THESE DE DOCTORAT DE L'UNIVERSITE DE PARIS VI

Spécialité :

Acoustique

Sujet de la thèse :

Techniques de spatialisation des sons pour la réalité virtuelle

Présentée par

Véronique Larcher

soutenue le 11 mai 2001

devant le jury composé de :

M. Jens Blauert	Rapporteur
M. Antoine Chaigne	Rapporteur
M. Marc Emerit	Examinateur
M. Jean-Marc Jot	Examinateur
M. Stéphane Natkin	Examinateur
M. Jean-Dominique Polack	Directeur de thèse
M. Jacques Prado	Examinateur

Résumé

Les techniques de spatialisation des sons permettent de suggérer une source sonore provenant de toute incidence, et constituent à ce titre un facteur important d'immersion pour les systèmes de réalité virtuelle. Cette thèse de concentre sur l'une d'entre elles, la synthèse binaurale, pour une reproduction sur casque d'écoute. Cette technique a pour objectif de restituer aux tympans de l'auditeur le champ sonore qui aurait été engendré dans une situation d'écoute réelle, et est à ce titre une technique de simulation idéale. Toutefois, elle rencontre deux principaux freins pour un déploiement à grande échelle de qualité, thèmes qui ont constitué le moteur du présent travail. Il s'agit tout d'abord de son coût de calcul, exhorbitant au regard des autres techniques de spatialisation. Nous en avons envisagé l'optimisation pour la mise en espace simultanée de nombreuses sources sonores. En outre, les mécanismes perceptifs en jeu dans la synthèse binaurale pour "duper" l'oreille dépendent fortement de l'auditeur, et nous avons mis en place plusieurs stratégies pour permettre l'adaptation individuelle de ces traitements.

Cette thèse aborde la synthèse binaurale sur des aspects théoriques et pratiques (relevés expérimentaux, implantation), et comprend la validation perceptive de plusieurs solutions proposées. Les résultats montrent que des formats binauraux multicanaux peuvent être définis sans perte de qualité de la reproduction par comparaison avec la structure bicanale traditionnelle, et permettent en outre d'optimiser le coût d'implantation dès que plus de quatre sources sont spatialisées.
On constate que les différences inter-individuelles peuvent être réduites à différents niveaux de l'implantation : par l'égalisation appropriée des filtres de la synthèse, par l'utilisation d'un format multicanal à encodeur universel, ou par un traitement spécifique. Le lien avec les paramètres morphologiques de l'auditeur ouvre la possibilité d'une adaptation individuelle automatique, donc très souple, s'appuyant par exemple sur la saisie vidéo des caractéristiques de la tête, du torse et des pavillons.

Notations et conventions de langage

BIR	réponses impulsionnelles binaurales,
HRIR	réponses impulsionnelles binaurales mesurées en chambre anéchoïque,
HRTF	transformée de Fourier d'une *HRIR*,
HRTF diffuse	HRTF en champ diffus,
base de données de HRTF	HRTF mesurées sur un même sujet pour un ensemble de positions,
tête	peut désigner une base de données de HRTF,
CASQUE	fonction de tranfert binaurale d'un casque d'écoute,
REF	filtre d'égalisation,
filtre FIR	filtre à réponse impulsionnelle finie,
filtre IIR	filtre à réponse impulsionnelle infinie,
H(z)	transformée en z,
$\|H\|$, *mag*	spectre d'amplitude de H,
ϕ	phase de H,
mph	phase minimale de H,
$eph = \phi$ - *mph*	excès de phase,
azimut	angle d'incidence mesuré dans le plan horizontal référencé par rapport à la direction frontale,
élévation	angle repérant l'altitude d'une source : en référence au plan horizontal (équivalent à l'angle de site), ou sur un cône de confusion (également noté en anglais "sagittal angle"),
BF	basses fréquences,
HF	hautes fréquences,
MF	fréquences intermédiaires entre BF et HF.

Sauf mention contraire, les réponses en amplitude sont affichées en dB, les phases en radians.

Table des matières

Introduction

Les systèmes de réalité virtuelle plongent les participants dans un espace suggéré artificiellement par des informations multisensorielles, notamment auditives, visuelles et tactiles. L'espace sonore est restitué selon deux attributs :

- la position des différentes sources sonores, le plus souvent distribuées dans les trois dimensions de l'espace,
- l'effet de salle, résultat des réflexions sur les murs ou les obstacles et de la réverbération tardive du lieu, et nécessaire pour la reproduction des effets de distance.

Cette thèse se concentre sur le premier point, et a pour objectif d'améliorer les techniques de reproduction sonore 3D sur écouteurs. Le positionnement des sources sonores dans l'espace contribue de façon essentielle au réalisme d'une visite virtuelle, d'une simulation, ou à l'ergonomie d'un service de collaboration à distance. Pour la visio-conférence par exemple, il renforce la discrimination des locuteurs distants et donc l'intelligibilité de leur message. Il favorise également le sentiment d'immersion pour les applications ne simulant qu'un champ visuel limité, et rend à l'audition sa spécificité de "sens d'alerte", mise à profit pour des applications telles que les jeux vidéos, les simulateurs de vol, de conduite.

Le système de reproduction sonore visé répond à un cahier des charges précis. Le cadre particulier de la réalité virtuelle justifie le choix de dispositifs limitant le couplage avec l'environnement extérieur : l'utilisation d'un casque d'écoute garantit la fidélité de la reproduction sonore en évitant le filtrage lié à la salle et supprime tout risque de bouclage électro-acoustique. A ce système de reproduction doit être associée une technique de codage des informations directionnelles, donnant une empreinte spatiale à un son. Ce codage peut être réalisé par enregistrement d'une scène sonore réelle, mais toute manipulation de la position relative des sources est alors compromise. Les applications visées réclament au contraire une forte interactivité et privilégient un codage directionnel par synthèse : la scène sonore virtuelle est construite par un algorithme de traitement du signal à partir de signaux élémentaires monophoniques. En outre, la simulation est mise à jour en temps réel dès lors que des éléments de la scène sonore sont modifiés par les actions ou les déplacements de l'utilisateur (déplacement de sources sonores ou systèmes de suivi de position par exemple).

Les premiers processeurs de spatialisation sonore orientés vers la réalité virtuelle ont été développés à la fin des années 1980, tirant profit d'études antérieures sur la localisation auditive ([Bla97], [Beg94]). Pour atteindre le plus grand réalisme, le codage directionel imite les mécanismes en jeu dans une situation d'écoute réelle : lors de sa propagation jusqu'au tympan, le son incident est diffracté par le corps de l'auditeur, notamment sa tête et son torse, et ainsi transformé, il livre au système auditif des indices caractéristiques de sa position. Pour chaque incidence, ces transformations peuvent être consignées sous forme de filtres audionumériques, qualifiés de Head-Related Transfer Functions, et la spatialisation, connue sous le nom de synthèse binaurale, est obtenue en filtrant un canal monophonique par la paire droite/gauche de HRTF appropriée. Ces techniques suscitent un intérêt grandissant mais présentent cependant certaines difficultés pour une utilisation "grand public". Ces difficultés résident essentiellement dans la lourdeur des traitements audio-numériques et la variabilité des HRTF en fonction de l'individu. L'objectif de cette thèse est précisemment de proposer des solutions pour repousser ces limites, par l'optimisation de l'implantation de la synthèse binaurale (minimisation du coût de calcul et de l'encombrement en mémoire) et par l'étude de techniques d'adaptation individuelle des HRTF. Elle se situe dans le prolongement d'études réalisées antérieurement à l'Ircam dans le cadre du projet Spatialisateur ([Lar94a], [JLW95], [Lar95], [Lar96], [Mar96a]). Le processeur d'acoustique virtuel *Spat~* a été développé afin de contrôler en temps réel les informations de localisation auditive et de réverbération. Notre travail bénéficie ainsi d'une

plate-forme d'écoute autorisant la confrontation systématique entre phases théoriques et phases d'expérimentation. Elle bénéficie en retour l'implantation des améliorations proposées. Les résultats obtenus pour la synthèse binaurale s'appliquent également à la reproduction dite "transaurale" sur haut-parleurs ([CB89]).

Dans une première partie, nous rassemblons les données sur lesquelles s'appuie l'ensemble de l'étude. Une première phase expérimentale est consacrée à la mesure de HRTF sur plusieurs sujets. Ces données contiennent certes les indices de localisation que l'on souhaite reproduire, mais elles renferment également des contributions parasites qui doivent être éliminées. Nous examinons plusieurs traitements permettant de les réduire à leur "substantifique moelle". Le premier d'entre eux consiste à simplifier l'information de phase en estimant le retard interaural, indice de localisation majeur. Nous proposons une méthode d'estimation à partir de la différence interaurale d'excès de phase, prolongeant les principes énoncés par Jot ([JLW95]). Dans une seconde étape, les contributions superflues sont éliminées des spectres d'amplitude mesurés. La méthode retenue est l'égalisation par rapport au champ diffus des HRTF et du casque d'écoute. Nous proposons une méthode de mesure du filtre d'égalisation rendant la procédure plus simple et plus rapide ([LJV98]). La représentation "minimale" de chaque HRTF est alors constituée d'un retard et d'un spectre d'amplitude.

La seconde partie est consacrée à l'implantation directe (ou bicanale) de la synthèse binaurale sous forme d'un retard fractionnaire en série avec un filtre à phase minimale, tel qu'elle est réalisée dans le *Spat~*. L'optimisation de cette implantation est faite sur trois plans. Nous cherchons tout d'abord à réduire le coût de calcul induit par le filtrage en temps réel. Nous nous penchons donc sur la conception des filtres audio-numériques à partir des HRTF mesurées. L'ordre de ces modèles ainsi que la structure retenue pour l'implantation constituent les degrés de liberté pour l'adaptation du coût de calcul. Les techniques d'interpolation de filtres sont également étudiées. Appliquées aux HRTF, elles permettent de synthétiser le filtre correspondant à une direction quelconque à partir d'une base de données de mesures réduite, et permettent ainsi de limiter l'encombrement mémoire ([LJ97]). Nous proposons une méthode pour l'interpolation des cellules d'ordre deux d'une structure transverse factorisée, et constatons les bonnes performances de la fréquence des raies spectrales et des Log Area Ratio pour l'interpolation de structures développées, respectivement transverse ou en treillis. Enfin, nous rappelons les techniques de commutation de filtres permettant de varier continuement la position apparente des sources, ou de compenser en temps réel les mouvements de la tête de l'auditeur, en minimisant les artefacts parasites (bruit impulsionnel).

En dépit des améliorations apportées, la structure bicanale demeure trop coûteuse en calcul pour autoriser la spatialisation de nombreuses sources sonores. Dans la troisième partie, nous cherchons une stratégie d'implantation plus adaptée. Elle repose sur un format d'encodage intermédiaire des indices directionnels, multicanal, couplé à un décodeur binaural pour la reproduction de ce format sur écouteurs. Ce décodeur est partagé par toutes les sources sonores, alors que l'encodeur, spécifique à chacune d'entre elle, est de complexité très réduite. L'ajout d'une nouvelle source n'engage donc aucun coût de filtrage supplémentaire. En outre, cette implantation offre un arbitrage supplémentaire, le choix du nombre de canaux, pour adapter la charge de calcul à la puissance disponible. Nous proposons un format d'encodage binaural s'appuyant sur l'analyse statistique des HRTF aux ordres supérieurs. Il est indépendant de l'auditeur et est à ce titre qualifié d"universel". De plus, son coût est optimisé pour que l'encodage de la position soit réalisée à l'aide d'un sous-ensemble réduit de canaux ([LJGW00]). Nous le comparons à d'autres formats d'encodage proposés dans la littérature, et notamment au format Binaural B proposé par Jot ([JWL98], [JLP99]).

La comparaison objective des différentes implantations de la synthèse binaurale est prolongée dans le chapitre 4 par une validation psycho-expérimentale. Un premier test de localisation est mené afin d'évaluer les performances d'une implantation bicanale pour une écoute non-individuelle, i.e. sans que les HRTF utilisées soient celles de l'auditeur, telle que le propose *Spat~* . Il met en évidence les défauts de localisation typiques, et souligne l'importance du choix des HRTF. En effet, les sujets ont jugé 17 bases de données de HRTF, et certaines d'entre elles augmentent de façon significative les performances de localisation. Une synthèse binaurale sans adaptation individuelle peut donc tirer profit du choix approprié de la base de donnée unique qu'elle utilise. Un second test perceptif soumet à comparaison plusieurs formats d'encodage multicanal de la synthèse binaurale. Pour la plupart d'entre eux, la qualité de localisation est comparable à celle de l'implantation directe, ce qui permet de conclure en leur faveur dès que quatre sources sonores au moins sont spatialisées.

Pour tirer le meilleur parti des ressources investies dans l'implantation et améliorer la qualité de localisation, il faut alors envisager l'adaptation individuelle de la synthèse binaurale, thème auquel nous consacrons le chapitre 5. Nous proposons tout d'abord plusieurs mesures des différences interindividuelles que nous souhaitons réduire. Cela nous conduit notamment à définir un protocole de relevé de caractéristiques morphologiques, éléments à l'origine de ces différences. Puis, nous étudions et mettons en oeuvre plusieurs stratégies d'adaptation, répondant à différents degrés de qualité. La première d'entre elles, que nous qualifions d"'adaptation discrète" consiste à offrir à l'utilisateur la possibilité d'ajuster le traitement à son écoute personnelle en choisissant un jeu de HRTF parmi plusieurs. Avec la seconde, l"'adaptation continue", cet ajustement est réalisé de manière continue par homothétie sur les valeurs de retard interaural et sur l'échelle des fréquences. Dans les deux cas, nous proposons des solutions pour automatiser la procédure en reliant les paramètres d'adaptation aux dimensions morphologiques de l'auditeur.

Chapitre 1

Spécification des filtres directionnels pour la synthèse binaurale

1.1 Introduction

La synthèse binaurale a pour objectif de reproduire et de contrôler la sensation de localisation du son. Il est pour cela important d'identifier les mécanismes en jeu dans l'écoute naturelle afin de les simuler efficacement. Dans ce chapitre, nous rappelons la nature des indices acoustiques de localisation que nous souhaitons synthétiser, et exposons le protocole de mesure que nous avons utilisé pour les caractériser. Cette caractérisation est réalisée sous forme de filtres, spécifiques à chaque incidence, qui consignent l'empreinte des transformations subies par le son lorsqu'il rencontre les obstacles formés par le torse, la tête, et les oreilles de l'auditeur. Nous les désignerons par Head-Related Transfer Functions (HRTF).

Afin de restreindre l'information qu'ils contiennent aux indices de localisation, nous appliquons deux traitements aux mesures. Le premier d'entre eux consiste à extraire le retard interaural des HRTF. Le second permet d'isoler dans les spectres d'amplitude les caractéristiques dépendant de la direction, par une égalisation par rapport au champ diffus. Certains éléments ont été publiés dans [LJ97] et [LJV98].

Chaque position est alors entièrement caractérisée par un retard et un spectre à phase minimal. Les données finales constituent les HRTF de référence sur lesquelles s'appuieront les chapitres suivants.

1.2 Indices acoustiques de localisation

1.2.1 Les indices interauraux

La théorie duplex de la localisation, développée par Lord Rayleigh dans [Ray07], met en évidence l'importance des différences interaurales de temps (ITD) et d'intensité (ILD) pour la latéralisation des sources sonore, i.e. pour la distance vers la droite ou vers la gauche avec laquelle le son s'écarte du centre de la tête (cf Figure 1.6). ITD et ILD sont "ambigus" de part l'existence d'un lieu d'indicences produisant des indices interauraux constants. Ce lieu a été mis en évidence pour l'ITD par Woodworth, à l'aide d'un modèle de la tête par une sphère, les deux points des oreilles étant choisis aux extrémités d'un diamètre. Il introduisit ainsi la notion de "cône de confusion".

1.2.1.1 Différence interaurale de temps (ITD)

L'ITD représente les décalages temporels observés entre les signaux parvenant aux oreilles. Il recouvre deux mécanismes du système auditif, actifs sur deux intervalles fréquentiels distincts :

- En basses fréquences ($f < 1500Hz$), l'ITD représente le retard de phase entre les signaux droite et gauche ([Kuh77]) :

$$ITD = \frac{1}{2\pi}\frac{\varphi_L(f) - \varphi_R(f)}{f} \tag{1.1}$$

  Les données expérimentales de Kuhn montrent que l'ITD basses-fréquences est indépendant de la fréquence, et peut être approché par :

$$ITD \simeq 3.\frac{a}{c}.\sin(\theta)$$

  où a désigne le rayon de la tête sphérique équivalente ($\simeq 9.3cm$ dans [Kuh77]), c la célérité du son dans l'air ($340m.s^{-1}$), θ l'azimut.
- Au dessus de $1500Hz$, l'ITD représente le retard d'enveloppe entre les signaux droite et gauche. Middlebrooks mentionne deux interprétations du traitement réalisé par le système auditif ([MG90]). D'après la première, l'information serait obtenue par le calcul du retard de groupe :

$$ITD = \frac{1}{2\pi}\frac{d(\varphi_L(f) - \varphi_R(f))}{df} \tag{1.2}$$

  Toutefois, cette interprétation accorde le même poids à chaque intervalle df, et traduit donc mal l'analyse fréquentielle "logarithmique" réalisée par le système auditif.
  Une alternative, adoptée par Middlebrooks, consiste à expliquer l'ITD perçu par l'intercorrélation des enveloppes droite et gauche, après filtrage passe bande des signaux, par exemple sur l'échelle des barks :

$$ITD = \max_{\tau}(corr(\Pi_{f_0} * env_L, \Pi_{f_0} * env_R))$$

  où Π_{f_0} désigne la réponse impulsionnelle d'un filtre passe-bande autour de la fréquence f_0.
  Les données expérimentales de Kuhn montrent que l'ITD hautes-fréquences, pour les fréquences supérieures à $3kHz$, peut être estimé par une valeur indépendante de la fréquence donnée par :

$$ITD \simeq 2.\frac{a}{c}.\sin(\theta)$$

D'après les équations (1.1) et (1.2), l'ITD basses fréquences apparait ainsi supérieur à l'ITD hautes fréquences. On a représenté en Figure 1.1 l'ITD basses fréquences mesuré sur une tête humaine pour plusieurs cônes de confusion. Pour cette tête, l'ITD atteint une valeur maximum d'environ $700\mu s$, légèrement inférieure à la limite supérieure mesurée par Blauert ($800\mu s$, [Bla97]). Si le modèle sphérique de la tête introduit par Woodworth était rigoureusement vérifié, l'ITD serait constant sur chaque courbe. Effectivement, il l'est approximativement jusqu'au cône de 0.35ms (à peu près à mi-chemin entre le centre de la tête et l'oreille). Au delà, le modèle sphérique a tendance à sous-estimer l'ITD pour les positions du sommet du cône (élévation autour de 100°). Une explication pourrait venir du fait que la tête est plus "ovale" que "ronde", notamment dans la dimension de hauteur de la tête. Cela justifierait l'occurence de cette sous-estimation au sommet des cônes. En outre, plus l'incidence est latérale, et plus l'écart entre tête sphérique et tête ovale est grand pour le chemin parcouru en "rasant" la tête (modèle de la propagation jusqu'à l'oreille contralatérale), ce qui expliquerait pourquoi le phénomène apparait aux positions les plus latérales. On peut ainsi attendre une bonne amélioration de la prédiction de l'ITD grâce au modèle de la tête en forme d'ovale, tel que le proposent Duda et al. ([AAD99]).

1.2.1.2 Différence interaurale de niveau (ILD)

L'ILD peut être exprimé pour chaque fréquence comme la différence des spectres d'énergie droite et gauche (en dB). Toutefois, on préfère souvent utiliser une version plus compacte, telle que celle proposée par Jot dans [JLW95] :

$$ILD = 10.\log_{10}\frac{\int mag_L^2}{\int mag_R^2} \tag{1.3}$$

L'intervalle d'intégration proposé est $[1kHz - 5kHz]$, domaine pour lequel l'ILD est un indice prépondérant (voir plus loin). Pour que l'intégration ait un sens perceptif, on ré-échantillonne les spectres d'énergie

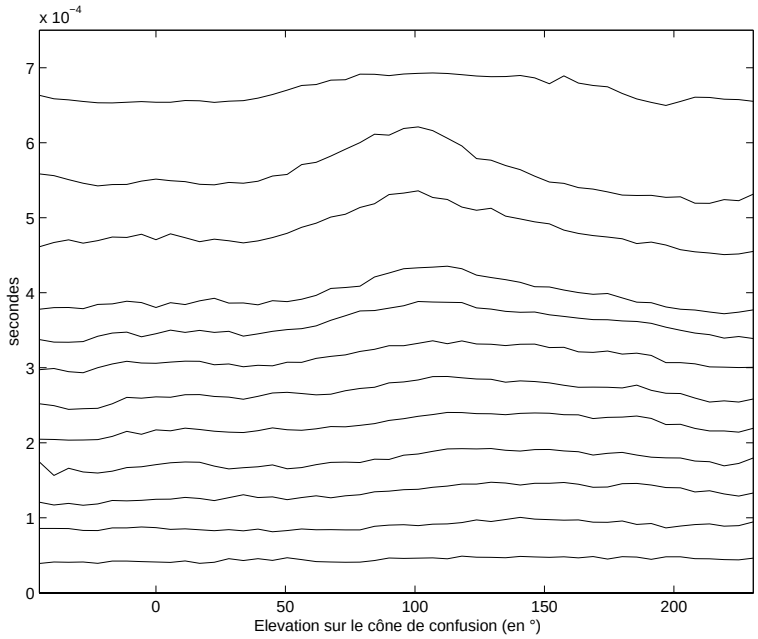


FIG. 1.1 – ITD sur des "cônes de confusion", pour des degrés de latéralisation régulièrement espacés, entre 0 et 0.98. Estimation réalisée par approximation linéaire de l'excès de phase interaural (voir section 1.4) à partir de mesures sur têtes humaines réalisées à UC Davis. En abscisse : "sagitallisation" de la Figure 1.6.

sur une échelle fréquentielle logarithmique avant de sommer les échantillons. On peut penser que l'analyse réalisée par le système auditif est néanmoins plus proche du modèle proposé par Gaik, pour lequel l'ILD est évalué pour chaque bande critique, l'intégration portant sur la largeur de la bande considérée ([Gai93]).
Nous avons estimé l'ILD à l'aide de la formulation (1.3) pour 17 têtes humaines, mesurées à UC Davis par Algazi et al. (voir section 1.3). On observe sur la Figure 1.2 que, comme pour l'ITD, la constance sur le cône de confusion n'est vérifiée que sur certaines plages d'élévation. Ainsi, pour les positions situées à l'arrière de la tête au dessus du plan horizontal (100°<élévation<180°), l'ILD mesuré est nettement inférieur à celui que l'on observe pour les positions frontales. Ainsi, comme l'observe Han ([Han94]), l'ILD constituerait un indice prépondérant pour la discrimination avant-arrière, .
En revanche, l'ILD des positions arrières au dessous du plan horizontal (élévation>180°) sont bien supérieures, ce que l'on pourrait expliquer par l'obstruction réalisée par l'épaule contralatérale. L'ILD moyen que nous observons atteint jusqu'à 27dB, ce qui est proche de la borne supérieure mesurée par Blauert, soit 30dB ([Bla97]).

1.2.1.3 Interaction entre ILD et ITD

Plusieurs expériences se complètent pour souligner d'une part l'autonomie de l'ITD et de l'ILD comme indice de latéralisation, et d'autre part leur très forte interaction pour le jugement de "plausibilité" du timbre perçu.

Dans [WK92], Wightman et Kistler proposent un test de localisation pour des signaux contenant un ITD synthétique en conflit avec les autres indices. Ils mettent en évidence la domination de l'ITD basses fréquences pour la latéralisation des sons large bande : un son parvenant aux oreilles avec un ITD de 90° d'azimut est perçu à cet azimut même si les autres indices correspondent à la position diamétralement opposée (270°) ! Toutefois, l'ITD perd sa prépondérance si la localisation est réalisée sur un son dénué de basses fréquences ($f > 2.5kHz$). Blauert propose plutôt la limite de $2kHz$ ([Bla97]). Au dessus de cette limite, donc, les indices spectraux sont les indices dominants. Ces résultats sont également confortés par l'étude de Kuhn, qui observe que l'ITD est minimum entre 1.4 et $1.6kHz$, et constitue à ce titre un indice de latéralisation peu robuste, tandis que l'ILD augmente fortement à partir de 2 à $3kHz$, et prend pour

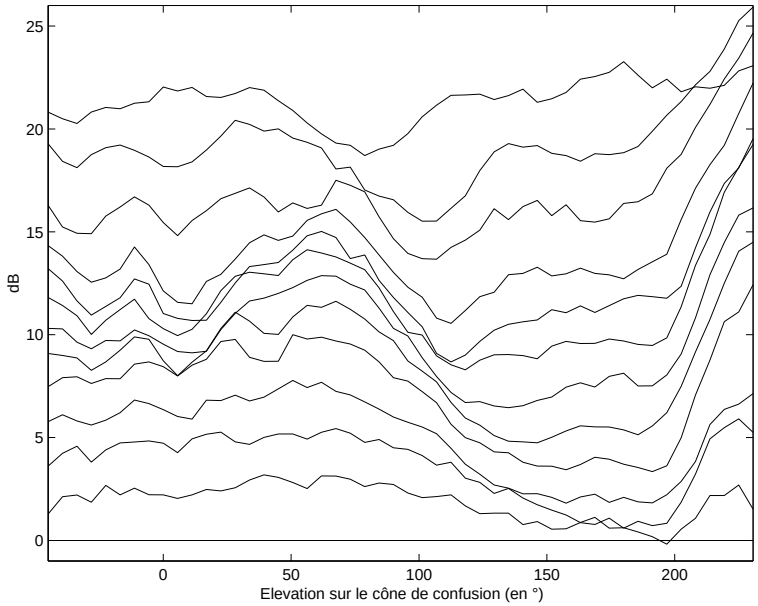


FIG. 1.2 – ILD sur des "cônes de confusion", pour des degrés de latéralisation régulièrement espacés, entre 0 et 0.98. Estimation réalisée par rapport en dB des spectres d'énergie, moyennés entre 1kHz et 5kHz. Mesures sur 17 têtes humaines réalisées à UC Davis (ILD moyennés sur les 17 têtes). En abscisse : "sagitallisation" de la Figure 1.6.

ainsi dire, le relais de l'ITD.

Gaik, quant à lui, mesure sur des têtes humaines les relations entre ILD et ITD en fonction de la position et de la fréquence ([Gai93]) : ces relations sont linéaires au dessous de 400Hz, et peuvent être approximées par un pôlynome d'ordre inférieur à 7 pour les fréquences supérieures. Il définit alors la notion de "relations naturelles" entre les indices interauraux, et vérifie que le non respect de ces lois (par exemple la "mise en conflit" du paragraphe plus haut) induit des artefacts de localisation dans les cas extrêmes (perception de sources multiples) mais surtout qu'il est facteur d'une faible plausibilité du son perçu.

ITD et ILD sont donc des indices de latéralisation autonomes : l'ITD est prépondérant pour un son large bande, mais l'ILD l'emporte sur l'ITD pour un son dénué de basses fréquences. Toutefois, la relation les unissant est un indice important pour la perception du timbre de la source.

1.2.1.4 Influence de la distance

Dans les études mentionnées précédemment, seules des sources sonores en champ lointain étaient considérées : les mesures présentées ont toutes été faites pour une distance source/récepteur, d, supérieure à 1m. Brungart et Rabinowitz mesurent et modélisent ITD et ILD en champ proche, pour des distances comprises entre 12cm et 1m ([BDR99a]). Deux principaux résultats peuvent être mentionnés :

1. l'ILD augmente fortement quand d diminue, et notamment lorsqu'elle est inférieure à 50cm. Cela s'explique par le phénomène de diffraction sur la tête apparaisant alors pour l'oreille ipsilatérale, tandis que l'oreille contralatérale est de plus en plus fortement masquée.
2. l'ITD varie de moins de 12% entre les valeurs établies pour le champ lointain et pour le champ proche.

Brungart et Rabinowitz présentent ainsi un nouvel attribut de l'ILD : il ne serait pas seulement un indice de latéralisation (faible, il est vrai, face à l'ITD), et de discrimination avant arrière ; il serait aussi un indice de la distance de la source.

1.2.2 Les indices monauraux

Même s'il ne s'agit pas réellement d'un cône, les Figures 1.1 et 1.2 mettent en évidence l'existence de positions présentant des différences interaurales identiques. ITD et ILD ne peuvent donc à eux seuls permettre de déterminer l'incidence d'une source sonore. Pour résoudre cette indétermination, le système auditif a recours à des indices dits monauraux, i.e. extraits des signaux droit et gauche indépendemment. Comme pour les indices interauraux, le système auditif interprète l'information monaurale en pratiquant conjointement une analyse dans le domaine temporel et une analyse dans le domaine fréquentiel.

1.2.2.1 Codage temporel des indices monauraux

L"'effet de précédence" désigne le phénomène par lequel le premier front d'onde conditionne la localisation d'une source sonore, indépendemment de l'incidence des réflexions lui succédant jusqu'à 50ms (parole) ou 80ms (musique) ([Gar68]). Ce phénomène permet donc de penser que le système auditif utilise des indices temporels pour la localisation. Batteau, puis Hiranika et Yamasaki se sont ainsi penchés sur l'évolution des réflexions sur le pavillon en fonction de l'indidence du son ([Bat67], [HY83]). Ils proposent les interprétations suivantes :

- l'information d'incidence de la source est codée sur un intervalle de $350\mu s$ après l'arrivée du son direct,
- une source provenant d'incidences frontales possède deux principales réflexions,
- une source provenant de l'arrière possède une seule réflexion,
- une source provenant du dessus ne possède pas de réflexion,
- l'intervalle de temps entre le son direct et ces réflexions augmente quand la source descend.

Ces résultats ne sont pas en accord avec ceux de Batteau, selon lesquels le nombre de réflexions ne dépend pas de l'incidence de la source.

1.2.2.2 Codage fréquentiel des indices monauraux

Les transformations subies par le son incident sur le corps de l'auditeur (oreille, tête, torse) engendrent également des phénomènes de résonances amplifiant certaines bandes de fréquences. Plusieurs auteurs ont cherché des caractéristiques objectives du codage fréquentiel des indices monauraux. Shaw par exemple propose un modèle de l'oreille sous forme de résonateurs en parallèle, que nous décrivons au chapitre 5 : chaque résonance, localisée en fréquence, serait active pour une zone d'incidences spécifique ([Sha80], [Sha82], [Sha97b], [Sha97a]). Hebrank et Wright, puis Han avancent l'hypothèse de "trajets" d'anti-résonances, toujours présentes mais de fréquence centrale variant en fonction de l'incidence ([HW74], [Han94]).

Pour valider ces interprétations "objectives", encore incomplètes, les études se réfèrent toutes aux résultats perceptifs de Blauert montrant l'existence de "bandes directionnelles" ([Bla97]). Ses expériences psycho-acoustiques, menées avec des bruits blancs à bande étroite diffusés de façon monaurale, ont mis en évidence des intervalles fréquentiels, qui, lorsqu'ils sont riches en énergie, induisent une localisation sans une zone d'incidence déterminée. C'est ainsi par exemple qu'un signal riche en énergie autour de 8kHz sera perçu comme provenant d'une incidence élevée (cf Figure 1.3). Le codage de l'incidence consisterait donc pour partie à régler le rapport d'énergie entre les bandes directionnelles.

1.2.2.3 Localisation et identification du timbre

Les indices monauraux présentés plus haut sont de même nature que les indices permettant d'identifier le timbre de la source (analyse temporelle, analyse spectrale). Il est donc important de comprendre comment le système auditif parvient à isoler les deux phénomènes.
Le modèle d'association de Theile, illustré en Figure 1.4, propose une interprétation ([The86]). Ce modèle suppose que nous possédons une base de données personnelle d'indices de localisation, acquise et mémorisée au cours de notre vie. Le son incident subit la transformation M lorsqu'il rencontre la tête, l'oreille et le torse de l'auditeur avant de parvenir au tympan. Le système auditif agirait alors en deux étapes couplées :

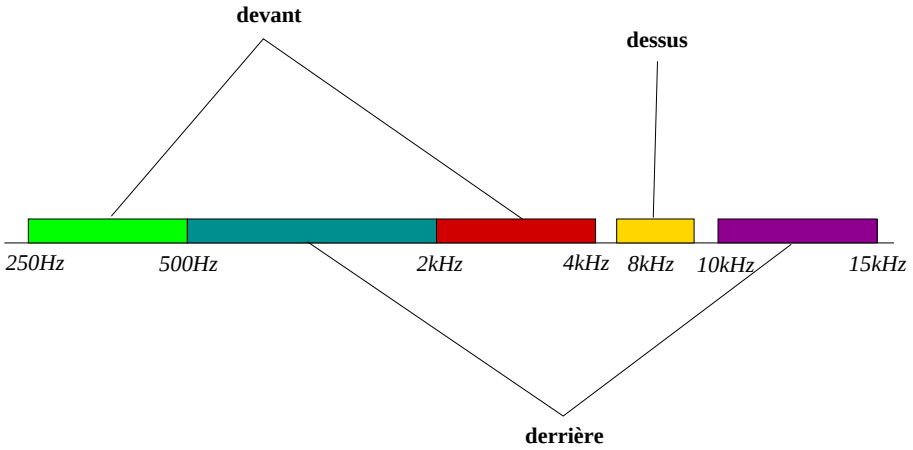


FIG. 1.3 – Bandes directionnelles proposées par Blauert dans [Bla97].

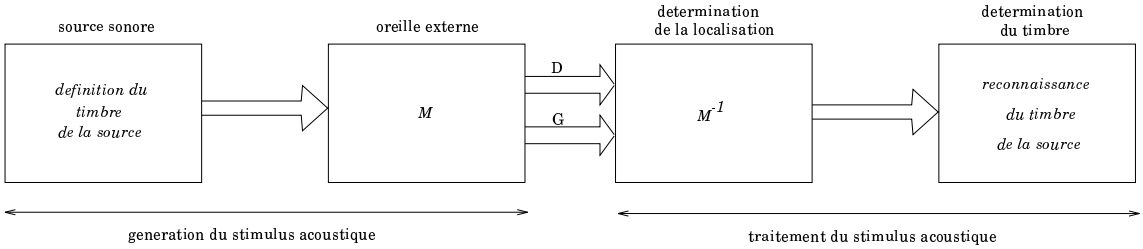


FIG. 1.4 – Principe du modèle d'association de Theile ([The80]).

1. il parcourt la base de données interne en appliquant au son la transformation inverse M^{-1} pour chaque position candidate (par exemple les positions du cône identifié par l'ITD).
2. pour chaque M^{-1} candidate, le son alors dénué des informations de localisation est traité par un processus de reconnaissance de forme afin d'identifier le timbre initial. Si le timbre n'est pas reconnu, alors on revient à l'étape 1 et on applique la transformation inverse d'une autre position.

Ce modèle montre ainsi qu'une bonne localisation est liée à la connaissance préalable du timbre de la source. En effet, Blauert a observé que de meilleures performance de localisation sont obtenues lorsque le stimulus est une voix familière plutôt qu'une voix inconnue ([Bla97]). Le modèle d'association prédit également l'altération du timbre de la source ou de l'incidence perçue lorsque les indices de localisation M ne concordent pas avec les indices de référence de la base de données interne : c'est le cas par exemple d'une synthèse binaurale non individuelle, que nous examinerons au chapitre 5.

1.2.3 Conclusion sur les indices de localisation

Les indices de localisation constituent des clés acoustiques permettant à notre système auditif d'identifier l'incidence d'une source sonore. Ils sont de deux natures (indices interauraux/indices monauraux) et possèdent dans les deux cas des attributs temporels et fréquentiels variant avec l'incidence. La connaissance de ces indices laisse entrevoir la possibilité de reproduire artificiellement une sensation de localisation. C'est l'objet de la synthèse binaurale qui constitue le coeur de cette thèse.

1.3 Caractérisation des indices de localisation

Connaissant la nature des indices acoustiques de localisation, nous souhaitons à présent quantifier leur variation en fonction de la position. Ils peuvent être caractérisés par la mesure des fonctions de transfert entre le point d'émission du son, à une incidence donnée, et le tympan de l'auditeur. Ces fonctions de

transfert contiennent l'empreinte des transformations subies par le son lors de sa propagation et plus précisemment les effets de diffraction, diffusion et réflexion sur le corps de l'auditeur. Elles dépendent de l'incidence du son et des caractéristiques morphologiques de l'individu, et sont couramment désignées par Head-Related Transfer Functions (HRTF).

Pour reproduire la sensation d'un son provenant d'une incidence donnée, il suffit alors de filtrer un son monophonique dénué d'information de localisation par les HRTF droite et gauche mesurées pour cette position. C'est le principe de la synthèse binaurale.

Dans cette section, nous décrivons les campagnes de mesure ayant permis d'obtenir les HRTF que nous utiliserons dans la suite du document. De plus, nous décrivons les modalités d'extraction de l'information utile, les indices de localisation, à partir des HRTF mesurées, qui nous permettrons d'optimiser le coût d'implantation de la synthèse binaurale.

1.3.1 Mesure de HRTF en champ libre

Les HRTF que nous utiliserons sont issues de trois campagnes de mesure :

- mesure de la tête artificielle KEMAR par Bill Gardner et Keith Martin au MIT ([GM94]),
- mesure de 24 têtes humaines et de la tête artificielle Head Acoustics HMSII à l'Ircam, dans le cadre de cette thèse,
- mesure de 17 têtes humaines par Ralph Algazi et al. à UC Davis ([AAT99]).

1.3.2 Protocoles de mesure

1.3.2.1 Mesures réalisées à l'Ircam

Les mesures ont été réalisées entre 1996 et 1999 dans la chambre anéchoïque de l'Ircam. Le matériel informatique se compose d'un PC muni d'une carte d'acquisition OROS Au22 et de convertisseurs A/N et N/A à 16 bits, que nous avons utilisés à la fréquence d'échantillonnage de 48kHz. Le signal de mesure est constitué des séquences de longueur maximale de 16383 points (Maximimum Length Sequences, [RV89]). Le rapport signal à bruit moyen est de 60dB.
Les séquences sont émises par une enceinte TANNOY System 8 NFM, qui, de par ses dômes concentriques, approxime une source ponctuelle. Les microphones à électrets Sennheiser KE4-211-2, utilisés à l'Ircam depuis 1993, ont été préférés aux sondes préconisées par exemple par Wightman et Kistler ([WK89a]), dans la mesure où ils ont une plus large bande passante. Les mesures sont réalisées en mode "conduits bouchés" tel que le décrit Moller ([Mol92]) : les microphones sont fixés à l'entrée du conduit auditif à l'aide de bouchons d'oreille perforés en leur centre pour que s'y fixe la capsule.

Pour des raisons pratiques, la distance source-récepteur n'a pu être choisie au delà de 1.40m, ce qui reste proche des conditions de Wightman et Kistler (1.38m), de Gardner (1.40m) ou d'Algazi et al. (1m). Les sujets sont assis sur une chaise pouvant tourner en azimut, selon un pas de rotation constant de 15°[1]. On mesure donc 24 positions par élévation. Pour les mesures sur tête artificielle, un opérateur fait tourner la chaise entre chaque mesure. Les autres sujets effectuent eux-mêmes l'opération, en essayant de maintenir la tête à une altitude constante.
Contrairement aux mesures pratiquées par Moller, Gardner, Wightman ou Algazi et al., nous n'avons pu fabriquer de support d'enceinte circulaire : le haut-parleur de mesure est fixé sur une barre métallique verticale, et le changement d'élévation est opéré par déplacement le long de cette barre. La distance source-récepteur varie donc en fonction de l'élévation, ce que nous compensons en égalisant les HRTF mesurées par une mesure de référence de l'enceinte réalisée à chaque élévation par un microphone omnidirectionnel (voir Figure 1.5). Cette mesure de référence permet également d'observer les pertes d'énergie en hautes fréquences lorsque l'enceinte est élevée, liées à sa forte directivité pour cette plage de fréquence. On constate que les mesures de références varient de moins de 3dB au dessous de 8kHz, et jusqu'à 5dB pour les plus hautes fréquences.

L'échantillonnage de l'espace que nous avons choisi s'inscrit dans un repère azimut/élévation (voir Figure 1.6). Le positionnement du haut-parleur à chaque nouvelle élévation étant long, la plupart de nos sujets

[1] La chaise a été construite dans le cadre de cette étude par Alain Terrier, de l'atelier mécanique de l'Ircam.

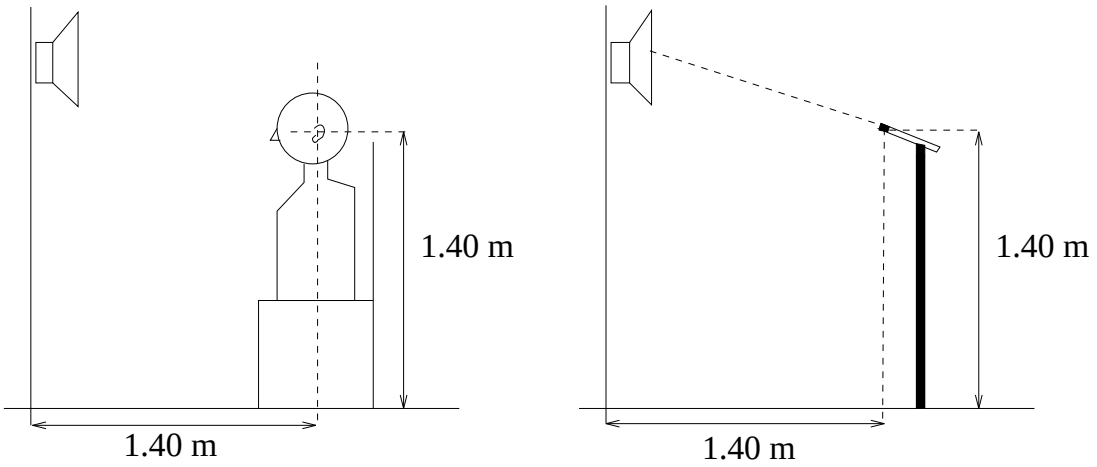


FIG. 1.5 – Système de mesure utilisé à l'Ircam pour les mesures de HRTF. Pour chaque élévation, une mesure de référence est réalisée avec un microphone omnidirectionnel placé au lieu du centre de la tête.

élévations	*pas en azimut*	*têtes mesurées*
0°	15°	8 têtes humaines
0°, 30°	15°	10 têtes humaines
0°, 30°, 60°, 90°	15°	1 tête humaine
0°, 10°, 20°, 30°, 40°, 90°	15°	5 têtes humaines et 1 tête artificielle (HMSII)

TAB. 1.1 – Liste des mesures de HRTF réalisées dans le cadre de cette thèse entre 1996 et 1999.

n'ont été mesurés que pour une ou deux élévations. Seuls la tête artificielle et les sujets directement impliqués dans l'étude ont été mesurés pour un ensemble plus complet d'élévations. Les mesures réalisées à l'Ircam dans le cadre de cette thèse sont recensées dans le Tableau 1.1. Pour la mesure au zénith (élévation 90°), on demande aux sujets de placer leur buste à l'horizontal, et on place le haut-parleur dans le prolongement de celui-ci.

1.3.2.2 Autres campagnes de mesure

Les campagnes de mesure de Gardner et d'Algazi se démarquent la notre par le nombre de positions recensées.

Le dispositif d'Algazi permet en effet de mesurer plus de 1000 positions en moins d'une heure ! Cette rapidité est obtenue grâce à l'automatisation du déplacement des hauts-parleurs. Contrairement aux sessions MIT et Ircam, les sujets sont ici immobiles. Un support en arc de cercle est équipé de 25 petits haut-parleurs. Le repère adopté pour l'échantillonnage de l'espace est un repère latéralisation/sagittalisation (cf Figure 1.6), et chacun de ces haut-parleurs correspond à une latéralisation différente. Par sa rotation, l'arc de cercle leur fait ainsi décrire un cône de confusion. Chacun de ces 25 cônes est caractérisé par 50 mesures, repérées par un angle d'élévation (ou de sagittalisation). Nous n'utilisons que les positions de la calotte supérieure, soient 825 mesures. Les mesures n'ont pas été réalisées en chambre anéchoïque, et, pour pouvoir éliminer par fenêtrage les réflexions engendrées par la salle de mesure, l'arc de cercle est situé très proche de l'auditeur.

Contrairement aux sessions UC Davis et Ircam, l'échantillonnage choisi par Gardner est tel que chaque mesure "occupe" un angle solide à peu près égal. Le nombre de mesures est ainsi moindre pour les fortes élévations que pour le plan horizontal. Quatorze élévations ont été recensées, de -40° d'élévation à 90°, pour un incrément en azimut allant de 5° pour les élévations près de 0°, à 30° pour l'élévation 80°, soit un total de 710 positions mesurées. Les mesures sont réalisées sur une tête artificielle, le KEMAR[2], munie de deux pavillons différents (pavillons DB061 et DB065).

[2]Les HRTF du mannequin KEMAR utilisées sont celles mesurées par Bill Gardner et Keith Martin, disponibles sur internet : http ://sound.media.mit.edu/KEMAR.html.

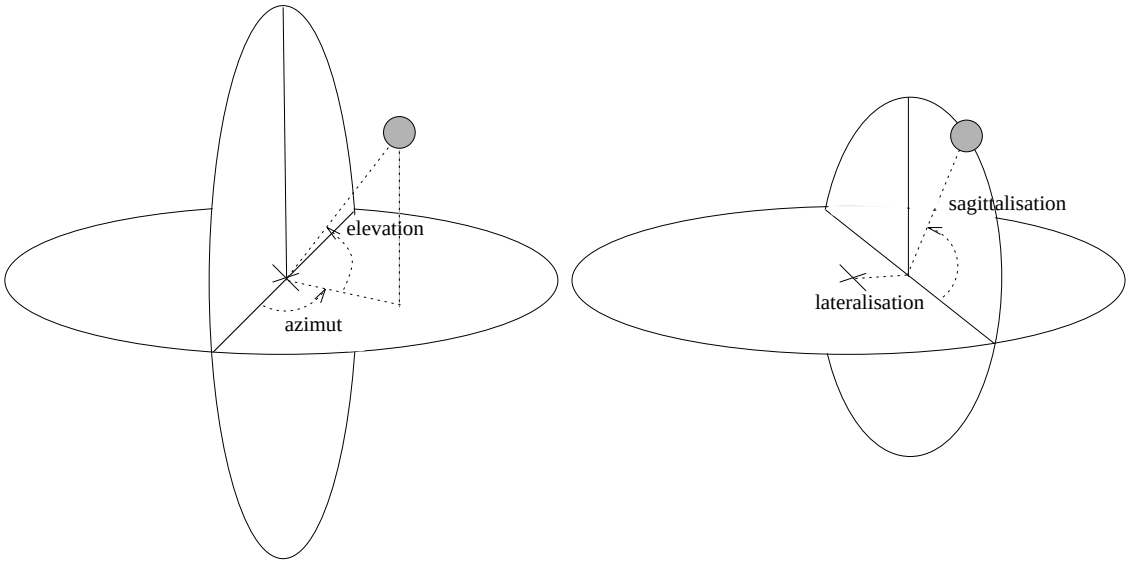


FIG. 1.6 – Repères choisis pour les mesures de HRTF : repère azimut/élévation pour les sessions Ircam et MIT, repère latéralisation/sagittalisation pour la session UC Davis.

1.3.3 Extraction de l'information dans les HRTF mesurées

Les HRTF mesurées contiennent beaucoup d'information. Pour concentrer cette dernière aux indices de localisation que nous souhaitons caractériser et reproduire, nous résumons les post-traitements pouvant être appliqués (voir aussi [JLW95]).

1.3.3.1 Séparation amplitude/retard pur

Comme tout filtre causal et stable, chaque HRTF peut être divisée en une composante à phase minimale, H_{min}, et une composante passe-tout, H_{exc} :

$$\begin{aligned} HRTF &= mag.\exp^{j.\varphi} = mag.\exp^{j.mph}.\exp^{j.eph} \\ &= H_{min}.H_{exc} \end{aligned}$$

avec :

$$\begin{aligned} H_{min} &= mag.\exp^{j.mhp} \\ H_{exc} &= \exp^{j.eph} \end{aligned}$$

La phase φ de chaque HRTF est ainsi divisée entre la phase de la composante minimale, ou "phase minimale" notée mph, et la phase de la composante passe-tout, ou "excès de phase" noté eph.

Le spectre d'amplitude est relié de façon univoque à la phase minimale, par la transformée de Hilbert [Opp74] :

$$mph = \Im\left(Hilbert\left(-\log\left(mag\right)\right)\right)$$

La composante à phase minimale des HRTF est ainsi entièrement caractérisée par la donnée des spectres d'amplitudes. Un exemple de phase minimale et de spectre d'amplitude est donné en Figures 1.8 et 1.9.

En outre, Mehrgart et Mellert ont souligné dans [MM77] le caractère linéaire de l'excès de phase jusqu'à 10kHz, et ont proposé un modèle simplifié des HRTF en remplaçant la composante passe-tout par un retard pur, variant avec l'incidence. On observe en effet sur nos données une quasi-linéarité de l'excès de phase, jusqu'à environ 8kHz, ce résultat se dégradant pour les positions contralatérales (Figure 1.7). La pente de la droite ainsi définie représente le retard τ avec lequel le son émis par la source parvient à

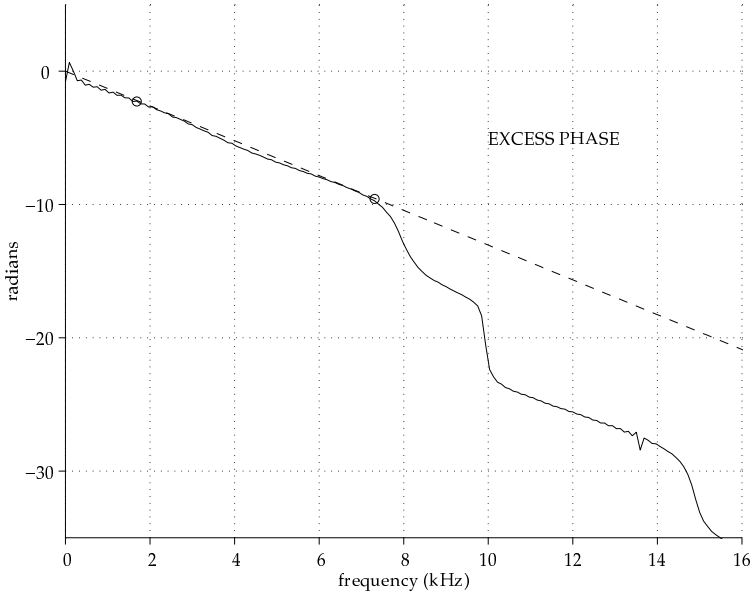


FIG. 1.7 – Excès de phase mesuré sur la tête artificielle HMSII (azimut=150°, élévation = 0°).

