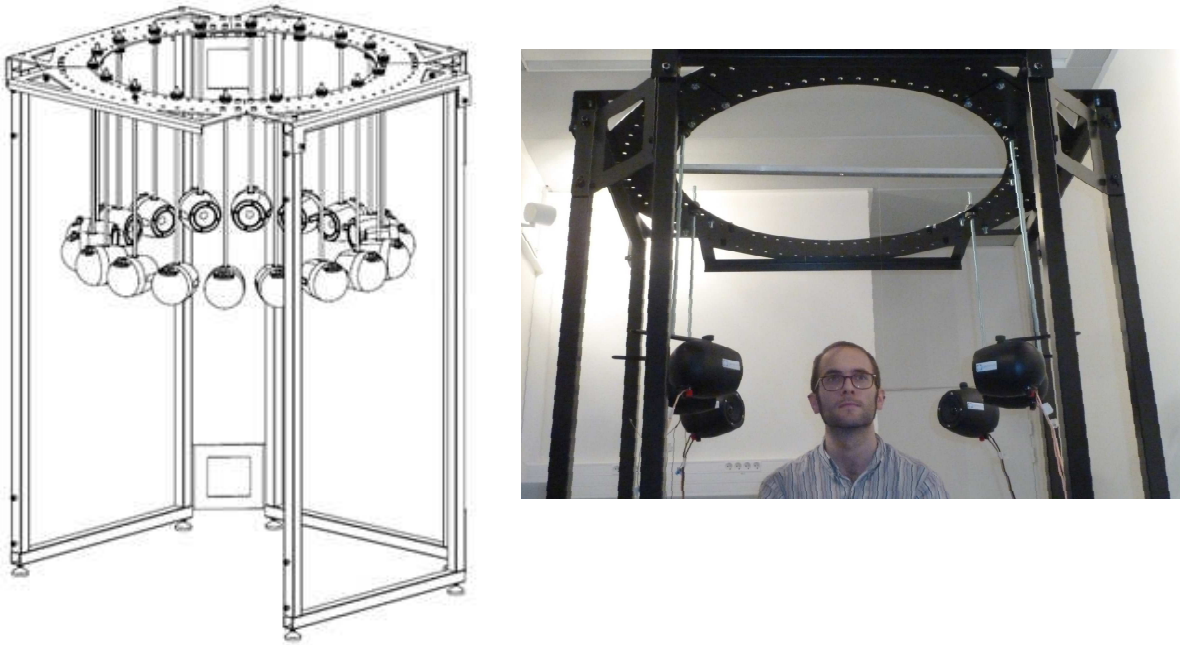


Figure 64 : Figure de gauche : plan du dispositif de fixation des haut-parleurs autour de l'auditeur. Les points de fixation permettent de les placer à 40 ou 50 cm de l'auditeur. Figure de droite : montage des trois configurations proposées dans un local peu traité en acoustique

VI.4 Conclusion du chapitre

Dans ce chapitre les indicateurs objectifs proposés dans les chapitres précédents ont été généralisés à d'autres configurations. De la même manière que pour les tests perceptifs, les différents effets ont été évalués indépendamment.

Le placement des sources en proximité de l'auditeur permet de limiter la nécessité d'un traitement acoustique coûteux, ceci étant donc un moyen efficace de réduire le coût d'un système de diffusion. En rapprochant les sources, le niveau sonore à reproduire est nécessairement plus faible et l'augmentation de débit des sources liée à l'obtention du contraste est limitée : le coût des sources dédiées à une telle reproduction est alors moindre que pour une restitution à grande distance.

Trois configurations ont été proposées en vue d'une évaluation en conditions réelles. Cette évaluation permettra de vérifier les résultats en conditions d'écoute plus complexes, prenant en compte les effets croisés de la salle d'écoute et du déplacement éventuel de l'auditeur. Par ailleurs, cette évaluation permettra de prendre en compte les effets liés aux HRTF des auditeurs. Un dispositif a été conçu pour disposer simplement plusieurs configurations à faible distance de l'auditeur, en permettant la cohabitation de plusieurs systèmes Ceci permet d'envisager de faire coopérer ces différentes solutions, pour réaliser un système à plus de deux canaux. Il est également possible grâce à ce dispositif de comparer la restitution du système proposé à d'autres systèmes, et dans plusieurs salles d'écoute différentes. Tout est donc maintenant prêt pour l'évaluation *in situ* du dispositif de reproduction résultant des travaux présentés.

Conclusion générale

Dans ces travaux de thèse, nous nous sommes intéressés à l'optimisation d'un système de reproduction sonore fidèle à moindre coût. L'aspect que nous avons considéré de plus grande importance est la fidélité du niveau et du timbre reproduits. Pour pouvoir effectuer une restitution spatialisée avec un faible nombre de canaux, nos travaux se sont basés sur la technologie transaurale, permettant de diffuser des signaux binauraux avec un nombre minimum de haut-parleurs. Afin de pouvoir implémenter ce type de système dans une salle d'écoute usuelle tout en préservant le timbre d'origine, nous avons proposé de rapprocher les sources de l'auditeur afin de maximiser le rapport du champ direct sur le champ diffus.

Dans le processus de diffusion transaurale l'influence des HRTF intervient deux fois : lors de la prise de son et lors du calcul de filtres transauraux. Notre étude concerne uniquement l'optimisation du système de reproduction pour un auditeur générique. Optimiser « la » reproduction sonore avec ce type de système est un problème différent pour lequel il faudrait certainement adapter les signaux binauraux en fonction de l'auditeur : selon de nombreux auteurs, la perception des sources virtuelles frontales est souvent intracrânienne lorsque les HRTF associées aux signaux binauraux ne correspondent pas à celles de l'auditeur. Les conclusions ci-après ne s'appliquent donc qu'à l'objectif d'optimisation du système de reproduction que nous nous sommes fixé pour ce travail.

La démarche adoptée pour cette thèse a été de quantifier l'effet du rapprochement des sources pour une restitution transaurale. Premièrement, l'effet de la distance sur les HRTF a été évalué, en étudiant aussi la possibilité de propager des mesures à faible distance pour obtenir une estimation des mesures à plus grande distance. Une décomposition en harmoniques sphériques a été employée, et une méthode d'inversion a été complétée par la sélection de termes séparables aux positions de mesures. Cette sélection permet de limiter le conditionnement numérique du problème inverse et simplifie ainsi la résolution. Quatre mannequins ont été caractérisés avec un même dispositif pour 72 azimuts et 7 élévations à 40 cm et 6 élévations à 2 m. Des mesures sur une sphère ont permis de quantifier les incertitudes de mesures, les réflexions liées au dispositif de mesure semblant plus importantes à grande distance. Pour tenter de réduire les erreurs d'estimation des HRTF, il semblerait ainsi qu'il soit plus efficace de mesurer les HRTF à faible distance pour les propager à plus grande distance. Les écarts entre les mesures sur quatre mannequins et sur une sphère se sont révélés être du même ordre de grandeur que les écarts entre mannequins : les mannequins censés approximer un auditeur moyen semblent donc différer autant entre eux qu'avec la sphère. Ce résultat nous a conduit à utiliser le modèle de sphère pour simuler les nombreuses situations d'écoute destinées à optimiser le système de reproduction dans les étapes suivantes.

Dans le troisième chapitre, l'impact du filtrage transaural sur les sources électro-acoustiques a été étudié. Il a été défini comme un « coût lié à l'obtention du contraste », se traduisant par une augmentation du débit demandé aux sources. Ce coût est prépondérant aux basses fréquences, alors que c'est à ces fréquences que les sources électro-acoustiques sont les plus limitées. Il a été

montré que ce coût est réduit pour les configurations dont le contraste « naturel » est déjà important. Les configurations de haut-parleurs à proximité de l'oreille sont donc celles dont le coût est le plus faible. Une source adaptée à notre dispositif a été conçue en conséquence, et est présentée en Annexe A.

Dans le quatrième chapitre l'effet de salle et son égalisation ont été abordés. Un test d'écoute de type MUSHRA a été utilisé pour évaluer la proximité entre un son anéchoïque et différents sons associés à plusieurs salles d'écoute, plusieurs distances et plusieurs méthodes d'égalisation. Le rapprochement des sources s'est révélé là aussi très avantageux car permettant d'améliorer significativement le rendu sonore. En revanche, les égalisations testées ont un intérêt plus limité : le rendu n'est amélioré que pour des salles de bonne qualité.

L'effet du placement non optimal de l'auditeur lors d'une restitution transaurale en conditions anéchoïques a été étudié au chapitre V, en utilisant le même type de test perceptif. Les sons évalués ont été construits par simulation à partir d'un modèle de sphère pour des déplacements de 5 cm. Une conclusion importante est ressortie de ce test : les dégradations liées à un placement incorrect de l'auditeur seraient moindres que celles liées à l'influence de la salle, pour les configurations testées. Contrairement à ce que l'intuition laissait penser, les configurations à faible distance ne sont pas toujours les moins robustes au déplacement de l'auditeur. En particulier, la configuration à 20 cm de distance et +/-60° d'écartement angulaire est la plus robuste de toutes les configurations testées.

Enfin, l'ensemble des résultats des chapitres précédents ont été généralisés et synthétisés dans le dernier chapitre. 1120 configurations ont été simulées, et des configurations optimales ont été identifiées à partir des critères proposés. En particulier, les configurations proches des oreilles (de l'ordre de 30 cm à 40 cm de distance avec un grand écartement angulaire) semblent être les plus performantes du point de vue des critères envisagés à ce stade de conception du système. Pour cette démarche d'optimisation, l'accent a été mis sur la précision avec laquelle le système peut reproduire des signaux binauraux en présence d'un auditeur « générique ». Les configurations proposées devraient ainsi être une bonne base de départ pour optimiser la reproduction transaurale.

Un seul stimulus a été employé pour les tests perceptifs, choisi car il est considéré comme discriminant pour évaluer des dégradations liées à l'effet de la salle d'écoute et représentatif de contenus audio (bruit rose modulé). Les indicateurs objectifs utilisés sont par contre calculés à partir de réponses impulsionnelles, et ne dépendent donc pas du contenu des signaux. Ils semblent donc généralisables à différents contenus, mais il serait intéressant de vérifier leur validité effective.

Le système proposé est aujourd'hui arrivé à l'étape de réalisation. La suite des travaux consistera à l'évaluer en conditions « réelles », en comparant les trois configurations *a priori* optimales : l'incertitude de placement de l'auditeur sera ainsi réaliste et le rendu respectera la morphologie réelle. Il est aussi maintenant possible de diversifier les situations de reproduction : plusieurs signaux pourraient être testés dans plusieurs environnements d'écoute. Une qualification

objective (mesures sur mannequin) et perceptive est donc possible à court terme, seul le temps nous a manqué pour la présenter dans ce mémoire.

Le dispositif de fixation et les sources électro-acoustiques ont été conçus de manière à pouvoir faire cohabiter plusieurs solutions dans l'optique de leur évaluation perceptive et de leur comparaison. Le système proposé est flexible : il peut être installé dans n'importe quelle salle d'écoute. Il permet aussi de faire cohabiter des systèmes basés sur d'autres principes (VBAP, HOA, WFS) et des sources physiques, pour les comparer entre eux.

Dans un système transaural, le rendu sonore est étroitement lié aux signaux à reproduire : l'information spatiale est codée lors de la prise de son. Le dispositif proposé permet de préserver le contenu spectral des signaux indépendamment sur chaque oreille, et donc de préserver l'information spatiale enregistrée. Si la prise de son est adaptée à l'auditeur, alors la restitution devrait être optimale.

Idéalement, les HRTF utilisées pour le calcul des filtres transauraux devraient alors correspondre à celles de l'auditeur. Plutôt que d'employer des HRTF génériques (modèles géométriques simples, tête artificielle) dans l'objectif d'un système adapté au plus grand nombre d'auditeurs, des solutions spécifiques à chaque auditeur pourraient alors être envisagées, la plus complexe consistant à mesurer *in situ* les HRTF de chaque auditeur. Une alternative consisterait à tirer profit des avancées récentes concernant l'individualisation des HRTF à partir d'une famille pré-établie.

Par sa précision de reproduction, le dispositif proposé permettra ainsi de conduire un test perceptif à plus grande échelle, afin de mettre en relation l'effet des différences inter-individuelles de morphologie avec l'influence de la salle et des déplacements de l'auditeur. L'objet du test pourrait porter sur l'effet croisé de l'individualisation des signaux binauraux et des filtres transauraux : leur importance relative pourrait alors être quantifiée pour des degrés différents d'individualisation.

Plus généralement, il serait envisageable de prendre en compte les particularités morphologiques de l'auditeur dans l'intégralité du processus (création de signaux et diffusion), en contournant les contraintes liées au binaural. Au lieu de synthétiser la pression uniquement en deux points, il serait par exemple possible de reproduire le champ de pression au voisinage immédiat des oreilles de l'auditeur. Ceci relierait la perception sonore aux petits mouvements de tête (qui contribuent à la localisation de sources sonores), tout en limitant la complexité du dispositif qui resterait focalisé sur deux zones peu étendues spatialement.

Références

Advanced Numerical Solutions LLC. (2016). Spherical wave scattering by a rigid sphere. Retrieved November 16, 2016, from http://ansol.us/Products/Coustyx/Validation/MultiDomain/Scattering/SphericalWave/HardSphere/Downloads/dataset_description.pdf

AES. (1996). *AES recommended practice for professional audio - Subjective evaluation of loudspeakers*. Audio Engineering Society.

Algazi, R., Duda, R., Thompson, D., & Avendano, C. (2001). The CIPIC HRTF database. Presented at the IEEE Workshop on applications of Signal Processing to Audio and Acoustics, New York (USA).

Andeol, G. (2012). *Localisation de sources sonores : caractérisation de la variabilité inter-individuelle et effet de l'entraînement* (Phd thesis). Université de Provence Aix-Marseille I.

Andreopoulou, A., Begault, D. R., & Katz, B. (2015). Inter-Laboratory Round Robin HRTF measurement comparison. *IEEE Journal of selected topics in signal processing*, *9*(5), 895–906.

ANSI S3.4. (2007). *Procedure for the computation of Loudness for steady sound*.

Aussal, M., Alouges, F., & Katz, B. (2013). A study of spherical harmonics interpolation for HRTF exchange. Presented at the ICA, Montreal (Canada).

Bahu, H. (2013). *Synthèse binaurale en champ proche : étude théorique et expérimentale* (Rapport de stage). IRCAM.

Bai, M., & Lee, C.-C. (2006). Objective and subjective analysis of effects of listening angle on crosstalk cancellation in spatial sound reproduction. *Journal of the American Society of America*, *120*(4), 1976–1989.

Bank, B. (2013). Combined quasi-anechoic and in-room equalization of loudspeaker responses. Presented at the 134th AES Convention, Rome (Italia).

Bauck, J., & Cooper, D. (1996). Generalized transaural stereo and applications. *Journal of the Audio Engineering Society*, *44*(9), 683–705.

Begault, D. R. (1994). *3-D sound for virtual reality and multimedia* (AP Professional.).

Begault, D. R., & Wenzel, E. (1993). Headphone localization of speech. *Human Factors*, *35*(2), 361–376.

Begault, D. R., Wenzel, E., & Anderson, M. R. (2001). Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *Journal of the Audio Engineering Society*, *49*(10), 904–916.

Berkhout, A. J., de Vries, D., & Vogel, P. (1993). Acoustic control by wave field synthesis. *Journal of the American Society of America*, *93*(5), 2764–2778.

Bertet, S. (2009). *Formats audio 3D hiérarchiques : caractérisation objective et perceptive des systèmes ambisonics d'ordres supérieurs* (Phd thesis). Institut National des Sciences Appliquées de Lyon.

Bertet, S., Daniel, J., Parizet, E., Gros, L., & Warusfel, O. (2007). Investigation of the perceveid spatial resolution of Higher Order Ambisonic sound fields: a subjective evaluation involving virtual and real 3D microphones. Presented at the AES 30th International Conference, Saariselka (Finland).

Bharitkar, S., Hilmes, P., & Kyriakakis, C. (2004). Robustness of spatial average equalization: A statistical reverberation model approach. *Journal of the American Society of America*, *116*(6), 3491–3497.

Bidondo, A., Vazquez, J., Vazquez, S., Arouxet, M., & Heinze, G. (2016). A new and simple method to define the time limit between the early and late sound fields. Presented at the 141th AES Convention, Los Angeles (USA).

Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization, Revised Edition*.

Boulandet, R. (2012). *Tunable electroacoustic resonators through active impedance control of loudspeakers* (Phd thesis). Ecole Polytechnique Fédérale de Lausanne.

Brinkmann, F., Lindau, A., Weinzierl, S., Geissler, G., & van de Par, S. (2013). A high resolution head-related transfer function database including different orientations of head above torso (pp. 596–599). Presented at the AIA-DAGA, Merano (Italia).

Bronkhorst, A. W. (1995). Localization of real and virtual sound sources. *Journal of the American Society of America*, *98*(5), 2542–2553.

Brungart, D., Rabinowitz, W., & Durlach, N. (1999). Auditory localization of nearby sources. II - Head-related transfer functions. *Journal of the American Society of America*, *106*(3), 1465–1479.

Bryk, P.-Y. (2014). *Optimisation de la restitution sonore en basses fréquences* (Rapport de stage). Genesis.

Bucklein, R. (1981). The audibility of frequency response irregularities. *Journal of the Audio Engineering Society*, *29*(3), 126–131.

Busson, S. (2006). *Individualisation d'indices acoustiques pour la synthèse binaurale* (Phd thesis). Université de la Méditerranée Aix-Marseille II.

Busson, S., Nicol, R., & Katz, B. (2005). Subjective investigations of the interaural time difference in the horizontal plane. Presented at the 118th AES Convention, Barcelona (Spain).

Carlile, S., Leong, P., & Hyams, S. (1997). The nature and distribution of errors in sound localization by human listeners. *Hearing Research*, *114*(1/2), 179–196.

Corteel, E. (2006). Equalization in an extended area using multichannel inversion and Wave Field Synthesis. *Journal of the Audio Engineering Society*, *54*(12), 1140–1161.

Craven, P., & Gerzon, M. (1992). Practical adaptative room and loudspeaker equaliser for hi-fi use. Presented at the UK 7th AES Conference : Digital Signal Processing, London (UK).

Damaske, P. (1971). Head-related two-channel stereophony with loudspeaker reproduction. *Journal of the American Society of America*, *50*, 1109–1115.

Daniel, J. (2000). *Représentation des champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia* (Phd thesis). Université Paris VI.

Defraene, B., Van Waterschoot, T., Diehl, M., & Moonen, M. (2012). Perception-based nonlinear loudspeaker compensation through embedded convex optimization. Presented at the European Signal Processing Conference (EUSIPCO 2012), Bucharest (Romania).

Duda, R., & Martens, W. (1998). Range dependence of the response of a spherical head model. *Journal of the American Society of America*, *104*(5), 3048–3058.

Duraiswami, R., Zotkin, D., & Gumerov, N. (2004). Interpolation and range extrapolation of Head-Related Transfer Functions. Presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Montreal (Canada).

Elliott, S. J., & Nelson, P. (1989). Multiple point equalization in a room using adaptative digital filters. *Journal of the Audio Engineering Society*, *37*(11), 899–907.

Everest, F. A., & Pohlman, K. C. (2009). *Master Handbook of acoustics - 5th edition*.

Fazenda, B., Stephenson, M., & Goldberg, A. (2015). Perceptual thresholds for the effects of room modes as a function of modal decay. *Journal of the American Society of America*, *137*(3), 1088–1098.

Fielder, L. (2003). Analysis of traditional and reverberation reducing methods of room equalization. *Journal of the Audio Engineering Society*, *51*(1/2), 3–26.

Friot, E., Alvarez, F., & Chatron, J. (2016). Caractérisation des transferts acoustiques entre banc de l'organiste et public. Presented at the Congrès Français d'Acoustique, Le Mans (France).

Fryer, P. A. (1977). Loudspeaker distorsions, can we hear them? *Hi-fi News & Record Review*, *22*, 51–56.

Gade, A. (2007). Acoustics in Hall for Speech and Music (pp 301-350). *Springer Handbook of Acoustics* (Springer New-York.). Thomas Rossing.

Galvez, M., & Fazi, F. (2015). Loudspeaker arrays for transaural reproduction. Presented at the International Congress on Sound and Vibrations, Firenze (Italia).

Galvez, M., Takeuchi, T., & Fazi, F. (2016). A listener adaptative optimal source distribution system for virtual sound imaging. Presented at the 140th AES Convention, Paris (France).

Gardner, W. G., & Martin, K. (1994). *HRTF Measurements of a KEMAR Dummy Head Microphone* (Technical Report No. 280). MIT: Media Lab Perceptual Computing.

Ghorbal, S., Séguier, R., & Bonjour, X. (2016). Process of HRTF individualization by 3D statistical ear model. Presented at the 141th AES Convention, Los Angeles (USA).

Guillon, P. (2009). *Individualisation des indices spectraux pour la synthèse binaurale (recherche et exploitation des similarités inter-individuelles pour l'adaptation ou la reconstruction de HRTF)* (Phd thesis). Université du Maine.

Guillon, P., Guignard, T., & Nicol, R. (2008). Head-Related Transfer Function customization by frequency scaling and rotation shift based on a new morphological matching method. Presented at the 125th AES Convention, San Francisco (USA).

Guillon, P., & Nicol, R. (2008). Head-Related Transfer Function reconstruction from sparse measurements considering a priori knowledge from database analysis: a pattern recognition approach. Presented at the 125th AES Convention, San Francisco (USA).

Harma, A., Karjalainen, M., Savioja, L., Valimaki, V., Laine, U. K., & Huopaniemi, J. (2000). Frequency-Warped signal Processing for Audio Applications. *Journal of the Audio Engineering Society*, *48*(11), 1011–1031.

Herzog, P. (2005). Procede et dispositif de correction active des proprietes acoustiques d'une zone d'ecoute d'un espace sonore (N° EP1941491A1).

Hosoe, S., Nishino, T., Itou, K., & Takeda, K. (2006). Development of micro-dodecahedral loudspeaker for measuring Head-Related Transfer Functions in the proximal region (pp. 329–332). Presented at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toulouse (France).

Huang, Y., Benesty, J., & Chen, J. (2007). On Crosstalk Cancellation and Equalization With Multiple Loudspeakers for 3-D Sound Reproduction. *IEEE Signal Processing Letters*, *14*(10), 649–653.

Huttunen, T., Vanne, A., Harder, S., Paulsen, R., King, S., Perry-Smith, L., & Karkkainen, L. (2014). Rapid generation of personalized HRTFs. Presented at the AES 55th International Conference, Helsinki (Finland).

ISO 532B. (1975). *Méthode de calcul du niveau d'isosonie*.

ITU-R BS.1116-3. (2015). *Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems*. International Telecommunication Union.

ITU-R BS.1534-1. (2001). *Method for the subjective assessment of intermediate quality level of coding systems*. International Telecommunication Union.

ITU-R BS.1534-3. (2015). *Method for the subjective assessment of intermediate quality level of coding systems*. International Telecommunication Union.

ITU-T P.58. (2013). *Head and torso simulators for telephonometry*. International Telecommunication Union.

Iwaya, Y. (2006). Individualization of head-related transfer functions with tournament-style listening test: Listening with other's ears. *Acoustical Science and Technology*, *27*(6), 340–343.

Jakka, J. (2005). *Binaural to multichannel audio upmix* (Master's Thesis). Helsinki University of Technology.

Karjalainen, M., Piirila, E., & Jarvinen, A. (1996). Loudspeaker response equalization using warped digital filters. Presented at the Norsig 96, Espoo (Finland).

Katz, B. (2001). Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. *Journal of the American Society of America*, *110*(5), 2440–2448.

Katz, B., & Noistering, M. (2014). A comparative study of interaural time delay estimation methods. *Journal of the American Society of America*, *135*(6), 3530–3540.

Kim, S.-M., & Choi, W. (2005). On the externalization of virtual sound images in headphone reproduction: a wiener filter approach. *Journal of the American Society of America*, *117*(6), 3657–3665.

Kirkeby, O., & Nelson, P. (1998). Local sound field reproduction using two closely spaced loudspeakers. *Journal of the American Society of America*, *104*(4), 1973–1981.

Kirkeby, O., & Nelson, P. (1999). Digital filter design for inversion problems in sound reproduction. *Journal of the Audio Engineering Society*, *47*(7/8), 583–595.

Kirkeby, O., Nelson, P., & Hamada, H. (1998). The Stereo Dipole - A virtual source imaging system using two closely spaced loudspeakers. *Journal of the Audio Engineering Society*, *46*(5), 387–395.

Kirkeby, O., Nelson, P., Hamada, H., & Orduna-Bustamante, F. (1998). Fast deconvolution of multichannel systems using regularization. *IEEE Transaction on Speech and Audio Processing*, *6*(2), 189–194.

Klippel, W. (1996). Compensation for Nonlinear Distortion of Horn Loudspeakers by Digital Signal Processing. *Journal of the Audio Engineering Society*, *44*(11), 964–972.

Kreuzer, W., Majdak, P., & Zhengsheng, C. (2009). Fast multipole Boundary Element Method to calculate HRTF for a wide frequency range. *Journal of the American Society of America*, *126*, 1280–1290.

Larcher, V. (2001). *Techniques pour la spatialisation des sons pour la réalité virtuelle* (Phd thesis). Université Paris VI.

Lavandier, M. (2005). *Différences entre enceintes acoustiques : une évaluation physique et perceptive* (Phd thesis). Université de la Méditerranée Aix-Marseille II.

Lentz, T. (2006). Dynamic crosstalk concellation for binaural synthesis in virtual reality environments. *Journal of the Audio Engineering Society*, *54*(2), 283–284.

Loudness Toolbox. (2016). . Retrieved November 16, 2016, from http://genesis-acoustics.com/sonie_en_ligne-32.html

Maazaoui, M., & Warusfel, O. (2016). Estimation of individualized HRTF in unsupervised conditions. Presented at the 140th AES Convention, Paris (France).

Majdak, P., Iwaya, Y., Carpentier, T., Nicol, R., Parmentier, M., Roginska, A., Suzuki, Y., Watanabe, K., Wierstorf, H., Ziegelwanger, H., & Noistering, M. (2013). Spatially Oriented Format for Acoustics: A data exchange format representing Head-Related Transfer Functions. Presented at the 134th AES Convention, Roma (Italia).

Makivirta, A., Antsalo, P., Karjalainen, M., & Valimaki, V. (2001). Low frequency Modal Equalization of Loudspeaker-Room Responses. Presented at the AES 111th Convention, New-York (USA).

Mertins, A., Mei, T., & Kallinger, M. (2010). Room impulse response shortening/reshaping with infinity- and p-norm optimization. *IEEE Transactions on Audio, Speech and Language Processing*, *18*(2), 249–259.

Meynial, X. (1999). Active acoustic impedance control for noise reduction (Patent N° W09959377A1).

Michaud, P.-Y. (2012). *Distorsions des systèmes de reproduction musicale : protocole de caractérisation perceptive* (Phd thesis). Aix-Marseille université.

Michaud, P.-Y., Lavandier, M., Meunier, S., & Herzog, P. (2015). Objective characterization of perceptual dimensions underlying the sound reproduction of 37 single loudspeakers in a room. *Acta Acustica united with Acustica*, *101*(3), 603–615.

Mitra, S. K. (1998). *Digital Signal Processing: A Computer Based Approach* (McGraw-Hill.).

Moller, H. (1992). Fundamentals of Binaural Technology. *Applied Acoustics*, *36*(3/4), 171–218.

Moore, A. H., Tew, A. I., & Nicol, R. (2007). Headphone transparification: A novel method for investigating the externalisation of binaural sound. Presented at the 123rd AES Convention, New-York (USA).

Moore, A. H., Tew, A. I., & Nicol, R. (2010). An initial validation of individualized crosstalk cancellation filters for binaural perceptual experiments. *Journal of the American Society of America*, *58*(1/2), 36–45.

Moore, B. C. ., & Glasberg, B. R. (1996). A revision of Zwicker's loudness model. *Acta Acustica united with Acustica*, *82*, 335–345.

Moore, B. C. ., & Glasberg, B. R. (2007). Modeling binaural loudness. *Journal of the American Society of America*, *121*(3), 1604–1612.

Moore, B. C. ., Glasberg, B. R., & Baer, T. (1997). A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Audio Engineering Society*, *45*(4), 224–240.

Morse, P., & Feschbach, H. (1953). *Methods of theoretical physics* (McGraw-Hill.).

Na, H. S., Lee, C. N., & Cheong, O. (2002). Voronoi diagrams on the sphere. *Computational Geometry*, *23*(2), 183–194.

Nagoya. (2016). *Nagoya HRTF databases*. Retrieved November 16, 2016, from http://www.sp.m.is.nagoya-u.ac.jp/HRTF/database.html

Neely, S., & Allen, J. (1979). Invertibility of a room impulse response. *Journal of the American Society of America*, *66*(1), 165–169.

Nguyen, K.-V., Carpentier, T., Noistering, M., & Warusfel, O. (2010). Calculation of HRTF in the proximity region using spherical harmonic decomposition comparison with measurement and evaluation. Presented at the 2nd International Symposium on Ambisonics and Spherical Acoustics, Paris (France).

Nicol, R., Gros, L., Colomes, C., Warusfel, O., Noistering, M., Bahu, H., Katz, B., & Simon, L. (2014). A Roadmap for Assessing the Quality of Experience of 3D Audio Binaural Rendering. Presented at the EAA joint Symposium on Auralization and Ambisonics, Berlin (Germany).

Nicol, R., Lemaire, V., Bondu, A., & Busson, S. (2006). Looking for a relevant similarity criterion for HRTF clustering: a comparative study. Presented at the 120th AES Convention, Paris (France).

Noistering, M., Carpentier, T., & Warusfel, O. (2012). ESPRO 2.0 - Implementation of a surrounding 350-Loudspeaker array for a sound field reproduction. Presented at the UK 25th AES Conference, York (UK).

Novak, A., Simon, L., & Lotton, P. (2009). Nonlinear System Identification Using Exponential Swept-Sine Signal. *IEEE Transactions on Instrumentation and Measurement*, *59*(8), 2220–2229.

Olive, S., & Toole, F. (1989). The Detection of Reflections in Typical Rooms. *Journal of the Audio Engineering Society*, *37*(7/8), 539–553.

Oppenheim, A., & Schafer, R. (1975). *Digital signal processing (pp 337-345)* (Prentice-Hall international Editions.).

Paquier, M., Koehl, V., & Jantzem, B. (2011). Effects of headphone transfer function scattering on sound perception. Presented at the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz (USA).

Parodi, Y. L. (2010). *A systematic study of binaural reproduction systems through loudspeakers: a multiple Stereo-Dipole approach* (Phd thesis). Aalborg university.

Parodi, Y. L., & Rubak, P. (2010). Objective evaluation of the sweet spot size in spatial sound reproduction using elevated loudspeakers. *Journal of the American Society of America*, *128*(3), 1045–1055.

Parodi, Y. L., & Rubak, P. (2011). Analysis of Design Parameters for Crosstalk Cancellation Filters Applied to Different Loudspeaker Configurations. *Journal of the Audio Engineering Society*, *59*(5), 304–320.

Parseihian, G. (2012). *Sonification binaurale pour l'aide à la navigation* (Phd thesis). Université Paris 6.

Parseihian, G., & Katz, B. (2012). Rapid HRTF adaptation using a virtual auditory environment. *Journal of the American Society of America*, *131*(4), 2948–2947.

Peeters, G. (2004). *A large set of audio features for sound description (similarity and classification) in the CUIDADO project* (CUIDADO IST Project report).

Périaux, B., Ohl, J.-L., & Thévenot, P. (2015). *Le son multicanal - De la production à la diffusion du son 5.1, 3D et binaural*. Dunod.

Pollow, M., Nguyen, K.-V., Warusfel, O., Carpentier, T., Muller-Trapet, M., Vorlander, M., & Noistering, M. (2012). Calculation of Head-Related Transfer Functions for arbitrary field points using spherical harmonics decoposition. *Acta Acustica united with Acustica*, *98*, 72–82.

Pulkki, V. (1997). Virtual Sound Source Positioning Using Vector Base Amplitude Panning. *Journal of the Audio Engineering Society*, *45*(6), 456–466.

Pulkki, V., & Hirvonen, T. (2005). Localization of virtual sources in multichannel audio reproduction. *IEEE Transaction on Speech and Audio Processing*, *13*(1), 105–120.

Qu, T., Xiao, Z., Gong, M., Huang, Y., Li, X., & Wu, X. (2009). Distance-dependent Head-Related Transfer Functions measured with a high spatial resolution using a spark gap. *IEEE Transactions on Audio, Speech and Language Processing*, *17*(6), 1124–1132.

Qu, T., Xiao, Z., Gong, M., Huang, Y., Li, X., & Wu, X. (2016). *PKU&IOA HRTF Database*. Retrieved November 16, 2016, from http://www.cis.pku.edu.cn/auditory/english/audio_publications_1.htm

Radlovik, B., & Kennedy, R. (2000). Nonminimum Phase Equalization and its subjective importance in room acoustics. *IEEE Transactions on Audio, Speech and Language Processing*, *8*(6), 728–737.

Ramos, G., & Lopez, J. (2006). Filter design method for loudspeaker equalization based on IIR parametric filters. *Journal of the Audio Engineering Society*, *54*(12), 1162–1178.

Rayleigh, L. (1907). On our perception of sound direction. *The London, Edimburgh, and Dublin Philosophical Magazine and Journal of Science*, *13*(74), 214–232.

Recanzone, G. H. (1998). Rapidly induced auditory plasticity: The ventriloquism aftereffect. *National Academy of Sciences of the USA*, *95*(3), 869–875.

Rivet, E., Karkar, S., & Lissek, H. (2016). Egalisation modale des salles avec des absorbeurs électroacoustiques. Presented at the Congrès Français d'Acoustique, Le Mans (France).

Robotham, T., Stephenson, M., & Lee, H. (2016). The Effect of a Vertical Reflection on the Relationship between Preference and Perceived Change in Timbre and Spatial Attributes. Presented at the 140th AES Convention, Paris (France).

Rui, Y., Yu, G., Xie, B.-S., & Liu, Y. (2013). Calculation of individualized near-field head-related transfer function database using boundary element method. Presented at the 134th AES Convention, Roma (Italia).

Santala, O., Delikaris-Manias, S., Ronkko, P., Azcoaga, E., Rekola, I., & Pulkki, V. (2014). Auditory perception of spatially distributes broadband pulse sequences. Presented at the AES 55th International Conference, Helsinki (Finland).

Santillan, A., Pedersen, C., & Lydolf, M. (2007). Experimental implementation of a low-frequency global sound equalization based on free field propagation. *Applied Acoustics*, *68*(10), 1063–1085.

Schroeder, M. R. (1969). Digital simulation of sound transmission in reverberant spaces. *Journal of the American Society of America*, *47*(2), 424–431.

Seber, G. A. ., & Wild, C. J. (2003). *Nonlinear regression*.

Shi, K., Zhou, T. G., & Viberg, M. (2007). Compensation for nonlinearity in Hammerstein system using the coherence function with application to nonlinear acoustic echo cancellation. *IEEE Transactions on Signal Processing*, *55*(12), 5853–5858.

Sivonen, V. P., & Ellermeier, W. (2008). Binaural loudness for artificial-head measurements in directional sound fields. *Journal of the Audio Engineering Society*, *56*(6), 452–461.