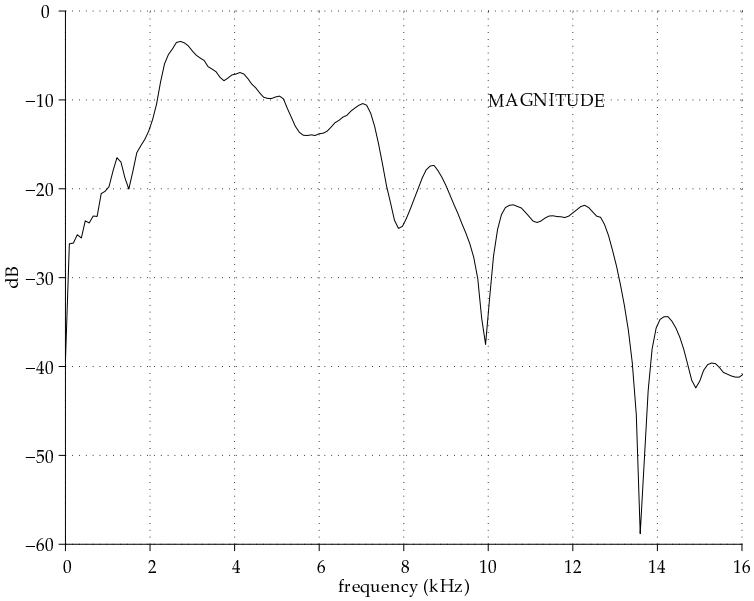


FIG. 1.8 – Spectre d'amplitude mesuré sur la tête artificielle HMSII (azimut=150°, élévation = 0°).

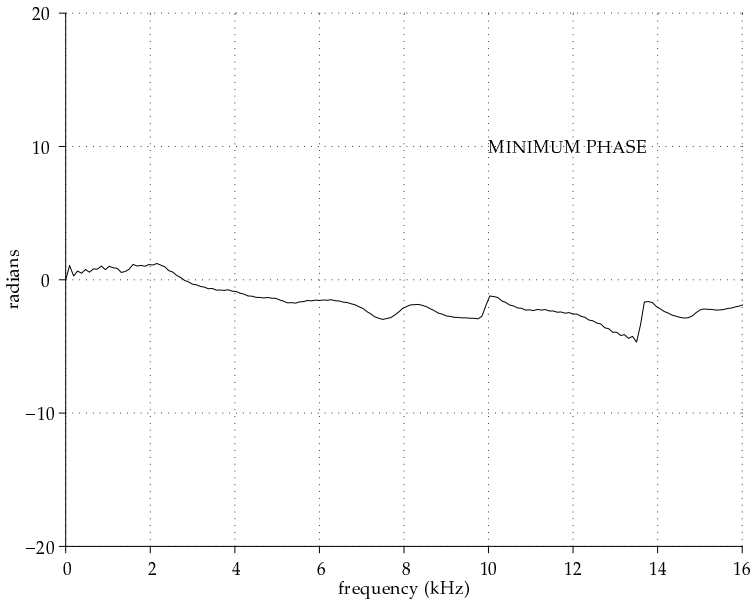


FIG. 1.9 – Phase minimale mesurée sur la tête artificielle HMSII (azimut=150°, élévation = 0°).

l'oreille. Il est qualifié de "retard monaural", la différence entre les deux oreilles donnant accès au retard interaural, ou ITD, défini en section 1.2.

Chaque HRTF est ainsi entièrement caractérisée par la donnée de son spectre d'amplitude et d'un retard.

1.3.3.2 Validité de ce modèle

Des validations perceptives de ce modèle ont été proposées par Wightman et Kistler ([WK92]) ou par Kulkarni et Colburn ([KC95b], [KC95a]). Elles ont montré que l'information de phase "sacrifiée", principalement située en hautes fréquences, n'est pas utilisée par le système auditif pour déterminer l'incidence du son. Plus récemment, Kulkarni montre dans [KIC99] que le système auditif extrait un ITD global des composantes basses fréquences de la phase : deux sons possédant un même ITD basses fréquences sont perçus comme identiques, quels que soient les détails de leur ITD respectif sur le reste de la bande audible. En outre, il montre que la modélisation de la phase des HRTF par une phase purement linéaire n'induit aucun artefact audible. Par conséquent, on peut envisager une implantation des HRTF sous forme d'un filtre à phase nulle d'amplitude *mag*, en série avec un retard pur, reproduisant le retard monaural basses fréquences mesuré.

D'autres études concernant l'audibilité de la composante passe-tout résiduelle des HRTF sont actuellement menées par Minnaar et Moller ([MCM+99]).

1.3.3.3 Réduction du spectre d'amplitude à l'information dépendant de la direction

Le spectre d'amplitude des HRTF contient des éléments indépendants de la direction, qui constituent à ce titre une information superflue. On peut mentionner les contributions suivantes :

- enceinte de mesure,
- chaîne de mesure (carte son, amplificateurs),
- microphones insérés dans les conduits,
- conduit auditif dans le cas d'une mesure en conduits ouverts.

Ces contributions peuvent être caractérisées puis éliminées par simple division de spectres d'amplitude dans le domaine fréquentiel. En effet, on fait communément l'hypothèse que les différentes contributions mentionnées ont une phase minimale, ou du moins que c'est cette composante qu'il est nécessaire d'éliminer. De plus, de part la linéarité de la transformée de Hilbert, le produit de filtres à phase minimale est

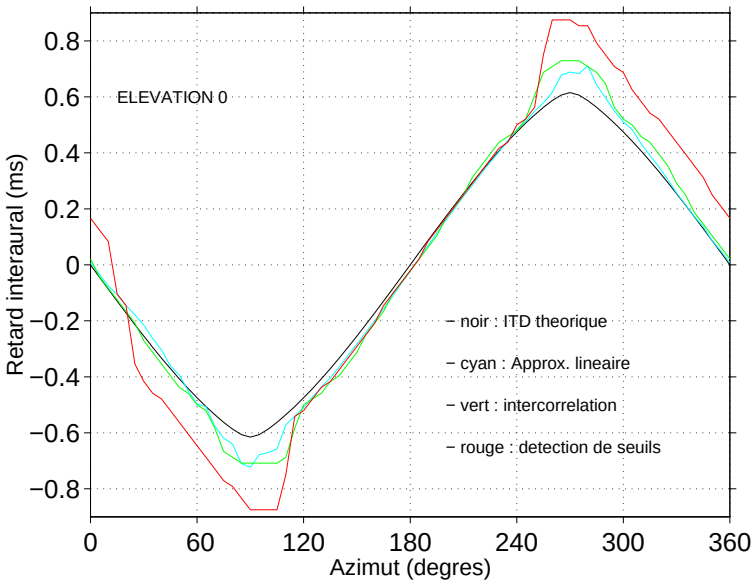


FIG. 1.10 – Approximations du retard interaural obtenues à partir de mesures réalisées tous les 5° en azimut sur la tête artificielle HMSII, à l'élévation 0°.

un filtre à phase minimale, propriété qui rend l'opération d'égalisation très simple. Nous en développons les modalités en section 1.5.

1.4 Estimation du retard interaural

Nous cherchons à estimer un ITD indépendent de la fréquence, correspondant à la valeur mesurées en basses fréquences. Ce chapitre reprend les éléments exposés au Confrès Français d'Acoustique 1997 ([LJ97]). Nous proposons en fin de section les résultats obtenus par Daniel pour l'estimation d'un ITD hautes fréquences ([Dan00]).

1.4.1 Estimation par approximation linéaire [JLW95]

Cette première méthode a été l'objet de développements spécifiques dans le cadre de cette thèse. Elle prolonge la réflexion faite en section 1.3 sur la linéarité de l'excès de phase : le spectre d'excès de phase peut être approximé par une droite dont la pente représente le retard monaural. Une estimation de l'ITD peut donc être obtenue en effectuant une régression linéaire sur la différence des excès de phase gauche et droite :

$$ITD = Lin\left(eph_L - eph_R\right)$$

Cette estimation utilisant conjointement l'information "des deux oreilles" est apparue plus efficace que le calcul indépendant de chaque retard monaural, suivi de la différence droite/gauche pour former l'ITD. Nous avons proposé plusieurs précautions pour garantir la robustesse de la méthode. En raison des ruptures de pente observées en figure 1.11, il convient en effet d'isoler les intervalles fréquentiels sur lesquels l'excés de phase est quasi-linéaire. Une solution consiste à détecter les pics de la dérivée de l'excès de phase, et à retenir comme intervalle de régression *le plus grand intervalle sans pic*, comme l'illustre la figure 1.12.

De façon pratique, l'application à nos données conduit toujours à une borne inférieure de l'intervalle d'estimation au dessous de 1kHz. L'ITD estimé contient donc toujours d'information de phase basses-fréquences. En fonction de la position, et donc de la linéarité de l'excès de phase, la longueur de cet intervalle oscille entre quelques centaines de hertzs à plus de 10kHz. Le résultat de l'approximation de la différence d'excès de phase est illustré en Figure 1.13.

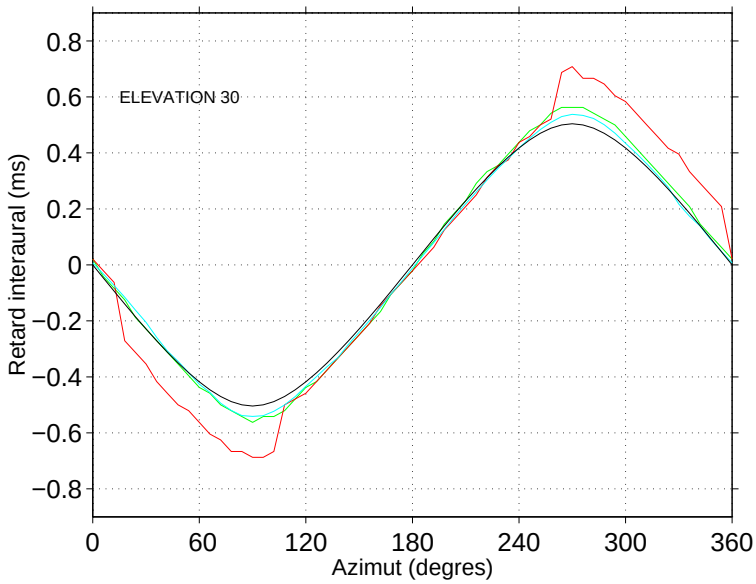


FIG. 1.11 – Approximations du retard interaural obtenues à partir de mesures réalisées tous les 5° en azimut sur la tête artificielle HMSII, à l'élévation 30°.

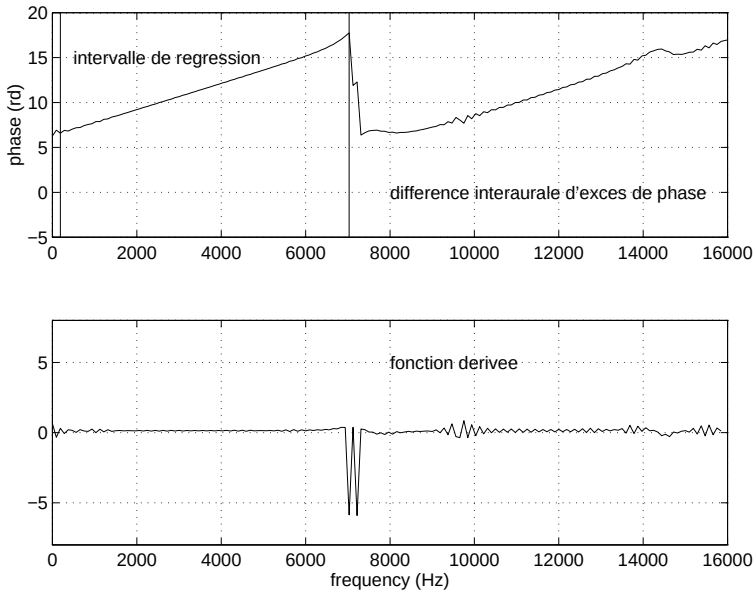


FIG. 1.12 – Différence d'excès de phase et sa fonction dérivée mesurées sur la tête artificielle HMSII à l'azimut 30°et à l'élévation 0°. Intervalle de régression retenu : [190Hz-7kHz].

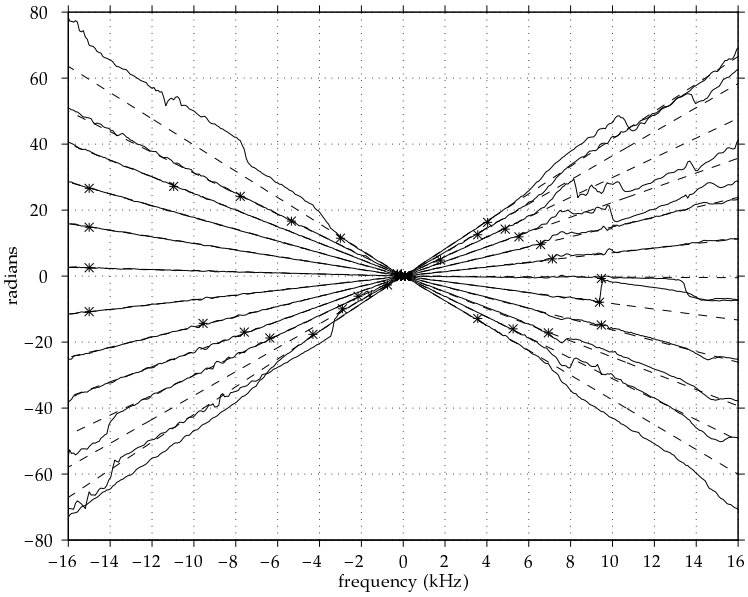


FIG. 1.13 – Différence interaurale d'excèse de phase dans le plan horizontal : tête artificielle HMSII, mesures par pas de 15°.

1.4.2 Estimation par corrélation en sous-bandes [WK92]

La méthode utilisée par Wightman et Kistler consiste à filtrer en sous-bandes les réponses impulsionnelles droite et gauche mesurées, contenant donc composante passe-tout **et** composante à phase minimale, et à réaliser l'intercorrélation de ces deux signaux. L'ITD est alors défini, pour chaque sous-bande, par la position du maximum de la fonction d'intercorrélation :

$$ITD(i) = \max\left(corr(h_L(i), h_R(i))\right)$$

Pour l'illustration présentée en figures 1.10 et 1.11, nous avons choisi un filtrage en bandes d'octave. L'ITD retenu a été obtenu pour l'octave [2.5Khz-5kHz] qui correspond à la bande la plus basse fréquences atteignant la courbe asymptote de l'ITD décrite par [CB89].
Les estimations obtenues par les méthodes d'approximation linéaire et d'intercorrélation ont des performances comparables. L'écart observé sur les figures, maximal aux azimuts 90° et 270°, est expliqué par le fait que Wightman et Kistler conservent la composante phase minimale dans les réponses intercorrélées. Pourtant, Kulkarni et al. montrent que le retard interaural entre les réponses impulsionnelles à phase minimale peut atteindre la valeur de $80\mu s$ aux azimuts 90° et 270° [KIC95]. Si l'on compare les performances des deux méthodes en utilisant les mêmes données, i.e. les composantes passe-tout droite et gauche, les résultats sont effectivement plus concordants (figure 1.14).

L'équivalence entre les deux méthodes peut être montrée de façon analytique. Soit r_k la fonction d'intercorrélation des réponses impulsionnelles droite et gauche de la composante passe-tout des HRTF pour une incidence donnée, filtrées par un filtre passe-bande, pour $k \in [0, N]$. Soit R_n sa transformée de Fourier, avec $n \in [0, nfft - 1]$. On a :

$$\begin{aligned} R_n &= H_L(n).\overline{H_R}(n) \\ &= \exp^{j.eph_L(n)}.\exp^{-j.eph_R(n)} \end{aligned}$$

soit :

$$R_n = \exp^{j.eph(n)}$$

avec :

$$eph(n) = eph_L(n) - eph_R(n)$$

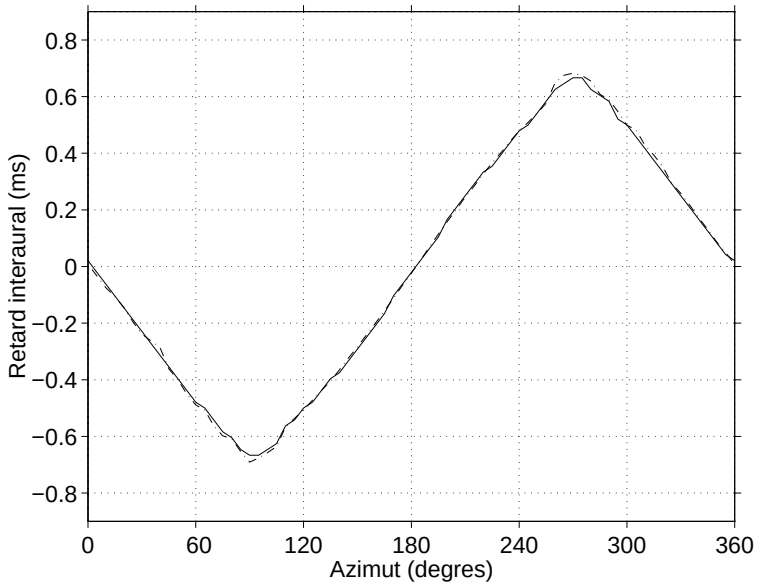


FIG. 1.14 – Estimation du retard interaural à partir de mesures réalisées sur la tête artificielle HMSII à l'élévation 0° : intercorrélation des réponses impulsionnelles "passe-tout" pour l'octave [2.5Khz-5kHz] (trait plein), approximation linéaire de la différence d'excès de phase (trait interrompu).

On a donc :

$$\begin{aligned} r_k &= \frac{1}{nfft}.\sum_{n=0}^{nfft-1} R_n.\exp^{\frac{2.j.\pi}{nfft}.k.n} \\ &= \frac{1}{nfft}.\{\sum_{n=n_0}^{n_0+\Delta n} R_n.\exp^{\frac{2.j.\pi}{nfft}.k.n} + \sum_{n=nfft-n_0-\Delta n}^{nfft-n_0} R_n.\exp^{\frac{2.j.\pi}{nfft}.k.n}\} \\ &= \frac{2}{nfft}.\{\sum_{n=n_0}^{n_0+\Delta n} \cos(eph(n) + \frac{2\pi.k.n}{nfft}) \end{aligned}$$

Le caractère quasi-linéaire de l'excès de phase peut se traduire par :

$$\forall \epsilon \quad \exists\, \delta(n) n \quad / \quad eph(n) = a.n + b + \delta(n)$$

avec :

$$|\delta(n)| < \epsilon$$

On a pour ce choix de $\delta(n)$ (vérifié si la bande n'est pas trop large : Δn petit) :

$$r_k \simeq \frac{2}{nfft}.\sum_{n=n_0}^{n_0+\Delta n} \cos(n.(a + \frac{2.j.\pi}{nfft}.k) + b)$$

Donc r_k est maximum pour :

$$\cos(n.(a + \frac{2.\pi}{nfft}.k) + b) = 1 \quad \forall n$$

Ce maximum est atteint si et seulement si :

$$n.(a + \frac{2.j.\pi}{nfft}.k + b) = 2\pi.p \quad p \in Z,\ \forall n$$

donc pour :

$$k = \frac{-nfft}{2\pi}.a$$

et :

$$b = 2\pi.p$$

Le retard interaural déterminé par la méthode de régression linéaire sur la différence d'excès de phase ($\frac{-nfft}{2.\pi}.a$) est donc égal à celui obtenu par la méthode d'intercorrélation en sous-bandes pour peu que la bande retenue pour cette dernière soit assez fine.

1.4.3 Méthode par détection de seuil [MSHJ95]

La méthode décrite par Moller et al. consiste à détecter le temps d'arrivée du son par une détection de seuil sur les réponses impulsionnelles complètes droite et gauche. Ils définissent le temps d'arrivée par le rang de l'échantillon dont l'amplitude atteint 10% de la valeur maximale de la réponse.

$$ITD(i) = \tau_{arrL} - \tau_{arrR}$$

Cette technique est dans son principe sensible au bruit de mesure. Ainsi la détection du temps d'arrivée du son sur l'oreille contralatérale, pour laquelle le rapport signal à bruit est inférieur de 15dB à celui obtenu sur l'oreille ipsilatérale, est moins fiable. Ce manque de robustesse est illustré sur les Figures 1.10 et 1.11 : les ITD estimés par cette méthode s'écartent de façon importante des autres courbes.

1.4.4 Estimation de l'ITD hautes fréquences [Dan00]

Daniel propose une méthode d'estimation de l'ITD s'appuyant sur l'écart des époques moyennes de groupe des HRTF droite et gauche. Prolongeant la théorie énergétique proposée par Mertens, il propose ainsi une estimation du retard monaural à partir de leur enveloppe temporelle $a(t)$:

$$\tau = \frac{\int_{-\infty}^{+\infty} a^2(t).t.dt}{\int_{-\infty}^{+\infty} a^2(t).dt}$$

L'ITD est alors obtenu par différence des retards droit et gauche. L'un des attrait de cette approche consiste en son extension naturelle, proposée par Daniel, pour estimer l"'indice de confiance" de l'estimation, σ_t : si l'enveloppe des signaux est large, étalée, alors σ_t est grand, et traduit l'ambiguïté de la détermination de l'instant "moyen" de l'enveloppe.

Daniel introduit également un raffinement supplémentaire, tirant profit de l'approche précédente mais traitant les informations droite et gauche simultanément, sans passer par un calcul isolé de chaque retard monaural. Cette solution est interprétée comme une moyenne du retard de groupe sur la bande utile ($f > 1kHz$ ou $2kHz$) avec pondération fréquentielle énergétique.

1.4.5 Conclusion sur l'estimation de l'ITD

Dans la suite de la thèse, nous utilisons l'estimation d'ITD "par approximation linéaire". Cette solution nous a en effet semblé la plus robuste aux artefacts de mesure, du fait de l'adaptation de l'intervalle d'estimation aux points "fiables". En outre, elle s'appuie sur l'information de phase en basses fréquences, qui semble avoir une importance perceptive prépondérante.

1.5 Egalisation des HRTF

Les techniques binaurales couvrent trois approches :

- l'enregistrement binaural d'une scène sonore constituée,
- la convolution d'un son monophonique par les réponses impulsionnelles binaurales d'un lieu,
- la convolution d'un son monophonique par les HRTF mesurées en champ libre.

Pour ces différentes approches, la fidélité de la simulation nécessite de compenser l'effet des transducteurs de mesure (microphones) et de restitution (écouteurs), opération que nous qualifierons d"'égalisation" ([Bla97], [JLW95], [BL95], [Mol92]).

Dans cette section, nous comparons tout d'abord l'égalisation par rapport au champ diffus et l'égalisation par rapport au champ libre. Des études antérieures ont déjà établi les avantages offerts par la première pour la compatibilité entre les enregistrements sur tête artificielle et les techniques stéréo traditionnelles ([The86], [Bla97]). L"'égalisation champ diffus" permet aussi d'éliminer les artefacts de mesure indépendants de la direction, et de réduire de façon significative les différences entre sessions de mesure ou entre

têtes. En outre, elle permet de réduire le coût de l'implantation de la synthèse binaurale.
Dans une seconde partie, nous comparons trois méthodes d'estimation de la "HRTF diffuse", filtre utilisé pour l'égalisation des enregistrements binauraux, des HRTF et des casques d'écoute. Nous proposons une nouvelle méthode d'estimation, nécessitant très peu de matériel spécifique (pas même un haut-parleur) et pouvant être réalisée dans une salle d'usage courant, par exemple dans la salle utilisée pour les enregistrements binauraux.

Ce chapitre reprend les éléments présentés dans l'article [LJV98].

1.5.1 Comparaison des égalisations "champ diffus" et "champ libre"

1.5.1.1 Notion d'égalisation découplée

La simulation binaurale repose sur des mesures ou enregistrements réalisés à l'aide de microphones insérés dans les conduits auditifs, et utilise un casque d'écoute pour la restitution. Ces deux transducteurs réalisent un filtrage du signal incident que l'on doit compenser pour que les caractéristiques spatiales du son ne soient pas altérées.

Une méthode pour réaliser cette compensation, que l'on qualifiera d'égalisation non découplée, s'appuie sur la mesure de la fonction de transfert du casque d'écoute, $CASQ(f)$, avec la même chaîne de mesure que celle utilisée pour les enregistrements ou la mesure des réponses impulsionnelles binaurales (BIR) ou des HRTF (notamment avec la même tête et les mêmes microphones). Elle est schématisée en Figure 1.15 : les canaux binauraux sont déconvolués par $CASQ(f)$ en amont de la reproduction sur le même casque d'écoute. Si le couplage casque/oreille reproduit les conditions du champ libre (casque "ouvert" selon la définition de Moller dans [Mol92]), cette égalisation compense tant l'effet des microphones que celui du casque. Si le récepteur considéré est une tête artificielle, ou tout autre dispositif pour lequel la position des microphones est fixe, la mesure du casque peut être réalisée à tout moment, éventuellement après la session de mesure. En revanche, si le récepteur est un individu, la position du microphone dans le conduit auditif change à chaque replacement. Le casque doit donc être mesuré lors de la même session de mesure, ce qui contraint dès lors le choix du casque à utiliser lors de la session d'écoute.

Une méthode d'égalisation alternative, que nous qualifierons d'égalisation découplée, a été présentée par Blauert dans [BL95] et [Bla97]. Elle repose sur un champ sonore de référence, utilisé pour égaliser les canaux binauraux et le casque d'écoute. Si le champ de référence peut être reproduit de façon robuste, cette approche présente la même efficacité que l'approche non découplée, avec une flexibilité accrue. En effet, l'égalisation du casque d'écoute est alors séparée de celle de la tête artificielle, et permet à tout moment de faire la restitution sur un nouveau casque, pour peu que ce casque soit, par construction ou par traitement informatique, égalisé par rapport au même champ de référence que les canaux binauraux. Cette méthode est illustrée en Figure 1.16. Le filtre d'égalisation du casque s'exprime par $REF(f)/CASQ(f)$, où $REF(f)$ représente la fonction de transfert de référence.

1.5.1.2 Compatibilité entre techniques binaurales et stéréophonie conventionnelle

Deux principaux champs de référence ont été préconisés par les constructeurs de casques et de têtes artificielles :

1. le champ libre, constitué d'une onde plane provenant d'une incidence donnée, le plus souvent l'incidence frontale,
2. le champ diffus, constitué d'ondes planes décorrélées provenant d'incidences uniformément distribuées autour du récepteur ([Ber49], [Sha88]).

Lorsqu'on enregistre une source frontale dans une chambre anéchoïque, une tête artificielle égalisée par rapport au champ libre produit des canaux droite et gauche identiques à celui qu'un microphone idéal aurait capturé. Toutefois, si le son ne provient pas de l'avant ou forme un champ diffus, une tête artificielle égalisée par rapport au champ libre modifie son spectre, différemment sur les canaux droite et gauche. De la même manière, une tête artificielle égalisée par rapport au champ diffus préserve le contenu spectral d'un champ sonore diffus, mais modifie celui d'une onde plane provenant d'une seule incidence.

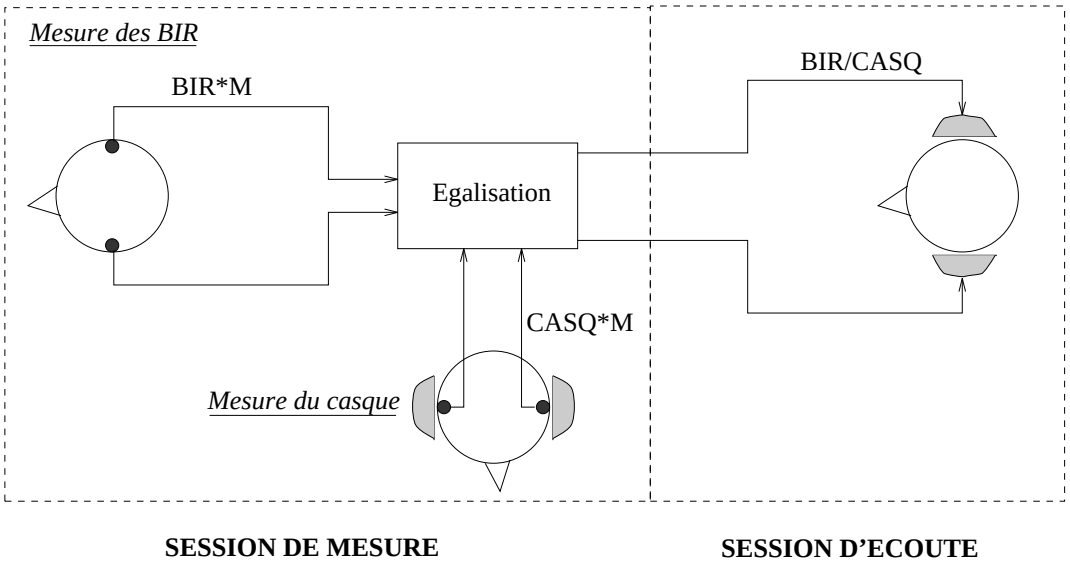


FIG. 1.15 – Egalisation non découplée. "BIR" désigne une réponse impulsionnelle binaurale, une HRTF ou un enregistrement binaural. "CASQ" désigne la réponse impulsionnelle du casque d'écoute. "M" désigne la contribution du microphone de mesure et du conduit auditif, indépendente de la direction. "*" représente la convolution et "/" la déconvolution.

Ces considérations permettent d'envisager l'écoute sur haut-parleurs d'enregistrements binauraux réalisés sur tête artificielle. Avec une tête artificielle égalisée par rapport au champ libre, les sons frontaux sont reproduits comme s'ils avaient été enregistrés avec une technique de prise de son stéréo non coïncidente standard. Toutefois, le timbre de sons non frontaux ou diffus (notamment la réverbération tardive) est altéré. Au contraire, une tête artificielle égalisée par rapport au champ diffus enregistre la réverbération tardive diffuse comme l'aurait fait un couple de microphones standard. Des expériences menées par Theile et reportées dans [Bla97] indiquent que si la scène sonore comporte un niveau important de réverbération les auditeurs préfèrent une prise de son égalisée par rapport au champ diffus.
De façon réciproque, dans [The86], Theile cherche l'égalisation optimale pour les casques d'écoute, pour la restitution d'enregistrements stéréo standard. Bien que l'égalisation champ libre soit théoriquement idéale pour reproduire les sources frontales de ces signaux, l'étude confirme, comme au paragraphe précédent, qu'il est préférable d'adopter une égalisation par rapport au champ diffus pour optimiser la fidélité globale de la diffusion.

On peut noter que dans les deux études précédentes, l'égalisation ne compense que les défauts de timbre de l'enregistrement, et ne corrige en aucun cas l'altération de la distribution spatiale du champ sonore liée à la dégradation des relations interaurales des signaux perçus par l'auditeur. En effet, pour garantir l'intégrité de l'information spatiale, le champ sonore devrait idéalement être capturé par la tête de l'auditeur (et subir une annulation des trajets croisés dans le cas d'une écoute sur haut-parleurs).

1.5.1.3 Compatibilité entre têtes à l'enregistrement et à la reproduction

Suivant l'hypothèse formulée par Blauert dans [BL95], nous souhaitons étudier la capacité de l'égalisation découplée à adapter les canaux binauraux à l'auditeur final, par l'introduction des caractéristiques individuelles au niveau de l'égalisation du casque d'écoute. En adoptant les mêmes notations que sur la Figure 1.16, et en utilisant les indices 1 et 2 pour différencier la tête utilisée à l'enregistrement et l'auditeur final, l'objectif est d'approximer le signal BIR_2 qui aurait été capturé aux oreilles de l'auditeur par :

$$BIR_2 \simeq BIR_1 * REF_2/REF_1$$

Evidemment, si les deux têtes sont identiques, les égalisations champ libre ou champ diffus garantissent toutes deux la meilleure fidélité pour toutes les incidences. En revanche, quand les têtes ne sont pas les mêmes, il n'est pas possible de reproduire toutes les composantes avec la même qualité : l'égalisation

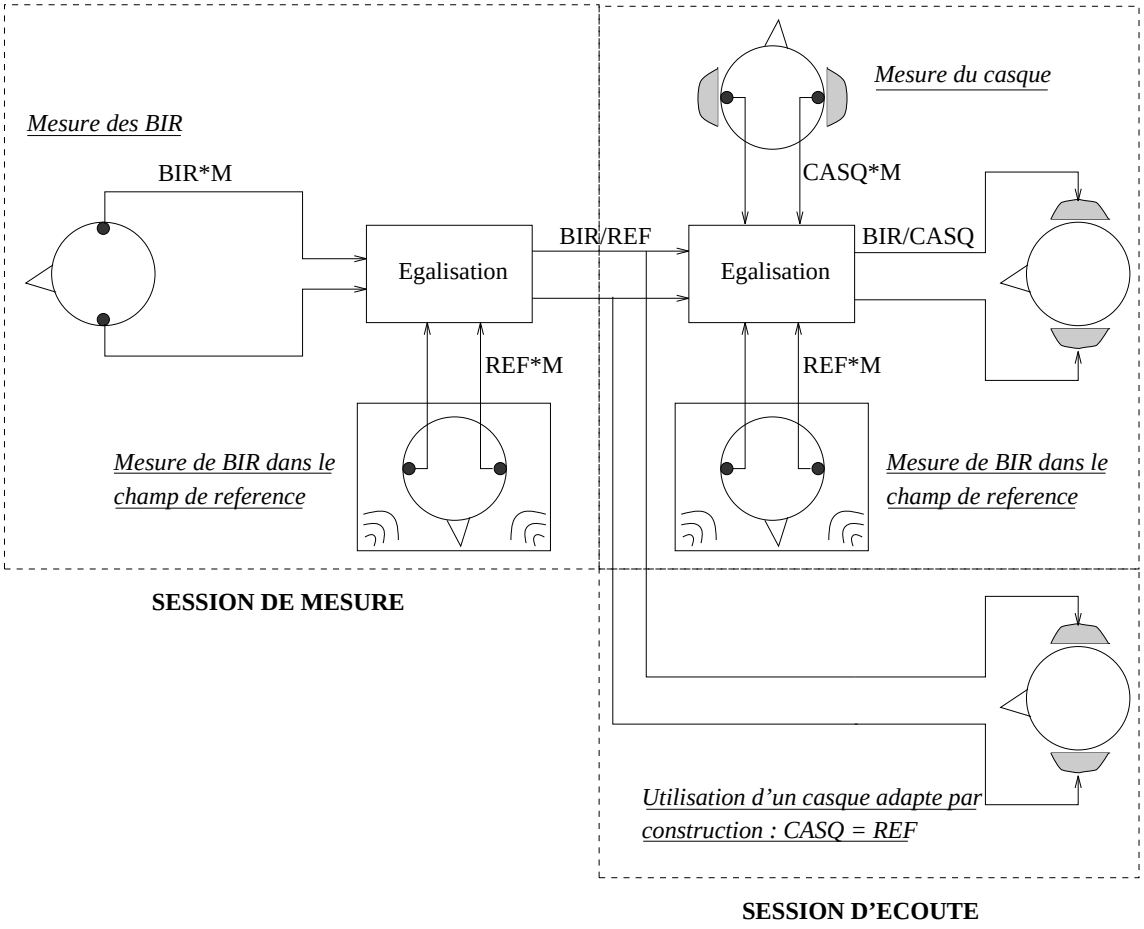


FIG. 1.16 – Egalisation découplée. Les systèmes d'enregistrement et de reproduction sont égalisés séparemment en référence au même champ sonore, "REF". Cela permet d'éliminer la contribution des transducteurs des canaux binauraux sans connaissance a priori du casque d'écoute utilisé à la reproduction. Les deux égalisations peuvent être réalisées lors de sessions différentes, avec un matériel différent (y compris la tête). Le choix du champ de référence, typiquement incidence frontale en champ libre, ou champ diffus, correspond à un standard pour les fabriquants de casque.

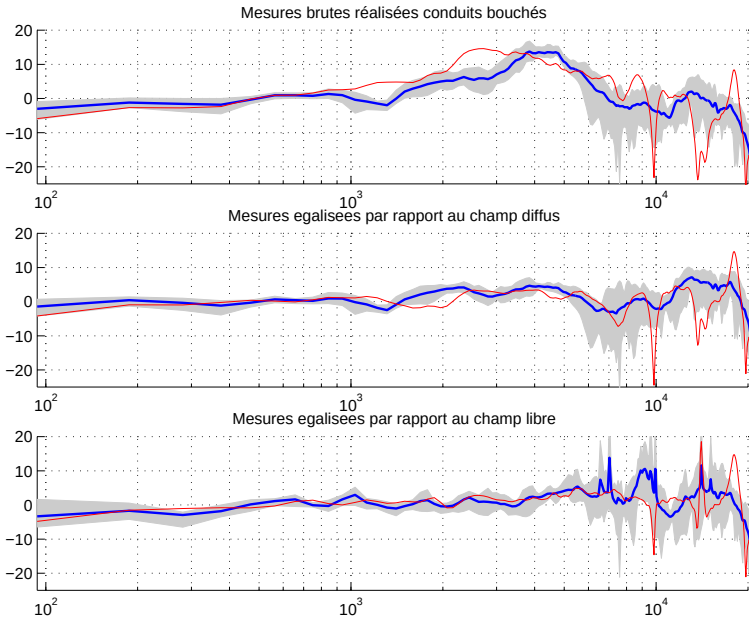


FIG. 1.17 – Spectres d'amplitude en dB de HRTF avec ou sans égalisation (azimut = 30°, élévation = 0°). Les réponses en fréquence ont été lissées au demi ton. Haut : mesures brutes ; Milieu : égalisation champ diffus ; Bas : égalisation champ libre. Trait fin : tête artificielle Head Acoustics HMSII (utilisée sans égalisation interne) ; grisé et ligne épaisse : enveloppe et moyenne linéaire de 7 sujets humains.

champ libre optimisera la reproduction des sons frontaux, tandis que l'égalisation champ diffus optimisera celle des sons diffus et de la réverbération. Les études rappelées en section 1.5.1.2 suggèrent que lorsque la scène sonore à reproduire contient un minimum de réverbération, la qualité perçue est meilleure pour une égalisation champ diffus.

Les Figures 1.17 et 1.18 montrent que l'égalisation par rapport au champ diffus est une technique robuste pour réduire les différences interindividuelles : les HRTF et fonctions de transfert de casque mesurés sur différents individus sont substituables jusqu'à 5kHz, après égalisation par rapport au champ diffus. Sur cet intervalle, on observe la réduction des différences entre têtes humaines. En outre, la tête artificielle HMSII, qui s'écarte nettement des autres avant égalisation, s'approche ensuite de la courbe moyenne.

Ces résultats viennent du fait que l'égalisation retire toutes les caractéristiques individuelles indépendantes de la direction, telles que la contribution des microphones binauraux, fixés dans le conduit à une position différente pour chaque sujet. Or le filtre d'égalisation champ libre contient lui-même des caractéristiques individuelles prononcées, telles que les résonances situées entre 8kHz et 10kHz, qui "contaminent" les spectres après égalisation. Au contraire, les caractéristiques individuelles contenues dans la HRTF diffuse sont très lissées et ne particularisent pas les réponses.

Ces observations nous permettent de déduire qu'entre 200Hz et 5kHz, on peut définir un filtre égalisateur standard pour les canaux binauraux et pour le casque. On a pu constater une cohérence satisfaisante entre la HRTF diffuse moyenne ainsi obtenue et la courbe normalisée (ISO389) reproduite dans [SB00]. L'égalisation par rapport au champ diffus permet donc de réduire le besoin de filtres mesurés sur l'auditeur final pour une simulation fidèle.

1.5.1.4 Réduction du coût d'implantation de la synthèse binaurale

On peut montrer que l'égalisation par rapport au champ diffus concentre l'énergie de la réponse temporelle des HRTF à phase minimale sur les premiers échantillons, ce qui présente un avantage pour une implantation FIR de la synthèse binaurale. Cette propriété est liée au fait que l'égalisation champ diffus

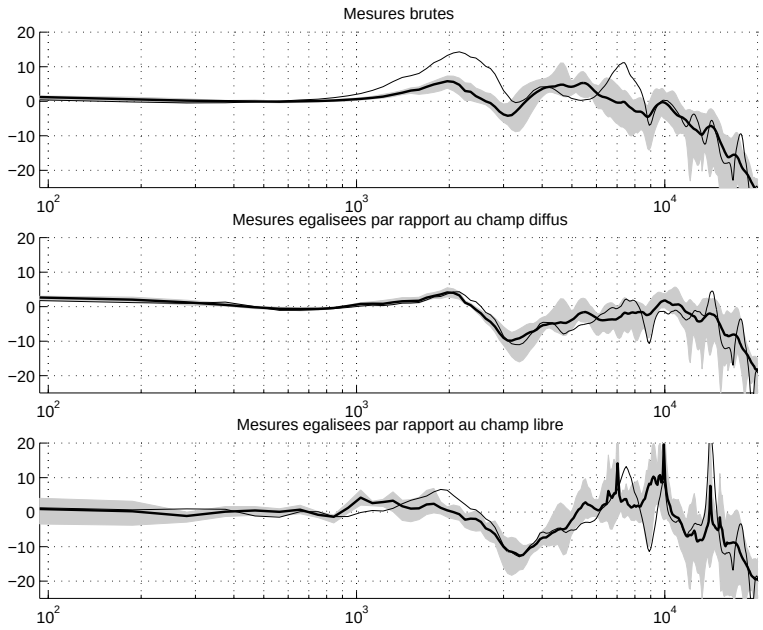


FIG. 1.18 – Spectres d'amplitude en dB du casque AKG K1000, avec ou sans égalisation. Les réponses en fréquence ont été lissées au demi ton. Haut : mesures brutes ; Milieu : égalisation champ diffus ; Bas : égalisation champ libre. Trait fin : tête artificielle Head Acoustics HMSII (utilisée sans égalisation interne) ; grisé et ligne épaisse : enveloppe et moyenne linéaire de 7 sujets humains.

σ_t (ms)	*HRTF brute*	*EQ champ libre*	*EQ champ diffus*
azimut 30°	0.13	0.11	0.09
azimut 60°	0.15	0.18	0.11
azimut 180°	0.17	0.22	0.11

TAB. 1.2 – Dispersion temporelle de l'énergie de la composante à phase minimale des HRTF, avec ou sans égalisation.

aplatit les spectres d'amplitude : aux fréquences correspondant à des caractéristiques indépendantes de la position, les spectres égalisés sont plats, et pour les autres fréquences, les caractéristiques dépendant de la direction sont lissées par la moyenne.
On peut établir une relation entre la dérivée du spectre d'amplitude $mag(f)$ et la dispersion d'énergie, σ_t, de la réponse temporelle de la composante à phase minimale ([Del95], pp.49 à 52) :

$$\sigma_t^2 = \frac{1}{4\pi}.\int_{-\infty}^{+\infty} \frac{d\left|mag(f)\right|^2}{df}.df$$

La remarque précédente implique que l'égalisation par rapport au champ diffus réduit l'aire délimitée par $\frac{d|mag(f)|}{df}$, donc par $\left(\frac{d|mag(f)|}{df}\right)^2$, ce qui conduit à de plus petites valeurs pour σ_t, donc à une distribution de l'énergie concentrée sur les premiers échantillons de la réponse temporelle des HRTF.

Ces résultats peuvent être observés en Figure 1.19. Le Tableau 1.2 recense les valeurs de la dispersion temporelle de l'énergie σ_t des HRTF avec ou sans égalisation. La formule adoptée pour le calcul de σ_t, associé à la réponse temporelle $x(t)$ dont la moitié de l'énergie est située de part et d'autre de l'instant t_0, est celle proposée par [Del95] :

$$\sigma_t^2 = \frac{\int_{-\infty}^{+\infty}(t-t_0)^2.\left|x(t)\right|^2.dt}{\int_{-\infty}^{+\infty}\left|x(t)\right|^2.dt}$$

Les données du tableau 1.2 montrent la concentration systématique de l'énergie pour l'égalisation par rapport au champ diffus. En revanche, si l'égalisation par rapport au champ libre réduit la dispersion

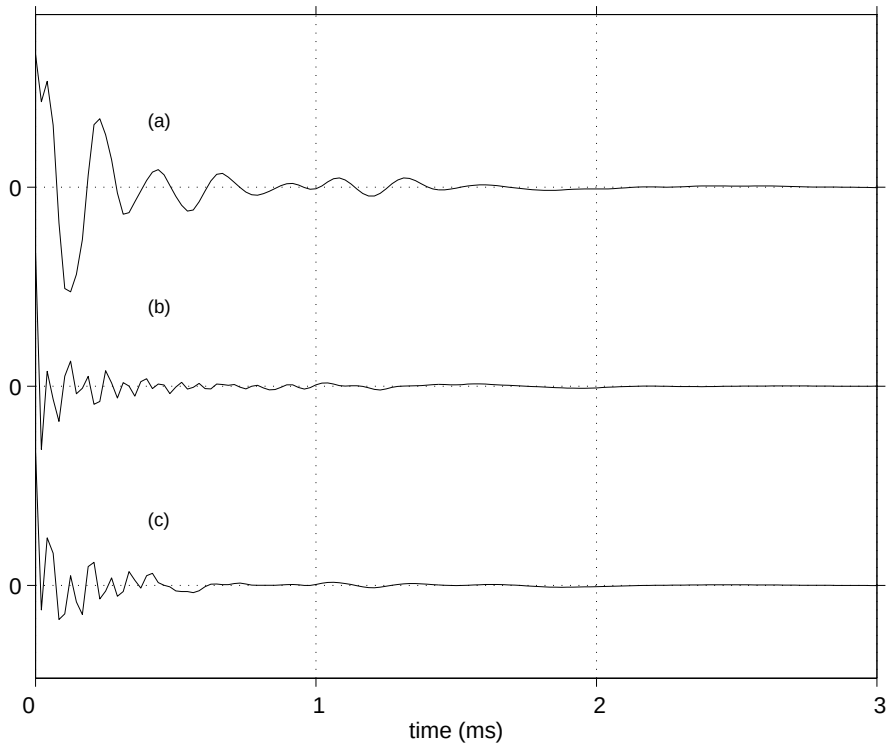


FIG. 1.19 – Réponse impulsionnelle à phase minimale de la HRTF (azimut = 30°, élévation = 0°) : réponse brute (a), réponse après égalisation par rapport au champ libre (b) et par rapport au champ diffus (c). Les réponses sont représentées après mise à plat des spectres d'amplitudes associés pour $f < 200Hz$ et $f > 16kHz$, et lissage au demi-ton.

temporelle de l'énergie pour l'azimut 30°, ces performances se dégradent d'autant plus rapidement que l'incidence de la HRTF modélisée est éloignée de l'incidence de référence (0° d'élévation, 0° d'azimut dans l'exemple choisi).

Par conséquent, l'égalisation par rapport au champ diffus permet de réduire l'ordre de la modélisation de la composante à phase minimale des HRTF sous forme de filtre FIR.

1.5.2 Méthodes d'estimation de la HRTF diffuse

Un champ sonore diffus est composé d'ondes planes décorrélées provenant d'incidences uniformément distribuées dans l'espace. C'est le cas de la réverbération tardive d'une salle. Le filtrage réalisé par le torse, la tête et les pavillons d'un auditeur sur un champ diffus incident peut être caractérisé par différentes méthodes, fournissant une estimation de son spectre d'amplitude. On se propose dans cette partie de comparer trois d'entre elles.

1.5.2.1 Excitation stationnaire en champ diffus (chambre réverbérante)

Cette méthode s'appuie sur la génération physique d'un champ sonore diffus, et consiste à reccueillir les réponses à une excitation en régime stationnaire dans une chambre réverbérante ([Kuh79] et [The86]).

Des mesures ont été réalisées sur la tête artificielle Head Acoustics HMSII et sur un individu, dans la chambre réverbérante de l'E.N.S.T. d'un volume d'environ $6m^3$ pour un temps de réverbération proche de 2 secondes. Une mesure de référence a également été réalisée avec un microphone omnidirectionnel placé au lieu du centre de la tête. Cette dernière réponse caractérise la contribution de la chaîne de mesure, qui doit être compensée dans les mesures binaurales. La HRTF diffuse est obtenue par l'estimation de la densité spectrale de puissance des signaux, par exemple par la méthode du périodogramme moyenné de Welch, puis par le rapport entre chaque "spectre binaural" et le "spectre de référence".

Afin de tester la robustesse de l'estimation, les mesures ont été répétées sous différentes conditions, illustrées en Figure 1.20. Jusqu'à 12kHz, la variance de ces estimations est supérieure à l'écart moyen entre les spectres d'amplitude, qui est inférieur à 2dB lorsque l'on déplace les transducteurs, et reste inférieure à 3dB lorsque l'on fait varier le signal d'excitation (bruit blanc, séquence MLS,...). Etant donnée la variance de l'estimation, on peut considérer que toutes les courbes de la Figure 1.20 sont identiques.

Cette méthode permet donc d'obtenir une estimation de la HRTF associée au champ sonore régnant dans la chambre réverbérante, par un protocole simple et rapide. Toutefois, ce champ sonore n'est diffus que pour une bande fréquentielle limitée, dépendant du volume de la salle. On ne peut obtenir une estimation précise en basses fréquences que si les modes se recouvrent ([Kuh79]). Pour la salle considérée, la fréquence de Schröder vaut approximativement 1kHz si on l'estime par $f_{schroeder}(kHz) = 2.\sqrt{\frac{Tr}{V}}$ ([SK62]). Cela signifie donc qu'en deçà de cette limite, il n'y a pas d'énergie à toutes les fréquences, et qu'aux fréquences où il y a de l'énergie, le champ sonore n'est pas diffus.
Une solution consisterait à utiliser une salle de plus grand volume, mais alors, l'absorption de l'air réduirait la limite en hautes fréquences de l'intervalles sur lequel le champ est uniforme ([GVM96]). Ce compromis à trouver sur le volume de la salle a conduit Kuhn a utiliser deux chambres réverbérantes afin d'obtenir une estimation de la HRTF diffuse sur toute la bande de fréquence utile : une première salle de $425m^3$ fournit l'estimation pour les fréquences inférieures à 6kHz, tandis que dans la seconde salle, de $49m^3$, l'estimation est valide pour l'intervalle [1kHz-10kHz] (voir [Kuh79]).