Small, R. (1972a). Closed-Box Loudspeaker Systems part I - analysis. *Journal of the Audio Engineering Society*, *20*(10), 978–808.

Small, R. (1972b). Direct radiator loudspeaker system analysis. *Journal of the Audio Engineering Society*, *20*(5), 383–395.

SOFA. (2016). *Spatially Oriented Format for Acoustics website*. Retrieved November 16, 2016, from https://www.sofaconventions.org/mediawiki/index.php/Files

Stevens, S. S. (1955). The measurement of loudness. *Journal of the American Society of America*, *27*(5), 815–829.

Takeuchi, T., & Nelson, P. (2002). Optimal source distribution for binaural synthesis over loudspeakers. *Journal of the American Society of America*, *112*(6), 2786–2797.

Thiele, A. (1971a). Loudspeakers in vented boxes: part I. *Journal of the Audio Engineering Society*, *19*(5), 181–191.

Thiele, A. (1971b). Loudspeakers in vented boxes: part II. *Journal of the Audio Engineering Society*, *19*(6), 192–204.

Toole, F., & Olive, S. (1988). The modification of timbre by resonances: perception and measurement. *Journal of the Audio Engineering Society*, *36*(3), 122–142.

Valimaki, V., & Reiss, J. D. (2016). All about audio equalization: solutions and frontiers. *Applied Sciences*, *6*(5).

Vannier, M. (2016). *Sonie de champ acoustiques stationnaires en situation d'écoute dichotique* (Phd thesis). Institut National des Sciences Appliquées de Lyon.

Vidal, A., Herzog, P., & Lambourg, C. (2016). Comparaison de méthodes d'égalisation pour une restitution en salle d'écoute. Presented at the Congrès Français d'Acoustique, Le Mans (France).

Vincent, E. (2005). MUSHRAM: A MATLAB interface for MUSHRA listening tests. Retrieved from http://c4dm.eecs.qmul.ac.uk/downloads/

Volk, F., Heinemann, F., & Fastl, H. (2008). Externalization in binaural synthesis: effects of recording environment and measurement procedure. Presented at the Acoustics'08, Paris (France).

Wang, Z., & Chan, F. (2014). Customization of Head-Related Impulse Response via two dimension common factor decomposition and sampled measurements. Presented at the 136 AES Convention, Berlin (Germany).

Ward, D., & Elko, G. (1999). Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation. *IEEE Signal Processing Letters*, *6*(5), 106–108.

Warusfel, O. (2003). Listen HRTF database. Retrieved from http://recherche.ircam.fr/equipes/salles/listen/index.html

Warusfel, O., Derogis, P., & Caussé, R. (1997). Radiation synthesis with digitally controlled loudspeakers. Presented at the 123rd AES Convention, New-York (USA).

Warusfel, O., & Misdaris, N. (2004). Sound source radiation synthesis: from stage performance to domestic rendering. Presented at the 116th AES Convention, Berlin (Germany).

Welti, T., & Devantier, A. (2006). Low frequency optimization using multiple subwoofers. *Journal of the Audio Engineering Society*, *54*(5), 347–364.

Wierstorf, H., Geier, M., Raake, A., & Spors, S. (2011). A free database of Head-Related Impulse Response measurements in the horizontal plane with multiple distances (Vol. 130). Presented at the AES Convention, London (UK).

Wilson, R. (1991). Equalization of loudspeaker drive units considering both on and off axis responses. *Journal of the Audio Engineering Society*, *39*(3), 127–139.

Yang, J., Gan, W.-S., & Tan, S.-E. (2003). Improved sound separation using three loudspeakers. *Acoustic research letters online-ARLO*, *4*(2), 47–52.

Yu, G.-Z., Xie, B.-S., & Rao, D. (2010). Characteristics of Near-field head-related transfer function for KEMAR. Presented at the AES 40th International Conference, Tokyo (Japan).

Zhong, X.-L., Xu, X., & Xie, B.-S. (2016). Auditory consistency of HRTF of KEMAR from different databases. Presented at the 23rd International Congress on Sound and Vibration, Athens.

Zhong, X.-L., Zhang, F.-C., & Xie, B.-S. (2013). On the spatial symetrie of Head-Related Transfer Functions. *Applied Acoustics*, *74*(6), 856–864.

Zotkin, D., Duraiswami, R., Grassi, E., & Gumerov, N. (2006). Fast Head-Related Transfer Function measurement via reciprocity. *Journal of the American Society of America*, *120*(4), 2202–2215.

Zwicker, E., Fastl, H., Widmann, U., Kurakata, K., Kuwano, S., & Namba, S. (1991). Program for calculating loudness according to DIN 45631 (ISO 532B). *Journal of the Acoustical Society of Japan*, *12*(1), 39–42.

Annexe A Conception d'une source adaptée à la diffusion en proximité

Les travaux de cette thèse concernent la reproduction sonore en proximité de l'auditeur. Ceci implique des besoins spécifiques en termes de sources électro-acoustiques. Le principe de fonctionnement d'un haut-parleur dynamique est brièvement rappelé et la conception d'une source adaptée à nos besoins est présentée.

Fonctionnement du haut-parleur

Le haut-parleur est un transducteur électro-acoustique : un signal électrique permet de créer un déplacement mécanique (de la membrane), et ce dernier permet de créer des variations de pression acoustique. Le type de haut-parleur le plus répandu est le haut-parleur électrodynamique : le mouvement de la membrane est généré par une bobine mobile parcourue par un courant électrique et placée dans le champ magnétique d'un aimant permanent.

Le mouvement de la membrane entraine des variations de pressions rayonnées par la face avant et par la face arrière. Lorsque le haut-parleur est nu, ces variations de pressions peuvent interférer créant un « court-circuit acoustique ». Pour l'éviter, les haut-parleurs sont montés dans une enceinte, la plus simple étant l'enceinte close. Une enceinte à évent ou « bass-reflex » est parfois employée, consistant à introduire une (ou plusieurs) ouverture(s) afin de créer un résonateur [Thiele, 1971a, 1971b]. Ces résonateurs sont généralement dimensionnés pour agir en basses fréquences et étendre la bande passante de fonctionnement.

Le champ de pression généré par un haut-parleur dans une enceinte close dépend à la fois du haut-parleur, et de l'enceinte qui contrôle sa face arrière.

Limites en bande-passante

Le comportement en basses fréquences d'un haut-parleur est analogue à celui d'un filtre passe-haut, caractérisé par une fréquence de coupure et un facteur de qualité. Pour une enceinte close, ce filtre est d'ordre deux et ses caractéristiques sont liées au volume d'air présent dans l'enceinte et aux propriétés du haut-parleur [Small, 1972a]. Ses propriétés sont généralement spécifiées par le constructeur et peuvent être estimées à partir de la mesure de son impédance électrique [Small, 1972b]. Elles dépendent principalement de la taille de la membrane et des matériaux du haut-parleur. Concernant l'encoffrement, plus le volume de l'enceinte close est réduit et plus la fréquence de coupure et le facteur de qualité sont élevés. Par exemple, la réponse en basse fréquence d'un haut-parleur Visaton FRS 8M est simulée pour plusieurs charges acoustiques dans la Figure 65.

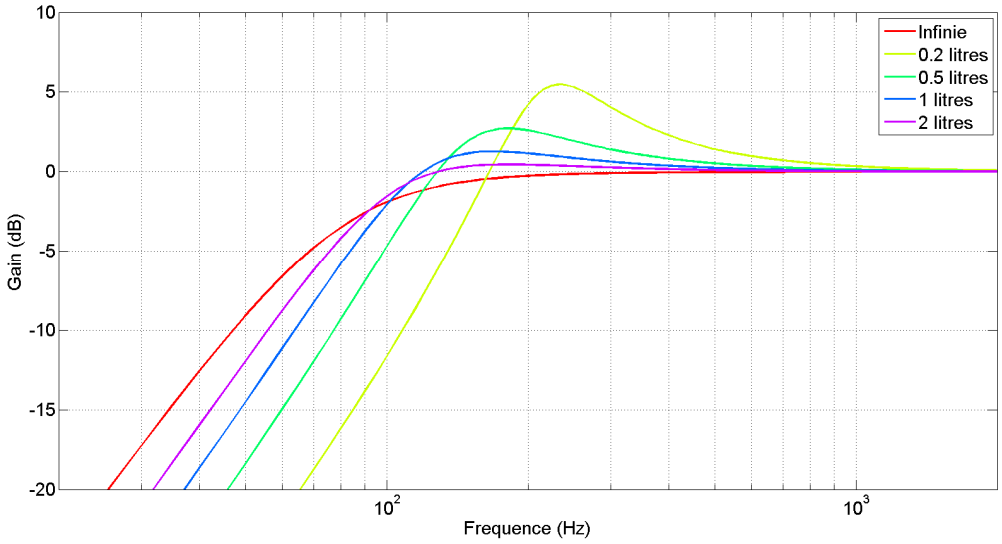


Figure 65 : modélisation de la réponse d'un haut-parleur Visaton FRS 8M pour différentes charges acoustiques

Limites en niveau : distorsion non-linéaire du signal

Le haut-parleur est limité en niveau de fonctionnement : il peut avoir un comportement non-linéaire lorsque le niveau d'excitation est trop important. Ces non-linéarités se traduisent par l'apparition de composantes qui n'étaient pas présentes dans le signal d'origine. Pour une sinusoïde pure, ces distorsions non-linéaires correspondent généralement aux harmoniques de cette sinusoïde, et sont souvent caractérisées par le Taux de Distorsion Harmonique totale (Total Harmonic Distorsion, THD) exprimé en % :

$$THD = 100.\left(\frac{\sqrt{\sum_{k=2}^{N} P_{xx}(k)}}{\sqrt{\sum_{k=1}^{N} P_{xx}(k)}}\right) \qquad (89)$$

Avec $P_{xx}(k)$ la puissance de la $k^{ième}$ harmonique (le calcul est limité à un ordre N d'harmoniques). Une valeur de 0 % correspond à l'absence de distorsion harmonique.

Ce phénomène augmente avec l'excursion de la membrane ; il apparait donc principalement pour les fréquences basses, introduisant des artefacts aux fréquences plus élevées. L'effet de la distorsion non-linéaire sur la perception auditive est complexe, et il est d'usage de tenter de la minimiser pour garantir une reproduction sonore de qualité. Cela se traduit par une limite du niveau de fonctionnement des sources électro-acoustiques.

Dimensionnement d'un haut-parleur

Le dimensionnement d'un haut-parleur doit se faire en fonction de la bande-passante et du niveau de reproduction désiré, mais doit également être adapté à la distance de restitution. Notamment, la surface de la membrane du haut-parleur conditionne sa capacité à reproduire des basses fréquences. Pour l'illustrer, le rayonnement d'un haut-parleur est assimilé à celui d'un monopole en champ lointain, et la pression à une distance r vaut :

$$|p(r,f)| = \frac{Q\rho f}{2r} \tag{90}$$

Avec Q le débit de la source, ρ la masse volumique de l'air (1.2 kg.m^{-3}), f la fréquence et r la distance émetteur-récepteur. Le débit est le produit du volume d'air déplacé par unité de temps, et pour l'estimer nous allons assimiler le haut-parleur à un piston plan.

A une fréquence donnée, le volume d'air déplacé correspond à celui déplacé par un aller-retour de la membrane du haut-parleur. En pratique, les haut-parleurs sont spécifiés pour un déplacement maximal de membrane afin que la reproduction puisse être considérée comme linéaire. Ce déplacement maximal désigné par x_{max} par les constructeurs, varie d'un modèle à l'autre et peut être de l'ordre de quelques millimètres. Le volume d'air maximal que peut déplacer la membrane d'un haut-parleur est alors le suivant :

$$V = \pi r_{HP}^2 \, x_{\max} \tag{91}$$

Où r_{HP} est le rayon de la membrane du haut-parleur. Le débit maximal est alors le suivant :

$$Q = V 2\pi f \tag{92}$$

Il est ainsi possible de déterminer un rayon minimal de membrane pour reproduire un niveau de pression donné, et respectant un déplacement maximal de la membrane :

$$r_{HP} = \sqrt{\frac{rp}{\rho f^2 \pi^2 x_{max}}} \tag{93}$$

Par exemple, pour reproduire un niveau de 100 dBSPL (2 Pa) aux distances de 20, 40 et 80 cm, et une excursion maximale de 1 mm, les rayons minimaux sont reportés dans le Tableau 14. A une distance donnée, la réduction du rayon de la membrane conduit inévitablement à une réduction des performances en basses fréquences. Pour une taille de membrane donnée, la gamme de fréquence qu'il est possible de reproduire est donc plus étendue à faible distance. Le dimensionnement de source est alors moins exigeant pour une diffusion en proximité de l'auditeur.

Toutefois, pour une reproduction sonore en proximité il faut tenir compte des composantes évanescentes du champ liées au comportement réel de la membrane et à la diffraction par l'enceinte.

	20 cm	40 cm	80 cm
100 Hz	5.8 cm	8.2	11.6
150 Hz	3.9 cm	5.5 cm	7.7 cm
200 Hz	2.9 cm	4.1 cm	5.8 cm

Tableau 14 : rayons minimaux de membrane de haut-parleur pour générer un niveau de 100 dB avec une excursion maximale de 1 mm.

Conception d'une source

Une enceinte a été conçue de sorte à répondre aux besoins spécifiques de la restitution audio en proximité. Ceci implique qu'elle doit être de petite taille et avoir une réponse en fréquence suffisante aux basses fréquences. Par ailleurs, cette source est employée pour mesurer des HRTF : la directivité de la source doit alors être limitée pour caractériser correctement l'effet de diffraction par des mannequins.

Pour cela, un haut-parleur Visaton FRS 8M (membrane de 6 cm de diamètre) a été encastré dans une enceinte close d'un volume de 1,2 L. Une photo de cette enceinte est visible sur la Figure 67. Nous considérons que l'effet de diffraction par la tête est prépondérant, et l'effet de diffraction par le torse des mannequins est secondaire. L'effet de la directivité de la source doit être négligeable pour les incidences englobant la tête, et doit être le plus faible possible pour les incidences englobant le torse. En considérant que les oreilles sont placées à 10 cm du centre de la tête (majorant des mannequins caractérisés), lorsque la source est placée à 40 cm de distance l'angle d'incidence entre l'axe du haut-parleur et l'oreille est de 14° comme l'illustre la Figure 66 (figure de gauche). Il est indispensable que la directivité de la source soit négligeable aux incidences inférieures à 14° pour que la mesure de l'effet de diffraction par la tête ne soit pas influencée par la directivité de l'enceinte.

En considérant que les épaules sont placées à 12 cm plus bas que les oreilles et d'une largeur de 22 cm (schéma simplifié en Figure 66 à droite), l'angle d'incidence entre l'extrémité de l'épaule et l'axe du haut-parleur est de 34°. Pour que l'effet de diffraction par les épaules ne soit pas influencé significativement par la directivité de l'enceinte, il faut qu'elle soit faible pour les incidences inférieures à 34°.

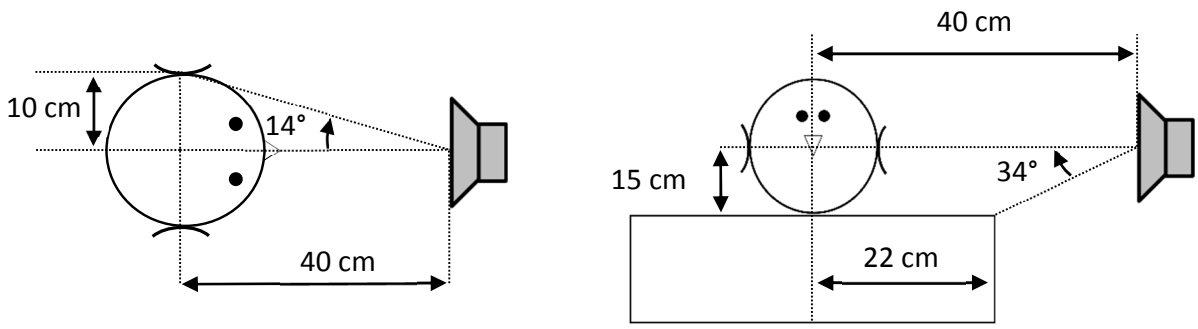


Figure 66 : représentation d'angles d'incidence avec le haut-parleur : à gauche angle entre l'oreille et le centre (14°), et à droite angle entre l'épaule et le centre (34°)

La directivité dans le plan horizontal de la source utilisée a été caractérisée. Pour cela, des mesures de fonction de transfert ont réalisées pour différentes positions de la source, avec une résolution de 5° entre -15° et +15° et une résolution de 15° sur le reste du cercle. Le diagramme de directivité par bandes d'octave est représenté en Figure 67. La directivité augmente avec la fréquence : en incidence arrière, l'énergie dans la bande de 250 Hz est atténuée de 4 dB et l'énergie dans la bande de 8 kHz est atténuée de 25 dB par rapport à l'incidence frontale. Pour toutes les bandes de fréquences représentées, l'atténuation liée à la directivité est inférieure à 1 dB pour les angles

d'incidence inférieur à 15°, et inférieure à 4 dB pour les angles d'incidence inférieurs à 30°. La directivité est considérée comme négligeable pour les angles inférieurs à 15°, donc la diffraction par la tête ne sera pas influencée significativement par la directivité de la source. La directivité pour les angles entre 15° et 30° n'est pas négligeable, mais est relativement faible : la diffraction par les épaules pourra être influencée légèrement par la directivité.