Pour s'affranchir de ces limites, certains auteurs ont étudié la possibilité de créer un champ diffus par superposition d'ondes planes en champ libre. Veit et al. ont ainsi proposé un montage en chambre anéchoïque de huit enceintes placées aux sommets d'un cube, qui, lorqu'elles sont alimentées par des générateurs de bruit à bande étroite décorrelés, créent un champ sonore diffus pour un large intervalle fréquentiel ([VS87]). Toutefois, cette approche ne constitue par une solution suffisante, puisqu'une mesure de l'intensité acoustique $\overrightarrow{I}$ régnant dans le cube d'enceintes montre que le champ n'est diffus ($\overrightarrow{I} = 0$) que dans des zones spatialement très limitées, a priori trop étroites pour y placer une tête artificielle.

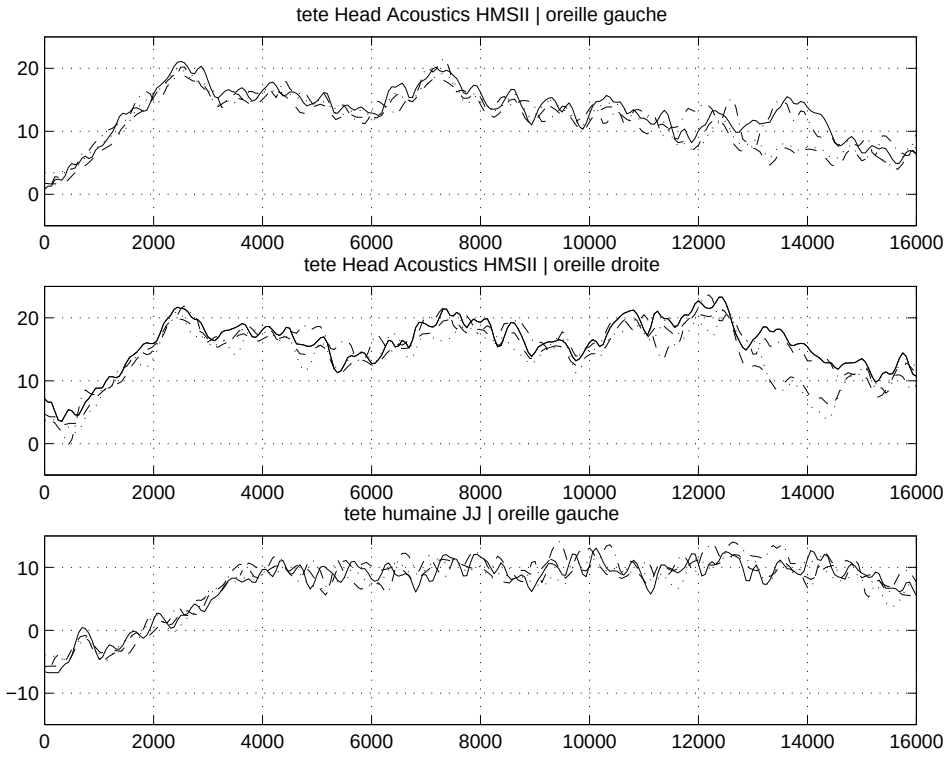


FIG. 1.20 – HRTF diffuse mesurée par une excitation en régime stationnaire en chambre réverbérante pour 3 oreilles et différentes conditions. Les spectres d'amplitude ont été lissés par bandes constantes de 250Hz. Trait continu : fréquence d'échantillonnage de 32kHz avec un bruit blanc comme signal d'excitation ; Trait interrompu : fréquence d'échantillonnage de 48kHz avec des séquences MLS comme signal d'excitation ; Trait pointillé : fréquence d'échantillonnage de 32kHz avec un bruit blanc comme signal d'excitation, pour une autre position des transducteurs.

1.5.2.2 Moyenne énergétique des HRTF (chambre anéchoïque)

Cette seconde méthode ne s'appuie pas sur le champ sonore diffus recréé dans une salle. Elle vise à reproduire la fonction de transfert en champ diffus par superposition des transformations subies par des ondes planes en champ libre "venant de partout" (voir par exemple [Mol92], [MJHS95], ou [JLW95]). Cette méthode s'appuie sur plusieurs caractéristiques du champ diffus.

La superposition est d'abord justifiée par le caractère décorrélé des ondes planes le composant. La pression p arrivant aux tympans d'un auditeur placé dans un champ diffus généré par un signal source de spectre S s'écrit :

$$\begin{aligned} p &= S \times HRTF_{diff} \\ &= S \times \sum_i HRTF_i \\ &= S \times \sum_i mag_i . \exp^{j.\varphi_i} \end{aligned}$$

où mag_i désigne le spectre d'amplitude de la HRTF mesurée pour la direction d'incidence i, et où φ_i en désigne la phase.
De plus, comme les ondes planes considérées proviennent d'incidences aléatoires et uniformément distribuées, on a :

$$\begin{aligned} \langle p^2 \rangle &= S^2 \times \sum_{i,j} mag_i . mag_j . \langle \exp j.(\varphi_i + \varphi_j) \rangle \\ &= S^2 \times \sum_i mag_i^2 \end{aligned}$$

où $\langle . \rangle$ désigne l'espérance statistique.

Cette méthode consiste à estimer la HRTF diffuse $\frac{p}{S}$ par $\sqrt{\sum_i mag_i^2}$, i.e. par la moyenne en énergie des spectres d'amplitude des HRTF mesurées en champ libre, le plus souvent en chambre anéchoïque, pour une distribution d'incidences uniformes aussi complète que le temps de mesure imparti et la souplesse d'installation du matériel de mesure le permettent.

Une estimation exacte de la HRTF diffuse par cette méthode requiert une distribution continue d'incidences. Toutefois, les HRTF ne peuvent être mesurées que pour un nombre discret d'incidences. L'influence du choix de la distribution a été étudiée en comparant la HRTF diffuse du mannequin KEMAR obtenue pour :

- élévations allant de -40° à 90° par pas de 10°, pour une distribution d'azimuts garantissant un angle solide constant, soit 710 incidences,
- élévations 0° et 30° pour un échantillonnage en azimut régulier de période 15°, soit 48 incidences.

Deux distributions intermédiaires sont aussi présentées en Figure 1.21. Elles correspondent à des échantillonnages réguliers en azimut et en élévation, ce qui peut présenter des avantages pratiques :

- hémisphère supérieur échantillonné par pas de 10° en élévation de 0° à 90°, pour un échantillonnage en azimut régulier de période 15°, soit 450 incidences,
- hémisphère supérieur échantillonné par pas de 10° en élévation de 0° à 40° pour un échantillonnage en azimut régulier de période 15°, plus une mesure réalisée à l'élévation 90°, soit 121 incidences.

Comme les directions mesurées pour ces deux dernières distributions n'échantillonnent pas la sphère uniformément, une pondération proportionnelle à l'angle solide "représenté" devrait théoriquement être appliquée. Cette correction n'a pas été réalisée. Les trois dernières distributions ont requis des HRTF non mesurées qui ont été reconstruites par interpolation linéaire sur les réponses impulsionnelles à phase minimale (cf chapitre 2).

La Figure 1.21 présente le rapport en amplitude entre la HRTF diffuse du KEMAR obtenue pour les 710 incidences, et le spectre d'amplitude obtenu pour des distributions partielles. Ce rapport peut être considéré comme le spectre d'amplitude d'un filtre de correction permettant de corriger la HRTF diffuse

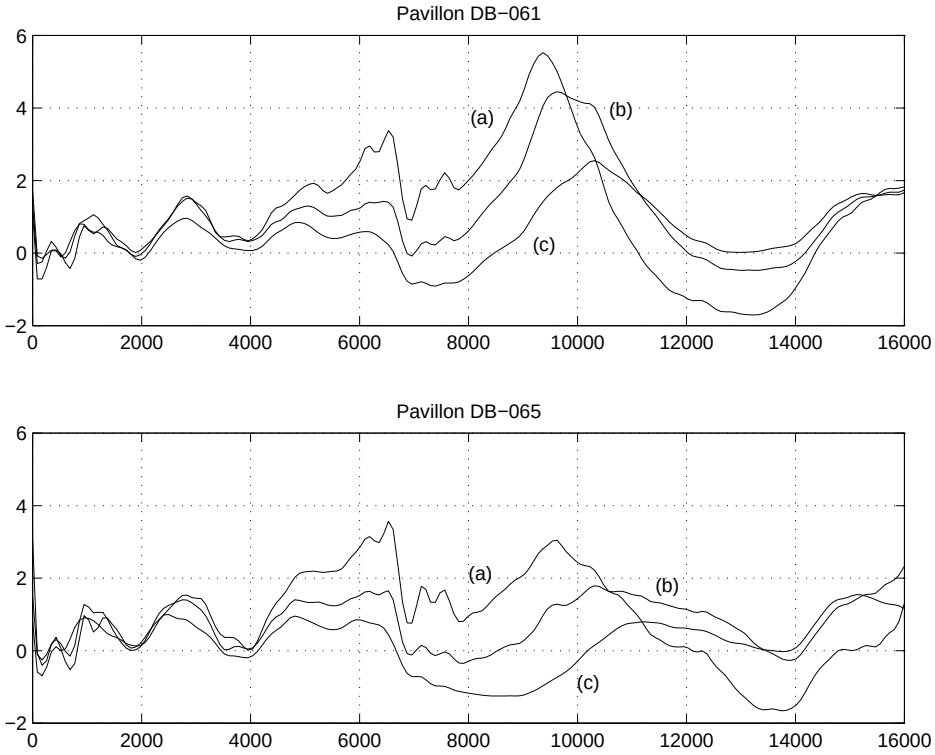


FIG. 1.21 – HRTF diffuse estimée par moyenne énergétique des HRTF de la tête artificielle KEMAR (deux pavillons), pour plusieurs échantillonnages. Rapport entre l'estimation obtenue avec les 710 positions et celle obtenue avec une distribution partielle : distance entre le spectre d'amplitude obtenu pour une distribution d'incidences complète et celui obtenu pour la distribution d'incidences partielle (a) : élévations 0° et 30° ; (b) : élévations 0°, 10°, 20°, 30°, 40° et 90° ; (c) : élévations de 0° à 90° par pas de 10°.

"partielle" pour la rapprocher de la HRTF diffuse "complète". On observe que l'erreur commise reste inférieure à 2dB jusqu'à environ 5kHz. Jusqu'à cette fréquence donc, seules 48 HRTF sont à mesurer pour avoir une estimation satisfaisante de la HRTF diffuse. Au delà de cette limite, l'erreur peut atteindre 5dB, notamment autour de 9kHz pour le pavillon DB-061.
Les filtres de correction pour les deux pavillons sont très semblables jusqu'à 8kHz, fréquence correspondant à une longueur d'onde d'environ 4cm. Ce résultat peut être justifié par le fait qu'en dessous de cette fréquence, il n'y a pas d'influence de la forme particulière du pavillon.

Cette méthode d'estimation de la HRTF diffuse s'appuie sur un nombre important de mesures, que l'on ne peut réduire qu'en sacrifiant la robustesse de l'estimation. Outre la longueur de sa mise en oeuvre, le protocole de mesure permettant le positionnement de la source autour du sujet peut également être encombrant et coûteux à mettre en place.

1.5.2.3 Excitation transitoire en champ diffus (salle "typique")

Nous proposons une nouvelle méthode d'estimation de la HRTF diffuse, s'appuyant sur l'excitation transitoire d'une salle typique, i.e. d'usage courant sans caractéristiques acoustiques notoires, par un signal de contenu fréquentiel "suffisamment riche" (impulsion, coup de pistolet, etc ...). La HRTF diffuse est obtenue par analyse spectrale de la réverbération tardive du signal recueilli. Cette nouvelle méthode s'affranchit des contraintes des deux précédentes méthodes, notamment l'utilisation d'une salle rare et la durée de la séance de mesures.

Comme pour la méthode en chambre réverbérante, une mesure de référence est nécessaire afin de compenser la contribution de la chaîne de mesure. Jot et al. ont proposé une méthode d'estimation éliminant toute contrainte sur la position du microphone omnidirectionnel de référence. Elle repose sur l'analyse de

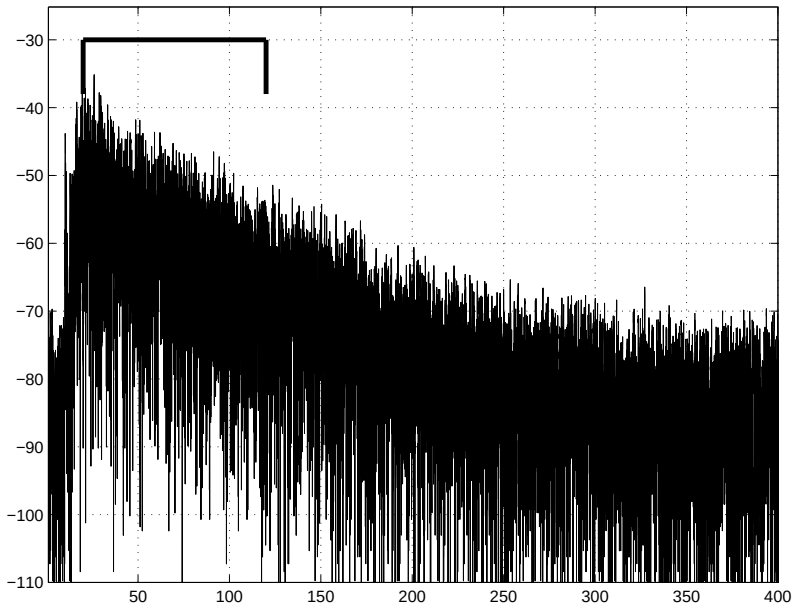


FIG. 1.22 – Echogramme d'une réponse impulsionnelle mesurée dans une salle typique, avec la fenêtre temporelle utilisée pour l'estimation de la HRTF diffuse.

l'enveloppe temps-fréquence des réponses impulsionnelles, dont l'isotemps $t = 0$ constitue le spectre de puissance en champ diffus du récepteur, mais nous ne l'utilisons pas dans le cadre de cette étude ([Van95], [JCV96]).

L'estimation de la HRTF diffuse étant réalisée sur la partie tardive des réponses, l'intervalle temporel d'étude doit respecter plusieurs conditions :

- la borne inférieure doit se situer après l'arrivée du son direct et des premières réflexions,
- la borne supérieure doit se situer avant que le signal soit noyé dans le bruit en hautes fréquences,
- l'intervalle doit être suffisamment long pour permettre une résolution fréquentielle satisfaisante.
  La salle considérée pour la Figure 1.22 est un bureau d'environ $35m^3$ dont le temps de réverbération moyen se situe autour de 300ms. La taille de fenêtre retenue est de 100ms. La borne inférieure est celle correspondant au premier intervalle sur lequel l'estimation est "stable"[3] , ne se situant ni sur les premières réflexions, ni en fin de réverbération tardive et a fortiori dans le bruit.

La Figure 1.23 présente les estimations obtenues pour plusieurs positions des transducteurs dans la salle, et montrent une robustesse satisfaisante de la méthode jusqu'à 10kHz. Cette méthode d'estimation de la HRTF diffuse présente deux atoûts majeurs :

- **une économie de moyen** : pour égaliser HRTF ou casques d'écoute, seules trois mesures sont nécessaires (2 mesures binaurales et 1 mesure avec un microphone omnidirectionnel), procédure équivalente à celle utilisant une chambre réverbérante. Il demeure que cette troisième méthode ne fait appel à aucune salle spécifique. La salle de mesure doit simplement ne pas être totalement "sourde" mais bien posséder une réverbération tardive. Pour l'égalisation de réponses impulsionnelles binaurales, la HRTF diffuse peut être estimée directement à partir de ces réponses et ne requiert en supplément qu'une seule mesure, celle du microphone de référence.
- **une sécurité maximale** : dans le cadre des deux premières méthodes, le risque de voir les microphones binauraux bouger à l'intérieur du conduit auditif subsiste, par exemple entre l'enregistrement binaural et les mesures en chambre réverbérante, ou au cours de la longue session de mesure de HRTF. Dans le cas particulier de l'égalisation de réponses impulsionnelles binaurales, cette éventualité disparaît puisqu'aucune mesure binaurale supplémentaire n'est requise.

[3]La stabilité est reliée à l'erreur L_∞ entre les estimations effectuées sur l'intervalle [t t+D] et [t t+2D] (t :borne inférieure de l'intervalle, et D : durée de l'intervalle, 100ms dans notre exemple.

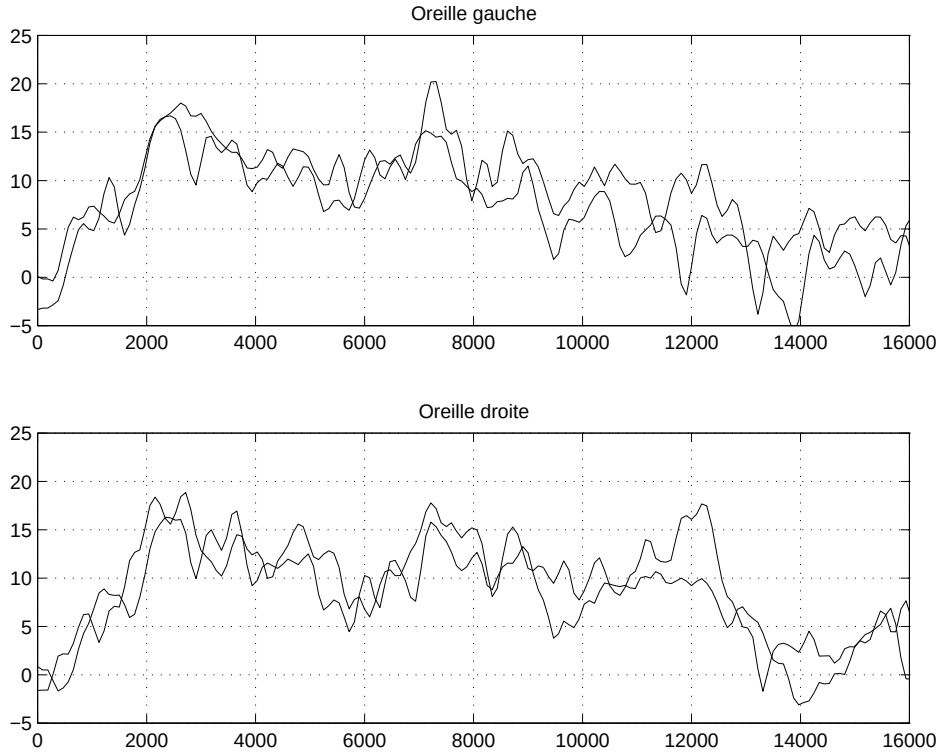


FIG. 1.23 – HRTF diffuse de la tête artificielle Head-Acoustics HMSII, estimée à partir de réponses impulsionnelles d'une salle typique pour deux positions du microphone de référence. Les spectres d'amplitude ont été lissés par bandes constantes de 250Hz.

1.5.3 Comparaison des trois méthodes

Pour comparer les 3 méthodes, il est nécessaire de corriger la fonction de transfert du microphone de référence. En effet, le microphone B&K utilisé (série 4145) a des caractéristiques différentes en champ libre et en champ diffus. La courbe de correction que nous appliquons aux mesures en champ diffus est représentée en Figure 1.24 d'après les données constructeur.
Les Figures 1.25 et 1.26 proposent une comparaison des trois méthodes d'estimation de la HRTF diffuse, pour la tête artificielle HMSII, et pour un individu. Pour chaque méthode, l'enveloppe des estimations obtenues pour différentes positions est représentée en grisé, la courbe en trait continu est obtenue par moyenne énergétique des estimations. Les trois méthodes donnent des résultats équivalents jusqu'à 12kHz.

1.5.4 Conclusion sur l'égalisation

Nous avons rappelé et complété les avantages d'une égalisation par rapport au champ diffus des enregistrements binauraux, des HRTF et des casques d'écoute :

- compatibilité des enregistrements binauraux avec l'écoute sur haut-parleurs ; compatibilité des enregistrements stéréo conventionnels avec l'écoute au casque,
- robustesse de la qualité de la simulation binaurale vis à vis des caractéristiques individuelles des auditeurs,
- optimisation de l'implémentation de la synthèse binaurale.

Ces résultats préconisent l'utilisation d'une équalisation par rapport au champ diffus au détriment de l'égalisation par rapport au champ libre, choix alternatif couramment adopté par les constructeurs de casques et de tête artificielle. Toutefois, l'égalisation par rapport au champ diffus ne peut s'imposer comme un standard si son estimation est malaisée, par exemple si elle requiert une salle rare ou une séance de

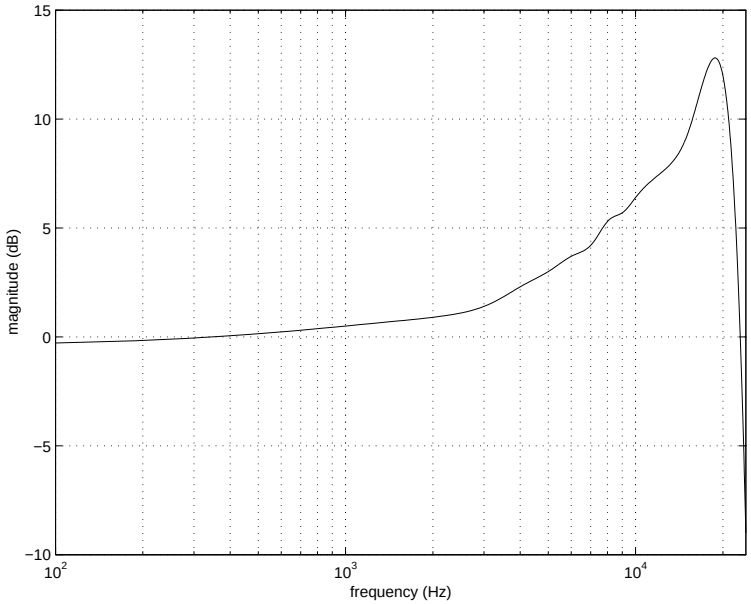


FIG. 1.24 – Rapport entre la réponse en champ libre et la réponse en champ diffus du microphone omnidirectionnel de référence, utilisé pour toutes les mesures binaurales (B&K modèle n°4145).

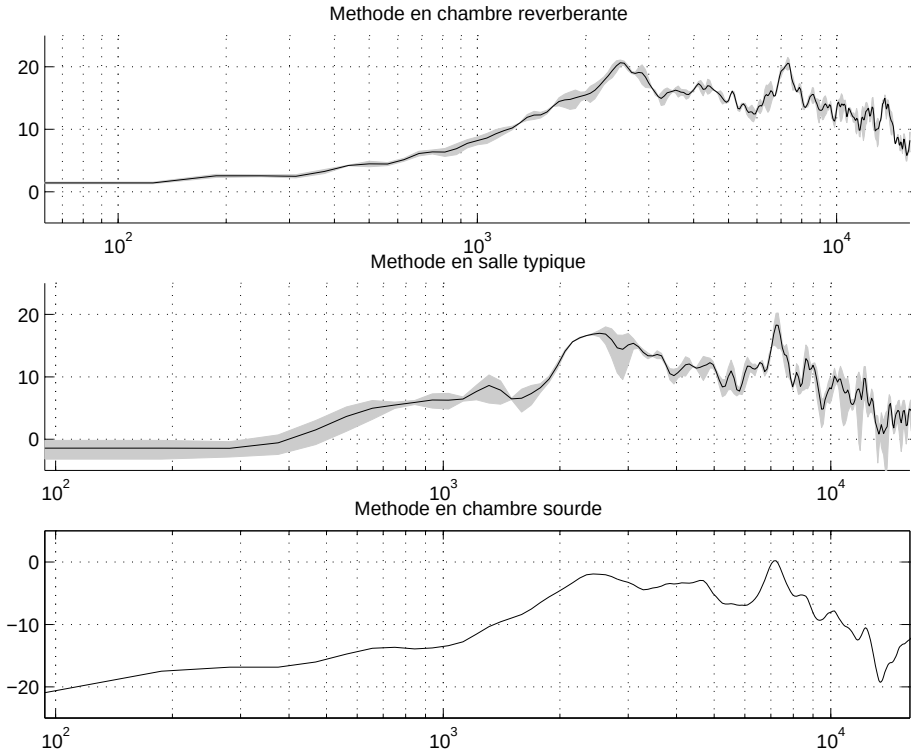


FIG. 1.25 – HRTF diffuse de la tête artificielle Head-Acoustics HMSII (oreille gauche), estimée par trois méthodes. Haut : excitation stationnaire en champ diffus (moyenne de 2 mesures) ; milieu : excitation transitoire dans une salle typique (moyenne de 2 mesures) ; bas : moyenne en énergie de mesures en champ libre.

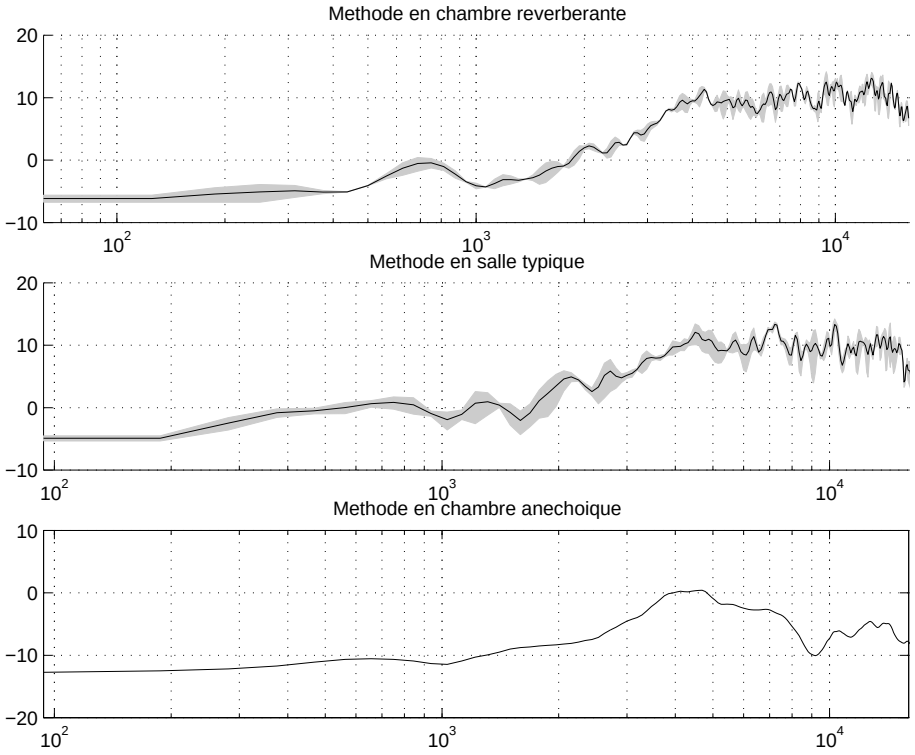


FIG. 1.26 – HRTF diffuse d'une tête humaine (oreille gauche), estimée par trois méthodes. Haut : excitation stationnaire en champ diffus ; milieu : excitation transitoire dans une salle typique ; bas : moyenne en énergie de mesures en champ libre. L'écart en hautes fréquences peut s'expliquer par le fait que les mesures n'ont pas été réalisée lors de la même session, ce qui suppose que les microphones ont été ôtés puis replacés dans les conduits et ont donc une contribution différente.

mesure longue, contraintes imposées par les deux méthodes d'estimation le plus souvent rencontrées. Une nouvelle méthode d'estimation a donc été proposée, permettant d'obtenir le filtre d'égalisation à partir de trois réponses impulsionnelles mesurées dans une salle d'usage courant.

Dans cette étude, nous nous sommes contentés de considérer les égalisations associées à deux champs sonores de référence : le champ libre et le champ diffus. Toutefois, certains auteurs ont proposé d'autres structures de champ, privilégiant un jeu de plusieurs incidences, dans l'objectif d'améliorer l'image frontale, fragile dans la simulation binaurale ([Bla97] p362, [MJHS95]). Les performances de l'égalisation par rapport à ces champs sonores "pondérés" mériteraient d'être étudiées plus avant.

1.6 Conclusion

Nous avons présenté et comparé plusieurs protocoles de mesure de HRTF, notamment celui que nous avons utilisé à l'Ircam. Ce dernier nous a permis d'entreprendre une vaste campagne de mesure, comprenant 24 têtes humaines et la tête artificielle Head Acoustics HMSII. Toutefois, des raisons pratiques nous ont contraint à limiter le nombre de positions mesurées. A contrario, les mesures gracieusement prêtées par Ralph Algazi et al., de l'université UC Davis, contiennent 1025 positions pour 17 têtes humaines.

Les mesures brutes ainsi réalisées contiennent plus d'information qu'il ne nous est nécessaire de reproduire. Afin de la concentrer sur les caractéristiques dépendant de la direction, nous avons appliqué deux traitements aux mesures brutes.
Nous avons tout d'abord comparé plusieurs méthodes d'estimation du retard interaural. L'une de ces méthodes, proposée par Jot, s'appuie sur l'approximation linéaire de la différence interaurale d'excès de phase des HRTF. Elle a fait l'objet d'optimisations dans le cadre de cette thèse, et constitue la méthode la plus fiable parmis celles envisagées. Par ailleurs, nous avons montré son équivalence avec une méthode d'intercorrélation en sous-bandes des composantes passe-tout des HRTF.
Afin d'éliminer les caractéristiques fréquentielles indépendante de la direction, nous avons étudié le processus d'égalisation des spectres d'amplitude des HRTF. Nous avons proposé de nouveaux arguments en faveur de l'égalisation par rapport au champ diffus. En outre, nous avons présenté une technique originale de mesure de la "HRTF diffuse", facilitant la mise en oeuvre de cette égalisation pour les HRTF mais aussi pour des réponses impulsionnelles binaurales de salles ou des enregistrements binauraux.

On aboutit ainsi à la représentation de chaque HRTF sous forme d'un retard et d'un spectre d'amplitude, modèle proposé par Mehrgardt et Mellert en 1977. La synthèse binaurale consiste alors à modeler un son "vierge" avec l'une de ces empreintes spatiales. Toutefois, nous ne devons sous-estimer les limites de notre approche, purement acoustique : la localisation est un mécanisme complexe faisant intervenir d'autres paramètres, dont la présence est éventuellement plus importante que la précision des indices acoustiques. C'est le cas par exemple des indices visuels, beaucoup moins ambigus et plus robustes ([SS80], [PW72]). Une amélioration peut également être apportée par la sur-impression d'autres indices acoustiques provenant de réflexions précoces, qui appuient les premiers et font converger la sensation vers l'incidence cible ([RH85], [Har83]). On peut également mentionner l'importance d'indices cognitiques, expliquant par exemple que l'on localise plus facilement un son familier qu'un son inconnu ([Bla97], [Col62]), ou les indices acoustiques dynamiques, provenant de la mise en mouvement de la source sonore ou de l'auditeur ([PM81], [WK99]).

Chapitre 2

Implantation bicanale de la synthèse binaurale

2.1 Introduction

Le chapitre précédent a permis de montrer que les HRTF peuvent être décomposées sous forme d'un retard, représentant le retard monaural, en série avec un filtre à phase minimale, pour chaque position et pour chaque oreille. Nous envisageons à présent l'implantation "directe" de cette décomposition, illustrée en Figure 2.1 : le signal monophonique incident est divisé en deux canaux, chacun d'entre eux subissant le filtrage en cascade des deux composantes des HRTF. Cette implantation est à ce titre qualifiée de "bicanale". Elle se distingue de l'implantation multicanale, que nous décrivons en chapitre 3.

Dans une première sous-partie, nous comparons plusieurs approches décrites dans la littérature pour réaliser un modèle paramétrique de la composante à phase minimale des HRTF, et pour implanter le retard sous forme fractionnaire.
Cette structure d'implantation étant établie, il est possible de simuler de façon statique les positions mesurées et stockées en mémoire sur le processeur. Il peut être utile de minimiser cette charge en mémoire en trouvant un moyen approprié de recréer les HRTF non stockées par interpolation en temps-réel. Nous étudions ainsi la problématique de l'interpolation locale des HRTF, qui rejoint plus généralement le problème de l'interpolation entre deux filtres numériques, interpolation entre formants par exemple. Cette section constitue le principal apport personnel de ce chapitre.
Enfin, la transition d'une position à une autre s'accompagne d'artefacts audibles liés au changement brutal des coefficients des filtres. Nous rappelons les principales méthodes permettant de gérer cette "commutation".

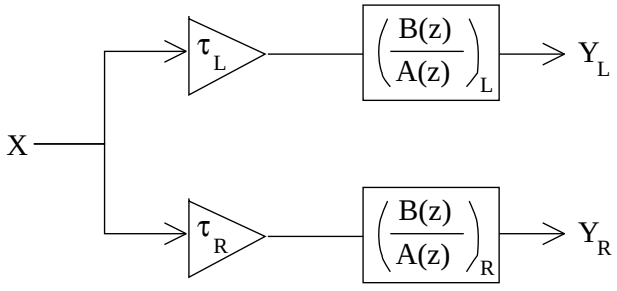


FIG. 2.1 – Schéma général d'implantation de la synthèse binaurale bicanale.

2.2 Implantation de la synthèse binaurale

2.2.1 Modélisation de la composante à phase minimale des HRTF

La composante à phase minimale des HRTF peut a priori être modélisée sous forme FIR ou sous forme IIR. Les propriétés de ces deux familles de modèles sont différentes : un modèle FIR essaie de répliquer aux mieux les échantillons de la réponse impulsionnelle et sacrifie les derniers échantillons en fonction de l'ordre choisi, tandis qu'un modèle IIR s'efforce de reproduire les résonances du spectre d'amplitude, sans contrainte sur la structure temporelle fine. Comme nous le voyons en chapitre 5, différentes interprétations physiques des caractéristiques des HRTF peuvent légitimer l'une ou l'autre de ces approches : le modèle de l'oreille comme réflecteur, introduit par Batteau ([Bat67]), privilégierait une implantation FIR, tandis que l'approche de Shaw et Teranishi, qui représentent les HRTF comme un réseau de résonateurs en parallèle ([ST68]), se prête naturellement à une implantation IIR.
Nous examinons des solutions pour les deux familles de modèles, et comparons leurs performances, pour un coût d'implantation fixé. Les méthodes envisagées sont issues d'études générales sur la modélisation de filtres numériques (Smith [Smi83], Steiglitz et Mc Bride [SMB65], Friedlander et Porat [FP84], Levi, et Belinczynski et al. [BKC92]), appliquées à la modélisation des HRTF, notamment par Jot et al. ([Jot92], [Pon92], [JLW95], [Tou96]), par Blommer et Wakefield ([BW94], [Blo96], [BW97]), et par Huopaniemi et al. ([HK97], [MHVK97], [HZK99], [Huo99]).

2.2.2 Choix d'une norme spectrale

2.2.2.1 Paramètres spectraux perceptivement pertinents

Dans [Smi83], Smith isole quelques caractéristiques de la perception auditive qu'il utilise comme critère de qualité pour la synthèse du spectre du violon. Il définit ainsi une mesure d'erreur de modélisation "perceptivement pertinente". Nous étendons ses critères à la modélisation des HRTF à phase minimale :

1. l'oreille est sensible à une échelle de fréquence non linéaire, approchant la division en bandes critiques de la membrane basilaire,
2. la dissimilarité entre deux spectres est corrélée avec l'écart en dB de leur amplitude.

Nous prenons en compte le premier de ces critères pour la spécification du modèle ainsi que pour l'observation des performances de la modélisation. En revanche, pour des raisons pratiques d'implantation, le deuxième critère n'interviendra que pour l'observation des performances.

2.2.2.2 Adaptation de la résolution fréquentielle du spectre d'amplitude pour la spécification du modèle

Smith définit un traitement préalable à la modélisation, permettant d'adaptater la représentation spectrale de départ à une échelle non-linéaire des fréquences, plus proche de la perception. L'application de ce pré-traitement aux HRTF réalise aujourd'hui un consensus dans la littérature. Il se divise en deux étapes :

1. Lissage du spectre de puissance des HRTF, mag^2, remplaçant chaque point du spectre par une moyenne de ses voisins contenus dans la même fraction d'octave.
   Nous utilisons une fenêtre de Hann dont la largeur est proportionnelle à la fréquence du point considéré, et appliquons typiquement un lissage par demi-ton. Lisser davantage conduirait à gommer des caractéristiques spectrales (résonances) qu'il semble important de reproduire.
2. Transformée conforme de l'axe des fréquences ("warping"), pour réequilibrer l'importance respective de chaque bande de fréquences et notamment donner une plus grande importance aux basses fréquences. Cette déformation de l'axe des fréquences est réalisée par un ré-échantillonnage du

spectre lissé. Si ω désigne les fréquences linéaires correspondant à l'échantillonnage de départ, et r le coefficient de warping, alors les fréquences warpées ω_r sont données par :

$$\omega_r = \arctan\left[\frac{(1-r^2).\sin(\omega)}{2r+(1+r^2).\cos(\omega)}\right]$$

Typiquement, l'approximation d'une échelle fréquentielle proche des Barks, Smith choisit $r = 0.67$. On peut noter que si le lissage des variations en amplitude n'est pas préalablement effectué, l'interpolation linéaire mise en jeu par le warping est susceptible de fortes imprécisions, liées à la concentration des échantillons en hautes-fréquences.

Le spectre d'amplitude, ainsi lissé et ré-échantillonné constitue le "gabarit" fourni à la routine de modélisation. L'effort de modélisation est ainsi réparti de façon perceptivement homogène sur les fréquences. Plusieurs algorithmes sont comparés en section 2.2.1. Si l'on note mag_Q^r le spectre lissé et warpé, sa phase mph est obtenue grâce à la transformée de Hilbert, et le modèle conduit à l'approximation :

$$mag_Q^r.e^{j.mph} \simeq \frac{B(z)}{A(z)}$$

Finalement, les coefficients du modèle approchant le spectre à phase minimale de départ sont obtenus après une transformée conforme inverse de l'axe des fréquences (substitution z par $\frac{z-r}{1-r.z}$). On note qu'étant donné que numérateur et dénominateur sont de degré 1, cette transformation n'augmente pas l'ordre du filtre.

2.2.2.3 Application de normes usuelles pour mesurer l'écart inter-spectres

Les techniques de modélisation de filtres cherchent à minimiser l'écart entre le spectre de départ, H_1, et le spectre modélisé, H_2, au sens d'une certaine norme. La modélisation est appliquée après lissage et warping de H_1. Nous noterons le spectre résultant $H_1^{Q,r}$ et la réponse impulsionnelle associée $h_1^{Q,r}$. Par analogie, $H_2^{Q,r}$ et $h_2^{Q,r}$ désignerons les représentations du spectre modélisé, bien qu'aucun lissage ne lui soit réellement appliqué. Comme le rappelle Smith, et plus récemment Huopaniemi et Smith ([HS97]), les méthodes de modélisation les plus courantes s'appuient sur trois normes usuelles :

1. Norme L_2, ou norme des moindres carrés :

$$E_2 = \sum_{n=0}^{\infty} |h_1^{Q,r}(n) - h_2^{Q,r}(n)|^2 = \frac{1}{N}\sum_{n=1}^{N} |H_1^{Q,r}(n) - H_2^{Q,r}(n)|^2$$

   Parmi les méthodes standards implantées dans Matlab et utilisées par plusieurs auteurs pour la modélisation IIR des HRTF, nous pouvons mentionner *prony.m* et *stmcb.m*, qui travaillent dans le domaine temporel, et *invfreqz.m* et *yulewalk.m*, qui travaillent dans le domaine fréquentiel. C'est également l'erreur minimisée par la modélisation FIR que nous utilisons, s'appuyant sur la troncature de la réponse impulsionnelle avec une fenêtre rectangulaire ([SH96]).

2. Norme L_∞, ou norme de Chebychev :

$$E_\infty = Max_n|H_1^{Q,r}(n) - H_2^{Q,r}(n)|$$

   Smith a proposé une méthode de pondération permettant de prendre en compte la résolution en log-magnitude de la perception. Toutefois, nous pensons que cette norme est mal adaptée à notre étude puisque le maximum sera systématiquement détecté en hautes-fréquences, où se produisent de forts écarts non structurels, mais bien plutôt liés à des résonances "conjoncturelles" mesurées avec peu de reproductibilité. Nous n'étudierons donc aucune méthode minimisant cette norme (par exemple *cremez.m*).

3. Norme L_H, ou norme de Hankel :

$$E_H = Max_i(\sigma_i)$$

   Cette norme est fréquemment utilisées pour l'étude des systèmes linéaires $F(z)$, et est évaluée comme la plus grande valeur propre de la matrice de Hankel associée à $F(z)$. Cette matrice est formée à

partir d'une représentation d'état du système, et ses valeurs propres mesurent la contribution de chaque état à l'évolution du système. En diagonalisant cette matrice, on décorrèle les différents états et on peut les classer par ordre d'importance. La minimisation de cette norme est notamment utilisée pour la problématique de "réduction du modèle" (balanced model reduction, ou BMR). Elle a été appliquée par Mackenzie et al. pour approximer la réponse impulsionnelle des HRTF sous forme d'un filtre IIR. Dans ce cas, on a : $F(z) = H_1^{Q,r}(z) - H_2^{Q,r}(z)$. Comme le montrent Beliczynski et al. , on peut obtenir un majorant de E_H sans avoir à calculer la matrice de Hankel associée à $F(z)$. Seule la matrice de Hankel associée à $H_1^{Q,r}$, calculée pour l'obtention du modèle, est utilisée, et l'on obtient :

$$E_H < 2. \sum_{k>k_0} \sigma_k$$

si k_0 est l'ordre choisi pour le modèle.
Il est difficile de donner une signification à la norme de Hankel en terme d'erreur spectrale. Toutefois, on peut la relier à la norme L2, et minimiser EH conduit à minimiser E2. Comme le démontrent Kung et Lin dans [KL81], on a :

$$E_2 < E_H < E_\infty$$

On constate que les normes que nous envisageons pour la modélisation ne tiennent pas compte de tous les paramètres de la perception rappelés en section 2.2.2.1, et notamment la sensibilité aux log-magnitudes plutôt qu'aux amplitudes linéaires. Pour évaluer l'efficacité de chaque méthode, nous utiliserons donc une autre mesure d'erreur, plus proches de la perception. Nous la définissons par :

$$E = \left| 20. \log_{10} \left| \frac{H_2^Q}{H_1^Q} \right| \right|$$

Cette erreur sera représentée sur une échelle logarithmique des fréquences. Lorsque nous souhaitons évaluer l'erreur moyenne sur un ensemble de m positions, nous évaluerons :

$$\overline{E} = \sqrt{\frac{1}{m}. \sum_\theta w(\theta). \left[20. \log_{10} \left| \frac{H_2^Q(\theta)}{H_1^Q(\theta)} \right| \right]^2}$$

Le poids $w(\theta)$ permet de pondérer position par l'angle solide qu'elle représente sur la grille d'échantillonnage spatial choisi pour les mesures. Une mesure plus rigoureusement calquée sur la perception a été proposée par Pulkki et al. ([PKH99]), et s'appuie sur la différence d'intensité perceptive (en phones) perçue dans chaque bande critique (ERB).

2.2.3 Comparaison des méthodes de modélisation

Pour une comparaison plus complète et systématique des méthodes de modélisation appliquées aux HRTF, le lecteur pourra se reporter aux travaux de Huopaniemi et al. ([HZK99], [HS97], [Huo99]) ou de Touzé ([Tou96]). Nous nous sommes concentrés sur la comparaison de certaines méthodes de modélisation IIR : *yulewalk.m*, utilisée pour les HRTF du *Spat˜* au début de la thèse, *stmcb.m*, semblant montrer de meilleures performances ([Tou96]), et la méthode utilisant une troncature de la représentation d'état équilibrée du système (BMR), la plus récemment proposée ([MHVK97]). L'implantation FIR pouvant sembler comme une alternative intéressante, notamment pour les problèmes d'interpolation que nous abordons en section suivante, nous avons également évalué les performances d'un modèle FIR tronquant la réponse impulsionnelle des HRTF avec une fenêtre rectangulaire.
Les modèles sont comparés, avec et sans prétraitement, pour trois ordres, à 44.1kHz :

- ordre 16, qui correspond à l'ordre utilisé dans le *Spat˜* au début de cette étude,
- ordre 12, qui est l'ordre utilisé aujourd'hui,
- ordre 20, utilisé pour le test perceptif décrit à la section suivante.

Pour le modèle FIR, l'ordre équivalent au même coût de calcul est double. Les résultats sont présentés en Figures 2.2, 2.3, 2.4.

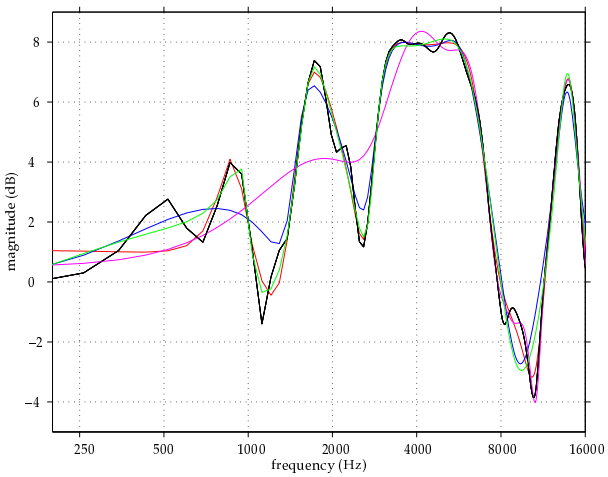


FIG. 2.2 – Modèle d'ordre 12 de la HRTF ipsilatérale (40°,0°) : yulewalk (bleu), stmcb (rouge), BMR (vert), FIR (magenta). Lissage au demi-ton, r=0.48.

Comme nous le voyions en chapitre 1, l'ensemble des traitements appliqués aux HRTF mesurées, et notamment l'égalisation par rapport au champ diffus, conduisent à des réponses impulsionnelles de moins de 1.5 ms. Typiquement, pour une implantation FIR à 44.1kHz, les HRTF peuvent être modélisées avec une faible erreur à l'aide de filtres d'ordre voisin de 60. La troncature à 24 coefficients est donc trop précoce (Figure 2.2). En outre, l'erreur de modélisation affecte principalement les basses fréquences, qui durent plus longtemps dans la réponse impulsionnelle et souffrent donc de la troncature. L'un des points faibles de l'approche FIR tient ainsi en l'erreur BF qui lui est inhérente, alors même que c'est la zone où nous souhaitons renforcer l'effort de modélisation. C'est ce que seule une modélisation IIR permet.

Comme on l'observe en Figure 2.2, la méthode yulewalk.m est nettement moins performante que les méthodes *stmcb.m* et BMR. Ce résultat est confirmé par les erreurs moyennes représentées en Figure 2.4, d'après lesquelles :

$$\textit{yulewalk.m} < \text{BMR} < \textit{stmcb.m}.$$

En abandonnant dans le *Spat˜* la modélisation IIR d'ordre 16 avec *yulewalk.m* pour un modèle d'ordre 12 avec *stmcb.m*, l'erreur a augmenté en moyenne sur toutes les fréquences, mais semble s'expliquer par des accidents localisés en fréquence (autour de 2kHz). Pour la grande majorité des fréquences, la réduction de l'ordre s'est accompagnée d'une réduction de l'erreur, du fait de l'efficacité de *stmcb.m*. En outre, on observe que pour un ordre 20, retenu pour la validation perceptive de la section 4.2, l'erreur est quasi-nulle, inférieure à 0.7dB.

La Figure 2.3 met en évidence l'effet du warping : les basses fréquences sont mieux modélisées, au point qu' un modèle non-warpé d'ordre 20 n'atteint pas les performances d'un modèle warpé d'ordre 12.

Chacune de ces méthodes de modélisation définit un filtre d'ordre n, supposé pair, dont la transformée en z notée $B(z)/A(z)$ est de la forme :

Afin de garantir la stabilité des filtres intermédiaires obtenus au cours de la transition d'une direction simulée à une autre, ce filtre est implanté sous forme de cellules d'ordre 2 mises en cascade.

2.2.4 Implantation du retard interaural sous forme de retard fractionnaire

Pour implanter l'ITD sous forme de filtre numérique, une solution consisterait à l'arrondir à un nombre entier d'échantillons à l'implanter avec une ligne à retard z^{-N}. Mais négliger la partie fractionnaire du retard peut entraîner une dégradation la qualité de localisation : pour une fréquence d'échantillonnage de 44.1kHz, une erreur d'un échantillon sur la valeur d'ITD peut entraîner une délocalisée de 2 à 3°. On

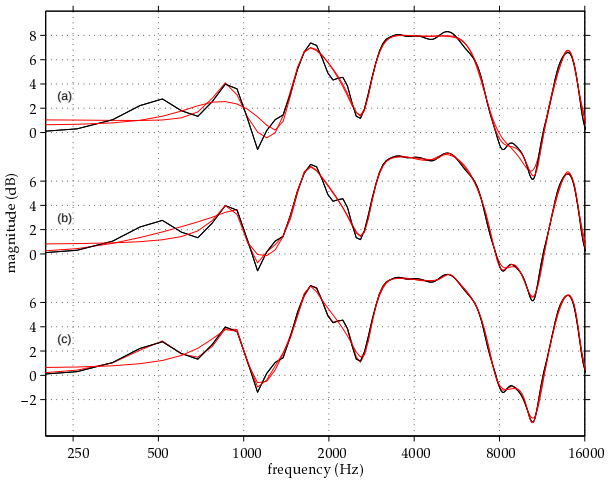


FIG. 2.3 – Effet du warping sur la modélisation (méthode *stmcb.m*) : (a) ordre 12, (b) ordre 16, (c) ordre 20. Lissage au demi-ton.

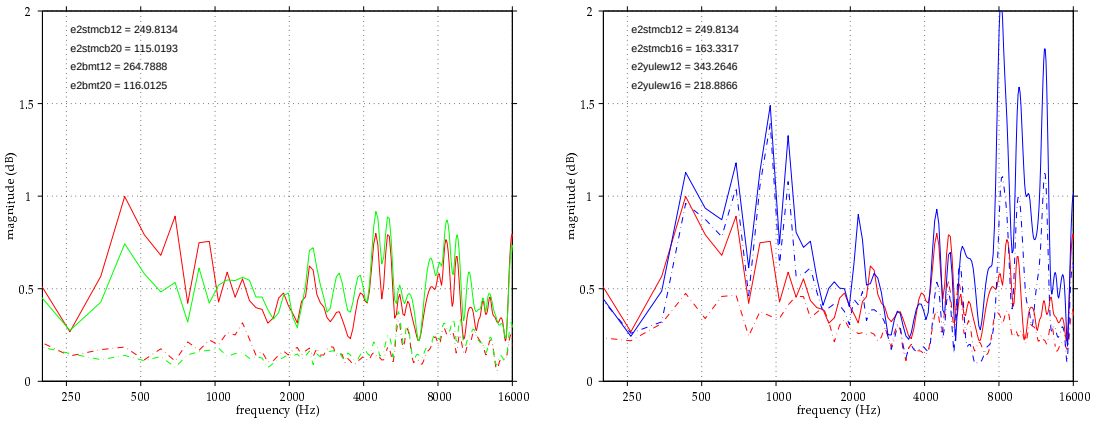


FIG. 2.4 – Erreur de modélisation moyenne pour les positions ipsilatérales du plan horizontal. Lissage au demi-ton et r=0.48.