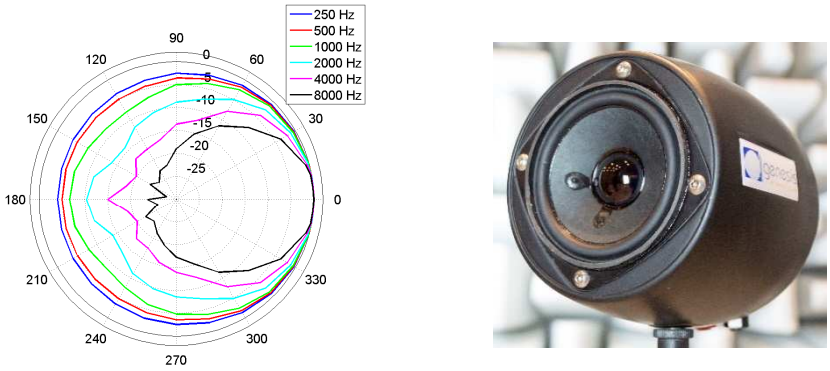


Figure 67 : diagramme de directivité par bandes d'octave et photo de la source conçue

La fonction de transfert de la source a été caractérisée à l'aide d'un microphone champ libre B&K 4190 ; elle est représentée par la Figure 68. La bande passante est de [200 Hz – 18 kHz] à +/- 10 dB, ce qui est compatible avec une diffusion de signaux audio, mais peut nécessiter éventuellement une égalisation.

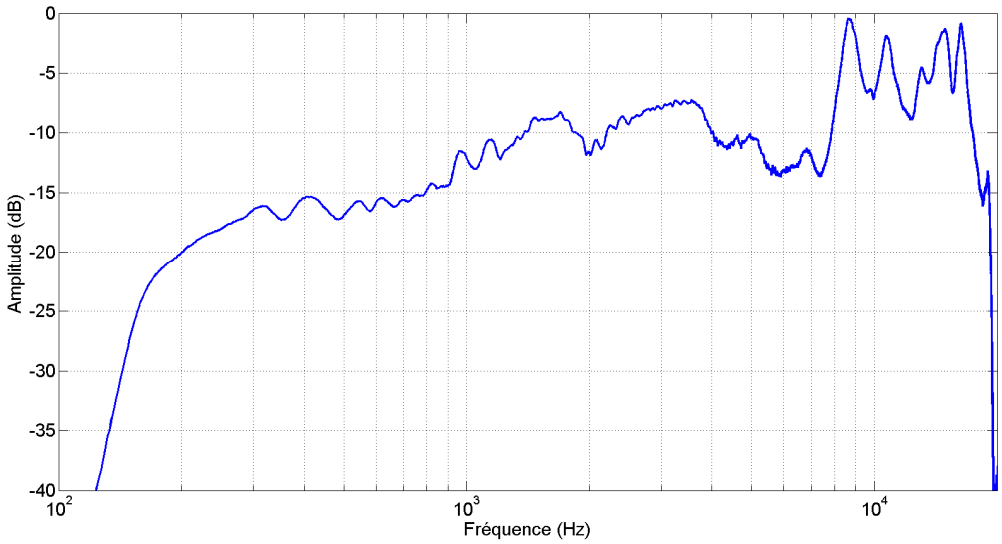


Figure 68 : Fonction de transfert de la source conçue à 40 cm de distance

Annexe B Diffraction par une sphère rigide

Cette annexe décrit le principe du calcul utilisé pour déterminer la réponse de sources monopolaires en présence d'une sphère. Il a été implémenté par Philippe Herzog à partir de travaux antérieurs [Advanced Numerical Solutions LLC, 2016; Morse & Feschbach, 1953]. Pour ce calcul, la sphère est supposée de rayon a et centrée à l'origine du repère $(O, \vec{X}, \vec{Y}, \vec{Z})$. Pour ce calcul, il est plus commode d'utiliser les coordonnées sphériques définies ici par :

$$\begin{cases} r = \sqrt{x^2 + y^2 + z^2} \\ \theta = \mathrm{acos}(z/r) \\ \phi = \mathrm{atan}(y/x) \end{cases} \tag{94}$$

Cette définition utilisée par Morse et Feschbach [Morse & Feschbach, 1953] n'est donc pas la même que celle employée dans la thèse : ϕ correspond à l'azimut, mais θ ne correspond pas à l'élévation. L'axe de rotation azimutal est alors $O\vec{z}$.

Une source monopolaire de débit unitaire ($Q = 1m^3/s$) est placé au point $S(r_0, \theta_0, \phi_0)$. La pression est calculée au point $M(r, \theta, \phi)$. La dépendance temporelle en $e^{+j\omega t}$ est sous-entendue ($\omega = 2\pi f$, avec f la fréquence en Hz). A noter que cette dépendance est inverse de celle utilisée par Morse et Feschbach [Morse & Feschbach, 1953] : elle est adoptée ici par compatibilité avec les autres calculs de la thèse.

Avec ces conventions, la fonction de Green en espace infini s'écrit [Morse & Feschbach, 1953] :

$$\begin{cases} G(P|P_0|\omega) = \dfrac{e^{-jkR}}{R} \\ \qquad = -jk \displaystyle\sum_{n=0}^{\infty} (2n+1) \sum_{m=0}^{n} \epsilon_m \frac{(n-m)!}{(n+m)!} \cos[m(\phi - \phi_0)] \, . \\ \qquad\quad P_n^m(\cos\theta_0)\, P_n^m(\cos\theta) j_n(kr_<) h_n(kr_>) \end{cases} \tag{95}$$

Où $R = \|\overrightarrow{PP_0}\|$, et k le nombre d'onde est $k = \omega/c$. Le facteur ϵ_m vaut 1 pour $m = 0$, et $\epsilon_m = 2$ sinon. Les symboles $r_<$ et $r_>$ désignent respectivement $\min(r, r_0)$ et $\max(r, r_0)$. Cette expression est valide pour $r_< \geq a$.

La fonction j_n est la fonction de Bessel sphérique de première espèce, et h_n est la fonction de Hankel sphérique de seconde espèce (elle serait de première espèce avec une dépendance fréquentielle $e^{-j\omega t}$). Ces deux fonctions sont reliées aux fonctions de Bessel cylindriques de première espèce J_n et de seconde espèce Y_n par les relations suivantes :

$$\begin{cases} j_n(\eta) = \sqrt{\dfrac{\pi}{2\eta}} J_{[n+1/2]}(\eta) \\ y_n(\eta) = \sqrt{\dfrac{\pi}{2\eta}} Y_{[n+1/2]}(\eta) \\ h_n(\eta) = j_n(\eta) - j y_n(\eta) \end{cases} \tag{96}$$

Les fonctions P_n^m sont les fonctions de Legendre associées (non normalisées) qui sont reliées aux polynômes de Legendre P_n par la relation :

$$P_n^m(z) = (1 - z^2)^{\frac{m}{2}} \frac{d^m}{dz^m} P_n(z) \tag{97}$$

L'expression (95) constitue un développement de la fonction de Green sur la série des fonctions obtenues par séparation de variables à partir de l'opérateur de propagation en espace infini exprimé en coordonnées sphériques. Ces fonctions constituent une base de solutions de cet opérateur, et sont orthogonales.

La pression rayonnée par une source monopolaire en espace infini dans un fluide de masse volumique ρ s'exprime facilement à partir de la fonction de Green :

$$\begin{cases} p_i(S|M|\omega) = C(\omega).G(S|M|\omega) \\ avec\ C(\omega) = \dfrac{j\omega\rho}{4\pi} \end{cases} \tag{98}$$

Cette pression constitue la pression incidente sur l'obstacle rigide constitué par la sphère. Cette pression est alors réfléchie totalement, donnant lieu à une pression diffractée p_d telle que la vitesse acoustique totale soit nulle sur la paroi. La projection de cette condition aux limites sur la base des fonctions orthogonales utilisées dans l'équation (95) revient à en multiplier chaque terme par un facteur R_n qui correspond au coefficient de réflexion radial d'une onde convergent sur la sphère :

$$R_n(ka) = \frac{\frac{d}{d_\eta} j_n(\eta)}{\frac{d}{d_\eta} h_n(\eta)} \qquad pour\ \eta = ka \tag{99}$$

La pression diffractée par la sphère s'écrit donc :

$$\begin{aligned} p_d = \quad & C(\omega).[-jk \sum_{n=0}^{\infty} (2n+1) R_n(ka) \sum_{m=0}^{n} \epsilon_m \frac{(n-m)!}{(n+m)!} \cos[m(\phi - \phi_0)]. \\ & P_n^m(\cos\theta_0)\, P_n^m(\cos\theta) h_n(kr_0) h_n(kr)] \end{aligned} \tag{100}$$

Cette expression correspond à celle utilisée dans l'article édité par ANSOL [Advanced Numerical Solutions LLC, 2016], à la différence que cet article utilise la dépendance $e^{-j\omega t}$, et fait appel aux fonctions associées de Legendre normalisées. A noter que dans cet article le coefficient $R_n(ka)$ est exprimé en fonction des fonctions j_n et h_n d'ordre $n-1$ et $n+1$, mais qu'il existe une autre expression légèrement plus simple, utilisée au LMA par J. Sageloli, et qui s'avère un peu plus rapide à calculer :

$$\begin{cases} \dfrac{d}{d_\eta} j_n(\eta) = \dfrac{n j_{n-1}(\eta) - (n+1) j_{n+1}(\eta)}{2n+1} = \dfrac{n}{\eta} j_n(\eta) - j_{n+1}(\eta) \\ idem\ pour\ y_n et\ h_n \end{cases} \tag{101}$$

L'expression (100) a deux inconvénients : d'une part le calcul des fonctions de Legendre est numériquement délicat, et toutes les versions de Matlab ne disposent pas d'une implémentation précise de ce calcul. D'autre part la sommation sur l'indice m ne se prête pas à une vectorisation, ce qui ralentit beaucoup l'exécution du programme.

Cette expression peut alors être simplifiée, en remarquant que les fonctions de Legendre ne décrivent que les variations azimutales de p_d. Un changement de variable peut alors être effectué de manière à ce que la source soit sur l'axe de rotation azimutal. En utilisant ainsi les coordonnées (r, θ', ϕ') telles que $\theta'_0 = \pi$, le champ incident est symétrique autour de cet axe et ne dépend donc pas de ϕ'_0, ni par conséquent de ϕ' :

$$p_d = C(\omega).[-jk\sum_{n=0}^{\infty}(-1)^n(2n+1)R_n(ka)P_n(cos\theta')h_n(kr_0)h_n(kr)] \qquad (102)$$

Cette expression est bien plus simple que la relation (100), mais nécessite de déterminer le cosinus de l'angle θ' pour chaque source S et chaque point M :

$$\cos\big(\theta'(S|M)\big) = \cos\left(\widehat{\overrightarrow{SO}\,\overrightarrow{OM}}\right) = \overrightarrow{OS}.\overrightarrow{OM} \qquad (103)$$

Ce calcul préalable ne fait appel qu'à des produits scalaires simples, et est donc très facile à vectoriser ; l'implémentation est donc bien plus efficace sous cette forme.

Annexe C Mise en forme des mesures de HRTF

Plusieurs étapes de traitement de données sont réalisées pour obtenir des HRTF correspondant à la définition donnée dans le Chapitre II et pour minimiser les biais liés au protocole de mesure. Les opérations sont réalisées à partir des HRIR, obtenues par transformée de Fourier inverse.

Fenêtrage temporel

La première opération appliquée sur les données est un fenêtrage temporel. L'objectif de cette opération est de supprimer les échantillons qui ne contiennent pas d'information, et aussi de limiter la présence d'écho dans les réponses impulsionnelles. Les premiers échantillons correspondants au temps de propagation et à la latence de la chaine audio sont préalablement supprimées (4 ms à 40 cm et 9 ms à 2 m). Ce premier fenêtrage est réalisé de telle sorte que le maximum d'amplitude intervienne autour de t=4 ms pour tous les cas.

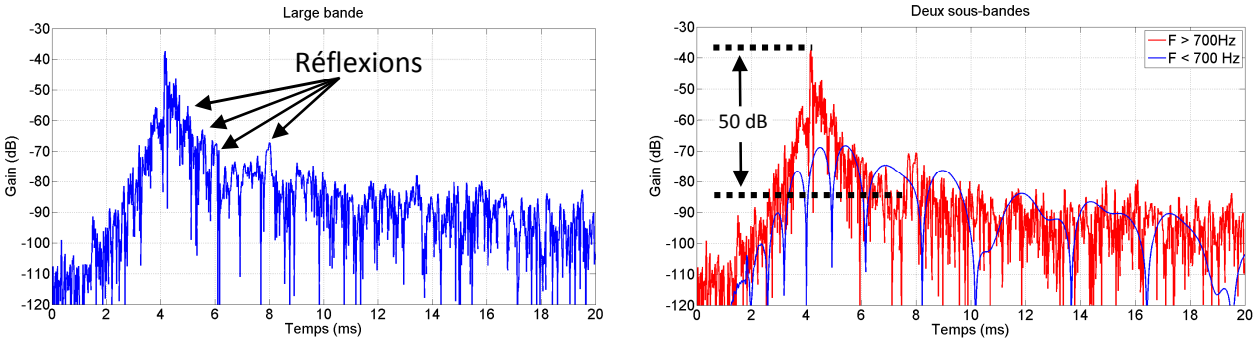


Figure 69 : ETC de la mesure du haut-parleur à 2 m à élévation nulle. Figure de gauche : ETC large bande, figure de droite : ETC en deux sous bandes

Pour mettre en évidence les réflexions, l'Energy Time Curve (ETC) de la mesure du haut-parleur à 2 m à élévation nulle est représentée dans la Figure 69. Outre le pic principal à t=4.2 ms, des pics secondaires peuvent être observés à t=4.6 ms, t=5 ms, t=5.9 ms et t=8 ms et sont assimilés à des réflexions. La réflexion qui émerge le plus est celle intervenant à t=8ms, soit environ 3.8 ms après le pic principal, ce qui correspond à une distance de propagation supplémentaire de 1.4 m pour une célérité égale à 344 m/s. Cette réflexion peut provenir du réseau de haut-parleurs en champ lointain, situé en cercle autour du centre du repère.

Un moyen de supprimer l'effet des réflexions consiste à fenêtrer la réponse impulsionnelle. Le choix de la fenêtre est un compromis entre une longueur de fenêtre suffisamment longue pour conserver toute l'information importante, et suffisamment courte pour supprimer les réflexions. Pour cette raison, il n'est pas possible de supprimer toutes les réflexions, mais moyennant quelques précautions il est possible d'éliminer la réflexion intervenant autour de t=8ms. Si le fenêtrage est effectué à cet instant, le support temporel de la réponse impulsionnelle est alors trop court pour les basses fréquences : une période de 3.8 ms correspond à une fréquence de 263 Hz. En supposant que le support temporel minimal soit de 2 périodes, le fenêtrage ne peut être réalisé pour les fréquences inférieures à 526 Hz. Selon ce critère, il n'est pas possible d'éliminer cette réflexion pour les fréquences plus basses et il est nécessaire de réaliser un filtrage en sous-bandes.

Deux bandes de fréquences sont considérées, séparée par un filtre passe-haut et un filtre passe-bas complémentaires. Un filtre sélectif en fréquence ne permet pas d'être sélectif dans le domaine temporel, et l'identification de la réflexion n'est alors plus possible : il faut trouver un compromis entre la sélectivité du filtre en fréquentiel et en temporel. Si le filtre est peu sélectif en fréquence, il est nécessaire que sa fréquence de coupure soit très supérieure à 526 Hz. Un compromis a été effectué en utilisant des filtres FIR passe-haut et passe-bas complémentaires d'ordre 316 à la fréquence de coupure de 700 Hz. Ainsi, à 500 Hz le gain du filtre passe-haut est de -20 dB et l'essentiel de l'énergie à cette fréquence est contenue dans la réponse impulsionnelle basses fréquences. La troncature n'influe donc pas significativement autour de cette fréquence. Les ETC de ces deux bandes de fréquences sont représentées dans la Figure 69. L'arrivée d'énergie autour de t=8 ms est toujours identifiable pour l'ETC hautes fréquences, et identifiable dès 7.7 ms. Cette différence est liée à l'étalement temporel du filtre FIR. A t=7.6 ms, l'énergie en hautes fréquences est atténuée de 50 dB par rapport au pic principal. Une dynamique de 50 dB permet de considérer que la majeure partie du champ direct est présente, et permet de fenêtrer la réponse impulsionnelle à cet instant en conservant l'essentiel du champ direct.

De plus, en considérant un rayon de tête sphérique de 9 cm le périmètre est de 57 cm. En considérant une célérité du son dans l'air de 344 m/s, l'onde sonore peut faire le tour de la tête en 1.7 ms. Nous considérons que l'essentiel de la diffraction est inclus dans cet intervalle.

Un fenêtrage de 7.4 ms (360 échantillons) a été appliqué aux hautes fréquences (f>700 Hz), et les échantillons de la dernière milliseconde (50 échantillons) sont apodisés par une demi-fenêtre de Hann. La réponse impulsionnelle fenêtrée est complétée par des échantillons nuls de telle sorte que sa durée soit de 43 ms (2048 échantillons). Pour les basses fréquences, un fenêtrage de 43 ms (2048 échantillons) a été appliqué, limitant simplement la durée sans éliminer d'éventuelles réflexions. De manière analogue aux hautes fréquences, la dernière milliseconde est apodisée. La réponse impulsionnelle totale est la somme terme à terme des réponses impulsionnelles basses fréquences et hautes fréquences.

Le résultat du fenêtrage est représenté dans le domaine temporel et dans le domaine fréquentiel en Figure 70. L'effet de cette opération est analogue à un lissage, en temporel les pics à partir de t=7.5 ms sont lissés, et en fréquentiel les oscillations caractéristiques de filtres en peigne sont atténuées essentiellement entre 800 Hz et 5 kHz.