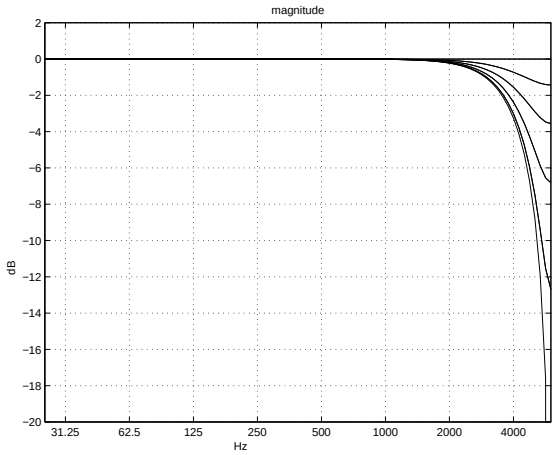


FIG. 2.5 – Réalisation d'un retard fractionnaire sous forme d'une filtre FIR d'ordre 3 : réponse en amplitude pour différentes valeurs de retard.

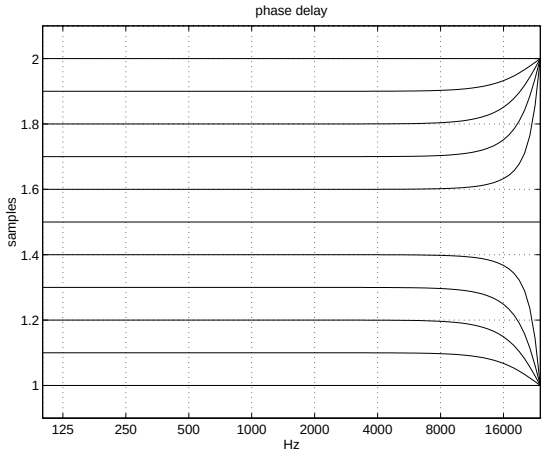


FIG. 2.6 – Réalisation d'un retard fractionnaire sous forme d'une filtre FIR d'ordre 3 : réponse en phase pour différentes valeurs de retard.

préfère donc implanter l'ITD sous forme de retard fractionnaire. Deux principales méthodes fournissent une expression analytique des coefficients du filtre en fonction du retard (voir [Val95]) et suggèrent :

- une implantation sous forme de filtre FIR (formule de Lagrange), dont les caractéristiques en amplitude et en phase sont données en Figures 2.5 et 2.6. On observe que les caractéristiques attendues ne sont obtenues que jusqu'à 8kHz environ.

- une implantation sous forme de filtre IIR passe-tout (formule de Thiran). L'implantation d'un filtre passe-tout garantit une amplitude unitaire, et permet ainsi de se soustraire pour partie aux artefacts de la méthode prédente. La réponse en phase obtenue présente par ailleurs des performances semblables à l'approche FIR. Afin de comparer des filtres de même coût, ces caractéristiques sont données en Figure 2.7 pour un filtre IIR d'ordre 2.

Pour le *Spat˜*, c'est l'objet Max *vd˜* qui réalise le retard variable, sous forme d'un filtre FIR. Chaque fois qu'une nouvelle valeur de retard parvient, les coefficients du filtre sont mis à jour à chaque période d'échantillonnage avec une progression linéaire vers la valeur cible qui n'est atteinte qu'au bout de 30ms.

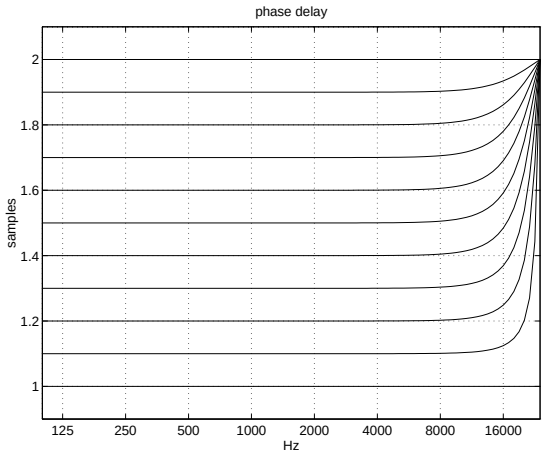


FIG. 2.7 – Réalisation d'un retard fractionnaire sous forme d'une filtre IIR d'ordre 2 : réponse en phase pour différentes valeurs de retard.

2.3 Interpolation locale des HRTF

Simuler toutes les directions d'incidence discernables par l'auditeur par la synthèse binaurale requiert a priori la mesure d'un nombre important de HRTF. Une première estimation de la résolution spatiale du système auditif, ou halo sonore, peut être obtenue à partir des données psycho-expérimentales fournies par Blauert ([Bla97]). Une extrapolation rapide de ces résultats indique qu'une base de données complète de HRTF nécessiterait 300 mesures par oreille (pour la calotte supérieure). On peut considérer le temps nécessaire à l'acquisition d'une telle base de donnés prohibitif. L'interpolation des HRTF peut donc avoir pour premier objectif la reconstruction de "mesures manquantes" à partir d'une base de données ne contenant que quelques HRTF. Cette interpolation peut être réalisée en temps différé.

En revanche, une contrainte temps réel est introduite lorsque l'objectif de l'interpolation est (aussi) de limiter le volume de données à stocker en mémoire. En effet, pour une représentation utilisant 60 coefficients codés sur 24 bits, une base de données de 300 HRTF par oreille occupe environ 80 kOctets, ce qui n'est pas négligeable en comparaison avec la mémoire disponible sur un DSP Motorola 56000 (2×64kOctets). L'interpolation permet alors de ne stocker en mémoire qu'un petit nombre de HRTF, en recréant à partir de ces dernières les positions intermédiaires, en temps réel.

On peut distinguer deux méthodes d'interpolation : l'interpolation **locale** s'appuie sur les mesures voisines pour reconstruire une donnée manquante, tandis que l'interpolation **globale** s'appuie sur l'ensemble des mesures. Nous nous concentrons sur la première approche, qui conserve un caractère général pouvant s'appliquer à tout problème d'interpolation entre un filtre de départ et un filtre cible.

2.3.1 Paramètres de l'interpolation

Pour étudier l'interpolation, nous proposons un formalisme, schématisé en Figure 2.8 : le filtrage est réalisée avec une structure définie par des paramètres d'implantation, l'interpolation est réalisée sur des paramètres de contrôle, et est validée par la continuité des paramètres d'observation. Nous reprenons en le prolongeant le formalisme introduit dans nos rapports antérieurs ([Lar94a], [LJ97]).
Pour une interpolation en temps différé, l'interpolation est effectuée directement sur les paramètres d'observation. Pour une interpolation en temps réel, en revanche, cette stratégie est trop coûteuse en calculs. On cherche donc les paramètres de contrôle réalisant le meilleur compromis : leur variation doit être aussi corrélées que possible à celle des paramètres d'observation ; le coût de l'interpolation doit être aussi faible que possible. Ce coût comprend le "coût d'implantation", i.e. requis pour le calcul d'un nouvel échantillon de signal, et le "coût de conversion", nécessaire pour la conversion des paramètres d'interpolation en paramètres d'implantation.

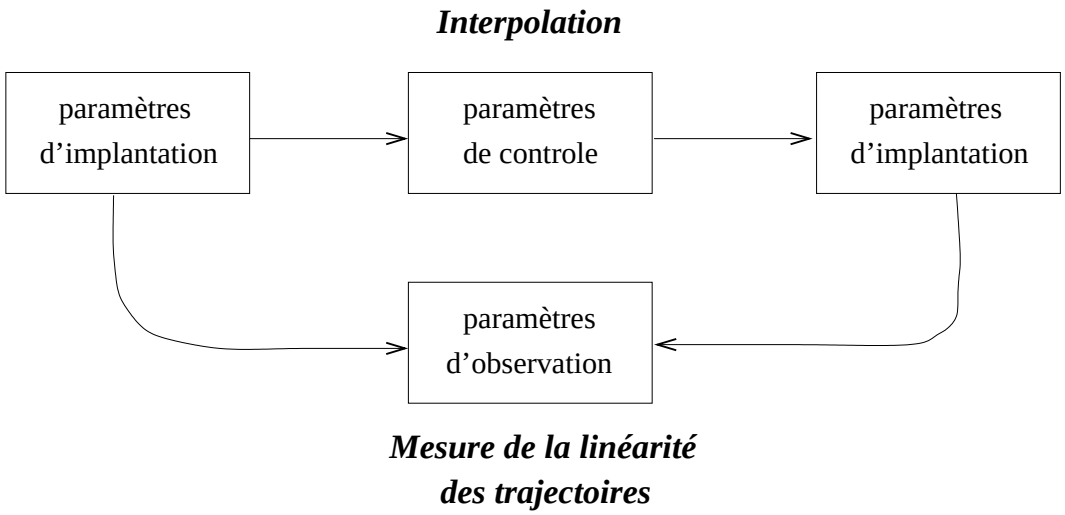


FIG. 2.8 – Paramètres définis pour l'optimisation d'une interpolation en temps-réel.

2.3.1.1 Paramètres d'implantation

Les paramètres d'implantation désignent les coefficients de la structure réalisant le filtrage. Les HRTF peuvent être modélisées sous forme de filtre à réponse impulsionnelle infinie, ou à réponse impulsionnelle infinie. Des structures adaptées à ces deux cas sont étudiées. Dans le cas IIR, on distinguera :

- la structure transverse directe (ou "développées"), par exemple avec la forme II transposée recommandée par Dattoro pour les applications audio ([Dat88]), et représentée en Figure 2.9.
- la structure transverse factorisée en cellules d'ordre 2,
- la structure en treillis dans laquelle le filtre est caractérisé par les coefficients de réflexion (Figure 2.10).

2.3.1.2 Paramètres d'observation

Les paramètres d'observation doivent fournir une évaluation de la qualité de l'interpolation, "objectivant" les défauts ou qualités perceptibles à l'écoute. Une évolution linéaire des paramètres d'observation s'interprète alors donc comme une évolution linéaire de la sensation auditive.

Comme nous le rappelions en section 2.2.1, la reconstruction du spectre de puissance en dB est un critère de qualité souvent adopté pour des signaux monauraux. Nous proposons donc les spectres d'amplitude en dB des HRTF, notés mag, comme paramètres d'observation. La validation la plus immédiate consiste à comparer les spectres interpolés avec les spectres mesurés aux mêmes positions. Cette validation est contrainte par la précision angulaire du protocole de mesure qui se limite dans notre cas à une résolution de 5° dans le plan horizontal. On mesure l'écart entre les HRTF interpolées et les HRTF mesurées par l'erreur aux moindres carrés définie en section 2.2.1.
Avec cet échantillonnage, on constate sur la Figure 2.11 que les spectres mesurés ont des variations très linéaires avec l'azimut, jusqu'à environ 8kHz. C'est d'autant plus vrai sur les spectres après modélisation IIR, représentés en Figure 2.12. Une bonne méthode d'interpolation sera donc celle qui, comme la mesure, génère des variations linéaires des spectres d'amplitude des HRTF en dB.

Pour quantifier cette linéarité, on peut utiliser une mesure de "sensibilité spectrale", $\frac{\partial S}{\partial c}$, exprimant la capacité à varier des spectres d'amplitude pour de petites variations du paramètre de contrôle c défini plus loin :

$$\frac{\partial S}{\partial c} = \frac{1}{\Delta c}. \int_{-f_e/2}^{f_e/2} 20 * \left| \log \frac{mag(c,f)}{mag(c+\Delta c, f)} \right| .df \tag{2.1}$$

Cette expression reprend la proposition de Viswanathan dans [VM75]. Son étude, concentrée sur la prédiction linéaire du signal de parole pour une implantation treillis, a pour objectif d'évaluer la robustesse

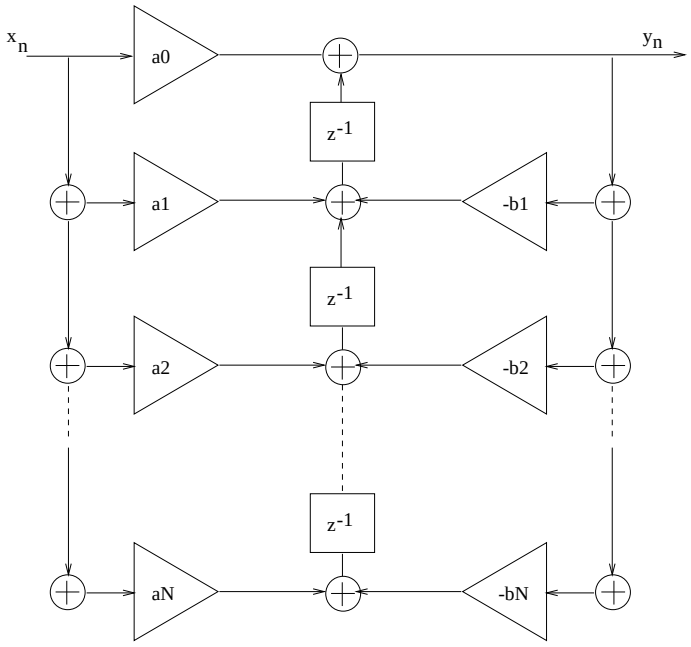


FIG. 2.9 – Firtre IIR d'ordre N implanté dans la forme directe transposée de la structure transverse, recommandée pour les applications audio ([Dat88]).

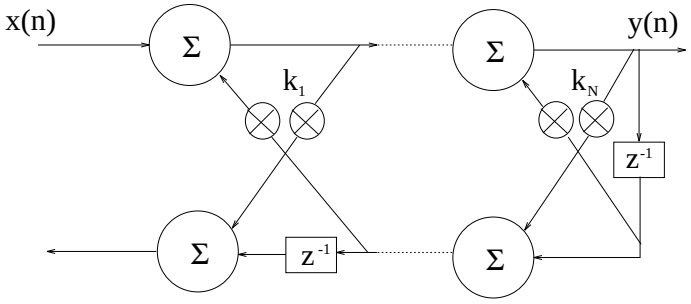


FIG. 2.10 – Filtre récursif tout pôle d'ordre N implanté en treillis.

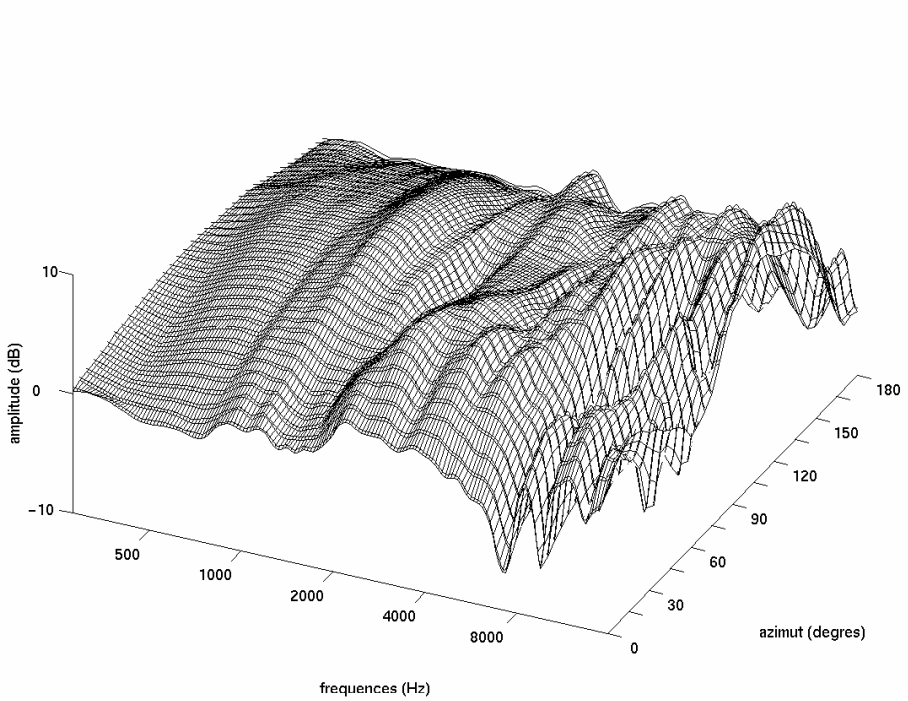


FIG. 2.11 – Spectres d'amplitude mesurés sur la tête artificielle HMS II, par pas de 5° dans le plan horizontal (oreille ipsilatérale : de 0° à 180°, lissage au demi ton).

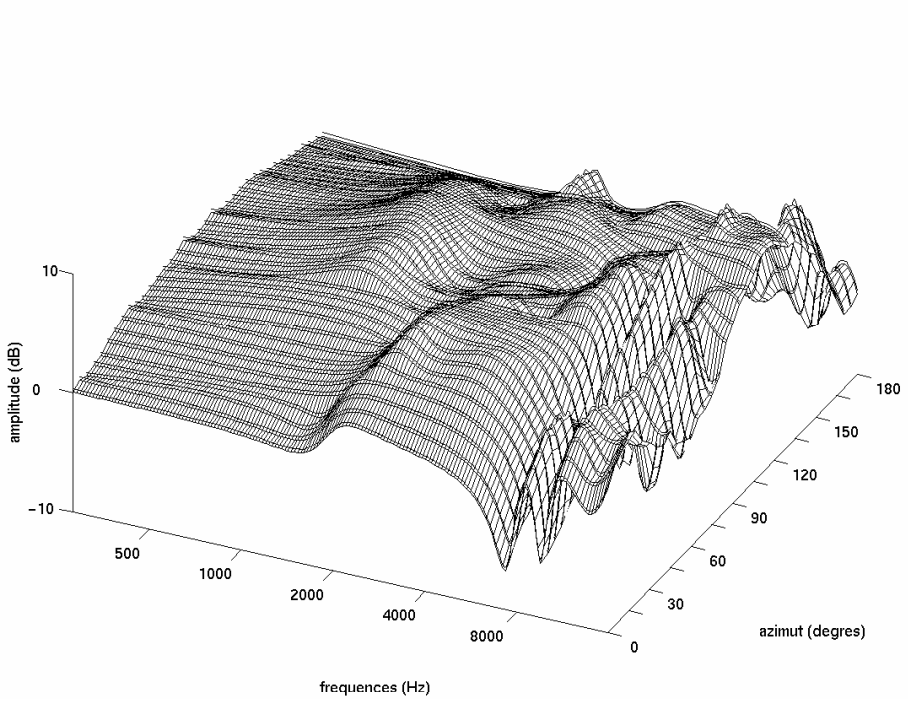


FIG. 2.12 – Spectres d'amplitude de la Figure 2.11 après modélisation IIR d'ordre 12 (lissage au demi-ton puis warping de 0.5).

des filtres "prédits" à la quantification scalaire des coefficients de réflexion. La sensibilité est evaluée pour plusieurs jeux de coefficients initiaux, en faisant varier chaque paramètre un à un. Puis, une courbe de sensibilité "globale" est obtenue par moyenne. Les paramètres présentent de bonnes propriétés pour la quantification si la sensibilité obtenue est plate, i.e. si la déformation spectrale induite par la quantification d'un coefficient est indépendante de sa valeur initiale, et n'est fonction que de l'amplitude de l'erreur de quantification Δc.
La problématique de l'interpolation se distingue de celle de la quantification scalaire puisque tous les paramètres évoluent simultanément, d'une quantité Δc différente, et que l'évolution de chaque paramètre peut interagir avec celle des autres pour constituer les spectres observés. L'étude de Gardner, menée sur le codage par quantification vectorielle, semble à ce titre mieux appropriée ([Gar94]). Il étend la notion de sensibilité spectrale de Viswanathan en définissant une "matrice de sensibilité". Les coefficients de la diagonale de cette matrice correspond à la sensibilité obtenue dans le cas de la quantification scalaire, tandis que les termes hors de la diagonale exprime l'interaction des paramètres 2 à 2. Toutefois, la formulation de cette matrice étant complexe, et n'étant fournie par Gardner que dans certains cas particuliers, nous illustrerons les performances de techniques d'interpolation à l'aide l'expression 2.1. L'objectif sera de chercher la sensibilité la plus plate afin de garantir une déformation spectrale indépendante de la valeur initiale de nos paramètres.

Autres paramètres d'observation empruntés au domaine de la parole : la trajectoire des pôles, qui, dans ce cadre, permet de suivre l'évolution des formants et revient donc à faire un "zoom" sur le critère précédent de continuité des spectres, en le limitant à certaines caractéristiques fréquentielles (e.g. [HKKM73]). Pour notre étude également, il semble important de respecter l'évolution douce des caractéristiques prononcées des HRTF telles que pics (résonances) et vallées (anti-résonances). En effet, comme nous le rappelions au chapitre 1, elles peuvent être attribuées aux phénomènes physiques de réflexion ou de résonance au contact du corps de l'auditeur, et constituent à ce titre des indices de localisation importants. Les pôles mesurés ou interpolés constituent donc de bons candidats comme paramètres d'observation, dont il s'agit de vérifier que l'évolution continue et douce en fonction de la position. Cette mesure s'applique au cas de l'interpolation d'un modèle IIR des spectres d'amplitude.
Nous avons représenté en Figure 2.13 les pôles donnés par une modélisation IIR de HRTF mesurées tous les 5° dans le plan horizontal. Sur certains intervalles, la représentation suggère des trajectoires : par exemple entre 200° et 300°, autour de la fréquence $5\pi/6$. Il est toutefois difficile de définir des trajectoires sur l'ensemble du plan horizontal, ce qui limite le recours à un critère de continuité pour la validation de l'interpolation. Ce problème est renforcé par l'existence de pôles réels, induits par le warping réalisé lors de la modélisation. Il est important de noter que ni la continuité des trajectoires, ni le nombre de pôles réels ne varient significativement avec l'ordre (2 à 4 pôles réels par azimut).

Enfin, lorsque l'interpolation sur les HRTF à phase mixte est envisagée, il est important d'évaluer l'ITD résultant. Ce dernier constituera un dernier paramètre d'observation.

Dans la suite, l'interpolation s'appuie sur l'échantillonnage du plan horizontal par pas de 15°, et reconstruit les états intermédiaires tous les 5°. Elle pondère linéairement les HRTF encadrant l'azimut cible, et choisissant une pondération "inversement proportionnelle à la distance" ("inverse distance weighting"), d'après la terminologie de Hartung ([HBS99]).

2.3.1.3 Paramètres de contrôle

Les paramètres de contrôle sont ceux sur lesquels est pratiquée l'interpolation. Ils doivent idéalement assurer la régularité des trajectoires des paramètres d'observation, elle-même garante de la régularité du résultat sonore. En particulier, une faible variation des paramètres de contrôle ne doit pas induire une forte variation des paramètres d'observation. L'interpolation idéale consiste donc à choisir les paramètres d'observation comme paramètres de contrôle. Néanmoins, il semble difficile de réaliser en temps réel l'interpolation des spectres d'amplitude puis d'en dériver les paramètres d'implantation associés. De la même manière, Gardner rappelle qu'une mesure directe de la distorsion spectrale n'est jamais choisie comme critère à minimiser en temps réel par un quantificateur, du fait de sa complexité de calcul ([Gar94]). On a recours à des mesures plus éloignées de la perception, mais plus proche de la structure de filtre utilisée,

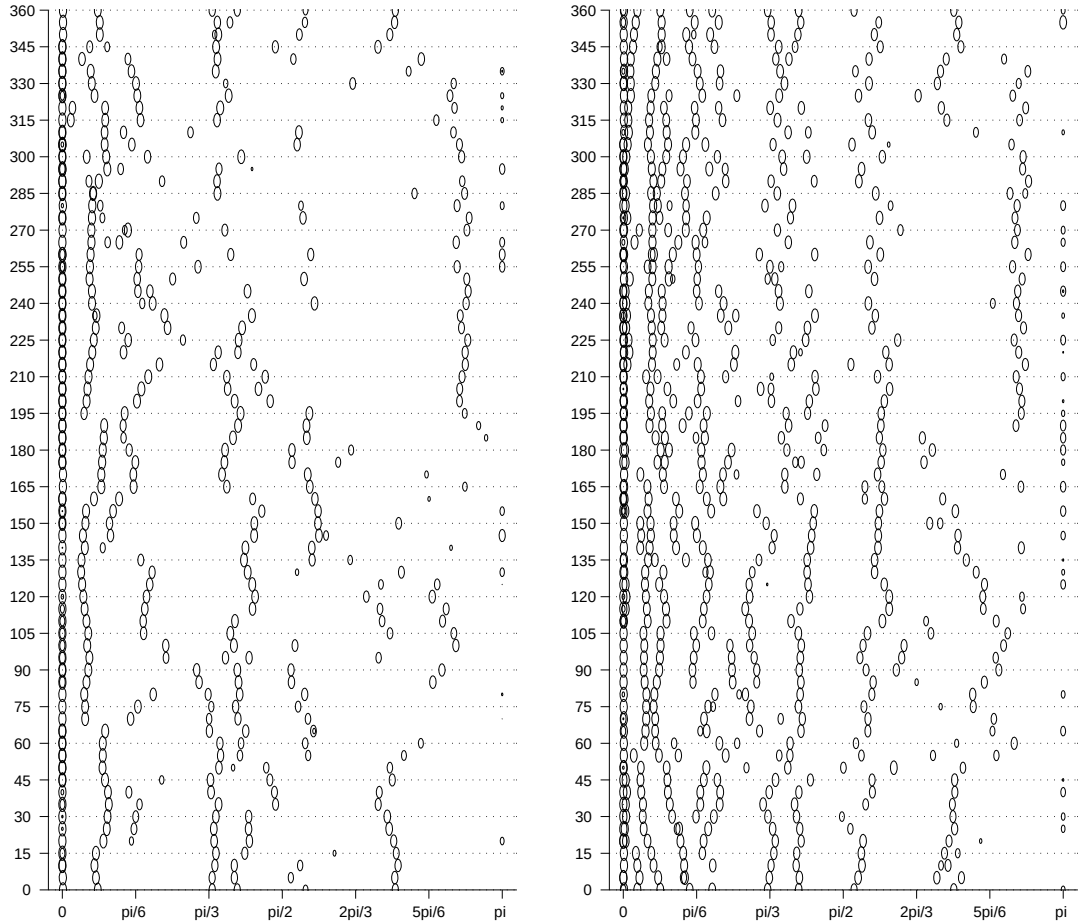


FIG. 2.13 – Pôles obtenus par une modélisation IIR des HRTF à phase minimale dans le plan horizontal, mesurées par pas de 5° sur la tête artificielle HMSII : fréquence repérée par l'axe des abscisse, module représenté par le rayon des cercles. Modélisation d'ordre 12 (a) et d'ordre 20 (b), avec un warping fréquentiel de 0.5.

comme l'erreur aux moindres carrés entre les coefficients de la prédiction et les coefficients quantifiés. Les paramètres de contrôle doivent donc être choisis comme des coefficients intermédiaires entre les paramètres d'observation et les paramètres d'implantation.

Ils doivent répondre à deux exigences :

- La stabilité des états interpolés doit être garantie par la stabilité des filtres servant à les créer. Notons qu'il s'agit ici d'une stabilité statique, à distinguer de la stabilité des filtres transitoires liée au phénomène de commutation.
- L'interpolation d'un vecteur de paramètres au suivant s'effectue coordonnées par coordonnées. Les paramètres de contrôle doivent donc être ordonnés.

2.3.2 Interpolation FIR

L'interpolation FIR des HRTF consiste à choisir les coefficients de la réponse impulsionnelle comme paramètres de contrôle.

2.3.2.1 HRTF à phase mixte

Comme nous l'avions déjà observé dans [Lar94a], l'interpolation sur les HRTF contenant le retard monaural conduit à une médiocre reconstruction des HRTF (Figure 2.14). Les "creux" apparaissant sur les états intermédiaires traduisent un effet de peigne. Ils sont localisés en hautes fréquences (f>5kHz) à cause du faible écart temporel entre les HRTF distantes de 15°.
Il existe un lien direct entre cette interpolation et le mécanisme de localisation en jeu avec les techniques de spatialisation multi-haut parleurs. En effet, imaginons un dispositif de haut-parleurs espacés tous les 15° sur un cercle autour de l'auditeur. Si un seul des haut-parleurs émet, le signal produit aux oreilles a subi le filtrage par la tête de l'auditeur spécifique à la position relative de ce haut-parleur. Une position intermédiaire est reproduite par une pondération du signal envoyé dans les deux haut-parleurs voisins. Le signal reconstruit au niveau des oreilles de l'auditeur correspond alors au signal émis filtré par une pondération des HRTF associées à la position des haut-parleurs. Plusieurs lois de pondération (potentiomètres panoramiques) sont étudiées dans [JLP99]. On a extrait de cet article la Figure 2.15. Elle correspond au cas de haut-parleurs espacés de 60°. On constate qu'au delà de l'altération des spectres d'amplitude, que nous observions déjà pour un écart de 15°, le retard interaural est mal reproduit : la source perçue reste "collée" au haut-parleur le plus proche, puis effectue un saut jusqu'au haut-parleur suivant. Le seul cas d'ITD fidèle est obtenu lorsque la position cible coincide avec celle d'un haut-parleur. Dans ce cas en effet, seul ce dernier est actif, et il constitue alors une source sonore "réelle".

2.3.2.2 HRTF à phase minimale

Dans le cas d'une implantation bicanale telle que nous l'avons décrite plus haut, le retard interaural et les HRTF à phase minimale sont séparées. L'interpolation doit alors être pratiquée sur les deux composantes. Nous avons examiné dans [LJ97] l'interpolation locale et globale de l'ITD. Nous en avons extrait la Figure 2.16, présentant :

- une interpolation locale réalisée par combinaison linéaire des valeurs d'ITD voisines de la cible,
- une interpolation globale s'appuyant sur la projection de l'ITD mesuré sur une fonction spatiale à la base du modèle de l'ITD que nous décrivons au chapitre 5. Cette projection permet de dériver les paramètres du modèle, et les valeurs intermédiaires sont données par la formule analytique explicitant le modèle.

Comme l'illustre la Figure 2.17, l'interpolation sur les coefficients de la réponse impulsionnelle à phase minimale donne des résultats satisfaisants, puisque les spectres interpolés semblent varier quasi- "linéairement en dB". Effectivement, l'erreur de reconstruction des spectres d'amplitude, présentée en Figure 2.18, montre que l'écart aux mesures reste inférieure à 1dB pour toutes les positions ipsilatérales (entre 400Hz et 9kHz), nettement plus faible que celle observée pour l'interpolation des HRTF à phase mixte.

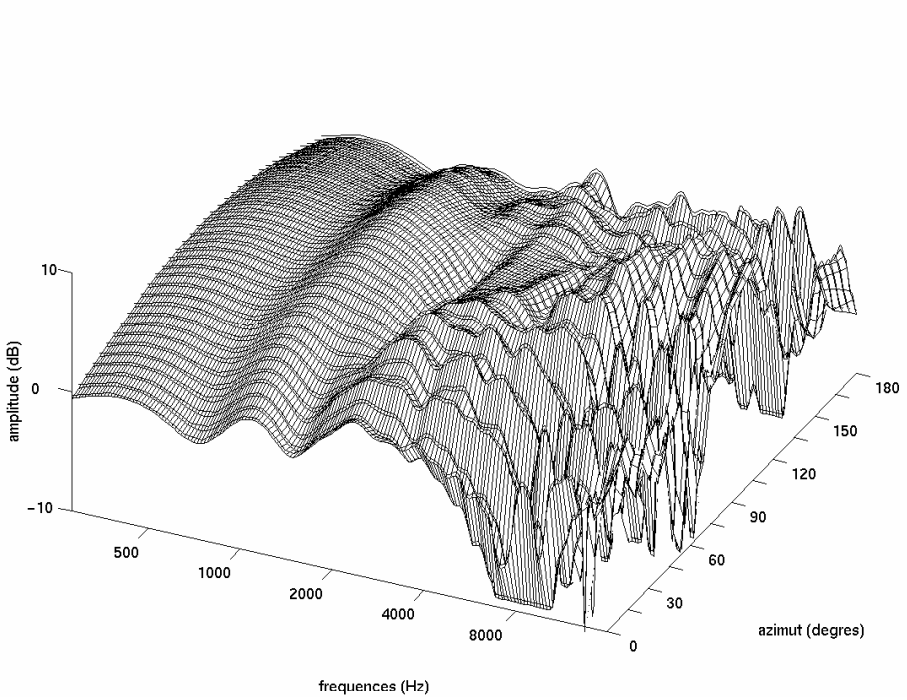


FIG. 2.14 – Spectres d'amplitude après interpolation sur les coefficients de la réponse impulsionnelle à phase mixte : échantillonnage de départ de résolution 15°, et reconstruction par interpolation d'un échantillonnage de résolution 5°.

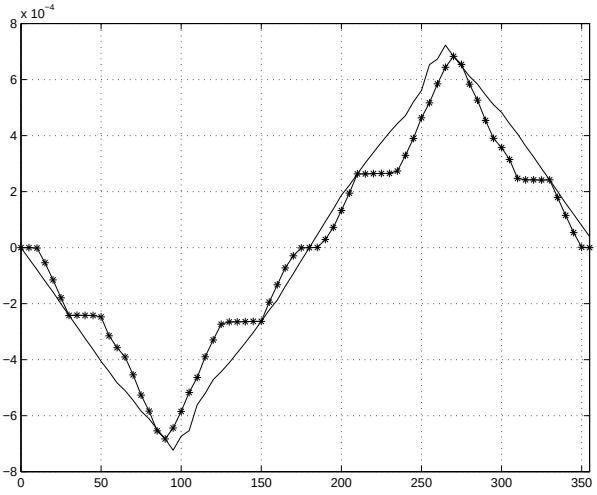


FIG. 2.15 – Retard interaural du plan horizontal mesuré sur une tête (courbe continue) ou reconstruit avec un dispositif de spatialisation multi-haut-parleur (courbe étoilée). Les haut-parleurs sont disposés à 0°, 60°, 120°, 180°, 240°, 300° et 360°. La loi utilisée est un potentiomètre panoramique d'amplitude (cf [JLP99]).

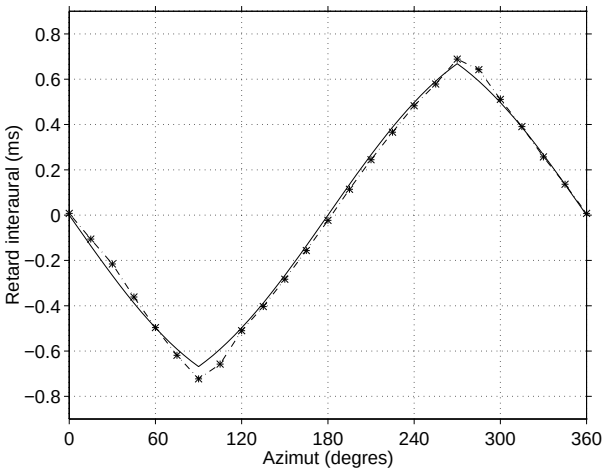


FIG. 2.16 – Retard interaural du plan horizontal mesuré tous les 15 degrés sur une tête (*) et tous les 5° : par interpolation locale (trait interrompu), ou par interpolation globale (trait continu) (cf [LJ97]).

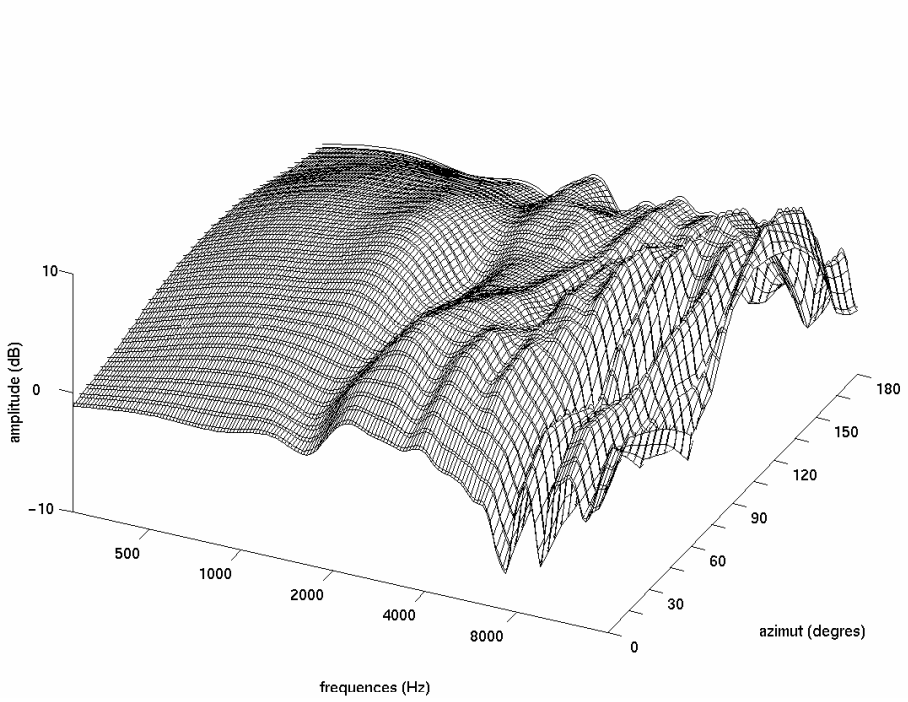
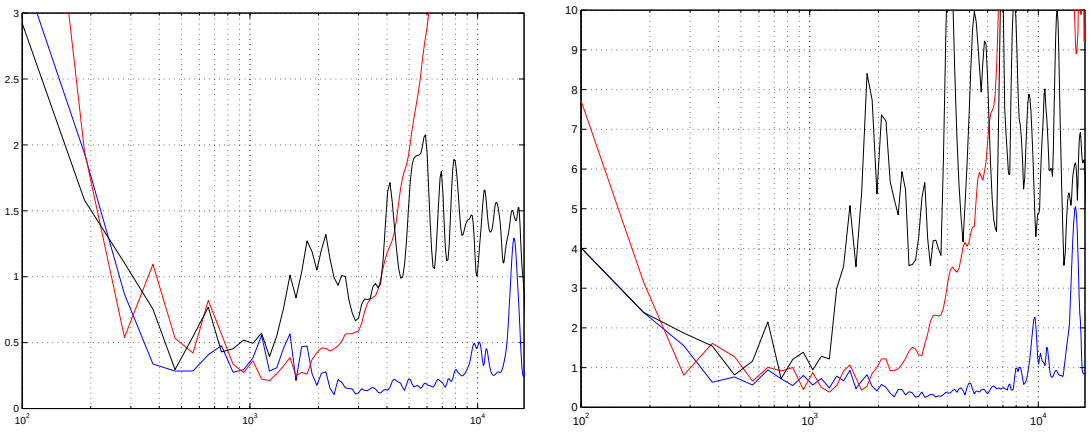


FIG. 2.17 – Spectres d'amplitude après interpolation sur les coefficients de la réponse impulsionnelle à phase minimale : échantillonnage de départ de résolution 15°, et reconstruction par interpolation d'un échantillonnage de résolution 5°.



(a) (b)

FIG. 2.18 – Comparaison de l'erreur aux moindres carrés (a) ou erreur L_∞ de reconstruction des spectres d'amplitude mesurés et lissés au demi-ton pour les positions du plan horizontal : interpolation sur les coefficients de la réponse impulsionnelle à phase minimale pour les positions ipsi (en bleu) et pour les positions ipsi et contra (en noir) ; interpolation sur les coefficients de la réponse impulsionnelle à phase mixte pour les positions ipsi (en rouge). Les HRTF sont considérées tous les 15° dans le plan horizontal et reconstruites par pas de 5°.

Les erreurs en basses fréquences sont liées à la modélisation et non à l'interpolation : la troncature des réponses impulsionnelles (60 échantillons pour les réponses à phase minimale, 110 pour les réponses à phase mixte) néglige les basses fréquences, comme nous l'avons vu en section 2.2.1. On constate néanmoins que les performances de l'interpolation sur les positions contralatérales sont nettement inférieures.

Dans [JLW95], Jot interprète ces bonnes performances de l'interpolation sur les HRTF à phase minimale :

1. interpoler sur les réponses impulsionnelles est équivalent à interpoler sur les spectres complexes.
2. l'interpolation idéale doit interpoler linéairement sur les spectres en dB, i.e. sur $\log(mag)$. Or, le spectre d'amplitude des HRTF est relié à sa phase linéaire par la transformée de Hilbert (cf chapitre 1), qui est linéaire. D'après cette relation, interpoler linéairement sur $\log(mag)$ conduit à interpoler linéairement sur la phase minimale mph, donc sur le logarithme du spectre complexe des HRTF à phase minimale $\log(H_{min})$:
$$\log(H_{min}) = \log(mag) + j.mph$$
3. Pour de petits écarts, H_{min} est assez proche de $\log(H_{min})$, donc interpoler linéairement les HRTF à phase minimale s'approche d'une interpolation idéale.

2.3.3 Interpolation pour une structure transverse développée

La structure transverse développée implante la fonction de transfert :

$$\frac{B(z)}{A(z)} = \frac{b_0 + b_1.z^{-1} + ... + b_n.z^{-n}}{1 + a_1.z^{-1} + ... + a_n.z^{-n}}$$

Le coût d'implantation de cette structure pour un filtre d'ordre n est de $2.n+1$ opérations par échantillon. Les coefficients a_i et b_i sont rarement utilisés directement, car on contrôle difficilement la stabilité du filtre lors de leur quantification. Dans [Ita75], la fréquence des raies spectrales (Line Spectral frequencies) a été proposée comme une représentation alternative plus résistante à la quantification scalaire. A l'origine, elles sont apparues pour la prédiction linéaire lorsque l'on a choisi de substituer aux conditions limites traditionnelles de la glotte deux conditions extrêmes : ouverture complète ou fermeture complète ([Dem99]). Ces deux nouvelles conditions conduisent aux paires de raies spectrales.

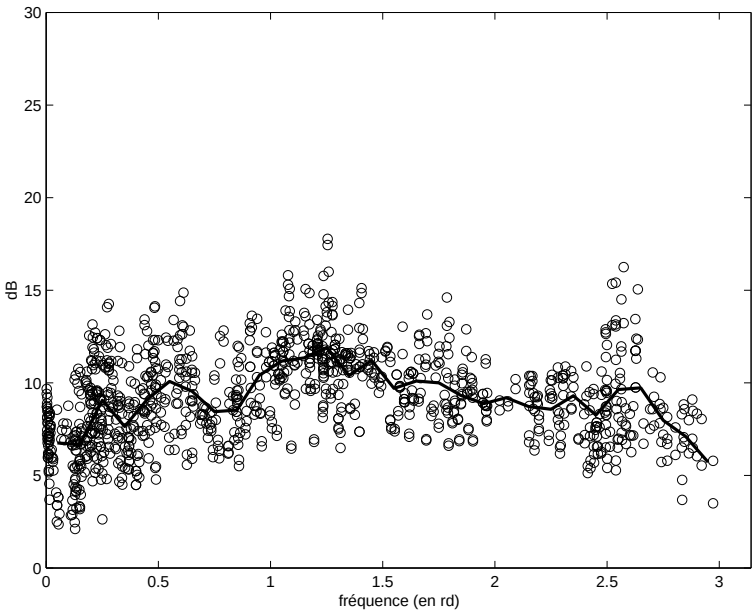


FIG. 2.19 – Courbe de sensibilité spectrale pour l'interpolation sur la fréquence des raies spectrales (LSF). On a superposé les courbes obtenues pour le numérateur et le dénominateur des modèles des HRTF.

L'expression des LSF est obtenue en considérant les polynômes $P(z)$ et $Q(z)$ associés au dénominateur $A(z)$ (même chose pour $B(z)$) :

$$\begin{aligned} P(z) &= A(z) + z^{-(n+1)}.A(z^{-1}) \\ Q(z) &= A(z) - z^{-(n+1)}.A(z^{-1}) \end{aligned}$$

La fréquence des raies spectrales est constituée de l'argument des racines de ces deux polynômes. Dans le prolongement du théorème de Schlüssler ([Sch76]), Soong et Juang montrent que si $A(z)$ est à phase minimale, le module de ces racines est unitaire et leurs arguments sont entrelacés, et réciproquement ([SJ84]). Ces propriétés sont valables pour les modèles des HRTF à phase minimale, et les LSF peuvent donc être étudiés comme paramètres de contrôle pour l'interpolation[1] :

- la stabilité des filtres interpolés est garantie puisque les fréquences créées sont entrelacées par construction et que leur module demeure unitaire.
- les LSF sont naturellement ordonnées de par leur entrelacement.

En outre, la transformation pour revenir aux paramètres d'implantation est simple :

$$A(z) = \frac{1}{2}\left(P(z) + Q(z)\right)$$

De façon pratique, le gain global des polynômes $A(z)$ et $B(z)$, que l'on "perd" lors de la conversion vers les LSF, est réalisée linéairement en dB. Plusieurs propriétés permettent de réduire le nombre de paramètres à interpoler. Les modèles ayant des coefficients réels, les coefficients des polynômes $P(z)$ et $Q(z)$ ont des racines réelles ou complexes conjuguées, auquel cas, seules les fréquences positives sont interpolées, les autres s'en déduisant directement. Qui plus est, dans le cas de modèles d'ordre pair, la configuration symétrique (resp. antisymétrique) des coefficients de $P(z)$ (resp. $Q(z)$) implique que π (resp. 0) est toujours une fréquence de raies spectrales. On se contente donc de pratiquer l'interpolation que sur les fréquences positives différentes de 0 et π.

La sensibilité spectrale présentée en Figure 2.19 a été calculée en prenant 0.1 comme valeur de l'écart Δc. Puisque l'interpolation que nous pratiquons vise à affiner la définition des données de départ d'un facteur 3 (on passe d'une résolution de 15° à une résolution de 5°), cette valeur a été choisie comme le

[1] Nous remercions Jacques Prado d'avoir porté à notre attention l'article de Schlüssler ([Sch76]) et d'avoir ainsi préssenti les bonnes propriétés des LSF pour l'interpolation des HRTF.

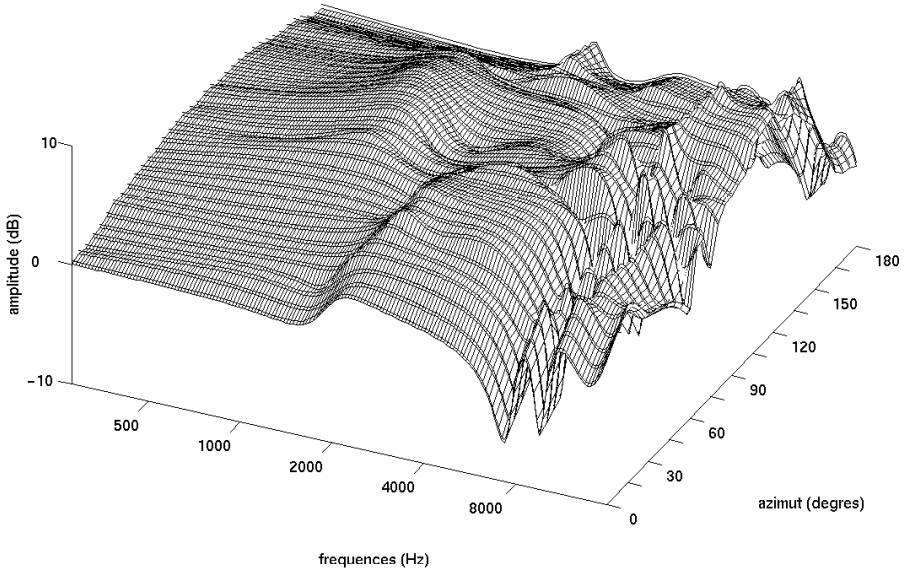


FIG. 2.20 – Spectres d'amplitude après interpolation sur la fréquence des raies spectrales (LSF) : échantillonnage de départ de résolution 15°, et reconstruction par interpolation d'un échantillonnage de résolution 5°.

tiers de la transition la plus large observée sur les 12 LSF pour le passage d'un azimut à l'azimut suivant espacé de 15°. Nous avons vérifié que la forme de la sensibilité ainsi obtenue reste semblable lorsque l'on augmente la valeur de l'écart. Par conséquent, nous calculons une sensibilité "prototype" par moyenne des différentes réalisations (courbe en trait épais de la Figure 2.19).
La sensibilité spectrale pour les LSF apparait plus faible aux fréquences proches de 0 et π, et maximale autour de $\pi/3$. Toutefois, entre $\pi/6$ et $5\pi/6$, la sensibilité spectrale semble assez constante, propriété que l'on recherche. En outre, Gardner montre que la matrice de sensibilité spectrale des LSF est diagonale, ce qui renforce la pertinence du résultat précédent pour une application à l'interpolation où tous les coefficients évoluent simultanément ([Gar94]).
Puisqu'une variation de la fréquence des raies spectrales induit une distorsion proportionelle sur le spectre d'amplitude des HRTF, on peut comprendre la linéarité observée sur l'évolution des spectres d'amplitude en Figure 2.20. Notons que celle-ci provoque parfois quelques défauts, comme par exemple entre 60° et 75°. Comme on l'observe en Figure 2.21, les spectres interpolés entre ces deux positions ne respectent pas le critère de "linéairité", qui nécessiterait qu'ils soient confinés entre les spectres ayant servi à les créer. Cette mauvaise interpolation s'explique par la forte distance entre 2kHz et 7kHz des spectres d'amplitude modélisés, qui se traduit par de forts écarts des LSF. Cet artefact est lié à l'ordre de modélisation, faible, et disparait dès lors que ce dernier est augmenté (ordre 14 pour notre exemple). L'ordre 14 permet d'individualiser les pics à 3kHz et 6kHz de la position 60°, et donne ainsi à ce filtre une "structure" semblable à celle du filtre cible à 75° , i.e. le même nombre d'extrema, ce qui semble permettre aux LSF de mieux reproduire la trajectoire de ces derniers.

La Figure 2.22 présente les pôles obtenus à partir des états interpolés. Sur tous les intervalles où les pôles des modèles décrivent une trajectoire, celle ci est reproduite par l'interpolation sur les LSF. En outre, il semble se dégager une capacité des LSF à "prolonger" ces trajectoires (se reporter par exemple aux zones indiquées par les flèches), propriété qui favorise la continuité des fortes résonances/antirésonances des HRTF.

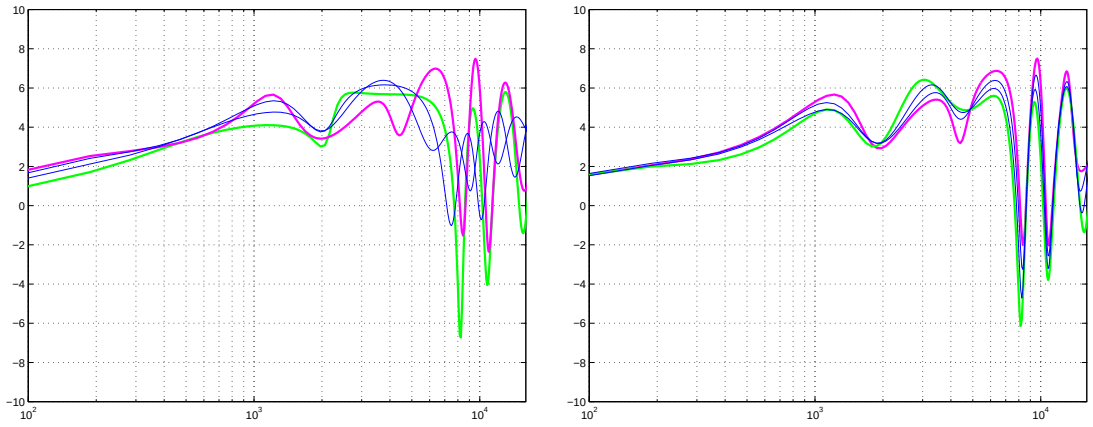


FIG. 2.21 – Spectres d'amplitudes modélisés à 60° et 75° (trait épais) et états interpolés à 65° et 70° en prenant les LSF comme paramètres de contrôle. Modélisation IIR d'ordre 12 (a), d'ordre 14 (b)

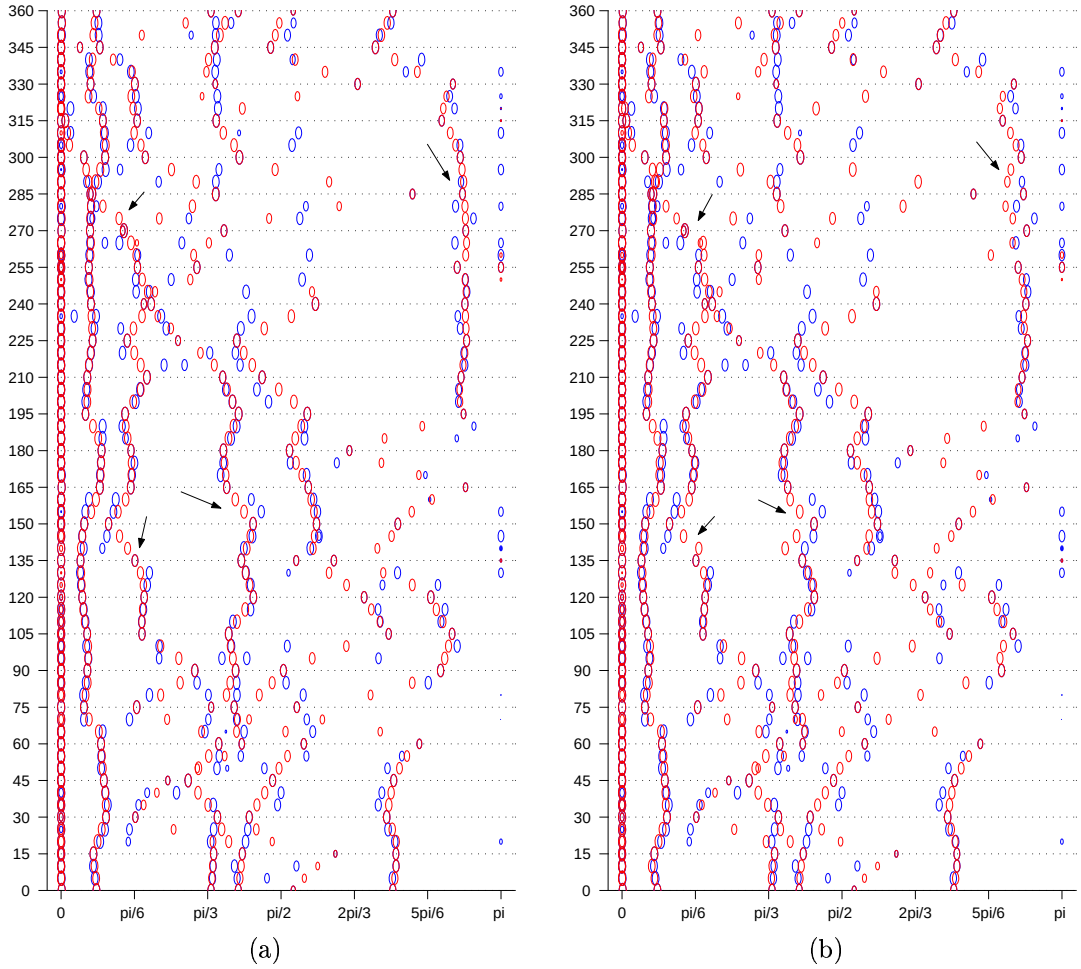


FIG. 2.22 – Pôles donnés par une interpolation sur les modèles d'ordre 12 des HRTF, pour reconstructuire une résolution de 5°, et pôles des modèles tous les 5° (en bleu) : interpolation sur la fréquence des raies spectrales (a), et sur les log area ratio (b).

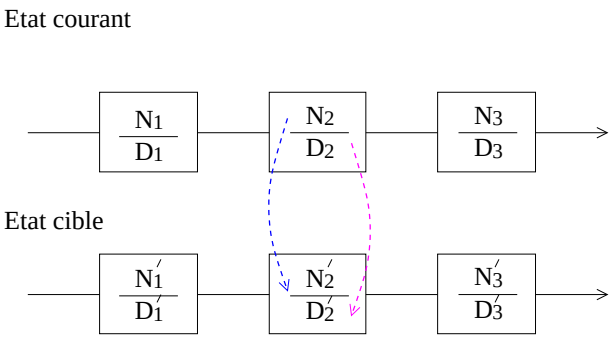


FIG. 2.23 – Problématique de l'ordonnancement des cellules d'ordre 2 : un ordre est défini pour les polynômes du numérateur et pour ceux du dénominateurs, puis les cellules sont constituées en appairant numérateur et dénominateur de même rang. N_i : numérateur, D_i : dénominateur.

2.3.4 Interpolation pour une structure transverse factorisée en cellules d'ordre 2

Comme nous le rappelions plus haut, l'implantation de la structure transverse directe (ou "développée") est délicate du fait de l'absence d'un critère simple permettant de vérifier la stabilité des filtres après quantification. Utiliser des paramètres transformés, les paires de raies spectrales par exemple, est une solution. Recourir à la factorisation de la forme développée du filtre en cellules d'ordre 2 en est une autre, que nous nous proposons d'étudier. En effet, la stabilité du filtre est assurée par celle des cellules d'ordre 2 pour lesquelles un critère peut être formulé analytiquement de façon très simple. Une cellule peut être représentée par les coefficients a_i et b_i :

$$H(z) = \frac{N(z)}{D(z)} = \frac{b_0 + b_1.z^{-1} + b_2.z^{-2}}{1 + a_1.z^{-1} + a_2.z^{-2}}$$

Le domaine de stabilité est obtenu en cherchant les valeurs des coefficients pour lesquelles les pôles de $H(z)$ sont à l'intérieur du cercle unité. Une représentation traditionnelle de ce domaine, un triangle, est présentée au coeur de la Figure 2.25. Puisque c'est un domaine convexe, l'interpolation entre deux cellule stables est une cellule stable.

Le problème majeur de la structure transverse factorisée réside en l'absence d'ordre naturel entre les cellules : la permutation de deux cellules ne modifie en rien le filtre global. Pourtant, l'interpolation requiert l'association d'une cellule de l'azimut cible à toute cellule de l'azimut courant. Il nous faut donc définit une méthode pour ordonnancer des cellules. Cet ordre ayant été introduit, on est alors capable d'associer un rang à chaque cellule de l'état courant et de l'état cible (cf Figure 2.23), et l'interpolation sur les paramètres de contrôle est alors effectuée rang par rang au numérateur et au dénominateur.

2.3.4.1 Ordonnancement des cellules d'ordre 2

Pour étudier l'ordonnancement, nous proposons de considérer séparemment numérateur et dénominateur des cellules d'ordre 2, pour ramener le problème à des polynômes du second ordre de la forme :

$$P(z) = 1 + c_1.z^{-1} + c_2.z^{-2}$$

Si nous savons ordonner les polynômes du dénominateur d'une part et ceux du dénominateur d'autre part, la cellule de rang i sera formée par le rapport des polynômes de rang i.

Si nous parvenions à définir des paramètres de contrôle produisant une interpolation linéaire sur les spectres en dB des polynômes, l'ordonnancement pourrait être défini de façon arbitraire. En effet, par leur mise en série, les polynômes interpolés engendreraient un spectre global évoluant lui-même linéairement en dB (cf Figure 2.24). Comme rien ne nous permet de faire cette hypothèse, nous proposons de définir un ordonnacement anticipant le travail de l'interpolation, en appairant des polynômes de forme voisine. Cette

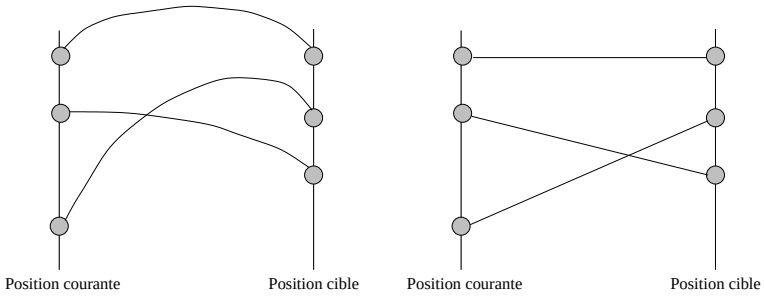


FIG. 2.24 – Interpolation sur les coefficients de polynômes d'ordre 2 en série : les gains en dB interpolés ne varient pas linéairement (à gauche) contrairement au cas idéal (à droite).

problématique peut être rapprochée des objectifs du suivi de partiels pour la synthèse additive tel que l'abordent Depalle et al. ([DGR93a], [DGR93b]) : chaque trajectoire est construite à l'aide de chaînes de Markov cachées, optimisant la continuité de l'amplitude et de la fréquence des partiels associés. Toutefois, notre objectif serait de proposer une méthode d'ordonnancement pouvant être implantée en temps réel : pour toute nouvelle position cible, les cellules de la HRTF sont ré-ordonnées en fonction de celles de la position courante.

Dans les études antérieures [Lar95] et [LJ97], nous avons étudié un ordonancement s'appuyant sur la fréquence des pôles et des zéros de la cellule. Cette approche se justifie par la relation entre cette fréquence, θ, et la fréquence du maximum du spectre, F_r, dans le cas d'une forte résonance (e.g. [Lar94b]) :

$$F_r = \arccos\left(\frac{1-\rho^2}{2.\rho}.\cos\theta\right)$$

où ρ désigne le module de la racine. Dans le cas d'une forte résonance ($\rho \simeq 1$), il y a coïncidence entre F_r et θ. Par conséquent, choisir un ordre favorisant la continuité de fréquence des pôles et des zéros de la cellule optimise la continuité des fortes résonances, que l'on observe notamment en hautes fréquences des HRTF (cf Figure 2.11). Toutefois, la présence de racines réelles, en nombre variable selon l'azimut, rend l'algorithme non trivial. En effet, les racines réelles positives apparaissent typiquement en remplacement d'une racine complexe de fréquence élevée. Un simple ordonnancement par fréquence croissante peut ainsi conduire à détourner une trajectoire rectiligne en basses fréquences afin de passer par une racine réelle de fréquence nulle. En outre, s'il y a plusieurs paires de racines réelles à une position, celles-ci se superposent (à la fréquence 0 ou à la fréquence π), ce qui souligne les limites d'une représentation de la seule phase des racines pour la définition de trajectoires. L'étude de Fillon ([Fil00]) montre que même en prenant compte du module en plus de l'argument des racines, par exemple en utilisant une représentation polaire, il est impossible de déterminer un ordonnancement optimal des racines. Il obtient ce résultat en empruntant la méthode du lieu de Evans au domaine de l'automatique. Celle-ci permet de vérifier la stabilité d'un système bouclé à retour unitaire, $\frac{Y(s)}{R(s)}$, en traçant l'évolution de ses pôles en fonction de K défini par :

$$\frac{Y(s)}{R(s)} = \frac{K.G(s)}{1+K.G(s)}$$

Nous proposons de définir un **ordonnancement par "forme spectrale"**, ne s'appuyant plus seulement sur les pics et les vallées, mais sur le spectre entier. A cette fin, nous étudions les polynômes du domaine de stabilité afin d'en déterminer une partition en classes de forme spectrale. Cette partition a pour principe la séparation des polynômes résonants des polynômes non résonants.
On montre par un calcul simple que le gain en fréquence de la cellule, $G(f)$, a un extremum si et seulement si :

$$\left|\frac{c_1.(c_2+1)}{4c_2}\right| < 1 \tag{2.2}$$

Cette inéquation définit deux courbes, représentées en jaune sur la Figure 2.25, qui délimitent quatre secteurs au sein du triangle de stabilité :

– deux secteurs de polynômes résonants :

1. le secteur D1 des spectres présentant une vallée,
2. le secteur D2 des spectres présentant un pic.

La fréquence de l'extremum est donnée par :

$$F_r = \arccos\left(\frac{-c_1.(c_2+1)}{4.c_2}\right) \tag{2.3}$$

et le gain de l'extremum s'exprime par :

$$G(F_r) = \frac{1-c_2}{2}\sqrt{\frac{4.c_2 - c_1^2}{c_2}} \tag{2.4}$$

– deux secteurs de polynômes non résonants (de type "shelving") :

1. le secteur D3 des spectres monotones croissants,
2. le secteur D4 des spectres monotones décroissants

Les extrema sont atteints aux fréquences 0 et π, et valent :

$$\begin{aligned} G(0) &= 1 + c_1 + c_2 \\ G(\pi) &= 1 - c_1 + c_2 \end{aligned}$$

On a représenté en trait bleu sur la Figure 2.25 la frontière entre polynômes à racines réelles et polynômes à racines complexes. On constate ainsi que D1 ne contient que des polynômes à racines complexes, et que D2 ne contient que des polynômes à racines réelles. D3 et D4 en revanche contiennent les deux familles. Pour les HRTF considérées (mesurées tous les 15° dans le plan horizontal), le nombre de représentants de chaque famille varie en fonction de l'azimut. La distribution des coefficients des polynômes du plan horizontal est représentée en Figure 2.26. Dans près de 65% des cas, un modèle de HRTF est composé de 5 cellules résonantes D1 et d'une cellule D3. A tous les azimuts, il y a au moins 3 cellules résonantes, 4 dans la grande majorité des cas (plus de 94%). Il n' a jamais 2 polynômes de D2 ou de D4 à un azimut donné.

Ainsi, appairer les polynômes d'ordre 2 selon la forme de leur spectre d'amplitude peut conduire à associer des polynômes de famille différente. On présente deux cas particuliers en Figures 2.27 et 2.28 où la recherche d'une plus grande proximité entre les spectres conduit "à l'oeil" à appairer un polynôme de D3 avec un polynôme de D2 (Figure 2.27, rang 6), ou un polynôme de D1 avec un polynôme de D2 (Figure 2.28, rang 4).
Avec cet ordonnancement, les réponses en fréquence des l'état courant et de l'état cible de l'interpolation sont les plus proches possibles. L'effort à réaliser par l'interpolation, i.e. la création de spectres intermédiaires variant doucement, est ainsi réduit au minimum. Une méthode pour évaluer les performances de paramètres de contrôle candidats consiste à vérifier qu'ils interpolent linéairement en dB le spectre des polynômes au sein d'une même famille, et entre deux familles différentes.

2.3.4.2 Interpolation sur les coefficients des cellules d'ordre 2

Les paramètres de contrôle les plus immédiats sont les coefficients des cellules, c_1 et c_2. En effet, comme on l'a vu, leur domaine de stabilité est convexe, ce qui garantit la stabilité de tout polynôme créé par interpolation linéaire entre deux polynômes stables.
On a représenté en Figure 2.29 les performance de l'interpolation sur les coefficients des cellules d'ordre 2, pour trois cas "critiques", mettant en avant les défauts de ces paramètres :

– interpolation entre les polynômes de rang 4 de la Figure 2.27. Tous deux appartiennent à la famille D1, mais ils diffèrent sensiblement quant à leur fréquence de résonance et leur facteur de qualité.
– interpolation entre les polynômes de rang 6 de la Figure 2.27. Ils appartiennent à deux familles différentes : l'un des deux polynômes a une forme de type résonante (D2), et l'autre de type shelving (D3).

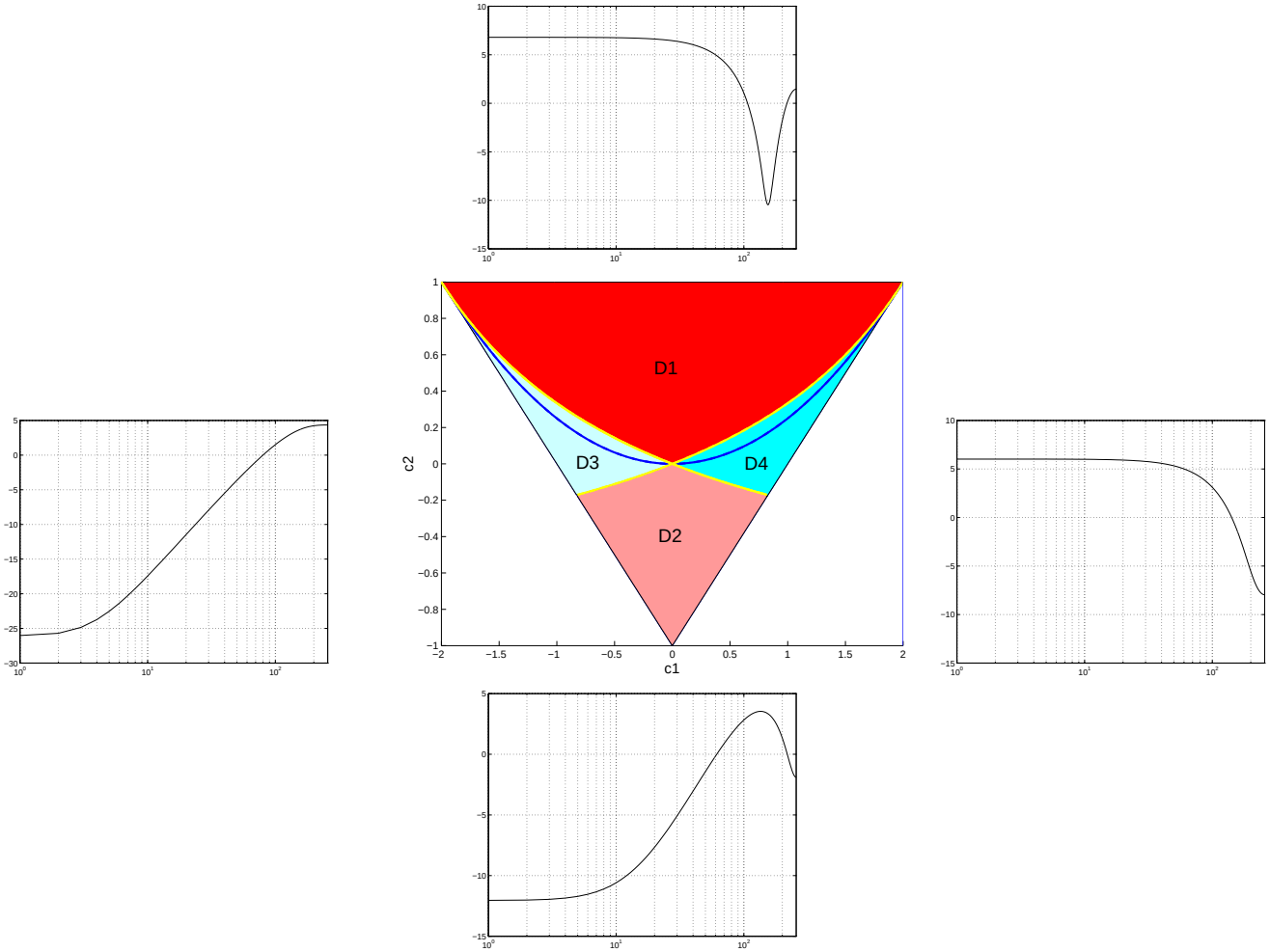


FIG. 2.25 – Familles de forme spectrale des polynômes d'ordre 2 au sein du domaine de stabilité (à côté de chacun des 4 domaines est présenté un exemple) : spectres présentant une vallée (D1), spectres présentant un pic (D2), spectres monotones croissants (D3), spectres monotones décroissants (D4). La courbe bleu délimite les polynômes à racines complexes (zone au dessus de la frontière), et les polynômes à racines réelles (zone au dessous de la frontière).

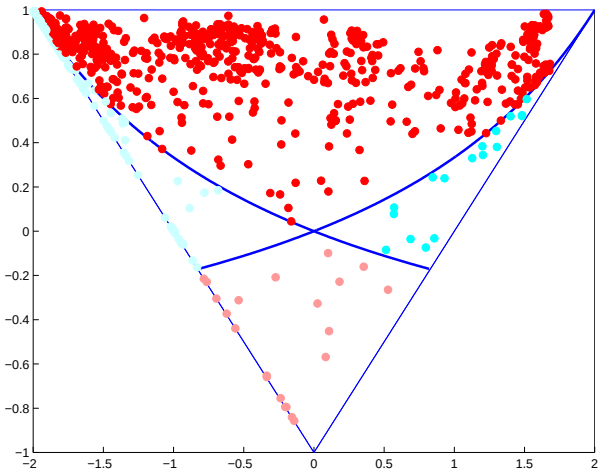


FIG. 2.26 – Coefficients de la structure transverse factorisée, obtenus par une modélisation IIR à l'ordre 12 des HRTF mesurées tous les 5° dans le plan horizontal.

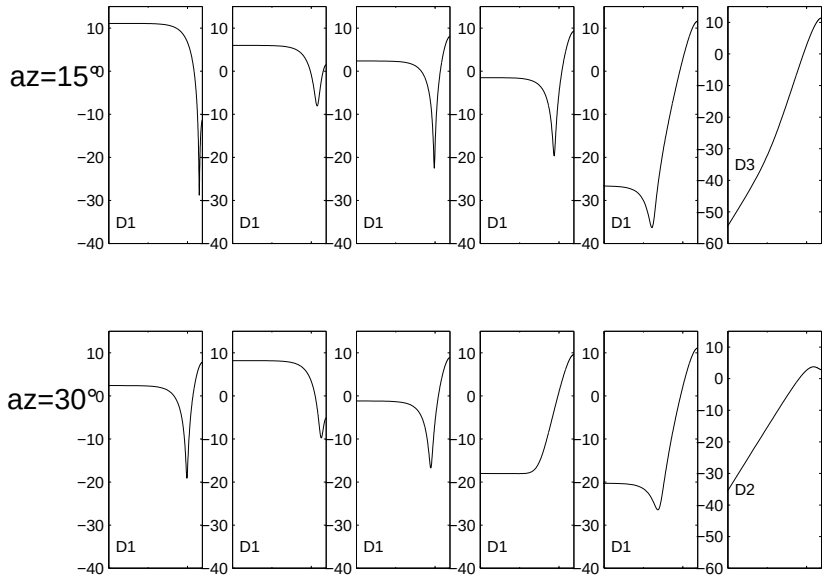


FIG. 2.27 – Appairage de polynômes entre les azimuts 15° et 30°. La famille d'appartenance du polynôme est notée D_i selon les conventions de la Figure 2.25.

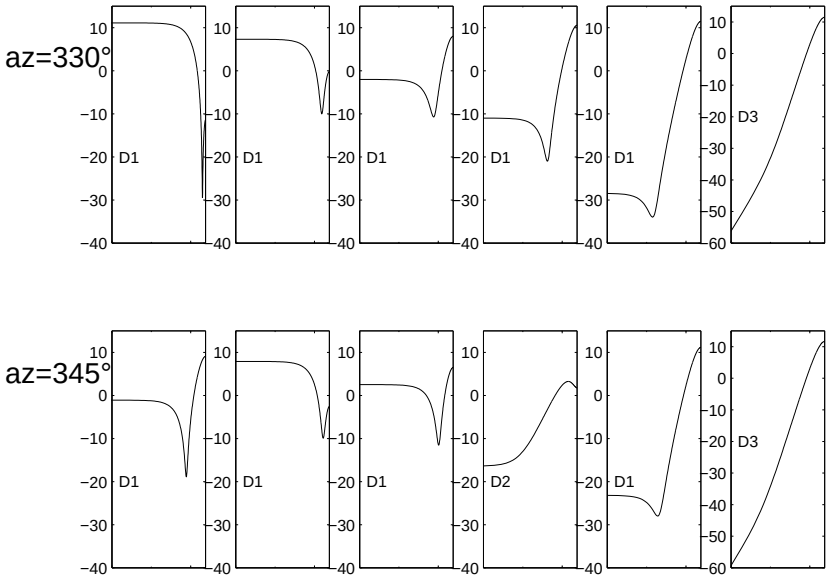
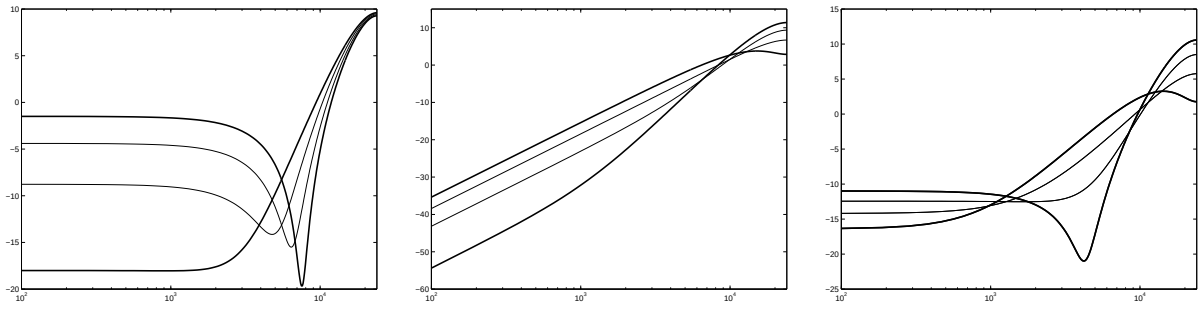


FIG. 2.28 – Appairage de polynômes entre les azimuts 330° et 345°. La famille d'appartenance du polynôme est notée D_i selon les conventions de la Figure 2.25.



(a) (b) (c)

FIG. 2.29 – Spectres reconstruits par interpolation sur les coefficients des cellules d'ordre 2 entre deux polynômes de même "forme spectrale" (a : $D_1 \leftrightarrow D_1$) ou non (b : $D_3 \leftrightarrow D_2$, c : $D_1 \leftrightarrow D_2$).

– interpolation entre les polynômes de rang 4 de la Figure 2.28. Ils appartiennent à deux familles différentes : l'un des deux polynômes a une forme de type résonante (D2), et l'autre de type anti-résonance (D1).

On constate une mauvaise interpolation des spectres en basses fréquences. On peut montrer que l'interpolation linéaire sur les coefficients c_1 et c_2 conduit le gain en 0 à rester "attiré" par une forte valeur initiale. En effet, le gain en 0 a pour expression $20.\log_{10}(1 + c_1 + c_2)$, et plus sa valeur est grande, moins il est sensible à une variation linéaire de $c_1 + c_2$. Théoriquement, cette observation vaut également pour le gain à la fréquence π, dont l'expression est $20.\log_{10}(1 - c_1 + c_2)$. Toutefois, dans les exemples de la Figure 2.29, caractéristiques des cas rencontrés dans notre étude, les valeurs à cette fréquence sont déjà élevées, sur un intervalle où le logarithme peut être approximé par une droite. Par conséquent, le gain à cette fréquence est le plus souvent convenablement interpolé.
Un autre défaut de l'interpolation sur les coefficients c_1 et c_2 s'observe sur les caractéristiques spectrales autour de la résonance, lorsqu'elle existe (Figure 2.29, a). Comme le notaient Ding et Rossum dans [DR95], la fréquence de résonance demeure "attirée" par les hautes fréquences. En outre, le gain à la résonance ne reste même pas dans l'intervalle délimité par les deux spectres de départ. Cette interpolation non linéaire s'explique par les relations entre les coefficients et fréquence/gain à la résonance, rappelées par les équations 2.3 et 2.4.

L'influence de ces différents artefacts sur les HRTF interpolées est illustrée en Figure 2.30. On constate notamment une surtension abusive apparaissant brutalement en basses fréquences. Les coefficients des cellules d'ordre 2 apparaissent ainsi comme de mauvais bons paramètres de contrôle pour l'interpolation des HRTF, observation déjà faite en [Lar95], et classique pour la quantification (e.g. [VM75]).

2.3.4.3 Interpolation sur les coefficients Armadillo

Les coefficients Armadillo, ν_1 et ν_2, sont définis dans [Ros91] et [DR95] pour l'interpolation des cellules d'ordre 2 résonantes, avec pour application l'interpolation d'égaliseurs paramétriques. Constatant l'échec d'une interpolation directe sur les coefficients des cellules, les auteurs adoptent pour stratégie de définir de nouveaux paramètres découplant le contrôle du module des racines, r, et de leur argument, θ. En effet, lorsque l'on interpole sur c_1, ces deux paramètres sont affectés simultanément, puisque : $c_1 = -2.r.\cos\theta$. Séparer les contrôle de r et θ permet donne la possibilité de leur imposer une variation logarithmique. Ils aboutissent au paramètres suivants :

$$\begin{aligned} \nu_1 &= -\log_2(1 + c_1 + c_2) + 2 \\ \nu_2 &= -\log_2(1 - c_2) \end{aligned}$$

Moyennant deux hypothèses, Ding et Rossum montrent que ν_2 est proportionnel au logarithme du gain à la résonance ($\log(G(F_r))$), proche de $\log(1-r)$, et ν_1 proportionnel au numéro d'octave musical ($\log(F_r)$), proche de $\log\theta$:

$$\log(G(F_r)) \simeq 6.\nu_2 + 6$$

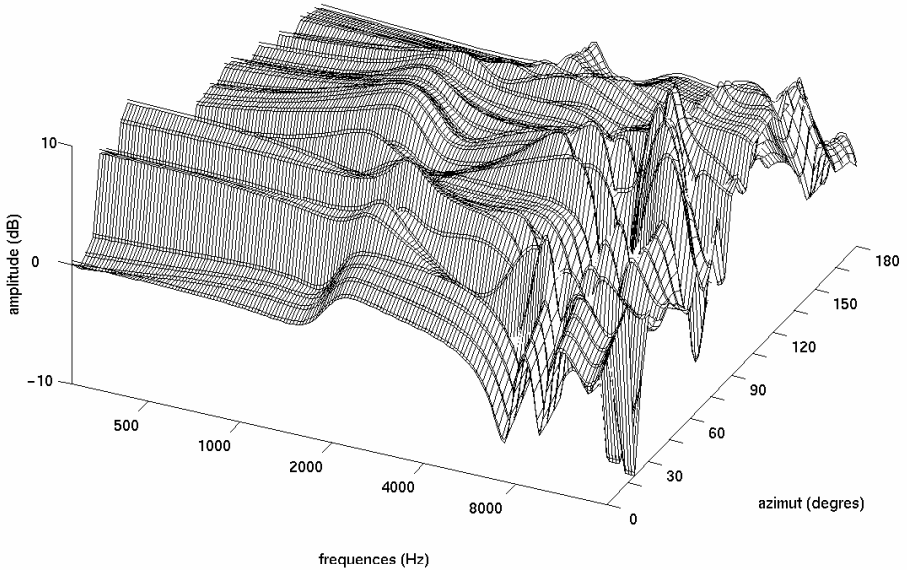


FIG. 2.30 – Spectres d'amplitude après interpolation sur les coefficients de la structure transverse factorisée : échantillonnage de départ de résolution 15°, et reconstruction par interpolation d'un échantillonnage de résolution 5°.

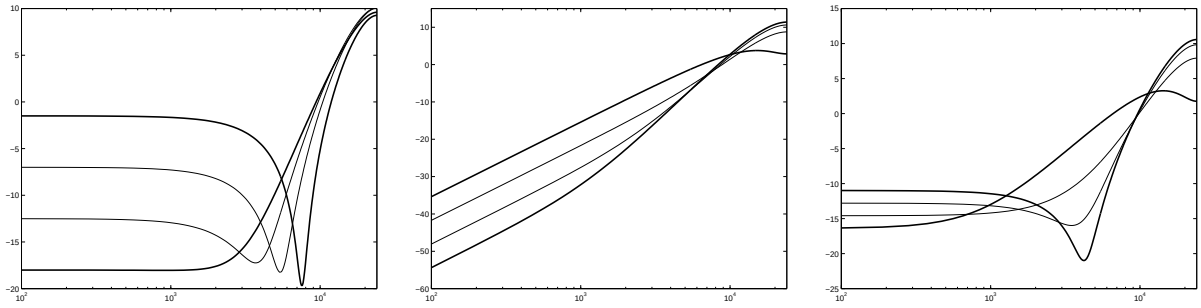
$$\begin{aligned} \Omega &= \log_2\left(\frac{F_r.F_e}{4\pi}\right) \\ &\simeq -\frac{1}{2}.\nu_1 + \log_2\left(\frac{F_s}{20\pi}\right) \end{aligned}$$

où Fe désigne la fréquence d'échantillonnage.
Ces approximations sont licites sous les deux conditions suivantes :

1. cas de fortes résonances,
2. résonances basses fréquences.

Nous utilisons ces coefficients comme paramètres de contrôle pour toutes les familles de polynômes, et non pour les seuls polynômes de D1, familles pour laquelle ils ont été définis. En effet, on peut remarquer que l'interpolation de ν_1 conduit naturellement à l'interpolation linéaire du gain à la fréquence nulle, ce qui doit permettre d'éviter les artefacts basses fréquences présentées par l'interpolation sur les coefficients des cellules. C'est ce que l'on observe sur la Figure 2.31, et plus généralement sur les HRTF interpolées en Figure 2.32. De même, il apparait également pour les polynômes (a) une interpolation continue sur la fréquence de résonance et le gain à cette fréquence. Comme on pouvait s'y attendre, les performances sont critiques en hautes fréquences, intervalle sortant du cadre de l'approximation.

Les coefficients Armadillo permment effectivement une interpolation "idéale" en basses fréquences, mais sont mal adaptés au cas de spectres de forme très différentes en hautes fréquences. Ils ne constituent donc pas une solution générale pour l'interpolation des cellules d'ordre 2. Toutefois, on peut penser qu'un ordonnancement plus adapté aux défauts de ces paramètres permettrait de réduire les artefacts, par exemple, un classement des polynômes minimisant l'écart des gains en π entre l'état courant et l'état cible.



(a) (b) (c)

FIG. 2.31 – Spectres reconstruits par interpolation sur les coefficients ARMADILLO (mêmes conventions qu'en 2.29).

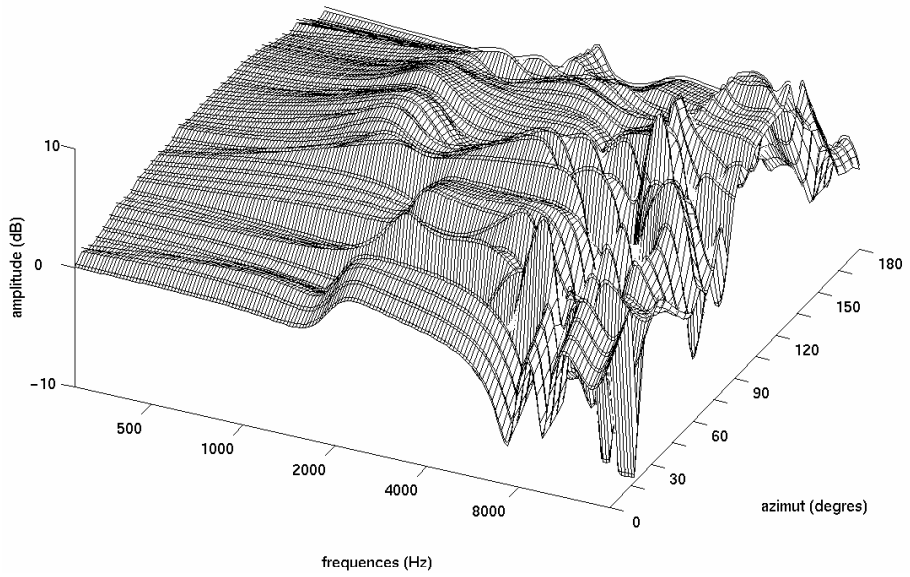
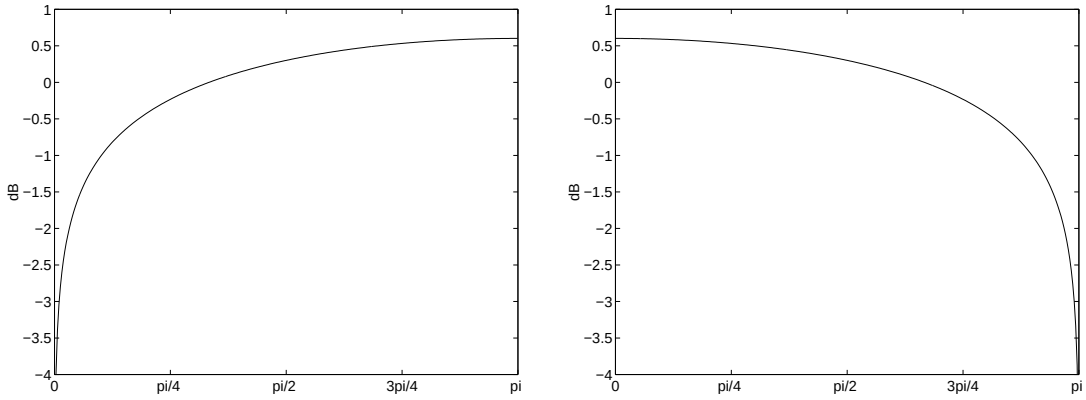


FIG. 2.32 – Spectres d'amplitude après interpolation sur les coefficients Armadillo : échantillonnage de départ de résolution 15°, et reconstruction par interpolation d'un échantillonnage de résolution 5°.



(a) (b)

FIG. 2.33 – Evolution du gain en dB aux fréquence extrêmes pour une interpolation linéaire pratiquée sur la fréquence des raies spectrales : gain à la fréquence nulle (a), gain à la fréquence π (b).

2.3.4.4 Interpolation sur la fréquence des raies spectrales (LSF)

Les LSF ont présenté de très bonnes performances pour l'interpolation de la structure transverse développée. Ce sont donc de bons candidats comme paramètres de contrôles pour la structure factorisée. En outre, pour de si faibles ordres des polynômes, il est aisé de manipuler les expressions analytiques des LSF, que nous désignerons par θ_1 et θ_2. Celles-ci sont associées aux polynômes $P(z)$ et $Q(z)$ suivants :

$$\begin{aligned} P(z) &= 1 + (c_1 + c_2).z^{-1} + (c_1 + c_2).z^{-2} + z^{-3} \\ Q(z) &= 1 + (c_1 - c_2).z^{-1} - (c_1 - c_2).z^{-2} - z^{-3} \end{aligned}$$

On obtient facilement que les racines de $P(z)$ (resp. $Q(z)$) sont $\left[-1, e^{i\theta_1}, e^{-i\theta_1}\right]$ (resp. $\left[1, e^{i\theta_2}, e^{-i\theta_2}\right]$), et l'on peut établir les relations suivantes :

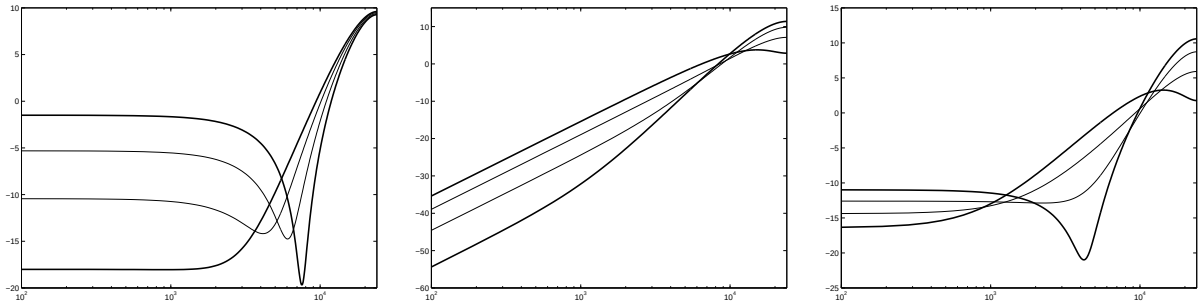
$$\begin{aligned} 2 - 2.\cos(\theta_1) &= 1 + c_1 + c_2 = g(0) \\ 2 + 2.\cos(\theta_2) &= 1 - c_1 + c_2 = g(\pi) \end{aligned}$$

L'interpolation linéaire de θ_1 et θ_2 fait donc évoluer indépendemment le gain en dB aux fréquences extrêmes. Toutefois, cette variation ne peut être linéaire en première approximation que pour les fortes valeurs de θ_1 et les faibles valeurs de θ_2, où les courbes peuvent être approximées par une droite (cf Figure 2.33). On peut vérifier que ces deux cas de figures apparaissent lorsque des coefficients (c_1,c_2) sont près des deux sommets supérieurs du triangle de stabilité. D'après la Figure 2.26, c'est un cas assez fréquent pour les HRTF. Toutefois, ce n'est pas suffisant pour garantir une interpolation satisfaisante des basses fréquences, comme le montre la Figure 2.35. En outre, les LSF n'offre aucun moyen de contrôle des extrema du spectres des polynômes. On retrouve ainsi les défauts de l'interpolation des coefficients des cellules, notamment sur le gain à la fréquence de résonance (Figure 2.34, a).

On peut alors conclure que la fréquence des raies spectrales ne constituent pas un paramètre de contrôle idéal pour l'interpolation des cellules d'ordre 2. Contrairement au cas de la structure transverse développée, pour laquelle les "défauts" d'interpolation demeuraient très faibles, les écarts induits sur les cellules se cumulent pour engendrer des artefacts majeurs sur le spectre global (Figure 2.35).

2.3.5 Interpolation pour une structure en treillis

La structure treillis, comme la structure transverse, offre la possibilité de pratiquer l'interpolation sur une forme "développée". Elle est toutefois deux fois plus chère, puisque le coût d'implantation est de $4.n$ opérations par échantillon, n représentant l'ordre du filtre IIR.



(a) (b) (c)

FIG. 2.34 – Spectres reconstruits par interpolation sur la fréquence des raies spectrales des cellules d'ordre 2 (mêmes conventions qu'en 2.29).

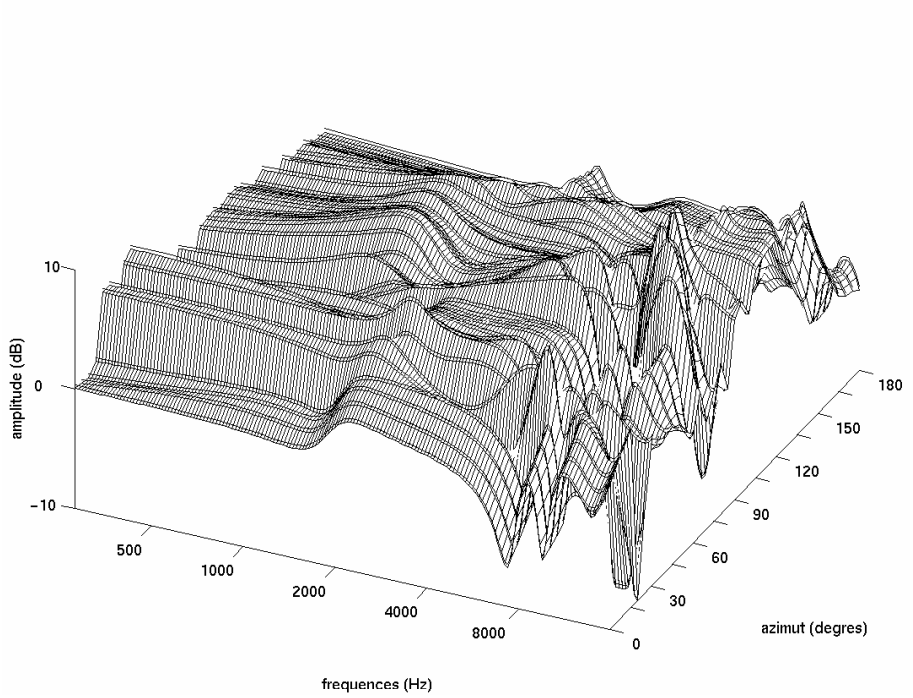


FIG. 2.35 – Spectres d'amplitude après interpolation sur la fréquence des raies spectrales associées aux coefficients de la structure transverse factorisée : échantillonnage de départ de résolution 15°, et reconstruction par interpolation d'un échantillonnage de résolution 5°.

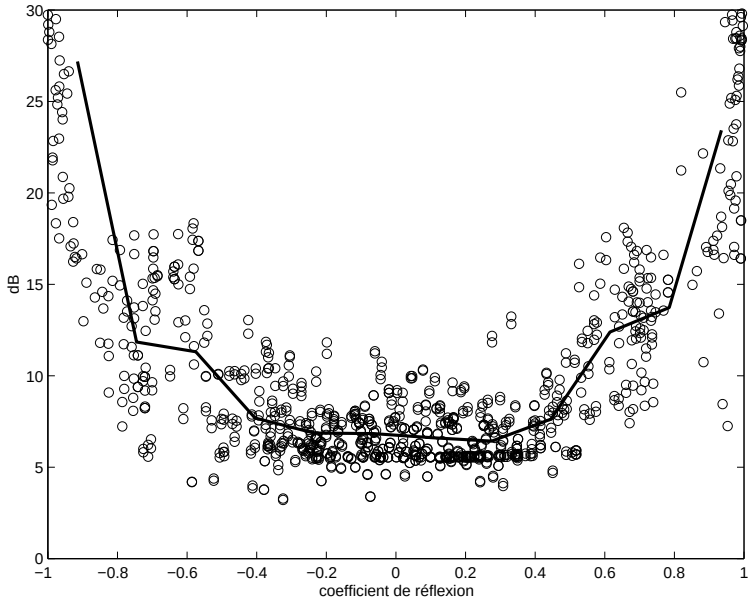


FIG. 2.36 – Courbe de sensibilité spectrale pour l'interpolation sur les coefficients de réflexion. On a superposé les courbes obtenues pour le numérateur et le dénominateur des modèles des HRTF.

2.3.5.1 coefficients de la structure treillis

Les coefficients de la structure en treillis, que nous noterons k_i, répondent aux contraintes que nous nous sommes fixées pour les paramètres de contrôle :

- la stabilité est garantie si les k_i ont un module inférieur à 1. L'interpolation entre deux jeux de coefficients définissant des filtres stables donne donc un filtre stable.
- les k_i sont naturellement ordonnés, et qui plus est cet ordre est "hiérarchique" : pour passer de l'ordre n à l'ordre $n+1$, le calcul d'un seul nouveau coefficient est nécessaire, les autres demeurant inchangés.

Une étude de la quantification scalaire des k_i a été proposée par Wisvanathan dans [VM75]. Comme lui (voir sa Figure 3), la sensibilité spectrale que nous obtenons pour les k_i présente une forme générale en U, indépendante de la valeur choisie pour l'écart Δc (Figure 2.36). Pour la représentation, nous fixons cet écart à 0.17, qui constitue l'écart moyen maximum observé sur les k_i pour un pas d'interpolation. La sensibilité prototype, présentée en trait épais, est symétrique par rapport à 0, et présente de fortes valeurs pour les valeurs de $|k_i|$ proches de 1, et de faibles valeurs autour de 0. Cela implique que la continuité des spectres d'amplitude interpolés dépend de la valeur des coefficients de réflexion de l'échantillonnage de départ. On remarque d'ailleurs que pour les HRTF, les fortes valeurs de $|k_i|$, facteurs donc de distorsion au niveau des spectres, sont confinées aux premiers rangs ($i = 1$ et 2). Ces défauts ne sont pas localisés en fréquence et se manifestent en Figure 2.37 par un affaissement des pics sur les états interpolés (par exemple entre 15° et 30° au dessus de 8kHz) ou bien par l'apparition de larges surtensions en basses fréquences (par exemple entre 90° et 105°). Afin de s'affranchir des défauts des coefficients de réflexion pour la quantification scalaire, Viswathan cherche une transformation non linéaire des k_i définissant des paramètres à sensibilité plate. Observant une bonne superposition de la sensibilité prototype des k_i et d'une parabole d'expression $10.\log \frac{1}{1-k_i^2}$, il aboutit à la transformation $f(k_i)$:

$$f(k_i) = \log \frac{1+k_i}{1-k_i}$$

Viswanathan rapproche ces nouveaux coefficients des "rapports d'aire" (area ratio) entre deux segments consécutifs du conduit vocal ($\frac{1+k_i}{1-k_i}$), paramètres déjà connus dans le domaine de la synthèse de parole. Il définit ainsi les Log Area Ratio, dont nous étudions les propriétés pour l'interpolation des HRTF en section suivante.

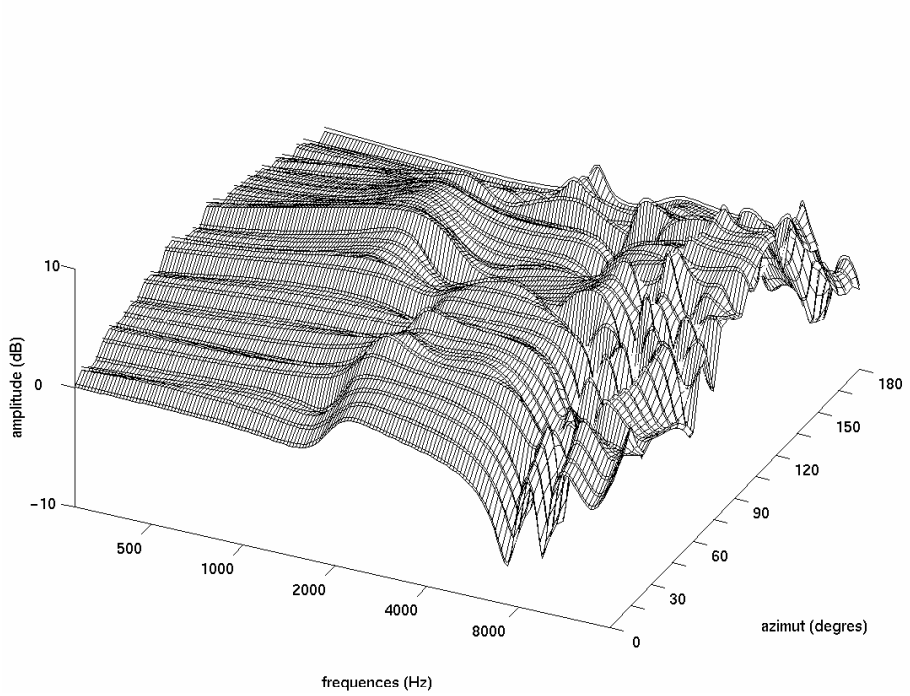


FIG. 2.37 – Spectres d'amplitude après interpolation sur les coefficients de réflexion : échantillonnage de départ de résolution 15°, et reconstruction par interpolation d'un échantillonnage de résolution 5°.