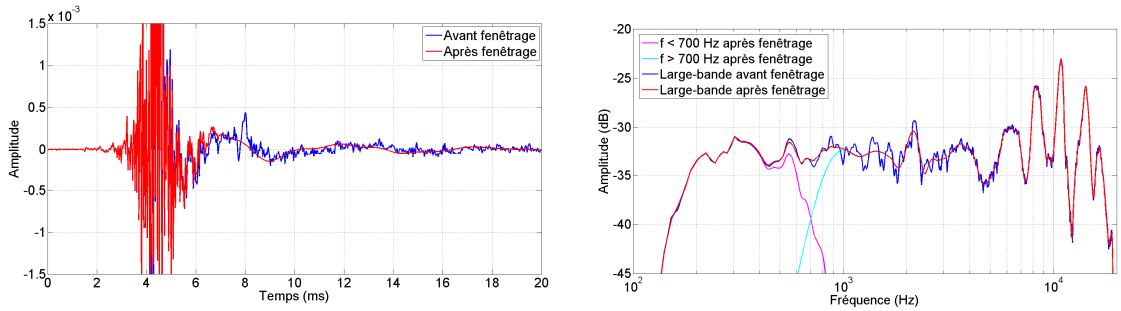


Figure 70 : Effet du fenêtrage sur la réponse impulsionnelle (gauche) et sur la réponse fréquentielle (droite)

Normalisation par la réponse en champ libre

La définition d'une HRTF employée correspond au rapport de la fonction de transfert entre une source et l'oreille, divisée par la fonction de transfert entre la même source et le point situé au centre de la tête (sans la tête). Cette opération permet notamment de supprimer toutes les contributions qui ne sont pas liées à la position spatiale, tel que la réponse de la source et de la chaine audio. Cette opération est réalisée en divisant les spectres complexes des fonctions de transfert. Les réponses impulsionnelles associées à ces HRTF ne sont donc pas systématiquement causales : pour tous les cas où l'oreille est orientée vers la source, la distance entre la source et le microphone de l'oreille est plus courte que la distance entre la source et le microphone au centre du dispositif. Cela se traduit par un temps d'arrivée de l'énergie négatif. De plus, elle ne permet pas de compenser l'influence de la résonance du conduit auditif lié à l'emplacement du microphone dans l'oreille.

Les sources sonores utilisées ont une bande-passante limitée, ce qui implique que l'estimation de la fonction de transfert en dehors de cette bande est bruitée, et tend vers 0. Dans ce cas, l'HRTF qui correspond au rapport des deux fonctions de transfert est une indétermination qu'il est nécessaire de traiter. Ces indéterminations apparaissent en basses et hautes fréquences, et font l'objet des prochains paragraphes.

Traitement des basses fréquences

Les valeurs mesurées en basses fréquences ne sont pas fondamentales pour l'estimation des HRTF. D'une part, en dessous de 250 Hz les longueurs d'ondes sont supérieures à 1.4 m, ce qui est supérieur à toutes les dimensions des mannequins caractérisés. Il en résulte que les HRTF ne varient que peu selon les mannequins. De plus, d'un point de vue perceptif, les indices de localisation interviennent peu en basses fréquences. A ces fréquences, la localisation repose principalement sur les différences de phase entre les deux oreilles, qui est principalement gouvernée par la position des oreilles. Nous considérons que les positions des microphones sont similaires pour tous les mannequins, ce qui se traduit par une phase similaire en basse fréquences.

Pour gérer l'incertitude de mesure en basses fréquences, plusieurs approches ont été proposées dans la littérature. Dans les mesures effectuées par Brinkmann [Brinkmann et al., 2013], l'amplitude des HRTF inférieure à 103 Hz a été fixée à 0 dB, en conservant la phase mesurée. Dans les mesures effectuée par [Yu et al., 2010], une méthode de modélisation (non détaillée) a été utilisée pour les fréquences inférieures à 400 Hz. Nous estimons que des différences de niveau peuvent avoir lieu selon l'incidence, et que fixer l'amplitude à 0 dB n'est pas une solution acceptable. Une modélisation par un modèle simple s'approchant des mannequins caractérisés a été préférée.

Pour cela, les HRTF sont calculées en basses fréquences à partir d'un modèle de tête sphérique. Le modèle choisi est commun à tous les mannequins, avec un rayon de 8.75 cm et les oreilles placées sur l'équateur à +/- 90° du plan médian. Pour les fréquences supérieures, les données mesurées sont conservées et la transition entre les deux bandes de fréquences est réalisée à partir

d'un filtre de cross-over de Linkwitz-Riley d'ordre 12 (mise en cascade de deux filtres de Butterworth d'ordre 6) à la fréquence de coupure de 230 Hz. Ce filtre a été choisi car la combinaison de deux filtres complémentaires (passe-haut et passe-bas) à la même fréquence de coupure permet une reconstruction sans distorsion d'amplitude (réponse en fréquence de la combinaison plate). L'ordre du filtre a été fixé à 12, car il permet d'avoir une pente de coupure raide dans le domaine fréquentiel, tout en ayant une réponse impulsionnelle dont le support temporel est inférieur à 43 ms (2048 points). La fréquence de coupure a été fixée à 230 Hz car autour de cette fréquence il y a suffisamment d'information dans l'HRTF mesurée pour pouvoir mettre en phase les deux HRTF.

Un exemple de réalisation de ce filtrage pour l'incidence d'azimut 90° et d'élévation 0° à 40 cm est présenté en Figure 71. Pour la zone autour de la fréquence de coupure (230 Hz) l'HRTF obtenue est régulière et correspond à la combinaison des HRTF calculée et mesurée. L'amplitude en basses fréquences de l'HRTF calculée est constante et vaut 3 dB, s'écartant ainsi significativement de la solution consistant à fixer le gain à 0 dB. Entre 100 Hz et 250 Hz, un léger écart entre l'HRTF calculée et l'HRTF initiale est visible, et d'autant plus grand que la fréquence est basse. Autour de 100 Hz la différence est de l'ordre de 3 dB, et probablement liée à l'incertitude de l'estimation de fonction de transfert (absence d'information). Autour de 150 Hz, cet écart provient probablement des différences entre le modèle choisi et le mannequin mesuré. Le modèle pourrait être peaufiné de façon à correspondre davantage aux mesures dans cette bande de fréquences, néanmoins l'approximation effectuée est plus acceptable que de fixer le gain à 0 dB.

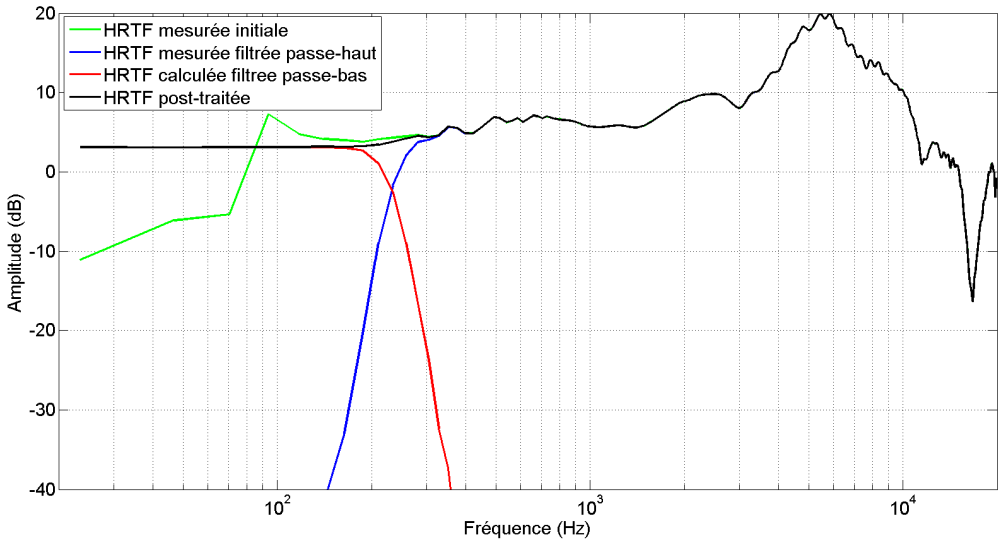


Figure 71 : exemple de combinaison HRTF calculée et HRTF mesurée pour le mannequin B&K à 40 cm d'azimut 0° et élévation nulle

Traitement des hautes fréquences

De la même manière que pour les basses fréquences, l'estimation des hautes fréquences (> 18 kHz) est bruitée. Dans cette gamme de fréquences où les longueurs d'ondes sont petites devant

l'objet caractérisé (elles sont inférieures à 2 cm), la pression diffractée est très variable selon l'objet caractérisé et la position du point de mesure. Il est alors délicat de modéliser simplement cette diffraction. Les mesures obtenues en hautes fréquences ont simplement été éliminées, en utilisant un filtre passe-bas de Linkwitz-Riley d'ordre 12 (cascade de deux filtres de Butterworth d'ordre 6) de fréquence de coupure à 19 kHz.

Egalisation des HRTF

Enfin, la dernière étape de traitement concerne l'égalisation des HRTF. Elle permet notamment de compenser l'effet des microphones et leur position dans le conduit auditif, et de compenser d'éventuels artefacts de mesures indépendants de la direction. Pour comparer les mannequins, ces influences ne sont pas souhaitables.

Il existe deux manières d'égaliser les HRTF : à partir d'une HRTF d'une incidence donnée (généralement l'incidence frontale) ou à partir d'une HRTF en champ diffus. Cette dernière méthode est plus avantageuse car elle ne privilégie pas de direction particulière [Larcher, 2001].

L'estimation d'une telle HRTF doit se faire en conditions de champ diffus, et nécessite l'emploi d'une chambre réverbérante. N'ayant pas accès à un tel équipement, l'HRTF en champ diffus a été estimée à partir d'une moyenne pondérée des mesures.

Pour prendre en compte la répartition inhomogène des points de mesure, nous les avons pondérés en fonction de l'angle solide représenté. Pour estimer cet angle solide, le diagramme de Voronoi sphérique correspondant aux points de mesure [Na et al., 2002] a été estimé. La sphère n'étant pas complète, cette manière de procéder attribue un poids important pour les mesures à l'élévation la plus basse. Nous avons modifié ce poids attribué à ces mesures, évitant de leur associer un poids excessif comparativement aux autres. Pour les mesures en champ proche le poids attribué aux mesures d'élévation -30° est alors identique au poids attribué à celles de l'élévation +30°, et pour les mesures en champ lointain le poids attribué aux mesures d'élévation 0° est égal au poids attribué à celles de l'élévation +15°. Les poids associés aux mesures en champ proches sont reportés en fonction de l'élévation dans le Tableau 15, et dans le Tableau 16 pour les mesures en champ lointain. Dans les deux cas, le poids associé à la mesure d'élévation 90° est nettement plus élevé que les autres. Il est toutefois nécessaire de rappeler qu'un seul azimut correspond à cette élévation, alors que 72 azimuts correspondent aux autres élévations.

Elévation	-30°	-15°	0°	15°	30°	60°	90°
Gain	1.30	0.45	0.45	0.45	1.30	1.95	6.67

Tableau 15 : gains appliqués aux mesures en champ proche en fonction de l'élévation

Elévation	0°	15°	30°	45°	60°	90°
Gain	0.80	0.80	0.74	0.65	1.86	11.84

Tableau 16 : gains appliqués aux mesures en champ lointain en fonction de l'élévation

Pour les quatre mannequins, l'HRTF champ diffus est estimée à partir des mesures en champ proche et en champ lointain. L'équivalent pour la sphère est calculé à partir de HRTF calculées, en utilisant les mêmes points de mesure et pondérations associées. Ces estimations sont représentées dans la Figure 72.

Les HRTF champ diffus des quatre mannequins sont similaires aux deux distances, avec des particularités liées au mannequin : une zone d'amplification apparaît entre 1 kHz et 7 kHz, suivie d'une zone d'atténuation. Dans la zone d'amplification, des différences caractérisent les mannequins : pour la Cortex, le maximum d'amplitude est localisé autour de 2.5 kHz et atteint 12 dB, pour la Head il est localisé autour de 3.5 kHz et atteint 13 dB, et pour la Kemar et la B&K il est localisé autour de 4.5 kHz atteignant respectivement 10 dB et 7 dB. Cette zone d'amplification est probablement liée à la résonance du conduit auditif, et à la position du microphone dans l'oreille.

La zone d'atténuation est variable, et d'autant plus marqué que la fréquence augmente. Cette zone est visible à partir de 7 kHz pour les mannequins Head et B&K et à partir de 10 kHz pour les mannequins Kemar et Cortex. L'atténuation est la plus marquée pour la tête Head et conduit à des écarts importants avec les autres mannequins : à 13 kHz l'écart avec le mannequin Kemar est de 23 dB. L'atténuation de la tête Head n'est pas régulière : les creux sont davantage marqués aux fréquences 8 kHz, 10 kHz et 13 kHz. Cette atténuation particulière pourrait provenir des caractéristiques des microphones, qui ne sont pas des microphones de mesure contrairement aux autres mannequins.

De légères différences sont visibles entre les estimations à 40 cm et à 2 m. Entre 300 et 500 Hz, des pics successifs de l'ordre de 1 dB sont visibles pour tous les mannequins à 40 cm qui sont absents à 2 m. Inversement, un creux d'amplitude est visible pour tous les mannequins à 2 m autour de 2.2 kHz. Ces particularités sont communes à tous les mannequins, et font penser à des particularités liées aux dispositifs de mesure. Les estimations des HRTF champ diffus sont sinon très similaires aux deux distances, ne mettant pas en évidence d'influence significative de la distance et de la répartition des points de mesure pour l'estimation.

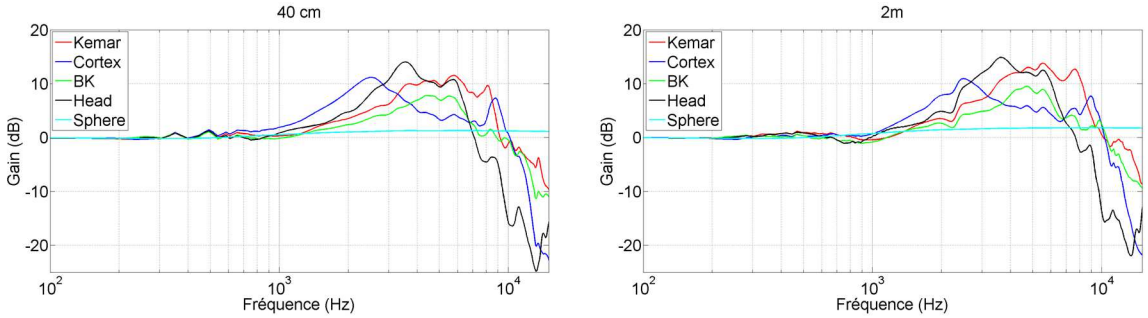


Figure 72 : HRTF champ diffus des quatre mannequins et de la sphère estimée à partir des données en champ proche (à gauche) et à partir des données en champ lointain (à droite)

Concernant la sphère, l'estimation de l'HRTF champ diffus est quasiment plate sur l'ensemble de la bande passante. Cela s'explique par le fait qu'il n'y a ni conduit auditif, ni pavillons contribuant

à l'amplification de certaines zones fréquentielles. De plus, les résultats sont issus de calcul, excluant l'influence de microphones.

Pour chaque HRTF champ diffus C, un filtre inverse associé est calculé en utilisant une régularisation de Tikhonov généralisée. Ce filtre dont la transformée de Fourier est H est obtenu de la manière suivante :

$$H = [C^{*}C + \beta]^{-1}.[C^{*}(A - CId)] + Id \qquad (104)$$

β le paramètre de régularisation fixé à $10^{\frac{-50}{20}}$, A la transformée de Fourier d'un retard pur de 1024 échantillons (21 ms pour Fe = 48 kHz), Id un vecteur unitaire, et * correspond à l'opération de transconjugaison. La composante à phase minimale du filtre est extraite à partir de la méthode du cepstre [Oppenheim & Schafer, 1975], et le filtre correspond à la transformée de Fourier inverse. Le filtre est fenêtré à une longueur de 10.6 ms (512 échantillons à la fréquence d'échantillonnage de 48 kHz). Pour appliquer le filtre aux HRIR, elles sont préalablement retardées de 4 ms par permutation circulaire des échantillons, les rendant nécessairement causales. Après filtrage, le retard introduit de 4 ms est compensé par permutation circulaire des échantillons de -4 ms. Par exemple, la FRF du filtre correspondant à l'HRTF champ diffus du mannequin B&K estimée à partir des mesures en champ lointain est représentée dans la Figure 73.

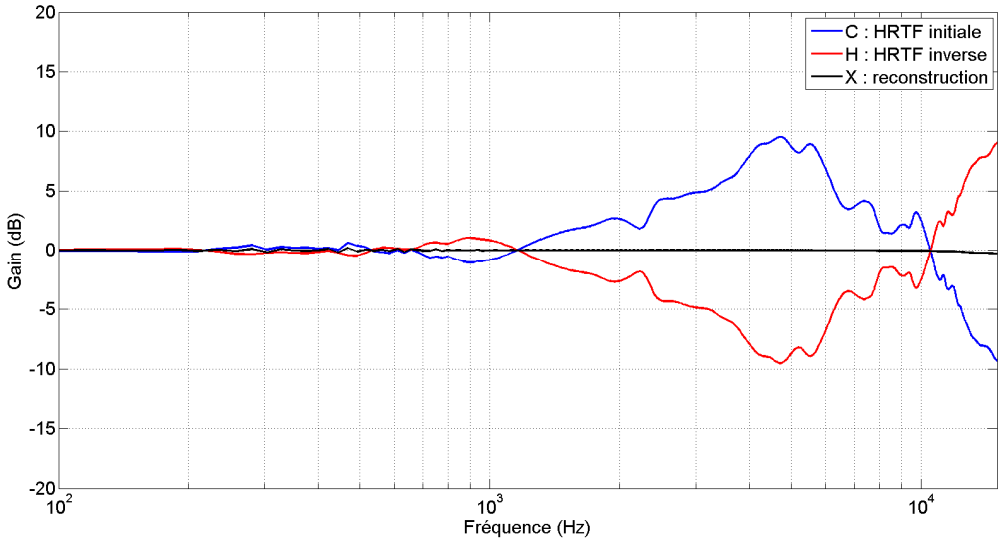


Figure 73 : HRTF champ diffus du mannequin B&K et son filtre inverse associée. En noir reconstruction réalisée à partir de la convolution de l'HRIR champ diffus et du filtre inverse

Annexe D Validation du dispositif de mesures et de post-traitement de HRTF

L'analyse préliminaire des données consiste à vérifier que le dispositif permet de mesurer ce qui est attendu. Le dispositif de mesures peut introduire des réflexions, il peut être décentré et la directivité de la source en champ proche peut influencer la mesure. Pour évaluer la présence de ces artefacts, des mesures sur une sphère sont comparées à un calcul analytique et la symétrie entre oreilles des mannequins est évaluée.