2.3.5.2 Log Area Ratio

Les Log Area Ratio (LAR) sont obtenus à partir des coefficients de réflexion k_i d'après la relation :

$$LAR_i = \log \frac{1 + k_i}{1 - k_i}$$

Ils prennent a priori leurs valeurs dans l'intervalle $[-\infty; +\infty]$, mais pour notre cas particulier, les LAR sont confinés au sein d'un intervalle borné, approximativement [-13.6 7.36]. En outre, comme l'illustre la Figure 2.38, la distribution des valeurs est très fortement concentrée entre -2 et 2 (près de 75%). C'est précisemment l'intervalle sur lequel la sensibilité est constante (Figure 2.39). Pour les valeurs plus extrêmes, correspondant aux k_i proches de 0, cette sensibilité est, par construction, très faible. Pour le calcul de la sensibilité, nous avons choisi 0.6 comme valeur d'écart, suivant la même démarche que pour les LSF ou les coefficients de réflexion.

On retrouve sur les spectres d'amplitude certains défauts observés pour l'interpolation des k_i, notamment en hautes fréquences. Toutefois, les artefacts aux plus basses fréquences ont presque disparu. L'interpolation des LAR semble ainsi plus efficace que celle des coefficients de réflexion. Elle demeure toutefois moins performante que l'interpolation sur les LSF.

Une première comparaison entre LAR et LSF peut être obtenue en observant le comportement induit sur les spectres par une interpolation entre 60° et 75° (Figures 2.41 et 2.21). Pour les deux familles de paramètres, l'interpolation sur les modèles d'ordre 12 engendre des spectres "débordants" des spectres de départ, et ne répond donc pas à nos critères de qualité. Le phénomène est plus prononcé dans le cas des LAR. Les performances sont nettement améliorée en augmentant l'ordre de modélisation. Toutefois, cette amélioration est observée sur toute la bande de fréquence dans le cas des LSF, alors qu'il est essentiellement localisé en hautes fréquences pour les LAR (au dessus de 6kHz).
Autre outil de comparaison : la trajectoire des pôles dans le plan horizontal (Figure 2.22). Les pôles introduits par les LAR ne s'écartent pas significativement des trajectoires tracées par les pôles de départ. Toutefois, comme on l'observe sur les intervalles indiqués par les flèches, une meilleure continuité est obtenue avec les LSF.

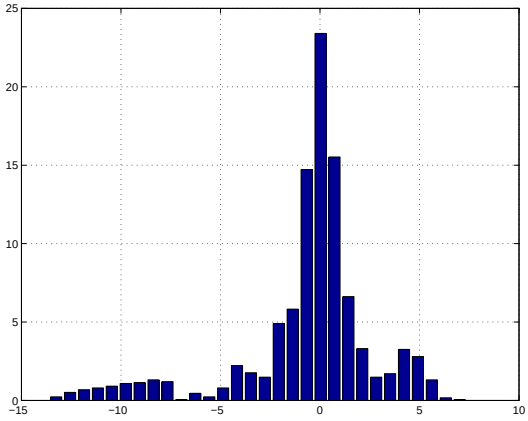


FIG. 2.38 – Distribution des Log Area Ratio obtenus par la modélisation IIR des HRTF dans le plan horizontal (pas de 5°).

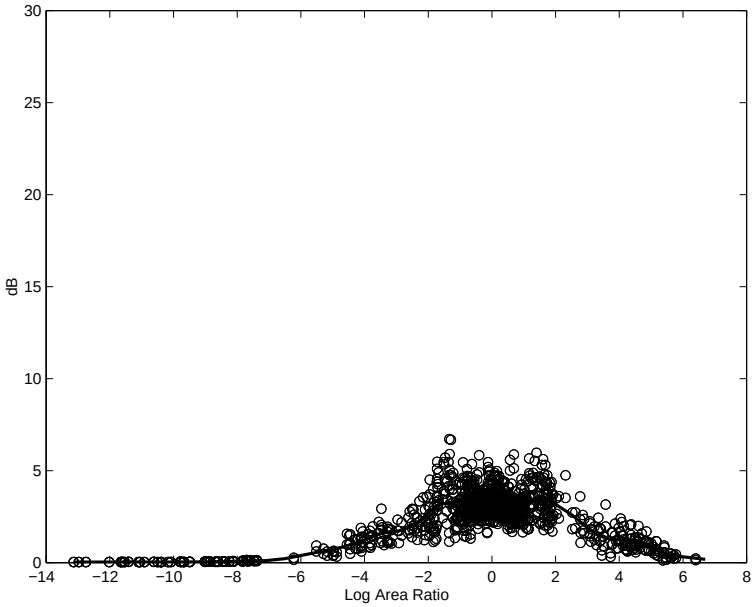


FIG. 2.39 – Courbe de sensibilité spectrale pour l'interpolation sur log area ratio. On a superposé les courbes obtenues pour le numérateur et le dénominateur des modèles des HRTF.

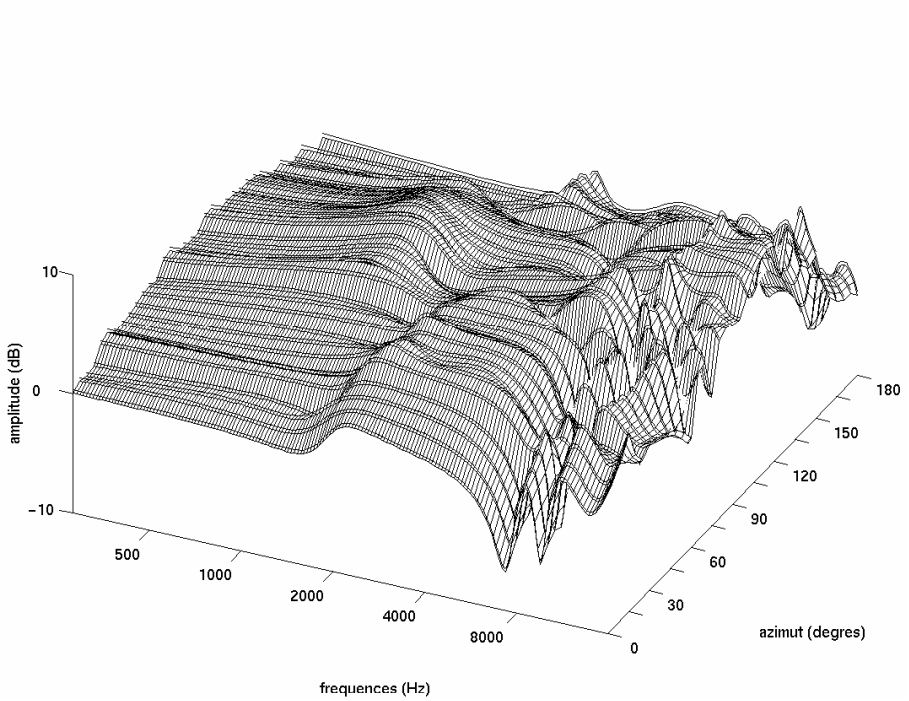
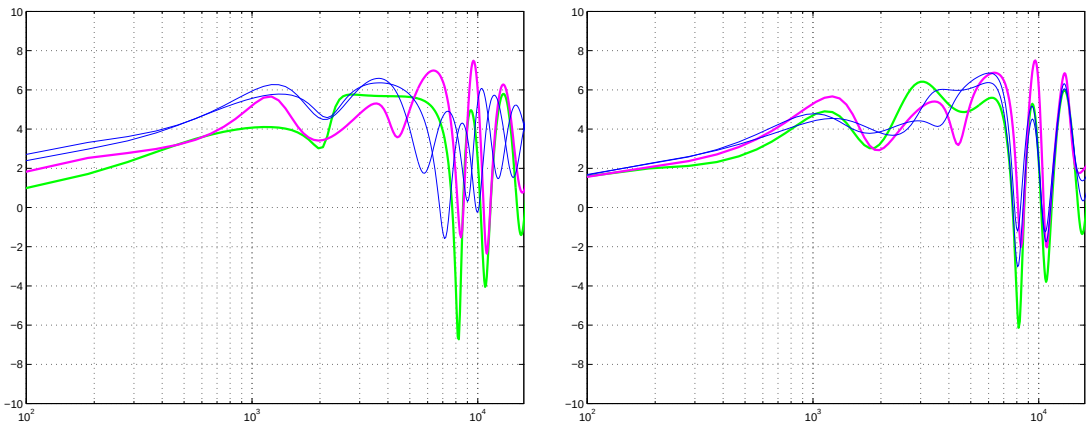


FIG. 2.40 – Spectres d'amplitude après interpolation sur les log area ratio : échantillonnage de départ de résolution 15°, et reconstruction par interpolation d'un échantillonnage de résolution 5°.



(a) (b)

FIG. 2.41 – Spectres d'amplitudes modélisés à 60° et 75° (trait épais) et états interpolés à 65° et 70° en prenant les Log Area Ratio comme paramètres de contrôle. Modélisation IIR d'ordre 12 (a), d'ordre 14 (b).

Une comparaison plus générale est menée pour les structures "développées", en observant l'erreur de recontruction aux moindres carrés des spectres d'amplitude (erreur L2). Cette erreur est calculée en référence aux HRTF mesurées aux mêmes position (après lissage au demi ton). Ce sont sur ces mêmes données qu'est pratiquée la modélisation IIR, dont l'efficacité est représentée en noir sur les Figures 2.42 et 2.43. Cette erreur de modélisation est celle que l'on entendrait en l'absence d'interpolation, i.e. si les positions mesurées et modélisées tous les 5° étaient utilisées. Le calcul est effectué pour les positions su plan horizontal : positions ipsi seulement, par cohérence avec les graphiques 3D représentés plus haut, et positions ipsi et contra, 72 en tout (Figure 2.42). L'erreur L∞ quant à elle n'est présentée que pour les positions ipsi, les fortes erreurs obtenues pour l'oreille contra pouvant masquer les performances réelles des méthodes.

De façon systématique, les performances des LSF sont meilleures que celles des LAR et des coefficients de réflexion. On remarque que l'introduction des positions contralatérales dans le calcul d'erreur introduit de fortes distorsions des spectres :

- sur toute la bande pour LAR et k_i,
- à partir de 1500Hz pour les LSF.

Il semble pourtant que le faible niveau des HRTF contralatérale rendre l'erreur à ces positions moins perceptible que celle des positions ipsilatérales. Nous nous concentrons donc sur les représentations les concernant.

D'une manière générale, l'erreur L2 est faible pour les trois méthodes (inférieure ou de l'ordre de grandeur de 3dB). Toutefois, cela ne doit pas nous faire oublier les artefacts "locaux" commentés aux paragraphes précédents, que l'erreur L∞ met en évidence. L'observation de cette erreur permet de généraliser certaines observations dégagées plus haut sur des cas particuliers.
Les LSF apparaissent comme les meilleurs paramètres de contrôle pour l'interpolation sur une structure IIR. L'erreur L∞ des LSF est de l'ordre de 1dB jusqu'à 3kHz, avec parfois de meilleures performances que la modélisation directe des mesures. Cela pourrait être le cas pour la mesure 60°, représentée en magenta sur les Figures 2.21 et 2.41 : le modèle d'ordre 12 agglomère deux résonances en une seule alors que l'interpolation LSF semble au contraire respecter la "structure" des filtres et pourrait donc par interpolation reconstruire les pics autonomes à 3kHz et 6kHz de la position 60°. Les performances des LSF transparaissent également à l'observation des pôles (Figure 2.13) : les trajectoires dessinées par les pôles introduits par l'interpolation LSF est parfois plus continue que celle définie par la modélisation. Des artefacts audibles (>2dB) apparaissent au dessus de 4kHz, limite qui est repoussée à 6kHz pour une modélisation à l'ordre 14. Pour cette dernière, l'erreur L∞ des LSF reste inférieure à ou de l'ordre de 3dB jusqu'à 14kHz, ce qui rend l'interpolation en temps-réel tout à fait envisageable : au delà de ces fréquences, des écarts du même ordre de grandeur peuvent provenir du mauvais positionnement de la tête lors de la mesure, de l'écart à la tête de l'auditeur dans le cas d'une simulation non individuelle, etc...
Les LAR sont moins performants que les LSF, sauf à l'odre 14 de modélisation, pour lequel LAR et LSF sont équivalent jusqu'à environt 4kHz. Toutefois, les LAR constituent systématiquement de meilleurs paramètres que les coefficients de réflexion. Cette amélioration est "localisée en fréquence". Elle porte en effet sur :

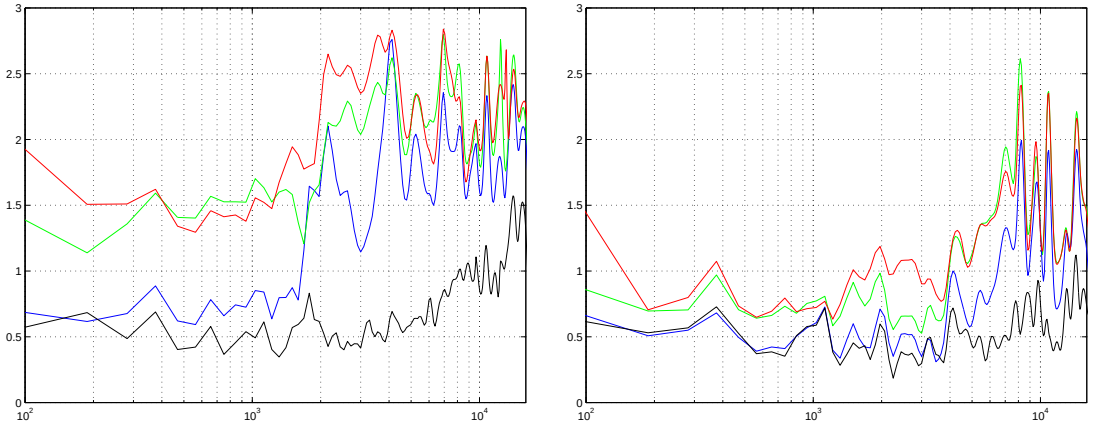
- les basses fréquences (au dessous de 200Hz), où les LAR offrent une reconstruction meilleure de 2dB environ,
- entre 2kHz et 4kHz, ou l'amélioration oscille entre 1 et 2dB.

Les performances en basses fréquences des LAR confirment les observations tirées plus haut de la Figure 2.40 : les surpressions avec l'interpolation sur les k_i sont fortement attenuées avec les LAR. En revanche, les artefacts situés en hautes fréquences sont très proches pour les deux familles de paramètres.

D'une manière générale, on obtient le classement : LSF $>$ LAR $>$ k_i, ces perfomances prenant une significativité différence en fonction de la fréquence et de l'ordre de modélisation. Ce classement est conforme avec le critère de sensibilité spectrale : les LSF ont la courbe de sensibilité la plus plate, suivie de celle des LAR, et enfin de celle des k_i.

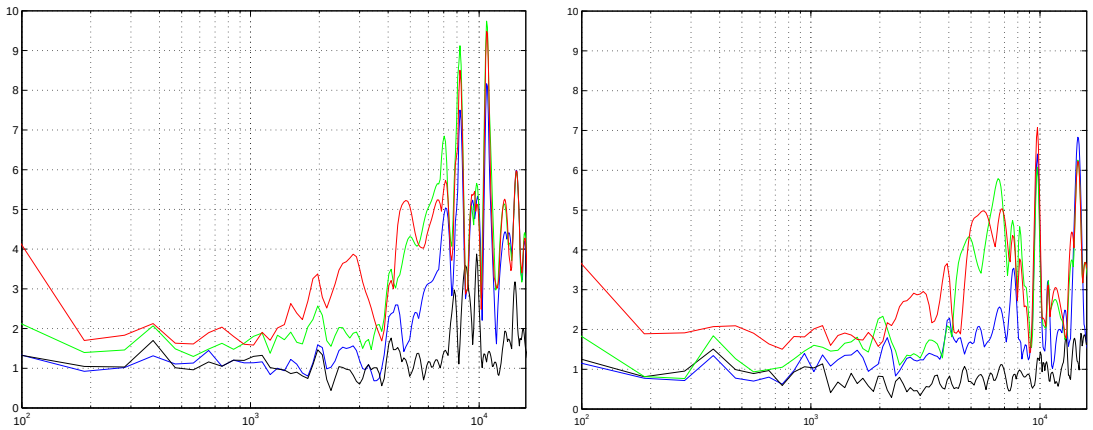
2.3.6 Conclusion sur l'interpolation locale des HRTF

Les coefficients de la représentation temporelle des HRTF à phase minimale constituent de bons paramètres de contrôle pour l'interpolation des HRTF, que celle-ci soit pratiquée "off-line" dans le seul but



(a) (b)

FIG. 2.42 – Comparaison de l'erreur de reconstruction aux moindres carrés des spectres d'amplitude mesurés et lissés au demi-ton pour les positions du plan horizontal (a), ou pour les positions ipsilatérales (b) : interpolation sur la fréquence des raies spectrales (en bleu) ; sur les coefficients de réflexion (en rouge) ; sur les Log Area Ratio (en vert). Erreur de modélisation IIR des spectres mesurés et lissé (en noir). La résolution angulaire est de 5°.



(a) (b)

FIG. 2.43 – Comparaison de l'erreur de reconstruction L∞ des spectres d'amplitude mesurés et lissés au demi-ton pour les positions ipsilatérales : mêmes conventions qu'en Figure 2.42. Modélisation IIR d'ordre 12 (a) ; modélisation IIR d'ordre 14 (b).

de recontruire une mesure manquante, ou en temps-réel avec une structure d'implantation FIR. Dans le cas d'une implantation IIR, on a pu montré les bonnes propriétés pour l'interpolation de la fréquence des raies spectrales pour une structure transverse, et, pour une structure en treillis, l'avantage des log-area-ratio sur les coefficients de réflexion. Ces résultats ont été évalués par l'observation de paramètres perceptivement pertinents (l'évolution linéaire des spectres en dB), ou de paramètres objectifs comme la comparaison à la mesure, ou le recours à des paramètres déjà utilisés pour la quantification scalaire. Nous ne sommes pas parvenus à proposer une méthode satisfaisante pour l'interpolation en temps-réel pratiquée sur une structure transverse factorisée, i.e. constituée de cellules d'ordre 2 en série. La difficulté provient de l'existence de racines réelles, expliquées par le warping appliqué à la modélisation. Elle est donc indépendente de l'ordre choisi pour la modélisation IIR.
On a néanmoins présenté une méthode pour ordonner les cellules dans l'objectif de minimiser le travail de l'interpolation. Cette stratégie a consisté à définir une partition des polynômes d'ordre 2 selon deux familles (et deux sous-familles) scindées sur des critères de "forme spectrale" (réponses en fréquence avec (anti)résonance ; réponses en fréquence monotones de type "shelving"). L'ordonnancement proposé est indépendant des paramètres de contrôle, ce qui permet une comparaison objective de différents candidats. Toutefois, comme on a pu le voir avec les coefficients armadillo, il pourrait être plus efficace d'étudier un ordonancement spécifique à chaque paramètre de contrôle.

2.4 Commutation des HRTF

Pour simuler des sources sonores en mouvement, il suffit de mettre à jour les coefficients des filtres de la synthèse binaurale en chargeant régulièrement les HRTF correspondant à la nouvelle position. Toutefois, une mise à jour brutale (ou commutation numérique) s'accompagne d'artefacts audibles sous forme de clics. Ce phénomène est caractéristique des filtres variants dans le temps, qui, à l'instant de la mise à jour engendrent des filtres transitoires instables. Il a été étudié par exemple dans le contexte du codage de la parole par prédiction linéaire ([VN86], [ZZ88]), celui de la synthèse sinusoïdale par filtres résonants ([Lar94b]) ou enfin celui de la synthèse par guide d'onde ([VLM95]).Pour notre application, la mise à jour des filtres doit assurer la cohérence entre le message sonore et l'interface visuelle. Cette cohérence est encore plus critique lorsque le mouvement est piloté par l'auditeur. Par exemple, avec un système de suivi de position de tête permettant decompenser les mouvement de l'auditeur pour simuler une source sonore immobile, la période de commutation doit être de l'ordre de 30ms (cas du *Spat~*).

Ces artefacts trouvent leur origine dans deux principaux mécanismes (e.g. [RC85]) :

- la transition entre les deux jeux de coefficients introduit une discontinuité dans le signal de sortie. Dans le cas d'un simple gain que l'on fait varier, cette transition se ramène à multiplier l'enveloppe spectrale du signal par un échellon de heaviside, ce qui se traduit par l'apparition d'une impulsion dans le signal temporel à l'instant de la commutation.
- dans le cas d'un filtre IIR, l'absence de ré-initialisation des mémoires internes du filtre lors de la transition entraîne la création d'échantillons de sortie éventuellement non bornés. Dans ces conditions en effet, les critères de stabilité "statique" des filtres ne peuvent plus être appliqués. En outre, le clic éventuellement généré est lui-même utilisé par la récurrence pour l'obtention des échantillons suivants, et devrait théoriquement se propager "à l'infini". Dans les faits, la durée du clic est de l'ordre de celle de la réponse impulsionnelle du filtre, soit dans notre cas environ 1ms.

L'effet produit est illustré en Figure 2.44. On observe que les deux types de filtrage, FIR ou IIR, génèrent des artefacts lors de la commutation. Dans le cas FIR, seul la discontinuité du signal de sortie est à mettre en cause. Dans le cas IIR en revanche, s'y ajoute la contribution liée aux mémoires de la partie récursive.

Plusieurs solutions ont été proposées dans la littérature pour s'affranchir de ces artefacts. Nous les regroupons en quatre familles :

1. ***Contraintes sur le filtrage***
   Une commutation idéale peut être obtenue en respectant certaines précautions. Pour les cellules du second ordre, par exemple, on peut montrer que la structure transverse couplée (Gold-Rader) et la structure en échelle normalisée ne peuvent générer de transitoires instables ([RC85], [Lar98]). Le

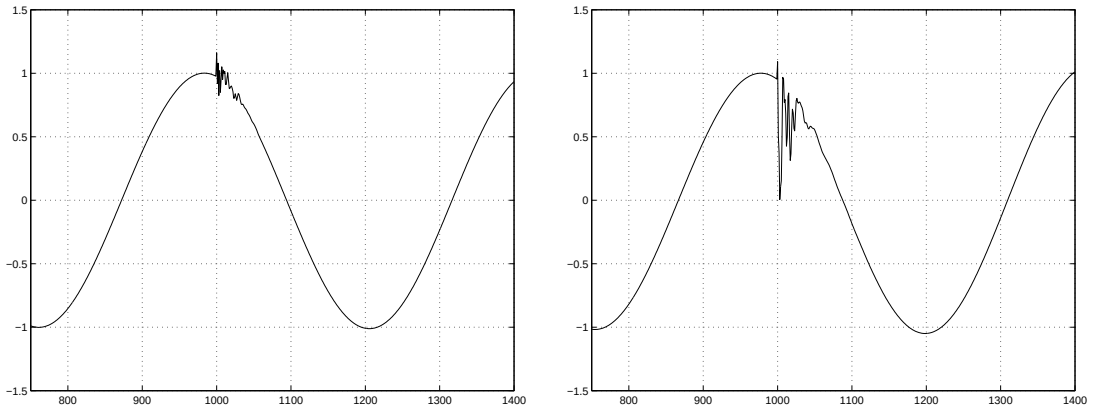
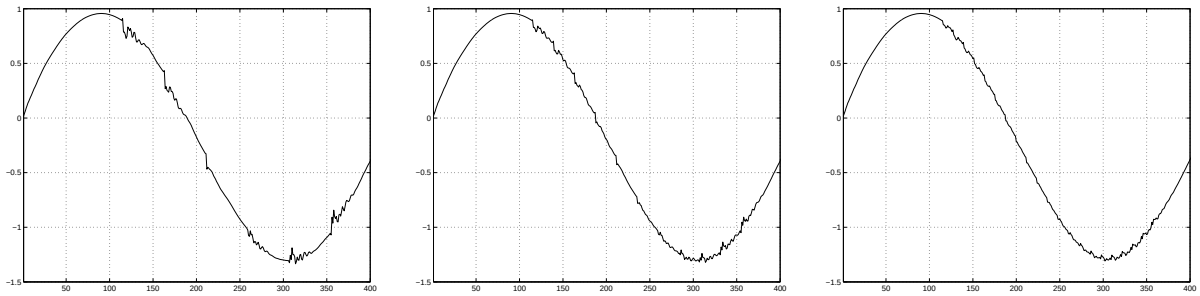


FIG. 2.44 – Artefacts crées par la commutation entre deux HRTF (positions 60° et 65° du plan horizontal) lorsque le signal d'entrée est une sinusoïde : implantation FIR (a) ; implantation IIR (b). Dans les deux cas, on considère une structure "Direct Form II Transposed", imposée par l'implantation de la routine de filtrage filter.m de Matlab.

coût de ces structures est toutefois beaucoup deux fois plus élevé que celui des structures transverse ou treillis traditionnelles.
Un autre type de contrainte peut porter sur les coefficients des filtres statiques utilisés. Un critère de stabilité "dynamique" a en effet été établi par Grenier dans le cas général d'un filtrage ARMA variant dans le temps ([Gre84]). Pour une famille de fonctions de transfert $\frac{B_n(z)}{A_n(z)}$, il définit le filtre tangent à $A_n(z)$, noté $\widetilde{A}_n(z)$, dont les pôles doivent être de module inférieur à 1 à tout instant n. Cette solution est indépendante de la structure de filtrage et du signal d'entrée, mais elle requiert la connaissance préalable de tous les filtres utilisés.

2. ***Commutation par remplacement de l'état interne***
Utilisant une représentation d'état des filtres, Zetterberg établit l'expression du vecteur d'état $x(n)$ lors d'un changement instantané des coefficients ([ZZ88]). Il isole ainsi la contribution de la réponse transitoire du système, qui dépend des échantillons d'entrée jusqu'à l'instant de transition et des matrices de la représentation d'état. La solution de Zetterberg consiste alors à modifier le vecteur d'état lors de la transition en lui ajoutant la quantité opposée. Välimäki adapte cette approche en réduisant les besoins en mémoire. Il estime un vecteur d'état $x_2(n)$ obtenu par le filtrage stationnaire avec les coefficients cibles ; puis il substitue $x_2(n)$ à $x(n)$. L'estimation de $x_2(n)$ requiert une durée d'observation égale à la longueur de la réponse impulsionnelle de ce deuxième filtre ([VLM95], [Vin97]).

3. ***Technique de superposition***
Cette approche, présentée dans [VN86], consiste à implanter deux structures de filtrages parallèles et à basculer le signal d'entrée d'un filtre à l'autre. A la commutation, le signal d'entrée est envoyé sur l'un des deux filtres, tandis que l'autre continue de tourner à vide et finit par s'éteindre. Lorsque les mémoires de ce filtre sont nulles, la structure est prête à accueillir de nouveaux coefficients, et à filtrer le signal d'entrée. La méthode n'est efficace que si la période de commutation est supérieure à la réponse impulsionnelle des filtres, ce qui est notre cas. C'est une technique très voisine qui est utilisée dans le *Spat˜* ([JLW95]) : au lieu de basculer soudainement de signal d'entrée (loi du tout ou rien), on utilise une rampe de transition de 30ms (une rampe descentante pour la structure contenant le filtre courant et une rampe montante pour la structure contenant le filtre cible), ce qui lui vaut le titre de ***Technique de fondu-enchaîné (cross-fade)***. La rampe peut également être placée en entrée des filtres. Dans les deux cas, on introduit ainsi des états intermédiaires, équivalents aux résultats de l'interpolation des HRTF à phase mixte entre la position courante et la position cible. Comme on l'a vu en section 2.3, les résultats de cette interpolation sont médiocres, tant sur les spectres d'amplitude que sur l'ITD. Toutefois, les HRTF utilisées dans le *Spat˜* étant écartées de 5° environ, et interpolées avec un grain très fin (30ms à 44100Hz conduisent à plus de 1300 états intermédiaires), ces artefacts ne sont pas audibles. Le problème de cette approche demeure son coût



5 états intermédiaires | 10 états intermédiaires | 20 états intermédiaires

FIG. 2.45 – Commutation "fractionnée" pour une transition entre deux HRTF du plan horizontal (azimuts 60° et 65°) sur une durée de 5ms, pour une fréquence d'échantillonnage 48kHz. Les paramètres de contrôle sont constitués de la fréquence des raies spectrales.

de calcul, puisqu'elle réclame le doublement de la structure utilisée dans le cas statique.

4. ***Technique de commutation "fractionnée"***
   Cette approche consiste à lisser la transition en interpolant les coefficients de la structure (et non plus les signaux comme dans le cas précédent). L'objectif est ici de répartir la nuisance pour qu'elle devienne inaudible ([MKBG90]). Cela suppose le choix approprié de paramètres de contrôle, et celui de la durée de lissage et du pas de transition. Ces deux derniers paramètres dépendent fortement du signal d'entrée : les transitoires instables s'entendent plus facilement sur une sinusoïde qu'avec des signaux larges bandes tels que la voix ou la musique, pour lesquels des phénomènes de masquage entrent en jeu. Nous illustrons en Figure 2.45 l'efficacité de cette méthode pour la commutation des HRTF. A la fréquence d'échantillonnage considérée, 48kHz, les différentes transitions comparées correspondent respectivement à un état intermédiaire tous les 12, 6 et 3 échantillons. On constate que pour que les artefacts disparaissent, il faudrait calculer un état intermédiaire à chaque nouvel échantillon, ce qui renvient à doubler le coût de calcul par rapport à l'implantation statique. Pour cet exemple, la fréquence des raies spectrales a été choisie comme paramètre de contrôle car les transitoires produits sont d'une amplitude moindre que dans le cas de la commutation sur les coefficients de la structure transverse.

L'implantation dynamique de la synthèse binaurale requiert donc la mise en place de l'une de ces méthodes pour éliminer ou masquer les artefacts de commutation, et revient dans le cas général à doubler le coût de calcul du filtrage statique. La charge induite pour chaque source supplémentaire nous conduira à chercher de nouvelles techniques d'implantation de la synthèse binaurale en chapitre 3.

2.5 Conclusion

Dans ce chapitre ont été examinées les différentes étapes nécessaires pour réaliser une implantation bi-canale dynamique de la synthèse binaurale. Certains résultats ont été intégrés à la version 1.2.2 du Spatialisateur de l'Ircam, notamment, la modélisation des spectres à phase minimale, réalisée sous forme IIR à l'ordre 12 par la méthode de Steiglitz-Mc Bride. Le retard variable est implanté sous forme FIR, par l'objet standard vd~ du langage Max . Enfin, la commutation des HRTF est gérée par superposition de deux structures de filtrage bicanales, entre lesquelles est réalisé un fondu-enchaîné.
Le choix d'une structure IIR s'appuie essentiellement sur des critères de coût de calcul ainsi que sur le contrôle qu'elle autorise sur la qualité de modélisation en basses fréquences. Toutefois, les comparaisons menées dans ce chapitre mettent en évidence les avantages d'une structure FIR pour l'interpolation, qui permettrait d'alléger la charge en mémoire nécessaire pour stocker les filtres.

Nous n'avons considéré ici qu'une interpolation globale, i.e. ne s'appuyant que sur les filtres à proximité de la position cible. Cette approche a le mérite d'être assez générale et de pouvoir s'appliquer à d'autres domaines, voire à d'autres objectifs tels que la quantification. Nous ne saurions néanmoins négliger l'efficacité des méthodes d'interpolation globale, s'appuyant sur toute la famille de filtres. Elles consistent

en général à décomposer les HRTF sur des fonctions de base, par des méthodes semblables à celles que nous examinons en chapitre 3. Les gains de pondération pour les positions manquantes sont extrapolées "par continuité" de leur forme globale (méthodes à base de "spline" : [CVVH96], [HBS99]), ou bien sont données par une expression analytique exacte, par exemple dans le cas d'une décomposition sur les harmoniques sphériques. Cette dernière méthode requiert le respect de critères d'échantillonnage stricts ([Lab00]), mais peut être implantée en temps réel ([Moh97]).

Chapitre 3

Synthèse binaurale multicanale

3.1 Introduction

Dans ce chapitre, nous envisageons la simulation binaurale de plusieurs sources sonores. Avec une structure d'implantation bicanale, telle que nous l'avons étudiée au chapitre 2, le traitement de chaque source supplémentaire requiert l'implantion de 4 nouveaux filtres. La charge de calcul engendrée devient rapidement prohibitive.
Une alternative repose sur l'implantation multicanale de la synthèse binaurale, proposée et décrite par Chen et al. dans [CVVH92]. Dans cette approche, les dépendances spatiales et fréquentielles des HRTF sont séparées grâce à une décomposition linéaire des HRTF. Celles-ci s'écrivent alors comme une somme de fonctions spatiales $C_i(\theta)$ et de filtres de reconstruction $L_i(f)$:

$$HRTF(\theta, f) = \sum_i C_i(\theta).L_i(f)$$

Cette représentation nous permet de définir un encodeur binaural multicanal, qui pondère un son monophonique incident par la valeur de chaque fonction spatiale pour la position cible. Dans le cas où la décomposition est réalisée sur les HRTF à phase minimale, l'encodeur comporte également l'implantation du retard interaural. Le décodeur associé consiste en un banc de filtres en parallèle, les filtres de reconstruction. Cette structure d'implantation est illustrée en Figures 3.1 et 3.2.
L'approche "pré-filtering", présentée en Figure 3.1 et décrite dans [Mar96b], consiste à commencer par l'étape de filtrage, pour créer des canaux audio qui sont éventuellement stockés avant la spatialisation, qui ne consiste alors qu'en quelques pondérations et mixages. L'approche "post-filtering", quant à elle, est avantageuse pour la synthèse binaurale multi-source puisque le traitement de chaque source supplémentaire ne requiert qu'un nouvel encodeur, consistué de gains et éventuellement d'un retard variable, tandis que le filtrage, réalisé par le décodeur, est partagé par toutes les sources. En outre, l'extrapolation des fonctions spatiales à des positions n'appartenant pas à l'échantillonnage de mesure des HRTF, permet de recréer des positions manquantes et constitue une solution au problème de l'interpolation global des HRTF. Enfin, cette décomposition linéaire fournit une représentation compacte des HRTF mesurées sur une tête, que nous mettrons à profit pour l'étude des différences interindividuelles au chapitre 5.

Différentes méthodes ont été proposées pour déterminer les fonctions spatiales et les filtres de reconstruction. Dans ce chapitre, nous étudions ces alternatives, et proposons une méthode originale, s'appuyant sur les statistiques d'ordre supérieure pour la recherche de fonctions spatiales discrètes. Les éléments exposés reprennent pour partie l'article [LJGW00].

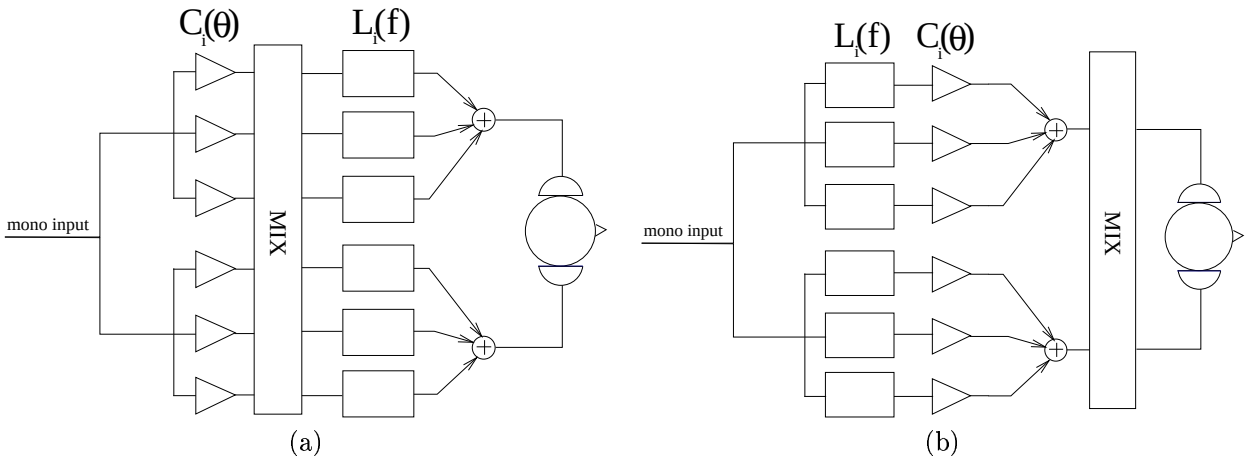


FIG. 3.1 – Implantation multicanale de la synthèse binaurale : décomposition linéaire des HRTF à phase mixte. Approche post-filtrage (a) et approche pré-filtrage (b).

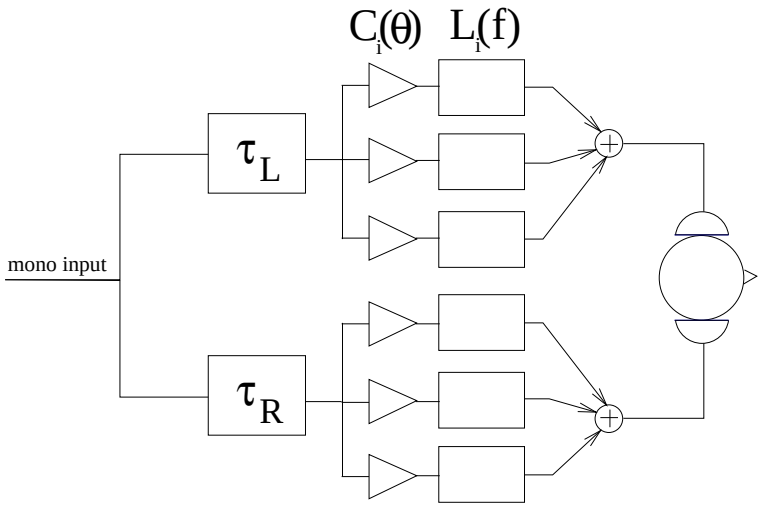


FIG. 3.2 – Implantation multicanale de la synthèse binaurale : décomposition linéaire des HRTF à phase minimale.

3.2 Formalisation du problème et des objectifs

3.2.1 Formulation matricielle

3.2.1.1 Notations

Soit H la matrice $N \times M$ contenant les spectres complexes des HRTF, mesurées pour N positions et M points fréquentiels. Nous pouvons indifféremment décrire H comme une matrice de N filtres ou comme une matrice de M directivités (fonctions spatiales).
Décomposer linéairement les HRTF consiste à identifier la matrice C des fonctions spatiales, de dimension $N \times r$, ainsi que la matrice L des filtres de reconstruction, de dimension $r \times M$ telles que l'on puisse approximer H par : $H = C.L$. Pour une représentation efficace, les fonctions spatiales (resp. les filtres de reconstruction) doivent être linéairement indépendantes, ce qui revient à imposer C et L de rang r.

3.2.1.2 Définition de produits scalaires

Les données que nous manipulons sont discrètes, échantillonnées dans l'espace ainsi qu'en fréquence.

Nous définissons le produit scalaire entre deux vecteurs colonnes a et b par :

$$[a|b] = b^{\dagger}.a$$

où l'opérateur † désigne la trans-conjugaison.
Ce produit scalaire définit la norme aux moindres carrés "usuelle" :

$$|a|^2 = [a|a]$$

Par extension, nous utiliserons la notation $[A|B]$ pour le produit matriciel $B^{\dagger}.A$, qui contient tous les produits scalaires possibles entre les colonnes de A et celles de B.

Nous définissons également un produit scalaire entre matrices :

$$[A||B] = \sum diag\,[A|B] = \sum diag(B^{\dagger}A)$$

où $diag(A)$désigne les éléments de la diagonale principale de la matrice A.
Ce produit scalaire conduit à la définition de la norme de Frobenius :

$$\|A\|^2 = \sum diag(A^{\dagger}A) = \sum_i \sum_j |a_{i,j}|^2 = \sum_i |A_i|^2 = \sum_j \left|A^{j,\dagger}\right|^2$$

où les $a_{i,j}$ désignent les éléments de A, les A_i $(i = 1....M)$ ses colonnes, et les $A^j (j = 1....N)$ ses lignes.
Dans la suite, nous désignons par A_i les colonnes de la matrice A et par A^i ses lignes.

3.2.1.3 Relations entre L, C et H

On fixe comme objectif de la décomposition linéaire des HRTF la minimisation de l'erreur aux moindres carrés $\|H - C.L\|^2$. Sous cette contrainte, L peut être définie à partir de H et de C : il suffit que les directivité reconstruites (colonnes de $C.L$) soient les projections orthogonales des directivités initiales (colonnes de H) sur l'espace engendré par les directivités de base, ou fonctions spatiales (colonnes de C). De manière analogue, C peut être obtenue par projection orthogonale de H sur L. Comme nous aurons souvent l'occasion d'utiliser ces relations, nous prenons le temps de les expliciter et de les interpréter.

1. **Déterminer L à partir de H et C**
   Nous disposons donc des fonctions spatiales, constituant les colonnes de C, et représentant une base de décomposition des fonctions de directivité de l'oreille, fréquence par fréquence (colonnes de H). Il est aisé d'exprimer la contrainte de minimisation de l'erreur aux moindres carrés pour chaque fréquence $|H_i - C.L_i|^2$: il est nécessaire et suffisant que l'erreur de décomposition soit orthogonale aux vecteurs de la base, ce qui s'exprime par :

$$\forall i \in [1...M] \, et \, \forall k \in [1...r] : \quad [H_i - C.L_i | C_k] = 0$$

   soit :

$$[H - C.L|C] = 0_{r \times M}$$

   ce qui conduit à :

$$L = G_C^{-1}.[H|C]$$

   où G_C désigne la matrice de Gram associée à C :

$$G_C = [C|C]$$

   Par construction, G_C est symétrique, de dimension $r \times r$. En outre, comme C est de rang r, la matrice de Gram associée est inversible. Si de plus les fonctions spatiales sont orthogonales, alors cette matrice est diagonale.
   Les filtres de reconstructions ainsi définis minimisent également la norme de Frobenius de l'erreur $\|H - C.L\|^2$, puisque l'on a la relation :

$$\|H - C.L\|^2 = \sum_{i=1}^{M} |H_i - C.L_i|^2$$

   On peut ainsi interpréter la norme de Frobenius comme la somme des erreurs élémentaires calculées pour chaque fréquence. Minimiser l'erreur au moindre carré de la décomposition de chaque fonction spatiale de l'oreille conduit alors naturellement à minimiser l'erreur de Frobenius.
2. **Déterminer C à partir de H et L**
   De façon très semblable au cas précédent, on peut montrer que nous obtenons :

$$C = \left[L^\dagger | H^\dagger\right] .G_L^{\dagger -1}$$

   où $G_L^\dagger$désigne la matrice de Gram associée à $L^\dagger$, et est définie par :

$$G_L^\dagger = \left[L^\dagger | L^\dagger\right] = [L|L]^\dagger$$

   On peut noter que si l'on stocke dans H et L les filtres avec leur symétrie hermitienne autour de la fréquence de Nyquist, alors, les fonctions spatiales obtenues sont réelles, de même que G_L.
   A nouveau, ces fonctions spatiales minimisent la norme de Frobenius de l'erreur. Cette fois, l'erreur élémentaire aux moindres carrés est calculée pour chaque position, et la norme de Frobenius somme cette erreur sur l'ensemble des positions.
   Par conséquent, les erreurs aux moindres carrés que nous représenterons par la suite dépendront ou bien de la fréquence, ou bien de la position. Une évaluation plus compacte pourra être fournie par la norme de Frobenius de l'erreur.

3.2.1.4 Invariance de la décomposition par transformation linéaire orthogonale

Les filtres de reconstruction obtenus par décomposition d'une transformée linéaire orthogonale de H, sont les transformées des filtres obtenus par décomposition directe de H. En outre, l'erreur de reconstruction aux moindres carrés, mesurée par la norme de Frobenius, est invariante par transformée linaire orthogonale de H. En effet, soit h l'image de H par la transformation linéaire A de dimension $M \times M$:

$$h = H.A$$

Ses filtres de reconstruction l sont donnés par :

$$l = G_c^{-1}.[H.A|C] = G_c^{-1}.C^\dagger.H.A = G_c^{-1}.[H|C].A$$

Soit :

$$l = L.A$$

En outre :

$$\forall j \in [1...N] \quad h^j = H^j.A$$

$$\|h - C.l\|^2 = \sum_{j=1}^{N} \left|(h^j - C^j.l)^\dagger\right|^2 = \sum_{j=1}^{N} \left|(H^j.A - C^j.L.A)^\dagger\right|^2$$

$$= \sum_{j=1}^{N} (H^j - C^j.L).A.A^\dagger (H^j - C^j.L)^\dagger$$

Si la transformation est orthogonale :

$$A.A^\dagger = 1_{M \times M}$$

Donc :

$$\|h - C.l\|^2 = \sum_{j=1}^{N} \left|(H^j - C^j.l)^\dagger\right| = \|H - C.L\|^2$$

Ce résultat montre notamment que décomposer une représentation complexe des HRTF ou bien la représentation temporelle associée conduit aux mêmes performances. Ce cas particulier correspond à une matrice A définie par :

$$A = \frac{1}{M}. \begin{bmatrix} 1 & ... & 1 & ... & 1 \\ ... & ... & ... & ... & ... \\ 1 & ... & e^{\frac{2j\pi}{M}(k-1).(m-1)} & ... & e^{\frac{2j\pi}{M}(k-1).(M-1)} \\ ... & ... & ... & ... & ... \\ 1 & ... & e^{\frac{2j\pi}{M}(M-1).(m-1)} & ... & e^{\frac{2j\pi}{M}(M-1).(M-1)} \end{bmatrix}$$

Cette propriété, démontrée pour une tranformation linéaire combinant les colonnes de H, s'étend à une transformation linéaire sur ses lignes.

3.2.2 Définition de contraintes pour la prise en compte de l'élévation

Les HRTF dépendent de la fréquence f, de l'azimut φ et de l'élévation, repérée par l'angle θ du repère sphérique. Nous souhaitons imposer à la décomposition linéaire des HRTF une contrainte pour privilégier la reconstruction dans le plan horizontal, au détriment des positions élevées. Deux principales raisons justifient ce choix : l'acuité de localisation du système auditif est supérieure pour les positions du plan horizontal, et, pour les applications multimédia visées, les sources se situent essentiellement dans le plan horizontal.
Si N désigne le nombre de canaux fixé pour la décomposition linéaire des HRTF, cette contrainte se formalise en 3 étapes.

1. la reproduction des HRTF du plan horizontal est assurée par $N-1$ canaux :

$$H(\varphi, \theta = 0, f) = \sum_{i=1}^{N-1} C_i(\varphi).L_i(f) = H_h(\varphi, f)$$

2. la représentation des HRTF dans le plan horizontal, $H_h(\varphi, f)$, contribue à la reconstruction des HRTF élevées :
$$H(\varphi, \theta, f) = Y(\theta).H_h(\varphi, f) + e(\varphi, \theta, f)$$
La fonction spatiale $Y(\theta)$ peut éventuellement être spécifique pour chaque canal : $Y_1(\theta)$,..., $Y_{N-1}(\theta)$. Cette flexibilité supplémentaire permet une meilleure reconstruction, mais peut être difficile à mettre en oeuvre.
3. le N^{eme}canal est consacré à la reproduction des positions élevées :
$$e(\varphi, \theta, f) = Z(\theta).H_z(f)$$

Finalement, on peut écrire :
$$H = Y(\theta).H_h(\varphi, f) + Z(\theta).H_z(f)$$
avec :
$$H_h(\varphi, f) = \sum_{i=1}^{N-1} C_i(\varphi).L_i(f)$$

On constate que les fonctions spatiales ainsi définies sont à variables séparées, ce qui présente un avantage pratique pour le stockage et l'accès aux tables en mémoire.

3.2.3 Critères de performances

Dans les sections 3.3, 3.4 et 3.5, nous étudions trois approches pour définir les fonctions spatiales et les filtres de reconstruction. Nous les comparons en section 3.6 sous trois aspects :

1. **Qualité de reconstruction des HRTF.** Une mesure de celle-ci est donnée par l'erreur aux moindres carrés définie plus haut. Par la suite, on calcule l'erreur de reconstruction complexe suivante :
$$e(f) = \left\{ \frac{|H_i - C.L_i|}{|H_i|} \right\}_{i \in [1 \ldots M]}$$
La moyenne sur les positions fait intervenir une pondération en proportion à l'angle solide de chaque position.
2. **Efficacité de l'implantation de l'encodeur.** Un encodeur est "efficace" si le nombre de canaux à transmettre au décodeur est minimum. Ce nombre dépend de deux paramètres :
   (a) Encode-t-on séparement les deux oreilles, i.e. encode-t'on explicitement l'ITD ?
   (b) quel est le nombre de gains non nuls pour chaque position ?

   Le cas idéal est obtenu dans le cas où l'on ne sépare pas les deux oreilles et où les fonctions spatiales sont discrètes, i.e. à support disjoint. La mesure de cette dicrétude ou compacité des fonctions spatiales sera définie à l'aide des moments d'ordre supérieurs en section 3.
3. **Universalité des fonctions spatiales.** Cette propriété est atteinte si les fonctions spatiales sont communes à toutes les têtes.

3.3 Optimisation conjointe des fonctions spatiales et des filtres de reconstruction

Les techniques d'analyse statistique, telles que l'Analyse en Composantes Principales (ACP) et l'Analyse en Composantes Indépendantes (ACI), fournissent la représentation compacte d'un ensemble de variables par un ensemble réduit de nouvelles variables. Cette réduction de données est obtenue en éliminant l'information partagées par les variables de départ, et en répartissant l'information résiduelle sur un nombre réduit de variables indépendantes. Les deux analyses diffèrent dans le niveau d'indépendance atteint : l'ACP assure une indépendance au second ordre et définit ainsi de nouvelles variables décorrélées, tandis

que l'ACI permet d'atteindre une indépendance aux ordre supérieurs.

La recherche de variables décorrélées pour engendrer les HRTF par combinaison linéaire garantit une minimisation de l'erreur aux moindres carrées de l'erreur de reconstruction. Les premières applications de l'ACP sur les HRTF visaient la décomposition des spectres d'amplitude ([Mar87], [MG92],[WK93]). Chen a été le premier à appliquer l'expansion de Kahunen Leove, nom fréquemment donné à l'ACP dans les communications, aux spectres complexes des HRTF à phase mixte ([CVVH92]). Cette décomposition conduit tout naturellement à la structure d'implantation présentée en Figure 3.1, qu'il a brevetée ([CVVH96], voir aussi [AF97], [APS97]).

L'attrait exercé par l'ACI repose sur le lien que nous établissons en section 3.3.2.1 entre indépendance statistique et compacité de support. Emerit l'a appliquée aux spectres d'amplitude des HRTF et a ainsi dégagé des filtres à supports fréquentiels disjoints ([EM95]). Dudouet applique l'ACI à la représentation temporelle des HRTF et dégage une base de réponses impulsionnelles à support compact ([DM98]). Dans la perspective d'optimiser le coût de calcul de notre encodeur multicanal, nous appliquons l'ACI sur les spectres complexes des HRTF, afin d'obtenir des fonctions spatiales discrètes.

3.3.1 Relation formelle entre ACP et ACI

3.3.1.1 Définition et mesure de l'indépendence

Soient $\{x_1, x_2, ..., x_M\}$ M variables centrées dont on possède N observations. Soient p_{x_i} la densité de probabilité de la variable x_i, et p_x la densité de probabilité conjointe de la variable $x = [x_1\, x_2 ... x_M]$. D'un point de vue pratique, on peut voir chaque variable x_i comme un vecteur de dimension N et x comme une matrice $N \times M$. Les variables x_i sont indépendentes si et seulement si :

$$p_x(u) = \prod_{i=1}^{M} p_{x_i}(u_i) \tag{3.1}$$

Dans la définition rappelée par Hyvärinen ([Hyv99]) : ***mener une ACI sur un vecteur aléatoire*** x ***consiste à rechercher une transformation linéaire telle que les variables*** y_i ***constituant le vecteur image*** $y = [y_1 ... y_M]$ ***soient aussi indépendentes que possibles, au sens d'une fonction*** $\phi(y)$ ***mesurant l'indépendence***. En effet, utiliser la relation 3.1 comme critère d'évaluation de la transformation linéaire considérée est difficile, puisque souvent on ne dispose pas de la densité de probabilité de nos variables. L'ACI est donc une démarche d'optimisation, qui requiert le choix d'une mesure de l'indépendence des variables, qu'il s'agit de maximiser. Comon ([Com94]) définit les propriétés d'une telle mesure, qu'il baptise "fonction de contraste". Une fonction de constraste ϕ est une application de l'espace des densités de probabilités $\{p_x,\, x = [x_1, ..., x_M]\}$ vers l'ensemble des réels, satisfaisant trois conditions :

1. $\phi(p_x)$ **est invariante par permutation circulaire des variables** x_i **:**

$$\phi(p_{A.x}) = \phi(p_x) \qquad \forall A \quad \textbf{permutation}$$

2. $\phi(p_x)$**est invariante par changement d'échelle :**

$$\phi(p_{A.x}) = \phi(p_x) \qquad \forall A \quad \textbf{matrice diagonale inversible}$$

3. **si** x **a des composantes indépendantes, alors :**

$$\phi(p_{A.x}) \leq \phi(p_x) \qquad \forall A \quad \textbf{inversible}$$

Plusieurs fonctions de contraste sont utilisées dans la littérature. L'une d'entre elles est définie à partir de la notion d'information mutuelle par Comon ([Com94]). Nous reprenons ici les principales étapes de son raisonnement, qui nous permettrons de souligner la relation entre ACP et ACI.

3.3.1.2 Minimisation de l'information mutuelle

La définition d'indépendance 3.1 peut être reformulée en exprimant la nullité de la distance entre les deux termes de l'égalité. La distance entre densités de probabilité est souvent définies à l'aide de la divergence de Kullback-Leibler :

$$\delta(p_x, p_z) = \int p_x(u).\log(\frac{p_x(u)}{p_z(u)})du$$

Nous obtenons ainsi :

$$\delta(p_x, \prod_{i=1}^{M} p_{x_i}(u_i)) = \int p_x(u).\log\frac{p_x(u)}{\prod p_{x_i}(u_i)}du$$

Cette "distance"[1] définit l'information mutuelle moyenne de x, $I(p_x)$. Celle-ci est nulle si et seulement si les variables sont indépendentes, et est strictement positive sinon. On peut ainsi définir la fonction de contraste ϕ^{IM} par :

$$\phi^{IM}(x) = -I(p_x)$$

L'information mutuelle $I(p_x)$ peut être décomposée en plusieurs contributions :

$$I(p_x) = J(p_x) - \sum_{i=1}^{M} J(p_{x_i}) + I(\phi_x) \tag{3.2}$$

L'équation3.2 fait intervenir la notion de négentropie $J(p_x)$ que l'on peut définir comme une mesure (positive ou nulle) de la proximité de la densité de probabilité p_x à une distribution gaussienne, au sens de Kullback. $I(\phi_x)$ représente l'information mutuelle d'une densité de probabilité gaussienne de mêmes moyenne et variance que p_x ; $J(p_x)$ désigne la négentropie de p_x ; et $J(p_{x_i})$ celle de p_{x_i}.
Minimiser l'information mutuelle $I(p_x)$ peut alors être obtenue en deux étapes :

1. recherche d'une transformation linéaire A telle que :

$$I(\phi_{A.x}) = 0 \tag{3.3}$$

2. recherche d'une transformation linéaire B telle que :

$$I(\phi_{B.A.x}) \quad \text{demeure nulle} \tag{3.4}$$

$$J(p_{B.A.x}) \quad \text{est constante} \tag{3.5}$$

$$\sum_{i=1}^{M} J(p_{B.A.x_i}) \quad \text{est maximum} \tag{3.6}$$

On peut montrer que l'égalité 3.3 est obtenue si et seulement si les nouvelles variables y_i sont décorrélées, i.e. si $y = A.x$ a une matrice de covariance $\frac{1}{N}.[y|y]$ diagonale. C'est l'opération réalisée par l'ACP. Nous pouvons ainsi dire que l'ACP réduit l'information mutuelle de nos variables de départ jusqu'à l'ordre 2. Cette première étape se ramène à chercher la base dans laquelle $[x|x]$ est diagonale, ce que l'on obtient par une décomposition en valeurs propres. Si l'on note k le rang de $[x|x]$, il existe une matrice diagonale définie positive de rang k, notée Σ, et une matrice orthogonale Q telles que :

$$\frac{1}{N}[x|x] = Q.\Sigma.Q^{\dagger}$$

Il nous suffit alors de choisir $A = Q$ et l'on obtient le résultat souhaité :

$$\frac{1}{N}[y|y] = \Sigma$$

[1] dénomination abusive de notre part puisque la divergence de Kullback-Leibler est non symétrique.

Les résultats de l'ACP consistent en k nouvelles variables décorrélées, associées aux k valeurs propres non nulles de $[x|x]$. Cette étape de décorrélation est couramment qualifiée de blanchiement. On peut éliminer les variables les moins impliquées dans la reconstruction de la variance des variables de départ, et ne garder que les r plus représentatives. Cette sélection est faite sur un critère minimisant l'erreur de reconstruction aux moindres carrés de l'ensemble des variables, et conduit à retenir les r variables associées aux r plus grandes valeurs propres de $[x|x]$ (voir [GVL96] pour les démonstrations). Pour la réalisation pratique de cette étape, on peut avoir recours à une décomposition en valeurs singulières de x, telle que nous la décrivons en section 3.3.2.3.

Dans son principe, l'ACI n'impose pas cette étape préalable de blanchiement. Cette contrainte de décorrélation est toutefois souvent retenue, car elle permet de simplifier l'expression des fonctions de contraste ([Car99b]). C'est dans le cadre de cette approche "orthogonale" de l'ACI que nous nous plaçons. On peut alors dire que l'ACI commence par une ACP conduisant à l'égalité 3.3. Si en outre on se contraint à rechercher une transformation orthogonale pour B, alors Comon montre que les relations 3.4 et 3.5 sont vérifiées. En omettant les termes invariants par la transformation B, la fonction de contraste à maximiser ϕ^{IM} s'exprime alors par :

$$\phi^{IM} = \sum_{i=1}^{M} J(p_{B.A.x_i})$$

Cette introduction aux principes de l'ACI s'est concentrée sur l'une des fonctions de contraste de la littérature, fondée sur la minimisation de l'information mutuelle de nos variables. D'autres fonctions de contraste très répandues sont définies à partir de la log-vraisemblance du vecteur aléatoire x ou encore par extension du principe infomax, souvent utilisé pour les réseaux de neurones ([BS95]). Ces deux dernières approches sont strictement équivalentes. Dans la discussion présentée dans [Car98], Cardoso souligne que seul le contraste défini à partir de l'information mutuelle ne fait aucune hypothèse sur la distribution de probabilité des variables. C'est la fonction de contraste que nous retenons donc pour la suite.
Enfin, il est important de mentionner une autre approche de maximisation de l'indépendance entre variables, ne faisant pas appel à la notion de fonctions de contraste. Cette approche, que Cardoso a défini comme "algébrique", étend en quelques sortes la formulation matricielle de l'ACP aux statistiques d'ordres supérieurs. Elle s'appuie sur la recherche de "matrices propres" d'un tenseur défini à partir des cumulants du vecteur aléatoire x (voir par exemple [Car98], [Car99b], [CS93], ou [Car94]).

3.3.1.3 Approximation de la fonction de contraste

La maximisation de la fonction de contraste requiert l'expression analytique de la densité de probabilité $p_{A.x}$, à laquelle nous avons rarement accès. Une solution est donnée par l'expansion d'Edgeworth, qui permet d'approximer une densité de probabilité "proche d'une gaussienne" en fonction des cumulants de $A.x$. Comon approche ainsi ϕ^{IM} par ϕ^{EE} ("Edgeworth Expansion") :

$$\phi^{IM}(B.A.x = y) \approx -\frac{1}{48} \sum_{ijkl \neq iiii} (Q^y_{ijkl})^2 = \phi^{EE}(y) \tag{3.7}$$

où les Q^y_{ijkl} désignent les cumulants d'ordre 4 de y, définis par :

$$Q^y_{ijkl} = Cum(y_i, y_j, y_k, y_l)$$

Puisqu'une propriété des inter-cumulants est de s'annuler si et seulement si les variables sont indépendantes, il est cohérent de chercher à minimiser la somme de ces inter-cumulants.
Cardoso ([Car99b]) utilise une autre approximation, obtenue sans faire appel à l'expansion d'Edgeworth[2] :

$$\phi^{IM}(y) \approx - \sum_{ijkl \neq iikl} (Q^y_{ijkl})^2 = \phi^{JADE}(y) \tag{3.8}$$

[2]La validité de cette expansion peut nous surprendre a priori : elle s'effectue "au voisinnage de la gaussiannité" alors qu'une consition pour pouvoir mener l'ACI, est au contraire de n'avoir que des variables non gaussiennes (sauf une au maximum).

Cette approximation constitue une mesure pour la diagonalisation des matrices de cumulants, définies dans l'approche algébrique que nous avons mentionnées en section 3.3.1.2. L'approximation ϕ^{JADE} est a la base de l'algorithme proposé par Cardoso[3] que nous avons utilisé pour pratiquer l'ACI sur les HRTF. La composition de ces deux approximations semblent différer par la famille de cumulants la mieux représentée ((ijkl), (iikl), (iijj), (ijjj)). Cardoso montre par un simple dépliement des expressions 3.7 et 3.8 que toutes ces familles sont représentées, et que seuls diffèrent le nombre de représentants. Ces deux approximations sont donc voisines mais diffèrent dans le "poids" accordé à chaque famille de cumulants. En conclusion, la maximisation du constrate, sous contrainte de blanchiement, donne accès à une transformation orthogonale A et à des fonctions spatiales indépendantes C_i telles que :

$$C = C_i.A$$

Les nouveaux filtres de reconstruction sont donc donnés par :

$$L_i = A.L$$

Et on a :

$$H = C_i.L_i$$

Il est important de remarquer, que, puisque la transformation subie par les fonctions spatiales est orthogonale, l'erreur de reconstruction donnée par $C_i.L_i$ est identique à celle obtenue en sortie d'ACP, i.e. correspond à la meilleure approximation aux moindres carrés accessible pour le nombre de canaux fixés.

3.3.2 Application à la décomposition des HRTF

Comme nous le montrions en section 3.2.1.4, appliquer la décomposition à une représentation temporelle des HRTF (réponse impulsionnelle) ou à sa transformée de Fourier conduit aux mêmes performances, mêmes fonctions spatiales, et à des filtres de reconstruction transformés les uns des autres. Sauf mention contraire, nous considérons dans cette section la décomposition des HRTF à phase minimale d'une oreille droite sur 7 fonctions de base.

3.3.2.1 Indépendance statistique et compacité de support

A première vue, il peut sembler surprenant d'avoir recours aux statistiques d'ordre supérieur pour analyser des données déterministes. L'objectif de compacité que nous recherchons peut se formuler de la façon suivante : une fonction spatiale est idéalement compacte si elle est nulle partout sauf pour une position, différente pour chaque fonction spatiale. Etant donnée la limitation que nous avons en nombre de canaux, nous ne pouvons octroyer un canal par position, et par conséquent notre objectif devient d'obtenir des fonctions spatiales constituées d'un seul lobe, très directif.
Les algorithmes utilisés pour maximiser la fonction de contraste peuvent suivre deux voies : optimisation par descente de gradient, ou optimisation itérative de Jacobi ([Car99a]). Cette deuxième approche identifie la transformation orthogonale recherchée en tant que composition de rotations sur chaque paire de variables. Sa description nous permet de faire le lien entre maximisation du contraste statistique et maximisation de la compacité du support des fonctions spatiales.
Considérons une étape élémentaire de l'optimisation de Jacobi (rotation de Givens décrite en [GVL96] pp.215-223), recherchant l'angle de rotation θ des variables y_i et y_j maximisant le contraste ϕ^{IM} entre ces deux variables. Cardoso montre en annexe A de l'article [Car99a] que cet angle est tel que :

$$\tan(4\theta) = \frac{E(\eta_y.\xi_y)}{\frac{1}{4}.E(\eta_y^2 - 4\xi_y^2)}$$

que l'on peut ré-écrire :

$$\tan(4\theta) = \frac{4.\sum_k \eta_y(k).\xi_y(k)}{\sum_k \eta_y^2(k) - 4\xi_y^2(k)}$$

[3] Jean-François Cardoso donne libre accès à l'implantation Matlab qu'il a réalisée : ftp ://sig.enst.fr/pub/jfc/Algo/Jade/jade.m

avec :

$$\begin{aligned} \xi_y(k) &= y_i(k).y_j(k) \\ \eta_y(k) &= y_i^2(k) - y_j^2(k) \end{aligned}$$

Les nouvelles variables sont alors obtenues par :

$$\begin{pmatrix} z_i \\ z_j \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} . \begin{pmatrix} y_i \\ y_j \end{pmatrix}$$

Pour interpréter la transformation $(y_i, y_j) \rightarrow (z_i, z_j)$ en termes de compacité de support, définissons deux nouvelles variables :

$$\alpha_y(k) = 4\eta_y(k).\xi_y(k) = 4.(y_i^2(k) - y_j^2(k)).y_i(k).y_j(k) \quad (3.9)$$

$$\beta_y(k) = \eta_y^2(k) - \xi_y^2(k) = y_i^4 + y_j^4 - 6y_i^2.y_j^2 \quad (3.10)$$

On démontre facilement les propriétés suivantes :

$$\begin{pmatrix} \beta_z \\ \alpha_z \end{pmatrix} = \begin{pmatrix} \cos(4\theta) & \sin(4\theta) \\ -\sin(4\theta) & \cos(4\theta) \end{pmatrix} . \begin{pmatrix} \beta_y \\ \alpha_y \end{pmatrix} \quad (3.11)$$

$$\tan(4\theta) = \frac{\sum_k \alpha_y(k)}{\sum_k \beta_y(k)} \quad (3.12)$$

$$\sum_k \alpha_z(k) = 0 \quad (3.13)$$

$$\beta_z^2 + \alpha_z^2 = \beta_y^2 + \alpha_y^2 \quad (3.14)$$

$$\left(\sum_k \beta_z\right)^2 = \left(\sum_k \beta_y\right)^2 + \left(\sum_k \alpha_y\right)^2 \quad (3.15)$$

La relation 3.14 traduit le caractère orthonormé de la rotation, et se déduit de 3.11. On peut interpréter l'expression développée de α et β, définies en 3.10 :

- β est notamment constituée des cumulants d'ordre supérieur au travers de y_i^4 et y_j^4, et c'est donc une quantité que l'algorithme maximise,
- si les fonctions spatiales y_i et y_j sont idéalement compactes, on doit avoir $y_i(k).y_j(k) = 0$ pour tout k, donc $\alpha(k) = 0$. Si cet objectif idéal n'est pas atteint, la compacité est néanmoins approchée si $|y_i(k).y_j(k)|$ est d'autant plus petit que $|y_i^2(k) - y_j^2(k)|$ est grand. En effet, le cas où ces deux quantités sont grandes correspond à deux maxima de signe opposé pour une même position. La compacité des fonctions spatiales est donc d'autant meilleure qu'$\alpha(k)$ prend de faibles valeurs.

Précisement, l'algorithme de Jacobi opère un transfert d'énergie de la composante moyenne α sur la composante moyenne β (relations 3.13 et 3.15). En particulier, si une position k présente une forte singularité en α et β (valeurs fortement négatives), alors $\alpha_y(k)$ et $\beta_y(k)$ ont un poids prépondérant pour le calcul de l'angle de rotation θ, et l'on peut montrer que la transformation annule $\alpha_z(k)$ et maximise $\beta_z(k) \approx \sqrt{\beta_y^2(k) + \alpha_y^2(k)}$.

Un autre cas particulier est constitué des fonctions spatiales trigonométrique $\sin(\theta)$ et $\cos(\theta)$. Il semble intuitivement impossible de maximiser le constraste de ces fonctions car leur symétrie n'autorise aucun angle privilégié de rotation. Les valeurs prises par α et β impliquent que $\theta = 0$, et que l'algorithme est effectivement impuissant.

3.3.2.2 Effet du centrage des données

L'application des principes statistiques rappelés en section 3.3.1 requiert de définir quelles sont les variables, quelles sont les observations. Etant donné que les HRTF sont fonction de la position et de la fréquence, deux alternatives peuvent être considérées :

1. les HRTF sont vues comme des filtres, et chacun d'entre eux constitue une variable observée M fois (i.e. à chaque échantillon fréquentiel). En d'autres termes, les variables sont les lignes de H.
2. les HRTF sont vues comme des directivités, ou fonctions spatiales, et chacune d'entre elles constitue une variable observée N fois (i.e. à chaque position). Cette fois, les variables sont les colonnes de H.

Le choix de la "dimension" définissant les variables (la position pour (1.), la fréquence pour (2.)) conduit à une adaptation de la structure d'implantation :

1. ***Chacune des deux options conduit à un centrage différent.*** Dans le cas (1.), centrer les variables consiste à retirer à chaque ligne de H sa moyenne. On obtient ainsi une valeur moyenne par position, ce qui définit une fonction spatiale $g_0(\theta)$.

$$\begin{aligned} HRTF(\theta, f) &= \sum_{i=1}^{r} C_i(\theta).L_i(f) + \sum_{k=1}^{M} HRTF(\theta, f_k) \\ &= \sum_{i=1}^{r} C_i(\theta).L_i(f) + g_0(\theta) \end{aligned}$$

   L'implantation qui s'en déduit comprend un canal supplémentaire, associé à la fonction spatiale $g_0(\theta)$ et sans filtre de reconstruction.
   Dans le cas (2.), la moyenne est calculée et retranchée pour chaque colonne, ce qui conduit à la définition d'un filtre moyen, $HRTF_0(f)$:

$$\begin{aligned} HRTF(\theta, f) &= \sum_{i=1}^{r-1} C_i(\theta).L_i(f) + \sum_{k=1}^{N} HRTF(\theta_k, f) \\ &= \sum_{i=1}^{r-1} C_i(\theta).L_i(f) + HRTF_0(f) \end{aligned}$$

   Nous obtenons ainsi un filtre supplémentaire, associé à une fonction spatiale constante, égale à 1.
2. ***Le choix de la famille de variables détermine sur quel paramètre la propriété de compacité sera obtenue avec l'ACI.*** Les études précédentes appliquant l'ACI aux HRTF recherchaient des filtres à supports fréquentiels disjoints. Notre objectif étant l'obtention de fonctions spatiales compactes, nous examinons par la suite l'alternative (2.).

3.3.2.3 Extension de la décomposition à des variables non centrées

Nous avons choisi de centrer les HRTF fréquence par fréquence, ce qui nous conduit à pratiquer l'analyse sur des fonctions spatiales centrées. On peut montrer que les résultats de l'analyse sont alors des fonctions spatiales elles-mêmes centrées, contrainte défavorable à l'obtention de fonctions spatiales compactes, i.e. nulles sauf sur un intervalle réduit. Aussi envisageons-nous l'application des méthodes d'analyse statistique à des données non centrées.
Comme nous le rappellions en section 3.3.1.2, l'ACP s'appuie sur la décomposition en valeurs propres de la matrice de covariance $\frac{1}{N}.[H|H]$. Si les variables constituant les colonnes de H, ne sont pas centrées, cette matrice n'est plus la matrice de covariance, mais ses valeurs propres peuvent néanmoins être utilisées pour choisir un nombre réduit r de variables représentant de façon optimale au sens des moindres carrés les variables initiales. Ces nouvelles variables de ne sont pas décorrélées, mais sont orthogonales. Nous appliquons la décomposition aux HRTF centrées ainsi qu'aux HRTF non centrées.

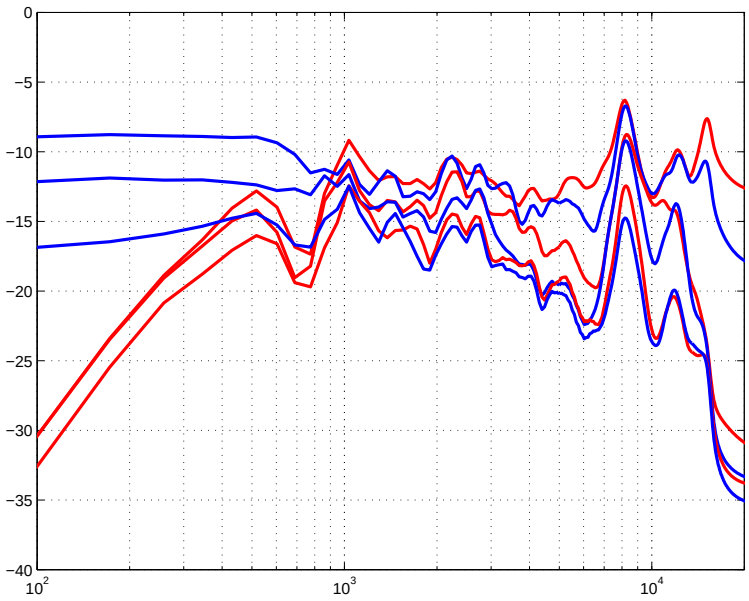


FIG. 3.3 – Erreur de reconstruction dans le plan horizontal pour une décomposition individuelle avec une ACP : données centrées (en bleu) ou non (en rouge), décomposition sur 7, 5 ou 3 canaux.

L'implantation pratique de cette décomposition peut utiliser la décomposition en valeurs singulières (SVD) de H :

$$H = U.\Sigma.D^\dagger$$

où U et D sont des matrices de dimension respective $N \times r$ et $M \times r$, telles que $U^\dagger U = Id$ et $D^\dagger D = Id$. Il suffit alors de choisir :

$$\begin{aligned} C &= U \\ L &= \Sigma.D^\dagger \end{aligned}$$

On peut vérifier qu'on a alors :

$$[C|C] = Id$$

En d'autres termes, les fonctions spatiales sont orthonormales. Si en outre elles sont le résultat d'une analyse de données centrées, elles sont même décorrélées.

La maximisation du contraste est alors pratiquée sur la matrice C donnée par la SVD, et conduit à une nouvelle matrice de fonctions spatiales, C', contenant r fonctions spatiales "indépendantes" stockées en colonne, ainsi qu'une matrice orthogonale B telle que :

$$C' = C.B$$

Les filtres de reconstruction associés sont obtenus par :

$$L' = B^\dagger.L$$

En conclusion, nous pratiquons l'ACP sur les données centrées, ou sur les données non centrées. Puis, la maximisation du contraste est réalisées sur l'ensemble des fonctions spatiales ainsi obtenues, contenant, dans le cas centré, une fonction spatiale "omni".

Les performances sont illustrées en Figure 3.3. On observe que la décomposition des données centrées permet une reproduction plus fidèle des basses fréquences, jusqu'à environ 1kHz. En effet, dans cette plage

de fréquences, les HRTF sont presque identiques pour toutes les positions, i.e. ont une directivité quasi-omnidirectionnelle. Le filtre moyen associé à la fonction spatiale omni contient donc toute l'information des HRTF au dessous de 1kHz. Dans le cas non centré, cette même information est disséminée sur l'ensemble des filtres de reconstruction, sans importance accrue par rapport aux autres fréquences. On constate d'ailleurs que l'erreur induite est alors plus homogène sur l'ensemble de l'intervalle fréquentiel. Les performances augmentent avec le nombre de canaux retenus pour la décomposition, présentant une amélioration d'environ 5dB par canal supplémentaire. Pour une décomposition sur 7 canaux, l'erreur reste environ 15dB au dessous du signal, et les données reconstruites contiennent 92% de la variance initiale (cas centré), ce qui nous laisse présager une bonne qualité perceptive de la synthèse binaurale ainsi réalisée.

3.3.2.4 Obtention du filtre d'élévation

Les spécifications de la section 3.2.2 peuvent être mises en oeuvre en trois étapes :

1. Déterminer la fonction spatiales $Y(\theta)$ par projection orthogonale à l'aide des relations rappelées en section 3.2.1.3 :
$$\forall i \in [1...N] \quad Y(\theta_i) = \left[H_h(\phi,f)^\dagger | H(\phi,\theta_i,f)^\dagger\right] .G_{H_h}^{\dagger -1}$$
Notre échantillonnage ne comprenant pas les mêmes azimuts pour chaque élévation, l'application de la relation précédente requiert l'interpolation des fonctions spatiales du plan horizontal, éventuellement pour le calcul de chaque tranche d'élévation. Nous réalisons ce réechantillonnage 1D par une interpolation cubique (routine *spline.m* de Matlab).
2. Evaluer l'erreur de reconstruction pour les positions élevées :
$$e(\phi,\theta,f) = H(\phi,\theta,f) - Y(\theta).H_h(\phi,f)$$
3. Déduire $Z(\theta)$ et $H_z(f)$ du premier vecteur propre donnée par une ACP menée sur $e(\phi,\theta,f)$.

La fonction spatiale $Y(\theta)$ est représentée en Figure 3.4, pour chacune des 17 têtes. On constate que la représentation du plan horizontal intervient encore fortement pour les positions très élevées ($\theta > 40°$) puisque le gain est minoré par 0.5. Ce résultat contraste avec la pondération en cosinus que nous rencontrerons pour les deux approches ultérieures (sections 3.3, 3.5). Contrairement à ces deux autres cas, la synthèse des positions très élevées ne repose pas sur le seul filtre d'élévation, et on peut ainsi s'attendre à une meilleure qualité de reconstruction. L'erreur, présentée en Figure 3.6, est néanmoins très élevée, puisqu'elle ne se situe qu'à 5dB du signal. Le filtre d'élévation $H_z(f)$, que l'on peut observer en Figure 3.5, présente une faible dynamique jusqu'à environ 5kHz, caractéristique des HRTF très élevées.

3.3.3 Optimisation des fonctions spatiales issues de l'ACI

3.3.3.1 Définition de fonctions spatiales universelles

Les fonctions spatiales issues de l'ACP sont a priori individuelles : elles sont associées aux HRTF d'une tête donnée, sur lesquelles a été pratiquée l'analyse. Une première solution pour définir un encodeur universel consiste à moyenner ces fonctions spatiales individuelles. C'est l'approche que nous avons présentée dans [LJGW00], suggérée par la constatation d'une nette ressemblance entre les fonctions spatiales de toutes les têtes. Toutefois, on peut lui reprocher de ne donner aucune prise sur l'erreur de reconstruction induite par le moyennage.
Une alternative consiste à appliquer l'analyse sur une gigantesque matrice de HRTF, constituée de bases de données individuelles concaténées. Bien sûr, l'encodeur ainsi obtenu ne mérite pas rigoureusement le qualificatif d"'universel", puisqu'il est obtenu à partir d'un certain ensemble de têtes. Nous le dérivons de la concaténation de 17 têtes, repérées par 1, 2, ... , p . L'ACP conduit alors aux filtres de reconstruction de chaque tête (L_1, ..., L_p) et on a :
$$[H_1\ H_2 ... H_p] = C.[L_1\ L_2 ... L_p]$$

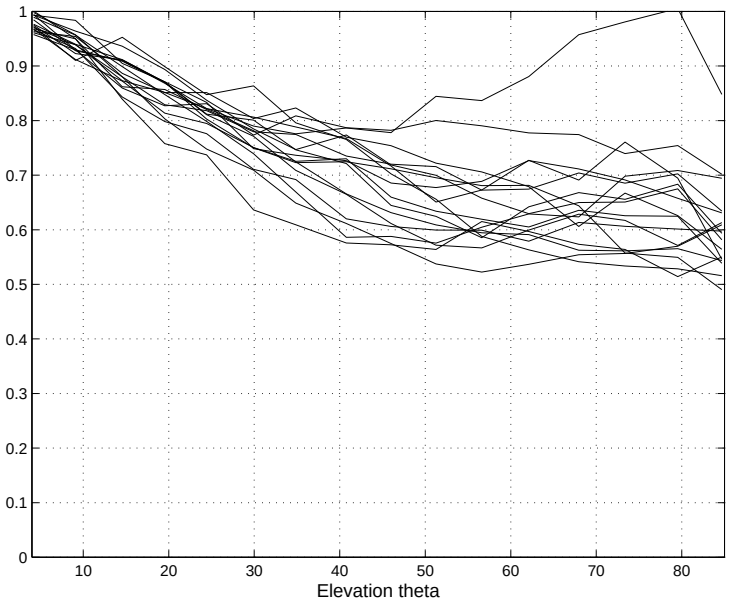


FIG. 3.4 – Fonction spatiale d'élévation $Y(\theta)$ associée à la décomposition horizontale individuelle des HRTF avec l'ACI.

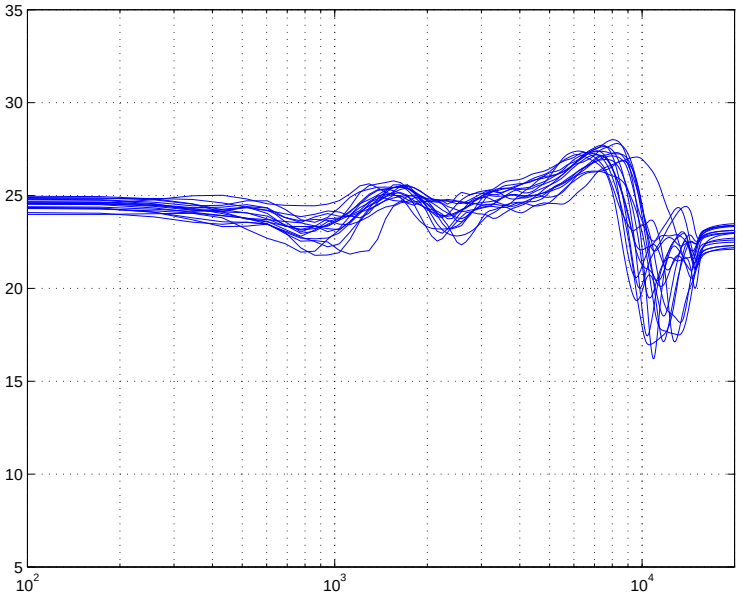


FIG. 3.5 – Filtre d'élévation pour l'ACI.

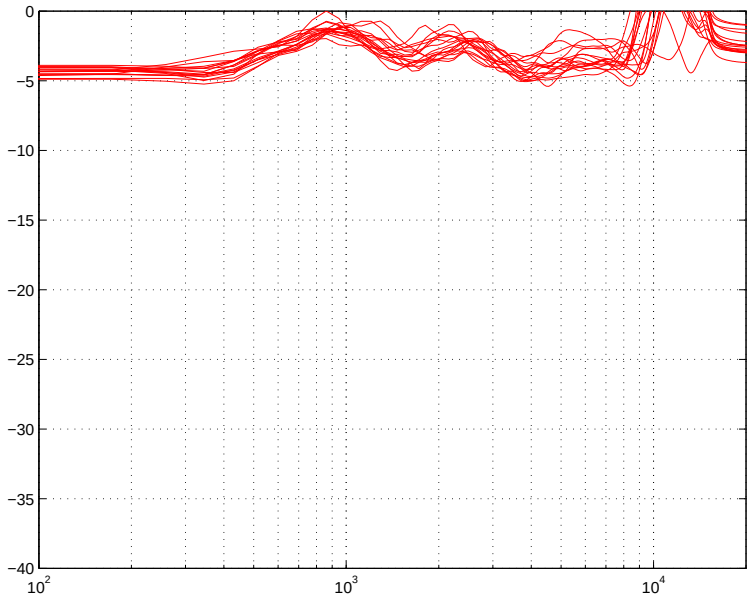


FIG. 3.6 – Erreur de reconstruction des HRTF en élévation, pour une ACI individuelle.

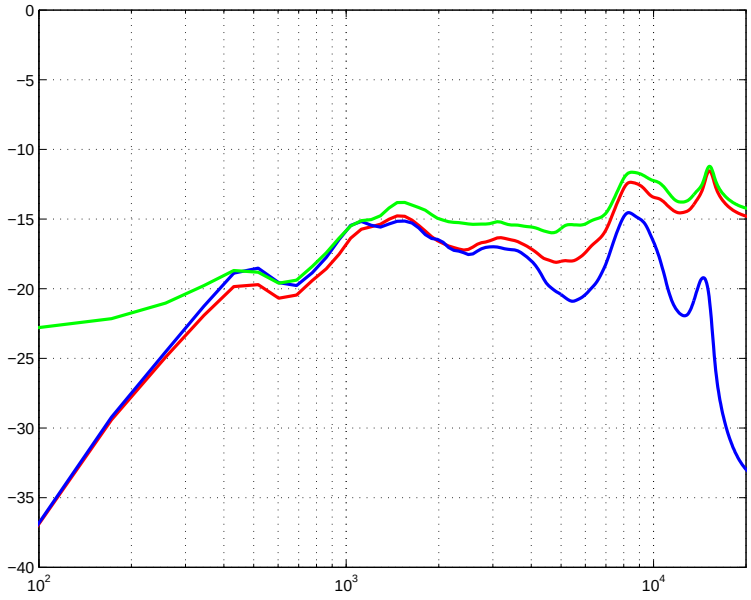


FIG. 3.7 – Influence de l'universalisation et de l'optimisation de la compacité des fonctions spatiales sur l'erreur de reconstruction dans le plan horizontal : décomposition ACP/ACI individuelle (en bleu), décomposition ACP/ACI "universelle" (en rouge), décomposition ACP/ACI "universelle" avec compacité accentuée (en vert). Les courbes représentent la moyenne des résultats obtenus pour les 17 têtes.

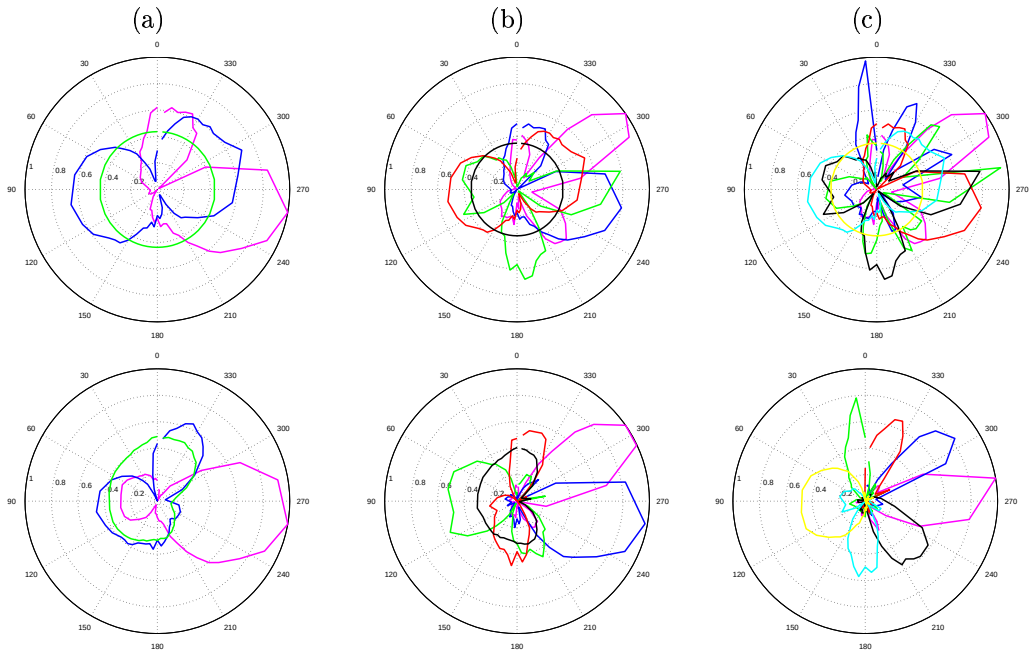


FIG. 3.8 – Robustesse des fonctions spatiales avec l'ordre de décomposition pour l'ACP (en haut) et l'ACI (en bas) : décomposition sur 3 canaux (a), 5 canaux (b) et 7 canaux (c). Cas d'une ACP pratiquée sur des données centrées.

L'ACI est ensuite pratiquée sur les fonctions spatiales communes C comme dans le cas "individuel". Cette universalisation de l'encodeur induit une perte de qualité pour la reconstruction, par rapport à celle obtenue pour une décomposition individuelle. En effet, comme on l'observe en Figure 3.7, cette augmentation de l'erreur est principalement localisée en hautes fréquences, intervalle sur lequel s'observent généralement les plus fortes différences inter-individuelles des HRTF.

3.3.3.2 Robustesse des fonctions spatiales avec l'ordre de décomposition

Les fonctions spatiales issue de l'ACP et de l'ACI sont représentées en Figure 3.8, pour une analyse pratiquée sur des données centrées. On observe que pour une décomposition sur 7 canaux, les fonctions spatiales de l'ACI présentent une compacité nettement supérieure à celles issues de l'ACP. Cet écart est moins visible, voir absent, pour les ordres inférieurs. En effet, la maximisation du contraste consiste à rechercher la "meilleure" combinaison linéaire des fonctions spatiales de l'ACP. On peut alors comprendre que cette optimisation soit moins performante avec un nombre réduit de degrés de liberté, i.e. avec un nombre réduit de fonctions spatiales.
Les lobes obtenus avec l'ACI sur 7 canaux réalisent un découpage hétérogène du plan horizontal. Le côté ipsilatéral est représenté par 6 fonctions spatiales, alors qu'une seule n'intervient pour la reconstruction des positions contralatérales. Pour ces positions en effet, les HRTF présentent une très faible énergie, et ont toutes pour gabarit moyen un filtre passe-bas, que synthétise le filtre de reconstruction contra-latéral que l'on peut observé en Figure 3.9. Par opposition, le lobe à -90°, représentant les positions les plus riches en énergie, est le premier à apparaitre, dès la décomposition à l'ordre 3, et est ainsi le seul lobe robuste à l'ordre de décomposition.
Pour les autres positions ipsilatérales, on observe que le demi-espace frontal possède une composante de plus que le demi-espace arrière. Le maximum des lobes est situé aux azimuts : 5°, -25°, -55°, -90°, -130° et 180°.

3.3.3.3 Régularisation de la compacité des fonctions spatiales

Par la maximisation du contraste entre les fonctions spatiales issues de l'ACP, nous souhaitons optimiser le coût de calcul de l'encodeur : idéalement, chaque position doit être reconstruite à l'aide de deux canaux actifs seulement (ou trois pour l'élévation). Les résultats de l'ACI optimisent une telle compacité de support pour une erreur de reconstruction donnée. Nous choisissons de sacrifier une part de cette qualité de reconstruction afin d'atteindre l'efficacité idéale de l'encodeur : nous forçons à zéro toutes les valeurs extérieures au lobe principal de chaque fonction spatiale. Cet algorithme rudimentaire suffit pour obtenir le résultat souhaité, puisque, comme on peut le constater dans le Tableau 3.1, le constraste est encore augmenté par cette procédure.

	données centrées	données non centrées
ACP	7.6	5.9
ACI	190	192
ACI avec discrétisation forcée	313	308

TAB. 3.1 – Contraste pour des fonctions spatiales universelles : issues de l'ACP, après maximisation du contraste, et après optimisation de la compacité.

L'erreur induite par cette manipulation est principalement localisée en basses-fréquences, comme on l'observe en Figure 3.7. Mais, en partant de données centrées, on voit qu'en dépit de cette détérioration, l'erreur demeure inférieure au cas d'une décomposition des données non centrées sans optimisation.

3.3.3.4 Relation entre filtres de reconstruction et HRTF

Par construction, et aux défauts d'orthogonalité introduits par l'accentuation de la compacité près, chaque filtre de reconstruction est obtenu par une somme pondérée de HRTF, voisines de la position du maximum de la fonction spatiale associée (cf section 3.2.1.3). Grâce à l'étroitesse du support des fonctions spatiales, propriété de compacité que nous avons optimisée, il est normal d'observer la proximité entre les filtres de reconstruction et les HRTF correspondant au maximum des fonctions spatiales (Figure 3.9). Toutefois, cette remarque ne s'applique pas à la fonction spatiale reconstruisant les positions contralatérales, puisque précisemment, son support couvre presque tout le demi-plan concerné.

3.4 Optimisation des filtres de reconstruction pour des fonctions spatiales fixées a priori

Imposer le choix des fonctions spatiales ne peut certes permettre la qualité de reconstruction optimale obtenues avec les méthodes statistiques de la section précédente. Cela constitue toutefois une manière simple de définir un encodeur universel. En outre, les harmoniques sphériques, fonctions spatiales que nous avons choisies pour la décomposition, présentent les intérêts suivants :

- elles sont définies analytiquement, et peuvent donc être évaluées pour tout échantillonnage spatial des HRTF.
- elles sont hiérarchisées : les ordres plus élevés permettent de décrire des détails plus fins (i.e. des variations spatiales plus rapides) de la directivité. Réduire le nombre de canaux utilisés pour s'adapter à la capacité de traitement est donc immédiatement obtenus en limitant l'ordre de la décomposition.
- l'encodage peut être réalisé à l'aide de microphones existants : micro soundfield seul si l'on considère la décomposition des HRTF à phase mixte, deux micros soundfield écartés d'une certaine distance pour la décomposition des HRTF à phase minimale ([JWL98]).
- elles forment une base génératrice des ondes spatiales de pression, en tant que solutions élémentaires de l'équation des ondes sphériques. Cette propriété, établie dans le domaine continu, peut s'étendre à un domaine spatial échantillonné, moyennant le respect d'un théorème d'échantillonnage spatial. Ces conditions étant remplies, la décomposition permet par interpolation globale de reconstruire les HRTF à des positions non mesurées.

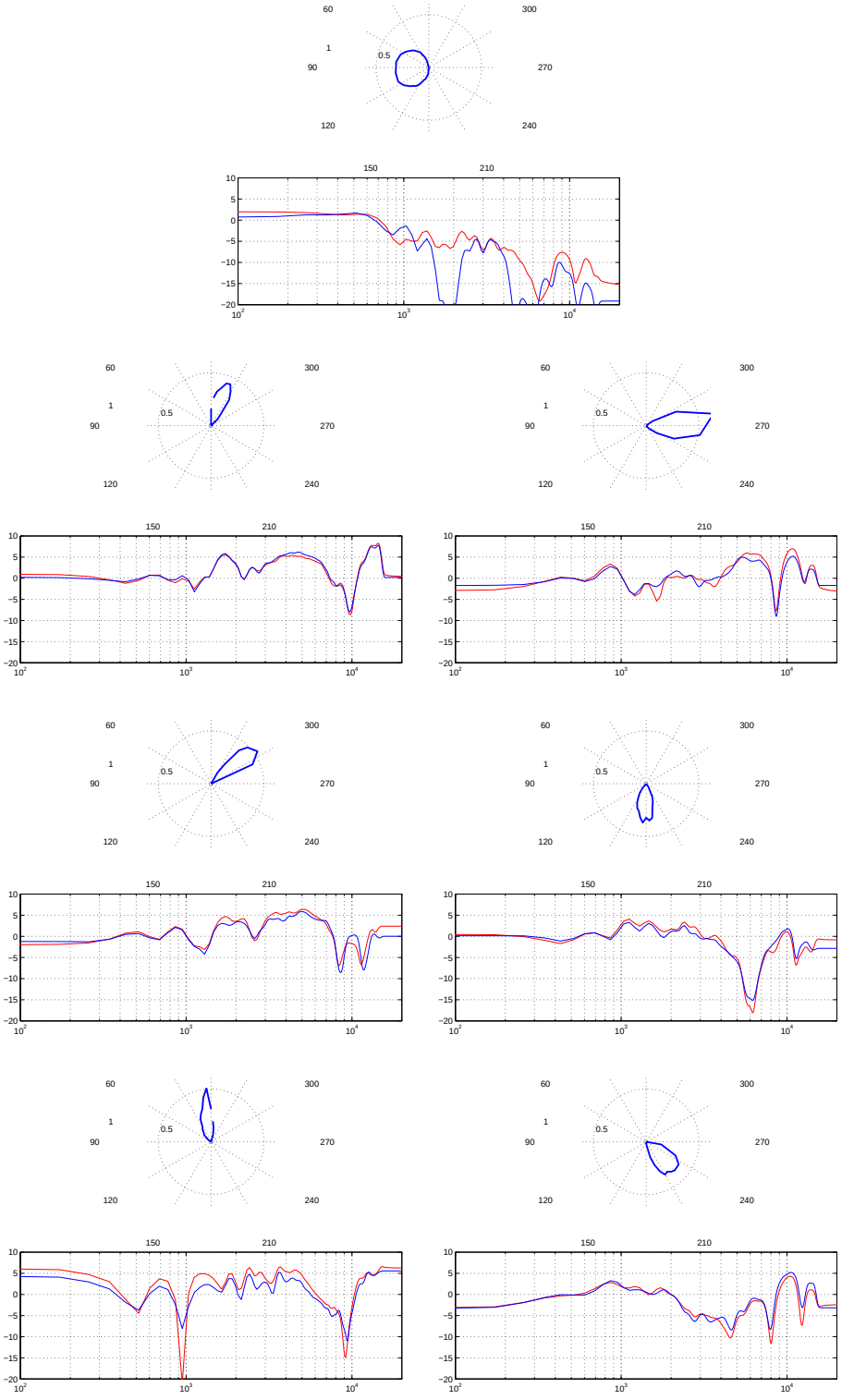


FIG. 3.9 – Fonctions spatiales dans le plan horizontal et filtres de reconstruction pour une ACI sur 7 canaux. On a superposé aux filtres de reconstruction d'une des tête la HRTF de la position pointée par le lobe de la fonction spatiale associée. ACP pratiquée sur des données centrées, maximisation du contraste avec omni incluse.

3.4.1 Choix d'une représentation des Harmoniques Sphériques

3.4.1.1 Expression analytique des harmoniques sphériques

Les harmoniques sphériques sont repérées par deux indices, notés l et m, qui représentent respectivement ordre et rang : à l'ordre l, il existe $2.l+1$ harmoniques sphériques, qui sont ordonnées en fonction de leur rang m. Ce sont des fonctions à variables séparées des angles θ et φ du repère sphérique. La dépendance en φ, angle coïncidant avec l'azimut, s'exprime par une composante $cos(m.\varphi)$ ou $sin(m.\varphi)$, tandis que la dépendance en θ, angle complémentaire de l'élévation, s'exprime à l'aide du polynôme de Legendre associé, $P_l^m(cos(\theta))$.
Plusieurs coefficients de normalisation sont proposés dans la littérature, afin d'obtenir une norme unitaire pour chaque harmonique. Etant donné le redressement que nous opérons sur la base des fonctions spatiales à l'aide de la matrice de Gram, ce choix nous est indifférent. Nous adoptons la solution proposée dans l'ouvrage de référence [Kap81], et utilisée par Evans et al. ([EAT98]). A l'ordre l, les harmoniques sphériques réelles de rang $m \geq 0$, c_l^m et s_l^m, ont pour expression analytique :
Pour $m = 1, ..., l$:

$$c_l^m = \sqrt{\frac{2l+1}{2\pi}.\frac{(l-m)!}{(l+m)!}}.P_l^m(cos(\theta)).cos(m.\varphi)$$

$$s_l^m = \sqrt{\frac{2l+1}{2\pi}.\frac{(l-m)!}{(l+m)!}}.P_l^m(cos(\theta)).sin(m.\varphi)$$

Et pour $m = 0$:

$$c_l^0 = \sqrt{\frac{2l+1}{4\pi}}.P_l^0(cos(\theta))$$

Avec :

$$P_l^m(x) = (-1)^m.(1-x^2)^{m/2}.\frac{d^m}{dx^m}\left[\frac{1}{2^l.l!}.\frac{d^l}{dx^l}(x^2-1)^l\right]$$

Les fonctions spatiales stockées dans les colonnes de C sont donc constituées de l'ensemble des c_l^m et des s_l^m pour un certain ordre l, et $m \in 0, ..., l$.

3.4.1.2 Décomposition du plan horizontal

La stratégie de projection adoptée consiste tout d'abord à réaliser la décomposition dans le plan horizontal : $\theta = \pi/2$. Le nombre d'harmoniques sphériques utilisables à un ordre donné est alors fortement réduit, d'une part du fait de la nullité de certains polynômes de Legendre associés, et, d'autre part, du fait de l'apparition de liaisons entre harmoniques sphériques d'ordre différents.

1. Les polynômes de Legendre associés sont nuls en $x = 0$ si $m + l$ est impair :

$$\forall l, m\, t.q.\, l + m\, impair \;:\; P_l^m(0) = 0$$

   En effet :

   (a) $(x^2-1)^l$ est un polynome contenant des puissances paires de la variable x :

$$[x^0\; x^2 ... x^{2l}]$$

   (b) le calcul du polynome de Legendre associé requiert la dérivation de l'expression précédente $m + l$ fois. On obtient immédiatement que :

      i. si $m + l$ est impair, le polynome résultant contient les puissances impaires de la variable x :

$$[x\; x^3 ... x^{l-m}]$$

      Par conséquent, 0 est racine du polynome, et $P_l^m(0) = 0$.

ii. si $m+l$ est pair, le polynome ne contient que les puissances paires de x, et $P_l^m(0) \neq 0$.