Mesures sur une sphère

Le modèle de sphère rigide permet de calculer des HRTF analytiquement. Nous avons construit une sphère en ABS avec deux microphones à l'emplacement des oreilles. La sphère utilisée a un rayon de 8.75 cm, et les microphones sont placés sur l'équateur avec un angle de 100° par rapport au plan médian. Contrairement aux mesures effectuées sur les mannequins, le centre de la tête ne correspond pas au point situé entre les deux oreilles. Le centre de la tête est le centre de la sphère, les microphones sont légèrement en arrière. Les microphones utilisés sont des GRAS 40PR (1/4'') et la mesure de $P_{champ\ libre}$ a été réalisée avec ces mêmes microphones pour compenser leur influence. La procédure de mise en forme des HRTF est la même que pour les mannequins, sauf l'égalisation champ diffus qui n'a pas été appliquée.

Les HRTF mesurées et calculées sont représentées pour les deux distances dans la Figure 74. Les mesures sur la sphère permettent d'obtenir des HRTF similaires à celles calculées, avec quelques différences. A 40 cm, les HRTF mesurées ont tendance à fluctuer autour des HRTF calculées, et pour les quatre incidences représentées les écarts sont inférieurs à 2 dB jusqu'à 8 kHz. Des différences sont communes à plusieurs incidences : pour les incidences 0°, 180°, 270° entre 350 Hz et 600 Hz et entre 900 Hz et 2 kHz l'HRTF mesurée est sous-estimée par rapport à l'HRTF calculée de l'ordre de 1 dB. Un pic autour de 2.3 kHz est visible pour les quatre incidences représentées, à l'origine d'un écart atteignant 1 dB pour l'incidence 90°. L'amplitude de ces écarts est variable selon les incidences, et ces écarts seraient liés à une particularité du dispositif de mesure. Ces oscillations s'apparentent à un filtrage en peigne : il pourrait s'agir de réflexions dont les effets dépendent de la localisation du microphone. Pour l'incidence 0° en hautes fréquences (> 11 kHz), l'amplitude de l'HRTF mesurée est plus basse que celle de l'HRTF calculée : les écarts atteignent localement 7 dB à 13 kHz et 9 dB à 14 kHz. Ces écarts peuvent se justifier par un effet de directivité du haut-parleur : en position 0°, les microphones sont décalés de l'axe du haut-parleur et les mesures sont plus sensibles à sa directivité. Il serait alors logique d'observer une différence équivalente pour la position 180°, ce qui n'est pas le cas ici : les HRTF issues de mesures et de calcul sont très proches. Une hypothèse est que le haut-parleur n'est pas parfaitement orienté vers le centre du dispositif, minimisant les effets de directivité d'un côté des mesures et les amplifiant de l'autre côté.

Les HRTF mesurées et calculées à 2 m ont également de nombreuses similitudes. Certaines différences systématiques sont aussi constatées : entre 800 Hz et 1800 Hz les HRTF mesurées sont atténuées de l'ordre de 1 dB par rapport aux HRTF calculées, cette différence étant commune avec

la mesure à 40 cm. Des oscillations sont identifiables pour les incidences 0°, 90° et 180° au-delà de 2 kHz. Les écarts observés dans cette plage de fréquence ne sont pas exactement les mêmes à 40 cm et 2 m, et cela viendrait confirmer l'hypothèse de réflexions créant un filtre en peigne dont les caractéristiques dépendent de la différence de marche entre le champ direct et le champ réfléchi. Une différence est assez marquée pour l'incidence 270°, où le minimum d'amplitude n'est pas localisé exactement à la même fréquence : il intervient autour de 8 kHz pour l'HRTF calculée et autour 7.5 kHz pour l'HRTF mesurée. Les différences induites entre 6 kHz et 8 kHz sont alors de l'ordre de 2 à 3 dB.

Il n'y a pas de différences systématiques à toutes les incidences et communes aux distances. Cela laisse présager que la sphère fabriquée n'a pas de différences majeures avec le modèle analytique, qui pourraient par exemple la faire résonner à certaines fréquences.

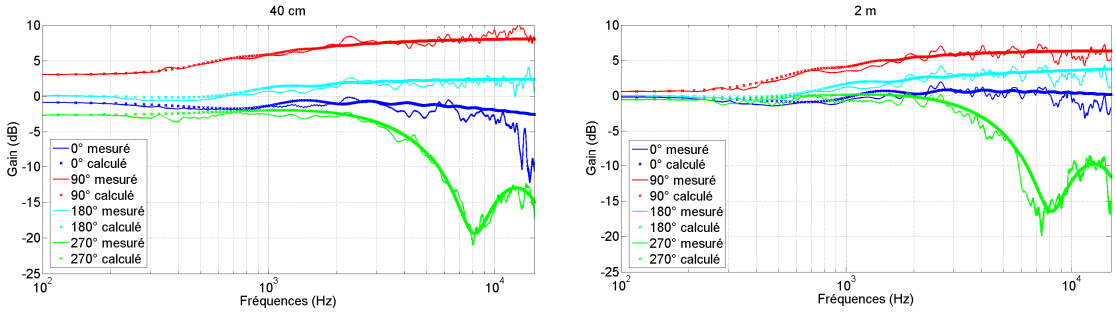


Figure 74 : HRTF mesurées et calculées sur une sphère rigide pour une sélection d'incidences à 40 cm de distance (figure de gauche) et 2 m (figure de droite).

Symétrie gauche droite

Les mannequins sont supposés symétriques, les HRTF mesurées sur les oreilles gauche et droite devraient être les mêmes. Ainsi, si des différences entre les deux oreilles sont observées, soit l'hypothèse est réfutée soit cela provient du dispositif de mesures (imprécision du positionnement et réflexions). Il peut également s'agir d'une combinaison des deux.

Pour évaluer les différences entre les deux oreilles, la métrique d'écarts spectraux moyens est employée :

$$ES(f) = \frac{1}{N}\sum_{\theta=0}^{N}\left|20\log_{10}\left(\frac{|HRTF_L(\theta,f)|}{|HRTF_R(360-\theta,f)|}\right)\right| \qquad (105)$$

$HRTF_L$ correspond à l'HRTF de l'oreille gauche à l'incidence (θ,ϕ) et $HRTF_R$ correspond à l'HRTF de l'oreille droite à l'incidence complémentaire $(360-\theta,\phi)$. Ces écarts spectraux sont calculés pour $N = 72$ azimuts du plan horizontal, et la moyenne sur tous les azimuts est représentée dans la Figure 75 pour les mesures à 2 m et 40 cm.

Une tendance commune aux deux distances est une augmentation de l'écart spectral avec la fréquence : il est toujours inférieur à 1 dB jusqu'à 3 kHz, et dépasse 4 dB à 10 kHz pour Head et Kemar à 2 m. Par ailleurs, l'ordre de grandeur est similaire à 40 cm et à 2 m mais des différences

sont visibles entre les deux distances. En particulier, un pic est visible pour la mesure à 40 cm pour tous les mannequins autour de 4 kHz. Cette différence proviendrait donc plutôt d'un artefact de mesure à 40 cm. Les écarts observés pour la sphère sont d'un ordre de grandeur similaire aux mannequins, constituant toutefois un minorant des mannequins. La géométrie de la sphère étant plus simplifiée que les mannequins, il serait attendu que les mesures obtenues soit les plus symétriques. Les écarts observés signifieraient que les artefacts de mesures seraient plus importants que les dissymétries des mannequins.

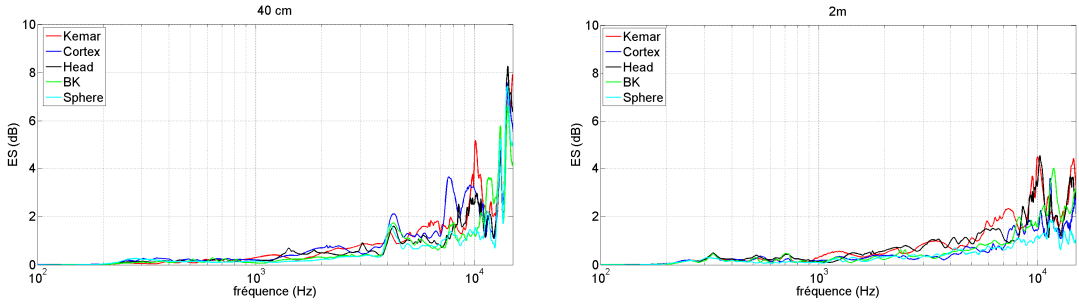


Figure 75 : Ecarts spectraux moyens entre les deux oreilles dans le plan horizontal pour les mesures à 2 m (figure de droite) et à 40 cm (figure de gauche)

Andreopoulou et al [Andreopoulou et al., 2015] ont également observés des dissymétries systématiques sur les mesures du mannequin KU-100 réalisées par 10 institutions différentes. Ces auteurs ont utilisés une métrique proposée par [Zhong et al., 2013] pour évaluer la dissymétrie entre deux oreilles. Ce calcul de dissymétrie entre les deux oreilles est effectué par bandes d'ERB selon la définition de Moore et Glasberg [B. C. . Moore & Glasberg, 1996]. Pour une bande d'ERB N, la dissymétrie est calculée de la manière suivante :

$$d(\theta,\phi,N) = 1 - \frac{|\sum_{k=1}^{K} H_L(\theta,\phi,f_k).H_R^*(360-\theta,\phi,f_k)|}{\sqrt{\sum_{k=1}^{K} H_L^2(\theta,\phi,f_k).H_R^2(360-\theta,\phi,f_k)}} \qquad (106)$$

Avec $H_L(\theta,\phi,f_k)$ et $H_R(360-\theta,\phi,f_k)$ les HRTF gauche et droite aux incidences (θ,ϕ) et $(360-\theta,\phi)$ respectivement à la fréquence f_k. Cette dissymétrie est calculée pour le mannequin Kemar à la distance de 2 m et élévation nulle, et représentée pour l'ensemble des azimuts en Figure 76. D'après la Figure 75, les mesures de ce mannequin représente un cas élevé de dissymétrie.

La dissymétrie n'est pas équivalente pour les deux demi-cercles autour de l'auditeur : pour les incidences ipsilatérales (azimut <180°) la dissymétrie est inférieure à 0.05 jusqu'à 10 kHz. Au-delà de 10 kHz, la dissymétrie est plus marquée, notamment à 10 kHz ou elle dépasse quasiment systématiquement 0.2. Pour les incidences contralatérales, trois zones différentes sont visibles. La première concerne les incidences entre 180° et 240° pour lesquelles la dissymétrie est similaire aux positions ipsilatérales. La seconde, pour les azimuts entre 240° et 300°, est une zone où la dissymétrie est toujours élevée au-delà de 10 kHz, mais également localement pour des fréquences supérieures à 2 kHz. Des pics de dissymétrie variant avec la fréquence et l'azimut sont visibles : par exemple un pic à 4 kHz pour l'azimut 240° se décale progressivement jusqu'à la fréquence 8.7 kHz pour l'azimut 285°. Enfin la 3ème zone concerne les azimuts supérieurs à 300° pour lesquels la dissymétrie est faible (<0.2 pour toutes les fréquences). La différence entre

positions ipsilatérales et contralatérales peut s'expliquer par la baisse du rapport signal à bruit pour les incidences contralatérales, liée à l'atténuation de l'onde par la tête. Pour ces incidences, le signal peut être davantage influencé par des réflexions précoces qui n'ont pas pu être éliminées par le fenêtrage. La dissymétrie systématiquement présente autour de 10 kHz pour l'ensemble des azimuts correspond au pic observé dans les figures précédentes. A 10 kHz, la longueur d'onde associée est de l'ordre de 3 cm, et correspond aux dimensions des pavillons des oreilles. Il est possible que les deux pavillons ne soient pas exactement les mêmes, introduisant des différences amplifiées pour certaines incidences.

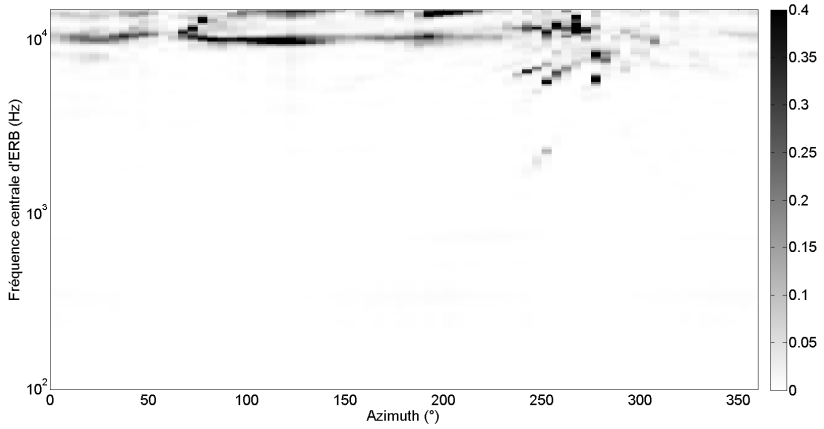


Figure 76 : dissymétrie oreille gauche / oreille droite pour le mannequin KEMAR à la distance 2 m et élévation 0°

La Figure 77 représente les HRTF gauche et droite des incidences 250° et 110° respectivement, qui représentent un cas particulier de dissymétrie d'après les remarques précédentes, avec une dissymétrie importante autour de 2.5 kHz, 6 kHz et 10 kHz. Le creux d'amplitude centré autour de 2.3 kHz n'est pas aussi prononcé pour l'oreille droite : un écart très localisé de 8 dB est visible. Entre 5 kHz et 9 kHz, les variations d'amplitude de l'oreille droite sont également moins prononcées que pour l'oreille gauche : les écarts atteignent 5 dB, couvrant des plages fréquentielles plus importante que pour l'observation faite à 2.3 kHz. Au-delà de 9 kHz, l'alternance de pics et creux d'amplitude est décalée entre les deux oreilles, conduisant à de larges écarts dépassant 10 dB.

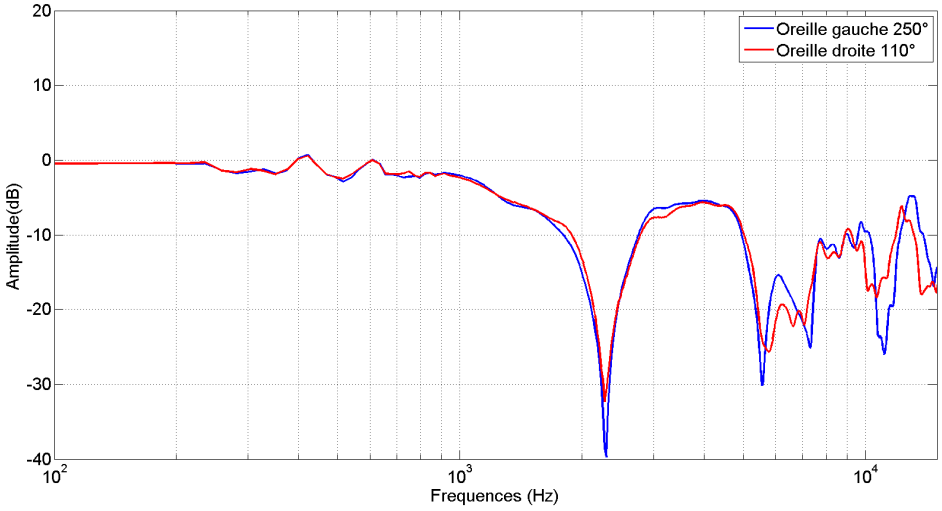


Figure 77 : Fonctions de transfert mesurées pour l'azimut 250°

Dans l'article [Andreopoulou et al., 2015], des dissymétries atteignant 0.25 ont été relevées dans la zone ipsilatérale aux fréquences inférieures à 9 kHz. Dans la zone contralatérale, des dissymétries atteignant 0.4 ont été signalées. Enfin, les différences entre incidences ipsilatérales et contralatérales ont également été relevées. Les ordres de grandeurs trouvés ici sont similaires, les valeurs de dissymétrie étant inférieures pour les incidences ipsilatérales. Le fait que les ordres de grandeurs soient similaires à d'autres travaux nous permet de valider le dispositif de mesures.

Pour minimiser l'effet de dissymétrie entre les oreilles, il serait envisageable de combiner les mesures issues des deux oreilles pour former une approximation moyenne des deux oreilles. La moyenne dans le domaine temporel n'est pas envisageable, le moindre décalage entre les signaux conduirait à un filtrage en peigne. Dans le domaine fréquentiel, la moyenne de grandeurs complexes est tout autant délicate, la moyenne des deux mesures pourrait conduire à des discontinuités de phase. Ce traitement n'a donc pas été appliqué.

Les dissymétries non négligeables constatées sur des mannequins spécifiquement construits pour être reproductibles conduit à penser que la diffraction par des auditeurs peut sans doute présenter des dissymétries marquées, éventuellement utilisées pour la localisation.

Annexe E Feuille de consignes du premier test perceptif

Bonjour,

Bienvenue dans cette expérience et merci pour votre participation. Le but de l'expérience est de comparer un panel de sons à un son de référence. Dans cette expérience, il n'y a ni bonne ni mauvaise réponse, et seul votre avis compte.

Déroulement de l'expérience

L'expérience va comporter une succession d'écrans sur lesquels 13 sons associés à 13 curseurs ainsi qu'un son de référence seront systématiquement présentés (voir interface). La première série de sons est destinée à vous familiariser avec l'expérience, et ne sera pas prise en compte dans notre analyse.

Pour chacune des séries, il faut déterminer si chacun de 13 sons à comparer est **proche ou non** du son de référence. Pour cela, vous disposez d'un curseur permettant d'évaluer cette proximité.

Pour chaque série de 13 sons, votre tâche va consister à :

- Trouver le son **le plus proche** et **le plus différent** de la référence et positionner les curseurs correspondant à ces sons respectivement au maximum et au minimum de l'échelle.
- Noter la ressemblance relative à ces deux extrêmes pour tous les autres sons

Vous pouvez attribuer la même note pour deux sons (y compris les notes maximales et minimales). Vous pouvez écouter tous les sons autant de fois que vous le désirez.

Lorsque vous pensez avoir correctement positionné chacun des curseurs correspondant à une série de sons, vous pouvez passer à la série suivante en cliquant sur le bouton « Suite ». Il y a un total de 9 séries de sons à évaluer.

Conseils pour le déroulement du test

Pour faciliter le déroulement du test, et vous permettre d'évaluer au mieux les sons, nous vous donnons les conseils suivants :

- Ecoutez d'abord tous les sons
- Lors de cette écoute, identifiez les extrema (le plus proche / le plus différent)
- Positionnez ensuite les autres sons par rapport à ces deux sons extrêmes, par écoutes successives avec la référence.
- Lorsque vous pensez avoir terminé, vous pouvez rejouer chacun des sons en allant du plus au moins proche de la référence, pour vérifier vos évaluations.

Merci et bonne écoute !

Annexe F Interface de test MUSHRA

Les tests employés dans ces travaux sont de type « MUSHRA » (Multiple Stimuli with Hidden Reference and Anchor). L'interface utilisée pour le test du chapitre V est représentée dans la Figure 78. Il est possible de jouer chacun des sons à évaluer à l'aide des boutons situés en-dessous de chaque curseur associé, et la référence à l'aide du bouton à gauche de l'interface. Les curseurs permettent alors d'attribuer des notes comprises entre 0 et 100.

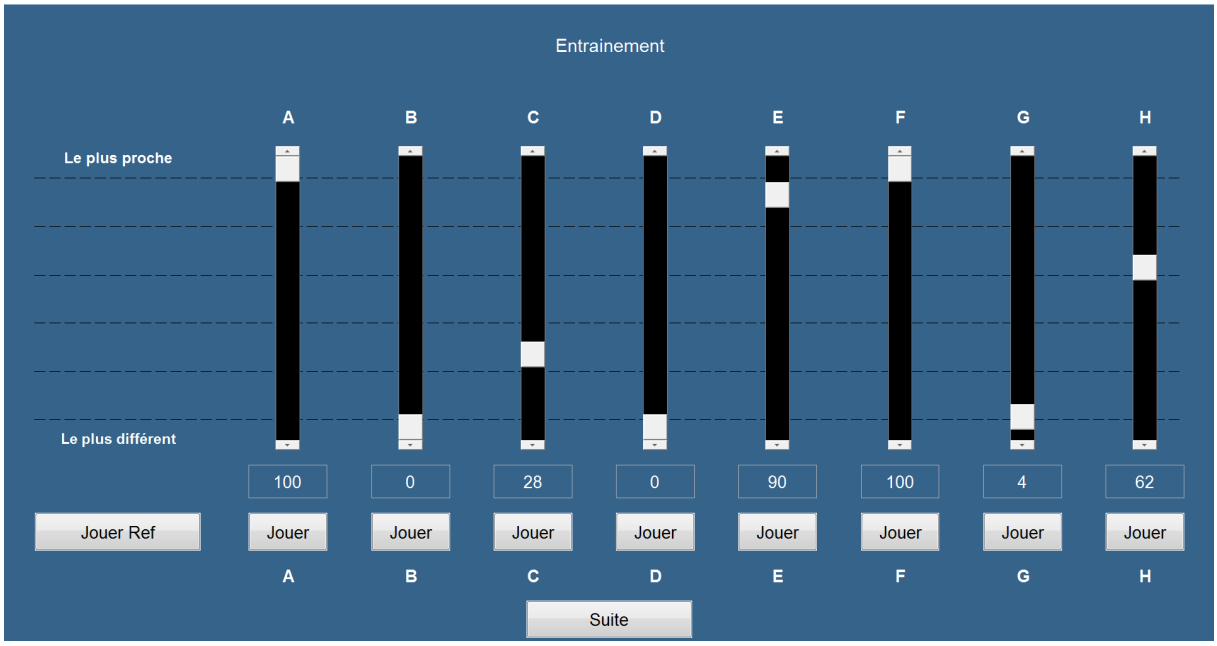


Figure 78 : exemple d'interface MUSHRA

Annexe G Caractérisation et égalisation de la Tannoy System 600

La source utilisée pour les mesures du Chapitre IV a été caractérisée dans quatre configurations : à 40 cm et 80 cm dans l'axe et avec le micro décalé de 10 cm par rapport à l'axe (de la même manière que dans les salles d'écoute). Les fonctions de réponses en fréquences associées sont représentées dans la Figure 79. Un accident significatif est visible autour de 1500 Hz pour l'ensemble des mesures : le niveau est atténué de l'ordre de 10 dB par rapport au reste de la bande fréquentielle. Cet accident intervient autour de la fréquence de crossover du haut-parleur, et avait déjà été identifié dans de précédents travaux utilisant cette enceinte [Michaud, 2012].

Les réponses mesurées à 80 cm dans l'axe et décalées sont très proches, et difficilement discernables. Les écarts entre les réponses mesurées à 40 cm dans l'axe et décalées sont plus importants, pouvant atteindre 2 dB entre 3 et 4 kHz. Cette différence s'explique simplement par le fait que la directivité d'une source dépend de l'angle d'incidence : à 40 cm, un décalage de 10 cm correspond à un désaxage plus important qu'à 80 cm et l'effet de la directivité est donc plus marqué.

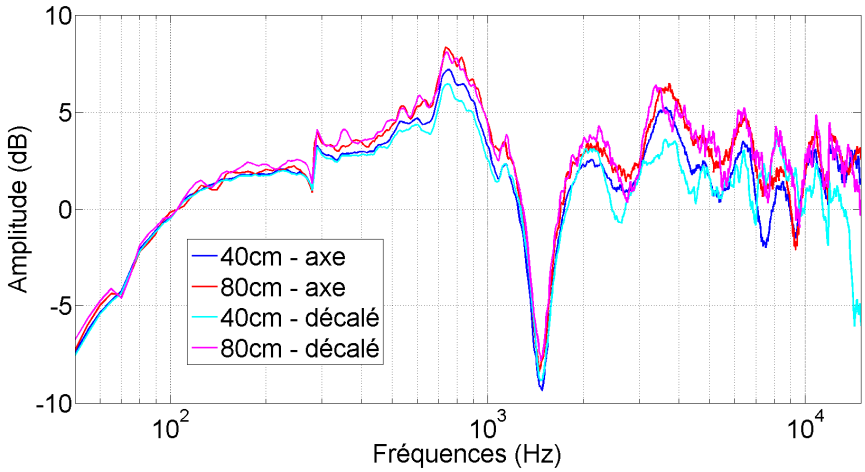


Figure 79: Réponses en Fréquences à 40 cm et 80 cm en champ libre, dans l'axe et en position décalée. Les réponses sont normalisées à 100 Hz

Filtre d'égalisation de la source

Un filtre destiné à égaliser la réponse de la source en champ libre a été calculé. Ce filtre est un FIR d'ordre 4096 (85 ms pour une fréquence d'échantillonnage de 48 kHz) calculé à partir des mesures anéchoïques dans l'axe, dont la transformée de Fourier H est obtenue en suivant la méthode présentée dans [Ole Kirkeby, Nelson, Hamada, et al., 1998] :

$$H = [C^{*}C + \beta]^{-1}.C^{*}A \qquad (107)$$

C est la réponse en fréquence de l'enceinte, β le paramètre de régularisation (dépendant de la fréquence), A la transformée de Fourier d'un retard pur de 2048 échantillons (43 ms pour une

fréquence d'échantillonnage $Fe = 48\ kHz$), et * correspond à l'opération de transconjugaison. La réponse en fréquence C de l'enceinte est normalisée à 0 dB à 200 Hz. Le filtre de compensation de la source est obtenu par transformée de Fourier inverse de H. Dans notre cas, seule la composante à phase minimale du filtre est utilisée pour compenser l'effet de la source. Son extraction est effectuée à l'aide de la méthode du cepstre [Oppenheim & Schafer, 1975].

La valeur du paramètre de régularisation β est variable avec la fréquence. Sa valeur dans la bande passante entre 80 Hz et 13 kHz est fixée à -50 dB par rapport au maximum de C, la réponse en fréquence de l'enceinte. En dehors de cette bande passante, la valeur de β augmente progressivement et atteint 10 dB par rapport au maximum de C aux fréquences extrêmes ($f = 0$ et $f = \frac{Fe}{2}$). Le gabarit de β est obtenu par interpolation en utilisant la fonction *fir2* de Matlab, et est représenté dans la Figure 80 en vert.

Deux filtres ont ainsi été calculés, l'un à partir de la réponse de l'enceinte dans l'axe à 40 cm, et l'autre à partir de la réponse de l'enceinte dans l'axe à 80 cm. Ces filtres ont été appliqués à l'ensemble des mesures correspondantes. L'effet du filtre est observé sur la réponse fréquentielle X d'un signal reconstruit, qui correspond à la transformée de Fourier de la convolution de la réponse de l'enceinte c et du filtre h calculé.

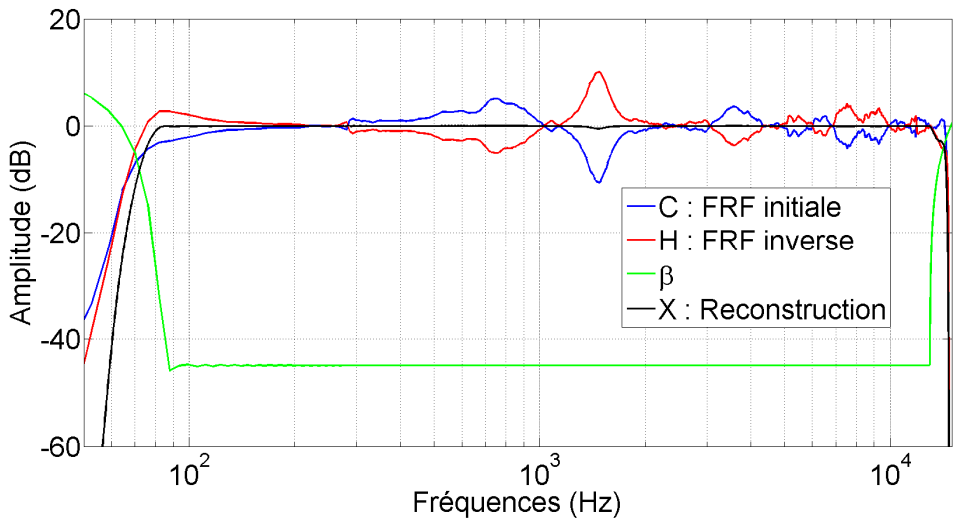


Figure 80 : FRF anéchoïque à 40 cm, FRF du filtre de déconvolution, paramètre de régularisation et reconstruction

La Figure 80 représente les caractéristiques fréquentielles du filtre d'égalisation pour la mesure à 40 cm, et de la reconstruction obtenue. La FRF inverse H est complémentaire de la FRF initiale C, permettant d'obtenir une reconstruction à +/- 1 dB dans la bande passante [80 Hz-13 kHz].

Annexe H Alternative à l'estimation du champ réverbéré

D'après les résultats du chapitre IV, la quantité de réverbération permet de décrire les résultats du test. Les indicateurs d'acoustique des salles permettent de l'estimer, et une autre approche est proposée ici.

Elle consiste à utiliser la réponse impulsionnelle anéchoïque du haut-parleur, qui correspond au champ direct. Après synchronisation adaptée, la réponse impulsionnelle du haut-parleur est soustraite aux réponses impulsionnelles en salle, afin d'obtenir idéalement uniquement le champ réverbéré de la salle d'écoute. Pour cette analyse, les RI brutes sont utilisées (pas de compensation de la source).

Synchronisation des RI

Pour réaliser les différentes mesures, les transducteurs ont été positionnés avec le plus de soin possible, mais il réside toujours une incertitude de positionnement. En outre, il n'est pas assuré que les conditions de température et d'humidité soient identiques lors des différentes mesures, et la célérité du son peut être différente. Ces incertitudes conduisent à une variabilité de la répartition temporelle de l'énergie sur les différentes RI. Or, pour supprimer la composante du champ direct dans les RI il faut que la RI anéchoïque correspondant au champ direct soit parfaitement synchronisée avec la RI mesurée en salle. Pour compenser un éventuel décalage, une procédure permettant d'estimer le retard optimal a été mise en place :

1) Un retard est appliqué à la RI anéchoïque dans le domaine de Fourier :
    a. TF de la RI
    b. Multiplication par la TF d'un retard : $e^{-2j\pi f\tau}$ avec τ le retard en *s.*
    c. TF^{-1} pour revenir dans le domaine temporel
2) Un gain est appliqué à la RI anéchoïque
3) Soustraction de la RI anéchoïque retardée et de la RI en salle
4) Calcul d'une erreur, correspondant au résidu sur un support temporel ne contenant *a priori* pas d'énergie issue de réflexions. Pour toutes les salles à l'exception de la cabine, la première réflexion possible est celle intervenant sur le sol. Cette réflexion intervient respectivement 5 et 6 ms après le direct à 80 cm et 40 cm de distance. Pour la cabine, il semblerait que la première réflexion intervienne 3 ms après l'arrivée du champ direct (voir Figure 83). L'erreur est calculée de la manière suivante :

$$e = 10\log_{10}\left(\frac{\sum_{n=n_0}^{N} RI_{residu}^2(n)}{\sum_{n=n_0}^{N} RI_{initiale}^2(n)}\right) \qquad (108)$$

Avec RI_{residu} correspondant au résultat de la soustraction, $RI_{initiale}$ correspondant à la RI mesurée dans la salle d'écoute. L'échantillon de départ n_0 correspond à l'échantillon précédent le maximum d'amplitude de 1 ms (50 échantillons) et l'échantillon d'arrêt correspond à l'échantillon suivant le maximum d'amplitude de 2 ms (100 échantillons). En ne considérant que 2 ms après le pic principal, aucune réflexion majeure n'est prise en

compte, même dans le cas de la cabine. Lorsque l'erreur vaut 0 dB, cela signifie que la procédure n'a pas permis d'éliminer de l'énergie.

5) Cette procédure est répétée pour plusieurs valeurs de retards (entre -3 et + 3 échantillons par pas de 0.01 échantillons) et pour plusieurs valeurs de gain d'amplitude (entre 0.8 et 1.2 par pas de 0.01). Le couple (retard, amplitude) ayant la plus petite erreur est retenu.

Pour une configuration particulière (Cab40), les résultats obtenus sont présentés dans les figures suivantes.

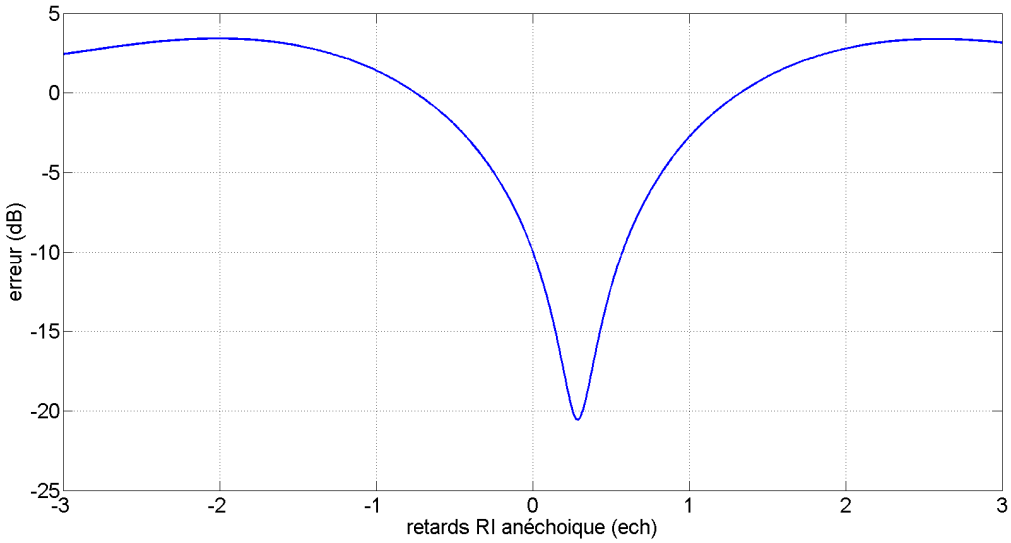


Figure 81 : erreur en fonction des différents retards pour la RI Cab40. Le gain est de 1.

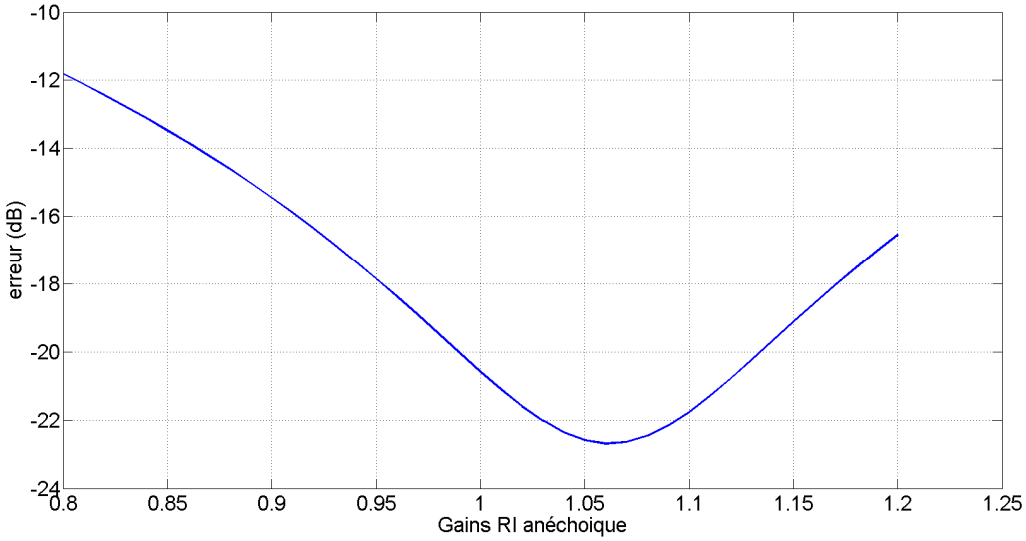


Figure 82 : erreur en fonction du gain appliqué sur la RI anéchoïque, retardée de de 7 µs pour la RI Cab40.

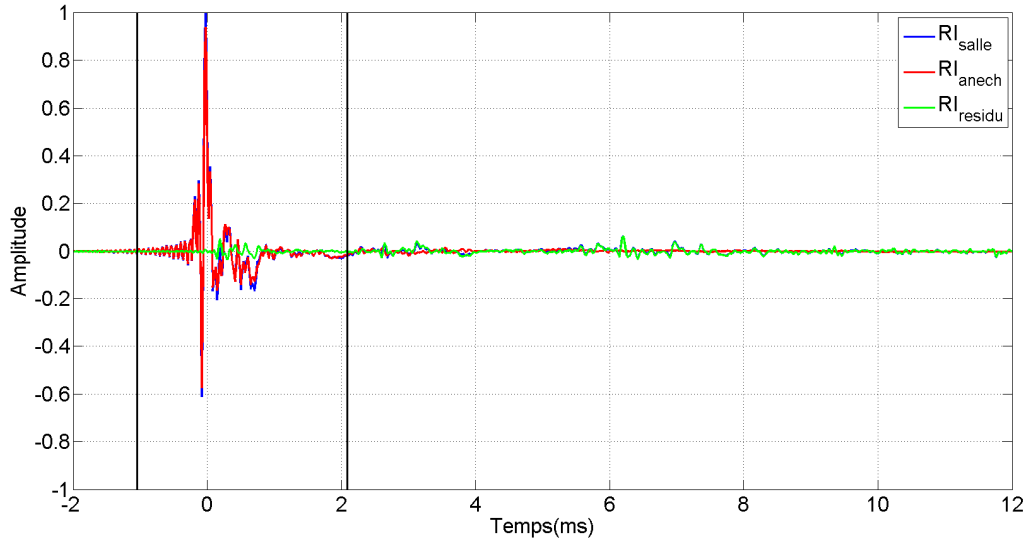


Figure 83 : Réponse impulsionnelle en salle, anéchoïque synchronisée (en temps et amplitude) et résidu pour la RI Cab40. Les traits noirs correspondent aux bornes pour le calcul d'erreur.

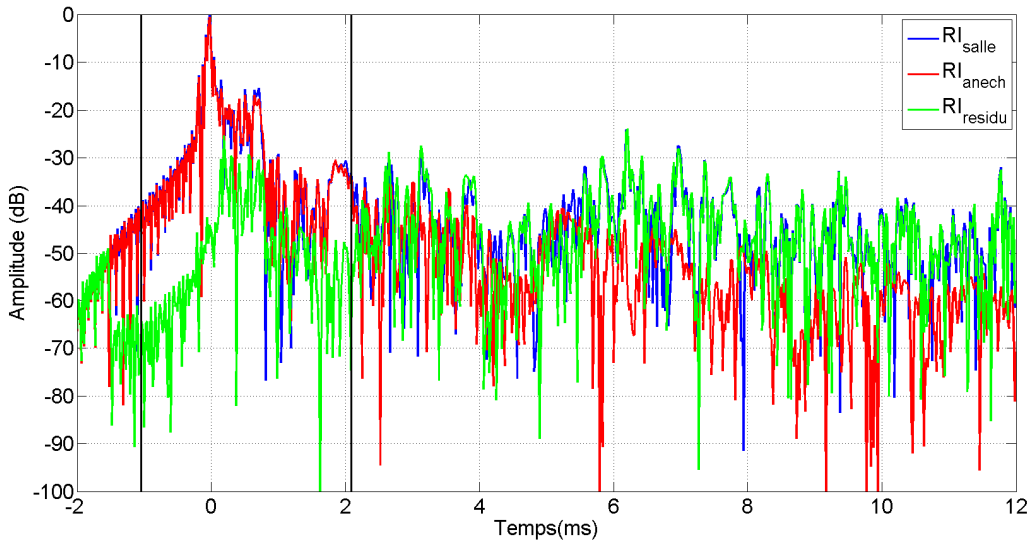


Figure 84 : ETC en salle, anéchoïque synchronisée (en temps et amplitude) et résidus pour la RI Cab40. Les traits noirs correspondent aux bornes pour le calcul d'erreur.

La soustraction n'est pas optimale, puisqu'il reste de l'énergie dans la zone correspondant au champ direct. Autour de 3.8 ms, l'amplitude du résidu est de l'ordre de -10 dB par rapport à l'amplitude du champ direct. Selon le Tableau 17, il s'agit pourtant de la configuration avec l'erreur résiduelle la plus faible.

Cette synchronisation est appliquée à l'ensemble des configurations, et les valeurs de synchronisation obtenues sont reportées dans le Tableau 17.

	BuP80	BuM80	BuP40	Reu80	Reu40	BuM40	Stu80	Stu40	Cab40
Retard (échantillon)	0.34	0.53	0.55	0.92	0.78	0.35	0.63	1.06	0.29
Retard (µs)	7.1	11.0	11.5	19.2	16.3	7.3	13.1	22.1	6.0
Gain	1.01	0.96	1.09	0.95	1.01	1.09	0.95	1.00	1.06
Erreur (dB)	-14.1	-13.8	-13.8	-14.6	-16.4	-17.4	-18.8	-19.6	-22.7

Tableau 17 : résultats de la synchronisation pour toutes les configurations

L'ensemble des retards est positif, et le plus grand retard est obtenu avec la RI Stu40 qui est de 22.1 µs. En considérant une célérité du son dans l'air de 330 m.s^{-1}, ce retard correspond à une erreur de positionnement de 7 mm, qui est donc relativement faible.

Les valeurs d'erreur diffèrent selon les configurations, allant de -13.8 dB à -22.7 dB. La synchronisation n'a pas eu la même efficacité selon les cas. Dans le cas où l'erreur résiduelle est la plus importante, le résidu obtenu est représenté dans la Figure 85. A certains instants correspondant au champ direct, l'amplitude du résidu est du même ordre que l'amplitude du champ direct.

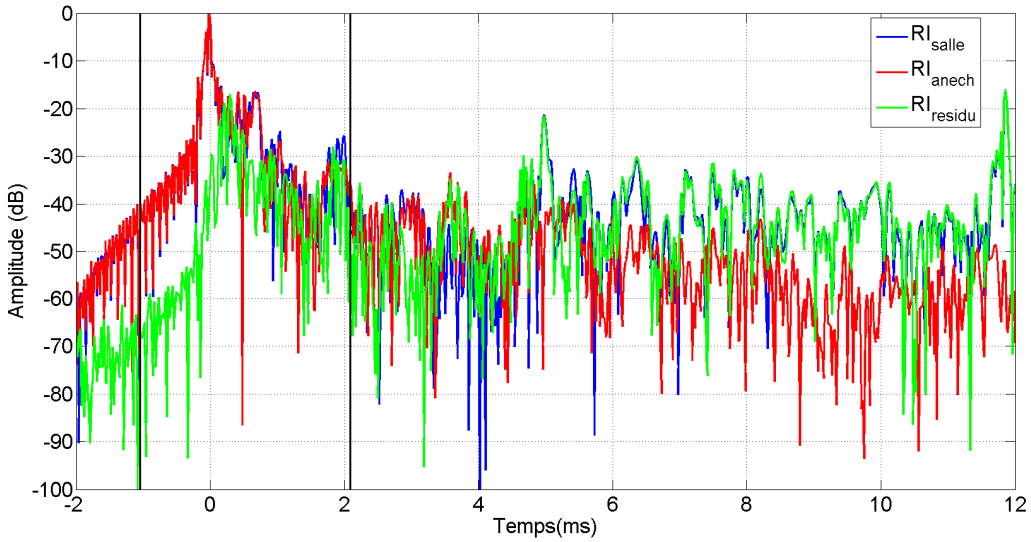


Figure 85 : ETC en salle, anéchoïque synchronisée (en temps et amplitude) et résidus pour la configuration BuM80. Les traits noirs correspondent aux bornes pour le calcul d'erreur.

Calcul d'indicateurs et corrélation avec les résultats du test

La procédure pour éliminer la contribution du champ direct dans les RI mesurées en salle est imparfaite, mais est toutefois utilisée pour calculer un rapport champ direct / champ réverbéré. Une clarté est alors calculée de la manière suivante :

$$Clarte = 10\log_{10}\left(\frac{\sum RI^2_{anechoique}}{\sum RI^2_{residu}}\right) \quad (109)$$

Les valeurs obtenues sont représentées en fonction du résultat du test perceptif dans la Figure 86. Le coefficient de corrélation vaut 0.96, le coefficient de rang de Spearman vaut 0.97 et l'erreur moyenne RMSE vaut 9.88.

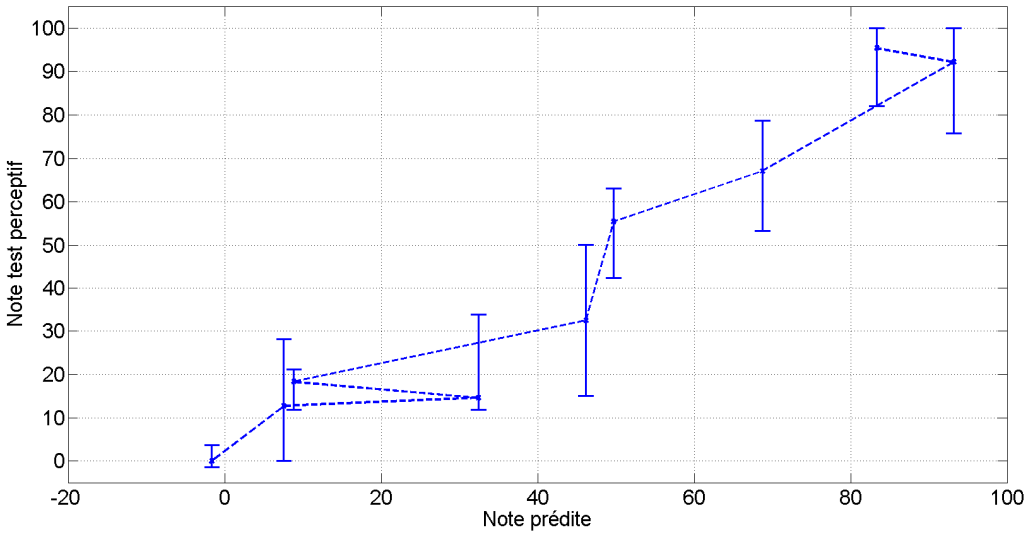


Figure 86 : notes attribuées en fonction des valeurs de l'indicateur clarté à partir du résidu. La régression linéaire du nuage de point est représentée en trait plein rouge.

Cette clarté est assez bien corrélée aux résultats du test, et les coefficients de corrélation obtenus sont du même ordre que les C20 et C30 calculées sur les RI de salle. L'utilisation de la RI anéchoïque ne permet cependant pas d'obtenir un indicateur plus corrélé que les indicateurs calculés directement sur les RI de salle.

Annexe I Calcul de sonie binaurale

Le calcul de sonie pour les sons monophoniques a été largement abordé dans la littérature, et plusieurs modèles de calcul ont été proposés. Le modèle le plus simple proposé par Stevens [Stevens, 1955] s'applique pour un son pur à 1 kHz de la manière suivante :

$$Sonie = k.P^{\alpha} \tag{110}$$

Avec P la pression en µPa, et $k = 0.01$ et $\alpha = 0.6$. Des modèles plus élaborés pour des signaux plus complexes ont ensuite vu le jour, et ont fait l'objet de normalisation des méthodes de calcul : la norme ISO 532B [ISO 532B, 1975] correspond au modèle de Zwicker et Fastl [Zwicker et al., 1991], et la norme ANSI S3.4 [ANSI S3.4, 2007] correspond au modèle de Moore et Glasberg [B. C. . Moore et al., 1997].

Ces méthodes de calculs permettent d'obtenir une sonie totale en sone, correspondant au niveau perçu d'un son. Il est également possible de définir une sonie spécifique, correspondant au niveau perçu par bandes de fréquences. La sonie totale ne correspond pas exactement à la somme des contributions par bandes de fréquences, des phénomènes de masquage d'une bande sur l'autre peuvent intervenir. Avec les conventions de la norme ISO 532B la sonie spécifique s'exprime en sone/Bark.

Les méthodes de calcul citées précédemment sont valables pour un signal monophonique. Pour un signal binaural, l'estimation de la sonie perçue est plus complexe et fait toujours débat à l'heure actuelle. Le terme « sonie binaurale » désignera dans la suite la sonie prenant en compte les effets des deux oreilles. Les prochains paragraphes s'inspirent des travaux réalisés dans la thèse de Michael Vannier [Vannier, 2016] sur le sujet.

Il est admis qu'un signal présenté aux deux oreilles est perçu plus fort que lorsqu'il est présenté à une seule oreille et on parle alors de sommation binaurale. Le rapport de sonie binaurale à monaurale entre la sonie d'un son diotique $Sonie_{diotique}$ (identique sur les deux oreilles) et la sonie de ce même son présenté de manière monaurale $Sonie_{monorale}$ (qu'une seule oreille) est défini par $R = \frac{Sonie_{diotique}}{Sonie_{monorale}}$. Le gain de sommation binaurale correspond au gain à appliquer à un signal présenté de manière monaurale pour qu'il produise la même sonie que s'il était diffusé de manière diotique :

$$g = \frac{20.\log_{10}(R)}{\alpha} \tag{111}$$

Plusieurs manières d'estimer cette sommation binaurale ont été proposées dans la littérature. La première approche proposée consiste à simplement sommer les sonies calculées indépendamment sur les deux oreilles, ce qui correspond à un rapport de sonie binaurale à monaural $R = 2$. Cette approche est celle qui est recommandée par la norme ANSI S3.4 [6] et désignée par sommation « parfaite ».

Un modèle plus élaboré a été proposé par Moore et Glasberg [B. C. . Moore & Glasberg, 2007], considérant que la sommation en sonie est « moins que parfaite ». Ce modèle consiste à calculer des termes d'inhibitions, modélisant un masquage d'une oreille sur l'autre : un fort niveau sur une oreille peut masquer le niveau de l'autre oreille. Ce modèle ne sera pas exploité dans le cadre de nos travaux.

A la même période, Sivonen et Ellermeier ont proposé un autre modèle [Sivonen & Ellermeier, 2008], considérant également une sonie « moins que parfaite ». Pour ce modèle, la sonie est calculée de la manière suivante :

1) Les spectres d'amplitude $P_L(B_N)$ et $P_R(B_N)$ des contributions gauche et droite sont calculés en tiers d'octave où B_N désigne la bande de tiers d'octave concernée. Les niveaux en dBSPL associés sont désignés par $L_L(B_N)$ et $L_R(B_N)$.
2) La sommation des contributions est réalisée selon la relation suivante :

$$L_{tot}(B_N) = g.\log_2\left(2^{\frac{L_L(B_N)}{g}} + 2^{\frac{L_R(B_N)}{g}}\right) \tag{112}$$

   Où g désigne le gain de sommation binaurale, une valeur de 3 dB a été retenue par les auteurs. Dans le cas d'un son diotique, on a $L_R(B_N) = L_L(B_N)$ et le niveau total est alors $L_{tot}(B_N) = L_L(B_N) + g$. Pour un son pur à 1 kHz et un exposant de fonction de sonie $\alpha = 0.6$, le rapport de sonie binaurale à monaurale est alors de 1.2. Plus le paramètre g est grand et plus il permet de s'approcher d'un modèle de sommation binaurale parfaite.
3) La sonie est alors estimée avec un modèle de sonie standard à partir de L_{tot}.

Les résultats obtenus avec ces modèles sont toujours sujets à discussion. Dans la thèse [Vannier, 2016] les différents modèles ont été comparés dans plusieurs situations et de manière générale le modèle de Sivonen et Ellermeier est mieux corrélé que les autres aux résultats des expériences réalisées. Une nuance est cependant appliquée, le paramètre g optimal pouvant varier selon les cas : un paramètre de 3.5 dB, proche de celui proposé par Sivonen et Ellermeier a été proposé pour des bruits à bande étroite. Dans le cas de sources bitonales, un paramètre de 5 dB s'est avéré davantage corrélé aux données des expériences. Ces expériences semblent par ailleurs confirmer que la sommation binaurale est moins que parfaite. Enfin, pour des sons complexes tous les modèles testés ont tendance à sous-estimer la sonie. L'auteur propose en piste d'amélioration de ces modèles de prendre en compte la corrélation entre les signaux gauche et droite. Ce dernier point confirme le fait que les modèles de sonie binaurale sont encore perfectibles.