2. Les composantes c_l^m sont liées aux composantes d'ordre $l' < l$ $c_{l'}^m$, et de même pour la famille des s_l^m. Pour un ordre $l > 1$, nous ne conservons donc que les harmoniques c_l^l et s_l^l.

Par conséquent, la projection dans le plan horizontal fait intervenir :

- à l'ordre 0 : $c_0^0 = \frac{1}{\sqrt{4\pi}} = c_0$,
- à l'ordre 1 : $c_1^1 = -\sqrt{\frac{3}{4\pi}}.cos(\varphi) = c_1$ et $s_1^1 = -\sqrt{\frac{3}{4\pi}}.sin(\varphi) = s_1$,
- à l'ordre 2 : $c_2^2 = \frac{3}{4}\sqrt{\frac{5}{3\pi}}.cos(2\varphi) = c_2$ et $s_2^2 = \frac{3}{4}\sqrt{\frac{5}{3\pi}}.sin(2\varphi) = s_2$,
- à l'ordre 3 : $c_3^3 = -\frac{5}{4}\sqrt{\frac{7}{10\pi}}.cos(3\varphi) = c_3$ et $s_3^3 = -\frac{5}{4}\sqrt{\frac{7}{10\pi}}.sin(3\varphi) = s_3$.

Une décomposition à l'ordre 1 impliquera donc une implantation sur 3 canaux, tandis qu'une décomposition à l'ordre 3 requiera 7 canaux. Pour une représentation graphique de ces figures de directivité échantillonnée, on pourra se reporter à la Figure 3.16.
Cette décomposition en harmoniques sphériques dans le plan horizontal s'apparente, on le constate, à une simple décomposition en séries de Fourier. L'intérêt du recours à des fonctions à double périodicité (en azimut et en élévation) prend tout son sens pour la construction du filtre d'élévation, auquel nous consacrons un canal supplémentaire.

3.4.1.3 Décomposition de l'élévation

Comme nous le décrivions en section 3.2.2, les fonctions spatiales utilisées pour la décomposition du plan horizontal interviennent pour celle de l'élévation, pondérée par les fonctions spatiales $Y_l(\theta)$, données par leur formule analytique :

- à l'ordre 1 : $Y_1(\theta) = sin(\theta)$,
- à l'ordre 2 : $Y_2(\theta) = sin^2(\theta)$,
- à l'ordre 3 : $Y_3(\theta) = sin^3(\theta)$.

En outre, pour la fonction spatiale consacrée à l'élévation, nous choisissons l'harmonique sphérique indépendante de φ d'ordre le plus faible :

$$Z(\theta) = \sqrt{\frac{3}{4\pi}}.cos(\theta)$$

3.4.2 Application à la décomposition des HRTF

C'est L'implantation des structures présentées en Figure 3.1 requiert la décomposition des HRTF à phase minimale et à phase mixte. La décomposition sur les harmoniques sphériques des HRTF à phase mixte permet de définir un décodage pour casque du format B, format multicanal utilisé pour la diffusion multi-haut-parleurs (e.g. [Dan00]). Celle des HRTF à phase minimale, proposée par Jot dans [JWL98], définit un format baptisé par extention "Binaural B". Puisque le champ sonore en un point peut être décrit par ses coordonnées sur la base des harmoniques sphériques, on peut fournir une interprétation de cette opération de décomposition :

- décomposer les HRTF à phase mixte sur les harmoniques sphériques consiste à approximer le champ sonore aux oreilles, connaissant le champ au lieu du centre de la tête. Plus exactement, nous ne disposons en ce point que de la représentation partielle du champ, représentée par le nombre limité d'harmoniques sphériques que nous considérons. L'absence de coïncidence entre le point observé et le point de mesure nous permet de prévoir qu'une approximation de qualité ne sera possible qu'en basses fréquences.
- en décomposant les HRTF à phase minimale, nous ramenons en quelques sortes la directivité de l'oreille au centre de la tête, ou plus exactement en un point très proche (la phase est minimale, et non pas nulle).

Evans et al. appliquent ces deux décompositions à une représentation temporelle des HRTF. Comme nous l'avons montré en section 3.2.1.4, la décomposition est invariante par transformée linéaire orthogonale des données. Nous avons choisi de projeter les spectres complexes des HRTF.
On peut remarquer que la décomposition sur les harmoniques sphériques s'appliquent à d'autres directivité que celle de l'oreille, notamment à la modélisation de la directivité des sources sonores ([Gir96], [Der97]).

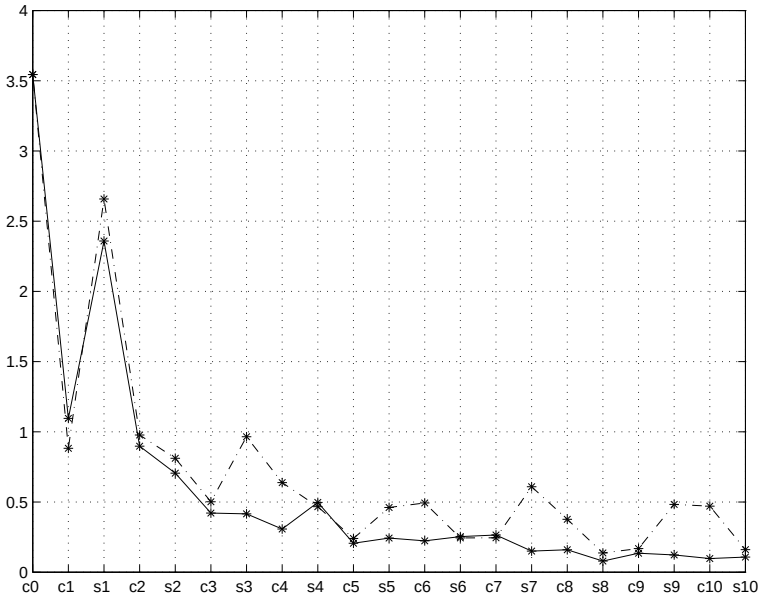


FIG. 3.10 – Spectre harmonique des HRTF pour une décomposition dans le plan horizontal jusqu'à l'ordre 10 : HRTF à phase minimale (trait continu) et HRTF à phase mixte (trait pointillé).

3.4.2.1 Poids de chaque harmonique sphérique dans la décomposition

Les harmoniques sphériques formant une base génératrice, toute fonction spatiale continue possède une représentation dans l'espace des harmoniques, baptisée "spectre harmonique". Evans et al. empruntent à [PHJ93] une définition de ce dernier, noté SH, qu'ils s'écrivent à l'ordre l :

$$SH(l) = \sqrt{\frac{1}{2l+1}\sum_{m}(|c_l^m|^2 + |s_l^m|^2)}$$

Par extension, nous appliquons la notion de spectre harmonique à la décomposition à un ordre fini des directivités discrétisées des HRTF. Notre objectif étant de localiser les harmoniques intervenant de façon prépondérante pour la décomposition, y compris au sein d'un même ordre, nous visualisons individuellement chaque coefficient de la décomposition, donné pour chaque fréquence par le module des filtres de reconstruction. Pour chaque rang, nous ne visualisons que la valeur maximale des coefficients sur l'ensemble des fréquences, ce qui nous conduit à une expression du spectre harmonique modifié, pour l'harmonique i :

$$SH(i) = \max_{f_k} \left| L^i(f_k) \right|$$

Comme on l'observe en Figure 3.10, c'est l'harmonique omni directionnelle (ou "composante continue") et la figure 8 orientée selon l'axe interaural qui dominent la décomposition. Ce seraient donc les deux fonctions spatiales que nous retiendrions si nous n'avions que deux canaux disponibles pour l'implantation. Le poids observé pour l'harmonique omni s'explique par la directivité des basses fréquences des HRTF, omnidirectionnelle jusqu'à environ 1kHz. La composante figure 8 oppose les directions principales de ses deux lobes, pour lesquelles elle prend deux valeurs maximales opposées. Le poids qui lui est accordé dans cette décomposition reflète l'écart de phase entre les HRTF ipsilatérales autour de 90° et les HRTF contralatérales autour de -90°. Cette opposition droite-gauche est naturellement plus importante que l'opposition entre les positions avant et les positions arrière, estimée par le coefficient c_1.
Comme l'on pouvait s'y attendre, cette opposition bipolaire de l'oreille est amplifiée dans le cas de la décomposition des HRTF à phase mixte, pour lesquelles le retard interaural augmente encore l'écart entre les spectres complexes. Dans ce cas, en effet, on observe que les coefficients des harmoniques en sinus $sin(m\varphi)$ sont supérieurs aux coefficients associés en cosinus. Cela est d'autant plus vrai que m est impair, puisque dans ce cas, l'opposition droite-gauche contient les lobes opposés 90°/-90°.

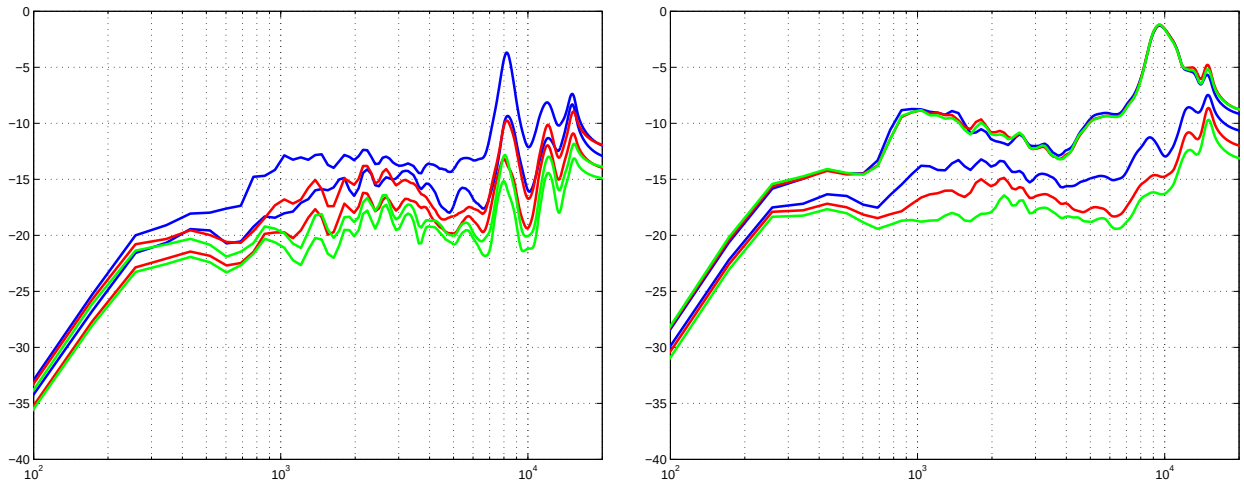


FIG. 3.11 – Erreur de reconstruction des HRTF à phase minimale pour une décomposition Binaural B : moyenne pour les positions du plan horizontal (a) et pour les positions hors du plan horizontal (b). Trois ordres de décomposition sont présentés (ordre 1 (bleu), ordre 3 (rouge), ordre 10 (vert)), ainsi que deux stratégie de décomposition présentées avec la même couleur (plan horizontal privilégié ou non).

Enfin, on observe qu'une projection à l'ordre 10 ne suffit pour reconstruire parfaitement les HRTF de notre échantillonnage. Cela est d'autant plus vrai pour les HRTF à phase mixte. Nous nous concentrons par la suite sur 3 ordres de décomposition représentatifs : ordre 1, ordre 3 et ordre 10.

3.4.2.2 Efficacité de la décomposition

L'erreur de reconstruction présentée en Figure 3.11 illustre les avantage de notre stratégie pour la décomposition des HRTF à phase minimale, privilégiant la reconstruction du plan horizontal. L'amélioration obtenue est de l'ordre de 3dB pour la projection aux ordre 1 et 3, mais s'estompe pour une projection aux ordre supérieurs. Pour la reconstruction des positions en élévation, l'ordre de décomposition du plan horizontal ne semble avoir qu'une faible influence puisque les courbes correspondant aux trois ordres sont superposées. Ce n'est bien sûr pas le cas lorsque la décomposition est réalisée sur l'ensemble des positions (les trois courbes sont bien distinctes).
Le passage d'une décomposition de l'ordre 1 à l'ordre 3 permet d'améliorer la reconstruction d'environ 6dB, score qui est tout juste dépassé par la transition de l'ordre 3 à l'ordre 10. On remarque que l'erreur minimum, atteinte pour l'ordre 10, est 20dB au dessous du signal jusqu'à environ 7kHz, et ne dépasse par -15dB sur toute la bande, ce qui nous permet de penser qu'elle est peu audible. Ce n'est pas le cas de la décomposition à l'ordre 1, voisine de -10dB. C'est également le niveau atteint par l'erreur de reconstruction en élévation, pour tous les ordres.

Les résultats pour la décomposition des HRTF à phase mixte sont illustrés en Figure 3.12. Contrairement au cas précédent, le gain en qualité de reconstruction apporté par la décomposition à l'ordre 10 est voisin de 10dB par rapport à la décomposition à l'ordre 1. Pour cet ordre d'ailleurs, jusqu'à 2kHz, les performances de décomposition des deux représentations des HRTF sont voisines, autour de -20dB. Pour l'ordre 3, on peut s'attendre à une erreur nettement perceptible au delà de 1kHz. Ce résultat est conforme à notre interprétation donnée en introduction.

Contrairement à ce qu'on observe pour la modélisation paramétrique des HRTF, abordée en chapitre 2, l'augmentation de l'ordre de la décomposition, ne se caractérise pas par la prise en compte de caractéristiques structurelles (anti/résonances) de plus en plus fines. En effet, comme on l'observe en Figure 3.13 , ces caractéristiques sont intégrées au filtre reconstruit dès l'ordre 1. En revanche, l'effet d'une augmentation de l'ordre réside en l'adaptation progressive du gain dans ces régions.

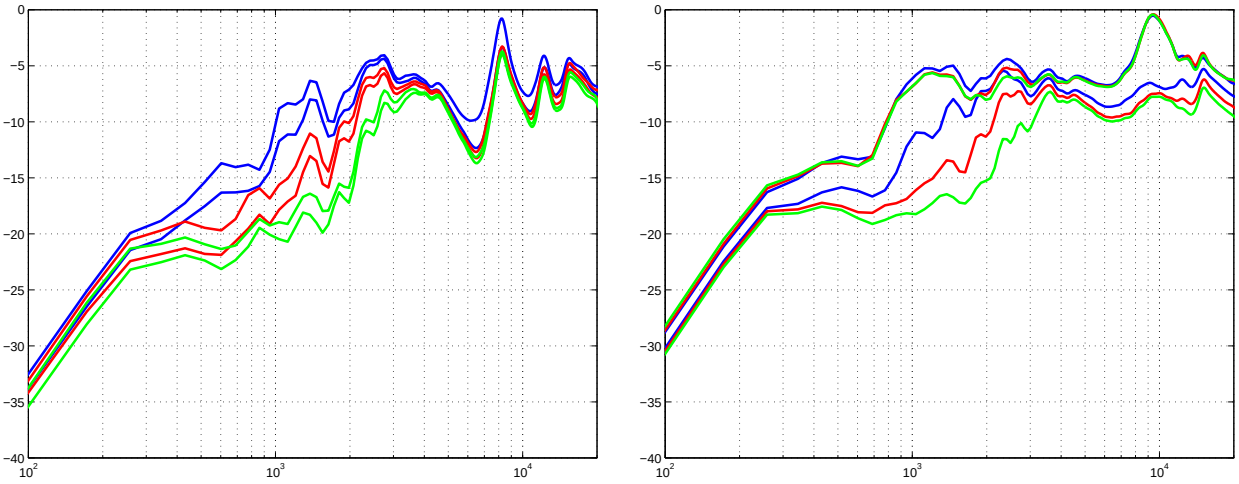


FIG. 3.12 – Mêmes convention qu'en Figure 3.11, pour la décomposition des HRTF à phase mixte.

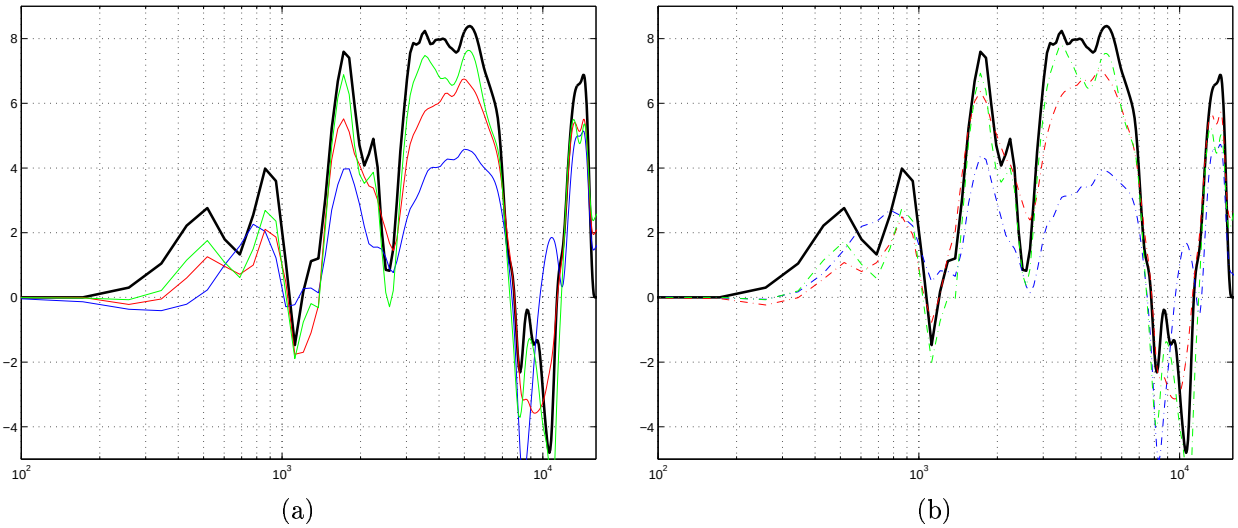


FIG. 3.13 – Reconstruction du spectre d'amplitude de la HRTF (40°,0°) : décomposition des HRTF à phase minimale (à gauche), décomposition des HRTF à phase minimale (à droite). Mêmes conventions de couleur qu'en 3.11.

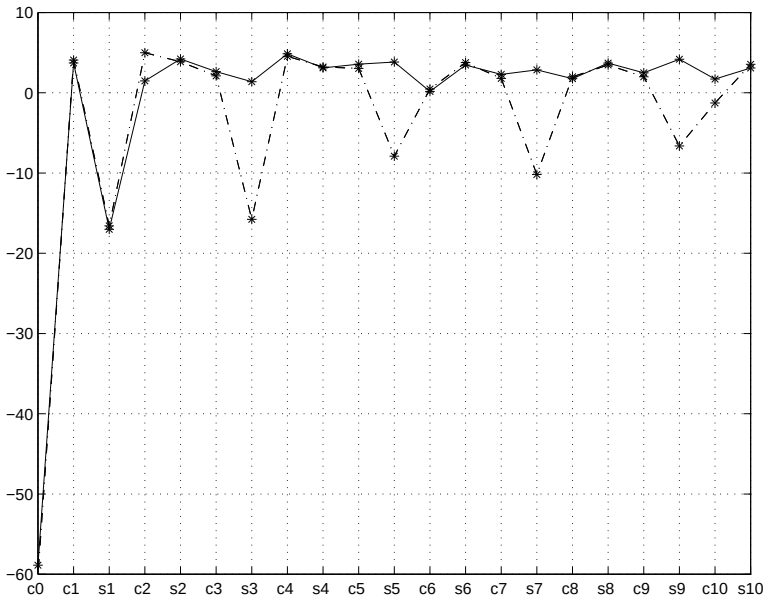


FIG. 3.14 – Mesure de la minimalité de phase des filtres de reconstruction : décomposition des HRTF à phase minimale (trait continu), à phase mixte (trait pointillé).

3.4.2.3 Filtres de reconstruction obtenus

Les filtres de reconstruction sont obtenus par projection orthogonale des HRTF H sur les harmoniques sphériques C (cf section 3.2.1.3). Cette opération, qui se ramène à des combinaisons linéaires de filtres à phase minimale (HRTF à phase minimale) ou à phase mixte (HRTF à phase mixte), ne laisse rien supposer à priori des propriétés de phase des filtres de reconstruction. Nous testons leur minimalité de phase avec le critère suivant :

$$crit = 20 * log_{10} \left| e^{j.ph} - e^{j.mph} \right|$$

où ph désigne la phase du filtre, et mph la phase minimale déduite de son amplitude à l'aide de la transformée de Hilbert. Les résultats sont présentés en Figure 3.14, pour une projection à l'ordre 10. On observe que les composantes participant de façon prépondérante dans la décomposition (voir Figure 3.10) ont une phase "quasi-minimale". Néanmoins, seul le filtre de la composante omni est véritablement à phase minimale.

Les filtres de reconstruction sont représentés en Figure 3.16. Par construction, le filtre associé à la composante omni représente la moyenne des HRTF sur l'ensemble des positions du plan horizontal. Comme on le mentionnait plus haut, ce filtre contient la composante basses fréquences commune à toutes les HRTF. Les autres filtres ont donc une faible énergie dans cette zone. Le filtre associé à figure 8 c_1 est construit de tel sorte qu'il participe à la décomposition avec la contribution $L_i(f)$ pour la position frontale, et avec la contribution de même amplitude mais de phase opposée, $-L_i(f)$, pour la position arrière.

On remarque que les filtres obtenus par projection à l'ordre 1 et par projection à l'ordre 3 ne sons pas rigoureusement identiques. Cela s'explique par le caractère non diagonal de la matrice de Gram associé à la décomposition d'ordre 3. L'inversion requise pour le calcul des filtres de reconstruction conduit ainsi à faire intervenir les harmoniques sphériques d'ordre 3 pour le calcul des filtres associés aux premières composantes.

3.5 Optimisation des fonctions spatiales pour des filtres de reconstruction fixés à priori

Dans cette section, nous cherchons à représenter un jeu de filtres par un nombre réduit d'entre eux, les plus représentatifs. Plus précisemment, nous suivons l'approche de Gardner développée dans [Gar99],

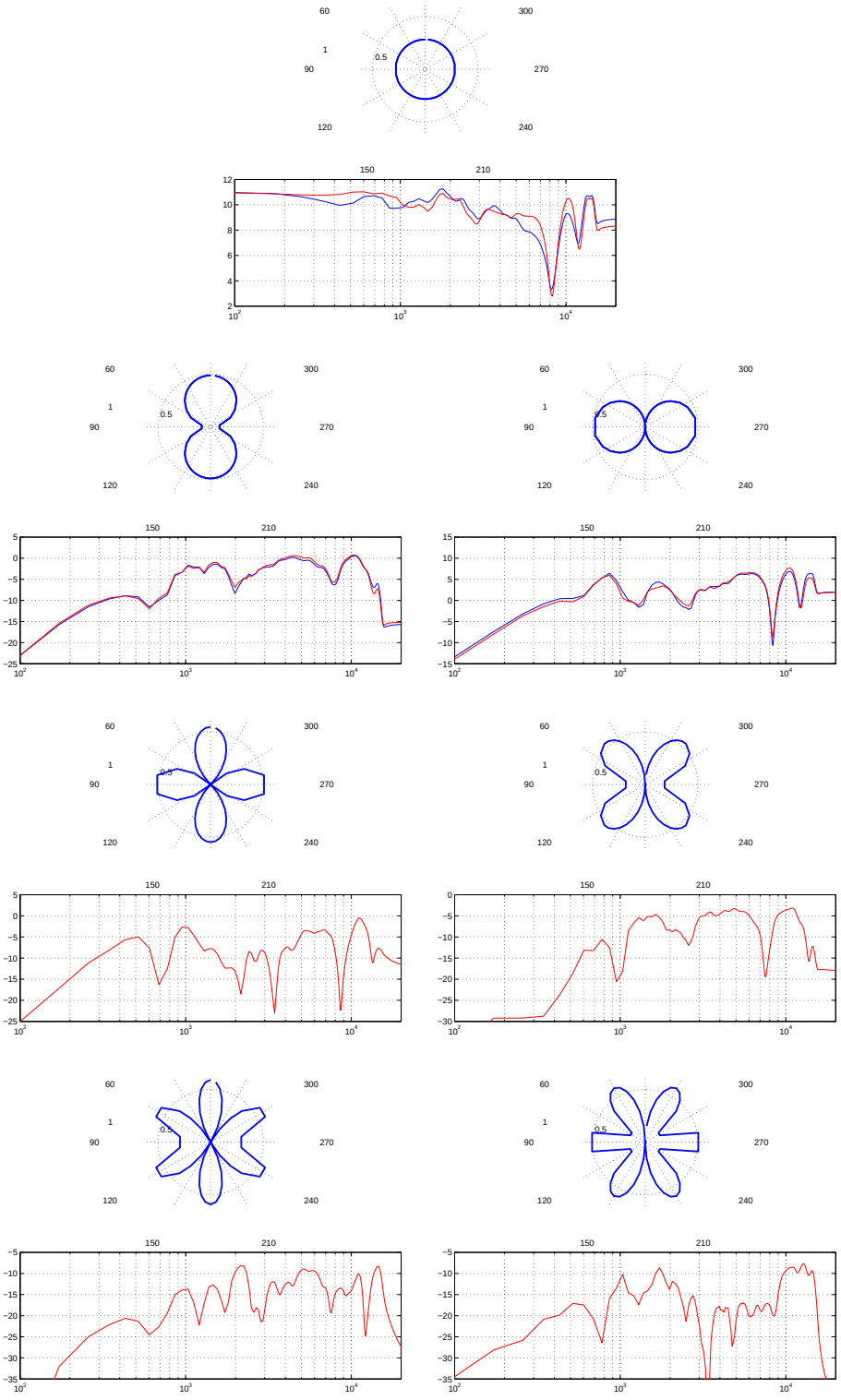


FIG. 3.15 – Fonctions spatiales dans le plan horizontal et filtres de reconstruction pour une décomposition à l'ordre 3. Pour les 4 premières composantes, on superpose les filtres obtenus pour une décomposition à l'ordre 1.

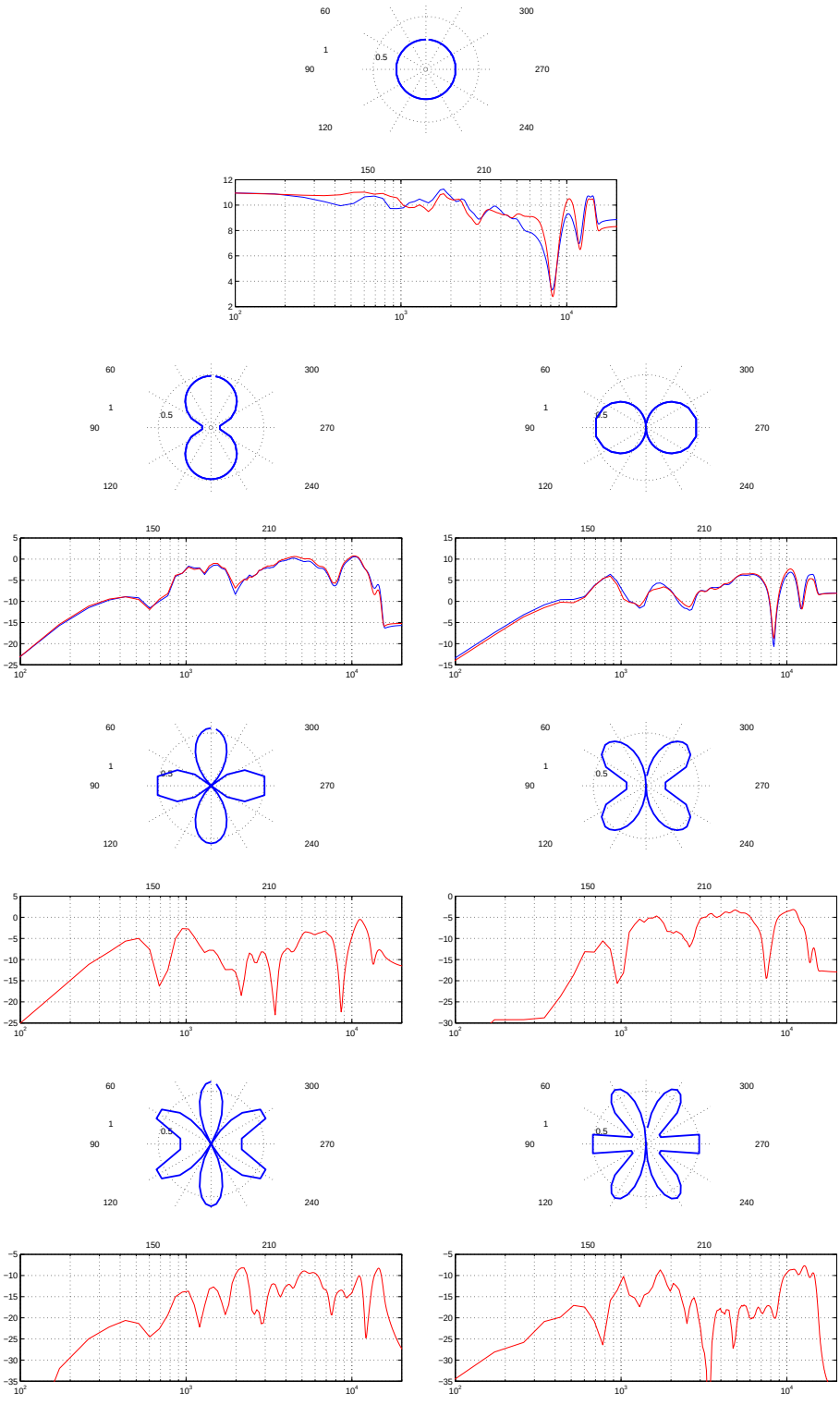


FIG. 3.16 – Fonctions spatiales dans le plan horizontal et filtres de reconstruction pour une décomposition à l'ordre 3. Pour les 4 premières composantes, on superpose les filtres obtenus pour une décomposition à l'ordre 1.

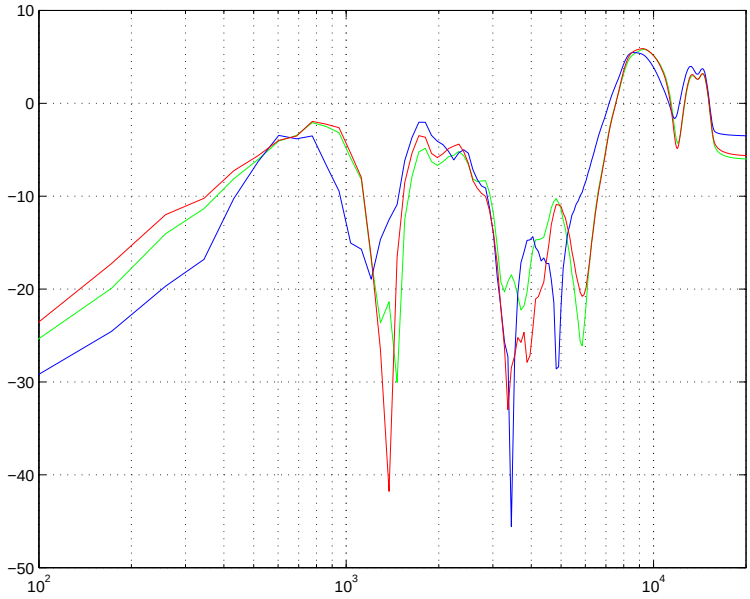


FIG. 3.17 – Filtre de reconstruction de l'élévation pour 3 ordre de décomposition (mêmes conventions de couleur qu'en Figure 3.11.

consistant à choisir un nombre réduit de HRTF à phase minimale pour engendrer toutes les autres. Cette méthode, que nous désignerons par "subset selection" présente l'intérêt de donner par construction des fonctions spatiales discrètes.
Nous rappelons la méthode utilisée par Gardner et l'appliquons à nos données. En outre, nous étendons cette approche au cas de HRTF à phase mixte, conduisant au concept bien connu des "haut-parleurs virtuels".

3.5.1 Application de la méthode "subset selection" à la décomposition des HRTF

3.5.1.1 Description de la méthode

La première étape consiste à appliquer une décomposition en valeurs singulières des HRTF (cf section 3.3.2.3). Les fonctions spatiales ainsi obtenues sont décomposées à l'aide d'une décomposition QR avec pivot de colonne, telle que décrite dans [GVL96] (pp.590-600). Cette décomposition conduit à une matrice de permutation permettant d'ordonner les lignes de C, i.e. les positions, selon leur norme $\left|[C_{i1}C_{i2}....C_{i7}]^t\right|^2$, où C_i représente la i^{eme} fonction spatiale de C. Cet ordonancement des positions est utilisé pour sélectionner les HRTF qui constitueront les filtres de reconstruction, minimisant par ce choix l'erreur de reconstruction aux moindres carrés (norme de Frobenius).
Les fonctions spatiales associées sont obtenues par projection orthogonale selon les relations 3.2.1.3. Comme l'observe Gardner, chacune d'entre elles atteint sa valeur maximale pour la position correspondant à la HRTF retenue, et pour laquelle toutes les autres fonctions spatiales sont nulles. Par construction, chaque fonction spatiale tend à concentrer son énergie autour d'une position donnée, et présente ainsi de bonnes propriétés de compacité, comme l'illustrent les résultats présentés dans [Gar99].
Nous appliquons cette méthode aux HRTF à phase minimale centrées et non centrées. Dans le premier cas, le centrage de H est réalisé par colonne, et les filtres de reconstruction ne sont plus sélectionnés parmi des HRTF, mais parmi des HRTF centrées.

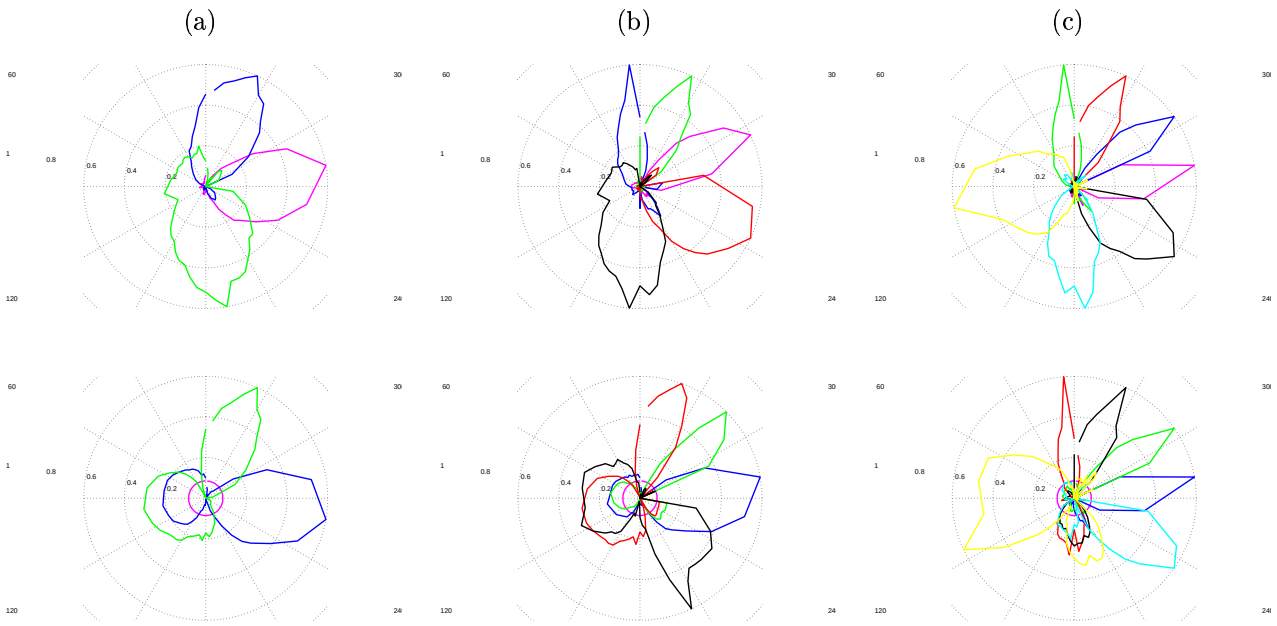


FIG. 3.18 – Robustesse des fonctions spatiales avec l'ordre de décomposition pour la méthode de subset selection : décomposition sur 3 canaux (a), 5 canaux (b) et 7 canaux (c). Cas d'une décomposition des HRTF centrées (en haut) et des HRTF non centrées (en bas).

3.5.1.2 Universalisation des fonctions spatiales

Comme pour l'ACP, les fonctions spatiales issues d'une analyse "subset selection" peuvent être universelles par construction, si la méthode est appliquée non aux HRTF de chaque tête, H, mais à leur ensemble concaténé, $[H_1\ H_2....H_N]$. Les fonctions spatiales ainsi obtenues sont présentées en Figure 3.18.
Dans le cas d'une décomposition non centrée, les fonctions spatiales sont, par construction, discrètes. Les lobes s'affinent en fonction de l'ordre de décomposition, et ce n'est que pour l'ordre le plus élevé qu'apparait un lobe pour les positions contralatérales. Pour la décomposition des données centrées, les lobes sont nettement moins compacts, et font apparaitre un lobe secondaire, orienté vers les positions arrières ou contralatérales. Ces fonctions spatiales sont donc de mauvaises candidates pour une accentuation de la compacité : la procédure pratiquée pour la décomposition sur 7 canaux induit une forte erreur pour les positions arrière où la résolution de notre système auditif est pourtant très précise.

Pour tous les ordres de décomposition, la qualité de reconstruction est meilleure pour une décomposition de données centrées au dessous de 1kHz. Contrairement au cas de l'ACP/ACI, l'erreur observée sur cette zone peut être significativement supérieure à celle obtenue pour les fréquences supérieures, par exemple pour la décomposition sur 3 ou 5 canaux présentée en Figure 3.19. En outre, dans le cas centré, l'augmentation de l'erreur est régulière en fonction du nombre de canaux, environ 1.5dB par canal supplémentaire.

3.5.2 Lien avec le paradigme des haut-parleurs virtuels

Le paradigme des haut-parleurs virtuels consiste à "binauraliser" les canaux d'un encodage multicanal destiné à une écoute sur haut-parleurs. Plus précisemment, chacun de ces canaux est considéré comme signal d'entrée d'un encodeur binaural bicanal statique, qui spatialise la source à la position du haut-parleur auquel le canal était initialement affecté. Ce paradigme constitue une approche supplémentaire pour la décomposition linéaire des HRTF à phase mixte, et peut s'implanter selon le schéma 3.1. Les lois de panpot que nous avons étudiées pour réaliser l'encodage directionnel, i.e. les fonctions spatiales, sont les panpots d'intensité et ambisonic. Quelques résultats sur la qualité de reconstruction sont présentés en [JLP99]. Nous utilisons par la suite un panpot d'intensité, qui offre des fonctions spatiales discrètes.

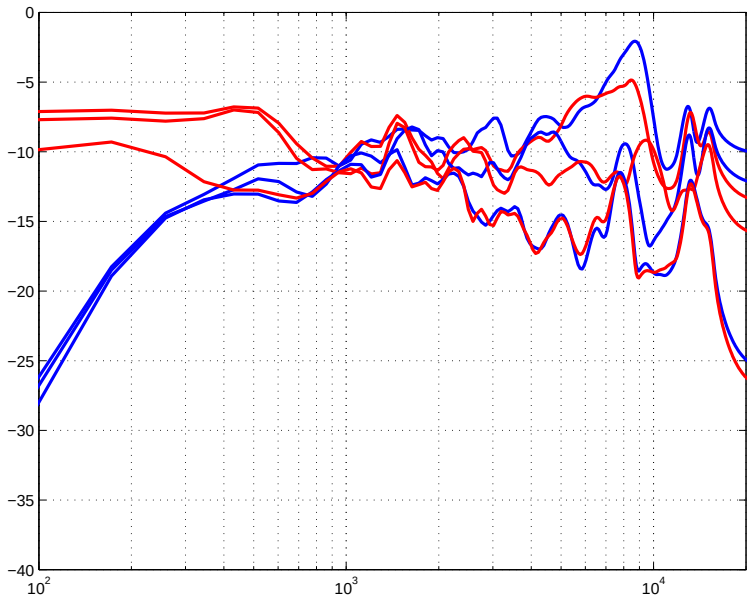


FIG. 3.19 – Erreur de reconstruction dans le plan horizontal avec la méthode de subset selection universelle : décomposition des HRTF à phase minimale centrées (en bleu) et non centrées (en rouge), sur 3, 5 ou 7 canaux.

Dans ce cas, l'erreur de reconstruction est par construction nulle lorsque la source cible coincide avec la position d'un haut-parleur virtuel, et est obtenue par interpolation linéaire de deux HRTF pour les positions entre deux haut-parleurs.
Le choix de la position des haut-parleurs virtuels est souvent choisie pour adhérer à des formats normalisés (format 5.1) ou bien pour réaliser une distribution régulière de l'espace. La méthode de "subset selection" appliquée à des HRTF à phase mixte, permet d'optimiser ce choix en terme de qualité de recontruction. Les fonctions spatiales associées peuvent ou bien être les lois de panpot pour les positions désignées, ou bien des fonctions spatiales obtenues par projection orthogonale, permettant une meilleure reconstruction au prix d'une moins forte compacité.

Par extension, on peut définir une décomposition des HRTF à base de "haut-parleurs virtuels à phase minimale", dans laquelle les filtres de reconstruction sont constitués de paires de HRTF à phase minimale, et les fonctions spatiales constituent une loi de panpot multi-haut-parleurs. Cette méthode diffère du paradigme standard des haut-parleurs virtuels, puisque l'ITD est synthétisé explicitement. Elle s'écarte également de la décomposition "subset selection" telle qu'appliquée plus haut, puisque les HRTF sont sélectionnées par paire droite/gauche. Toutefois, la méthode peut être utilisée, mais cette fois, à un ordre deux fois inférieur.

3.6 Comparaison objective des différentes approches

Les résultats que nous avons présenté pour chacune des trois approches nous conduisent aux conclusions suivante :

1. Erreur de reconstruction aux moindres carrées (Figure 3.21)
   (a) Par définition, erreur ACI = erreur ACP.
   (b) Toutes les méthodes conduisent à une erreur similaire sur l'intervalle [1-6kHz].
   (c) Le binaural B est moins précis que les autres méthodes en hautes fréquences.
   (d) Centrer les données avant la décomposition permet une meilleure reconstruction en basses fréquences. Ce résultat s'explique parce-que le filtre moyen, qui devient alors filtre de reconstruction, contient l'information commune à toutes les positions, et ainsi l'information en basses-fréquences.

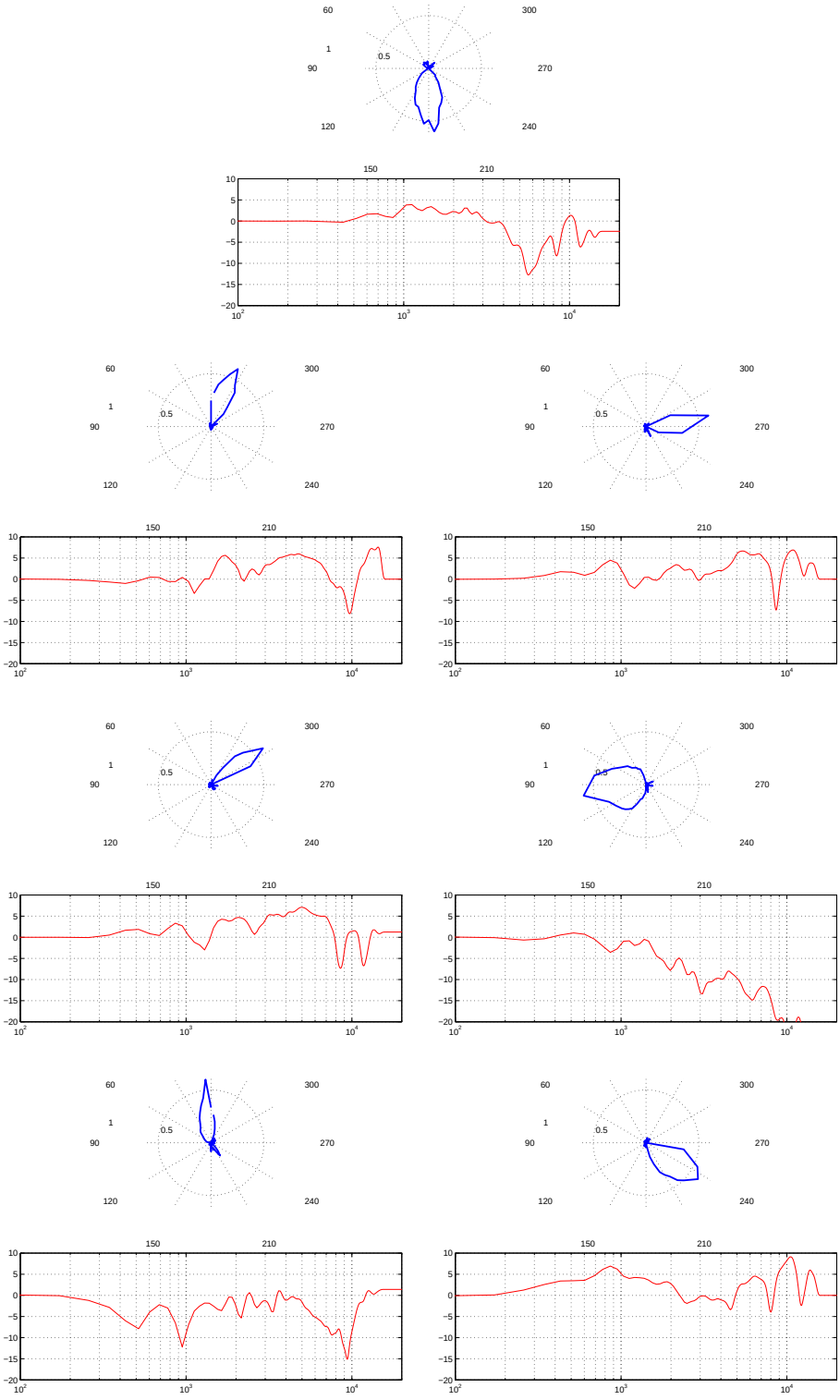


FIG. 3.20 – Fonctions spatiales dans le plan horizontal et filtres de reconstruction pour une décomposition "subset selection" non centrée sur 7 canaux.

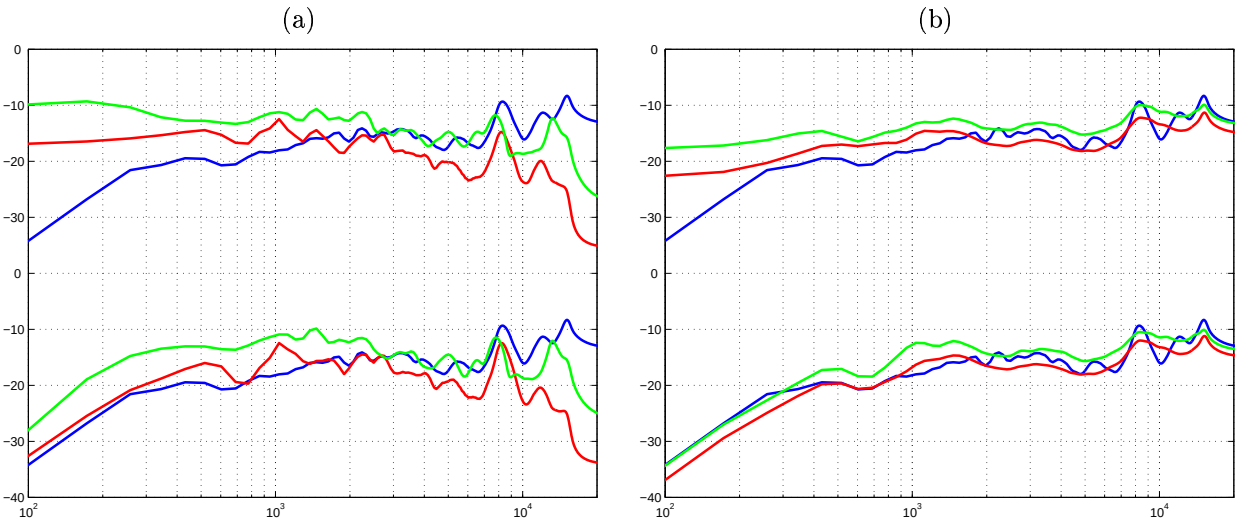


FIG. 3.21 – Erreur de reconstruction des HRTF à phase minimale dans le plan horizontal pour une décomposition individuelle (a) et universelle (b) : Binaural B (en bleu), ACP/ACI (en rouge), subset selection (en vert). Données non centrées (en haut) et centrées (en bas).

(e) La méthode de "subset selection" conduit à une erreur légèrement supérieure à l'analyse statistique du fait de la contrainte imposée sur les filtres de reconstruction.

2. Différences inter-individuelles de l'implantation (Figures 3.21 et 3.23)
   (a) Avec toutes les méthodes, il est possible de dériver des fonctions spatiales universelles.
   (b) L'universalisation des fonctions spatiales augmente l'erreur de reconstruction, si bien que toutes les méthodes présentent alors une erreur semblable à celle du Binaural B.
   (c) Cette augmentation de l'erreur est localisée au dessus de 9kHz.
   (d) Les différences interindividuelles sont conservées par la décomposition, et se situent 10dB au dessus de l'erreur de reconstruction.
3. Efficacité de l'encodeur (Figure 3.22)
   (a) les décompositions appliquées sur des données centrées ne conduisent pas à des fonctions spatiales compactes (du fait de la fonction omnidirectionelle), mis à part l'ACI.
   (b) ACP et Binaural B conduisent à des fonctions spatiales peu compactes.
   (c) ACI d'une part et "subset selection" de données non centrées d'autres part, conduisent à des fonctions spatiales à support très compact, privilégiant les mêmes positions.
   (d) Pour un faible ordre de décomposition (inférieur à 7), la méthode de "subset selection" sur données non centrées conduit à des fonctions spatiales plus compactes que l'ACI.
   (e) L'accentuation de la compacité, ou discrétisation "forcée", augmente l'erreur de reconstruction, en basses fréquences. Les performances de l'ACI et "subset selection" sont semblables, mais sont à présent inférieures à celles du Binaural B jusqu'à 2kHz.

3.7 Conclusion

Dans ce chapitre, nous avons présenté et comparé plusieurs approches réalisant une décomposition linéaire des HRTF à phase minimale permettant de les décrire avec un nombre réduit de fonctions spatiales et de filtres de reconstruction :

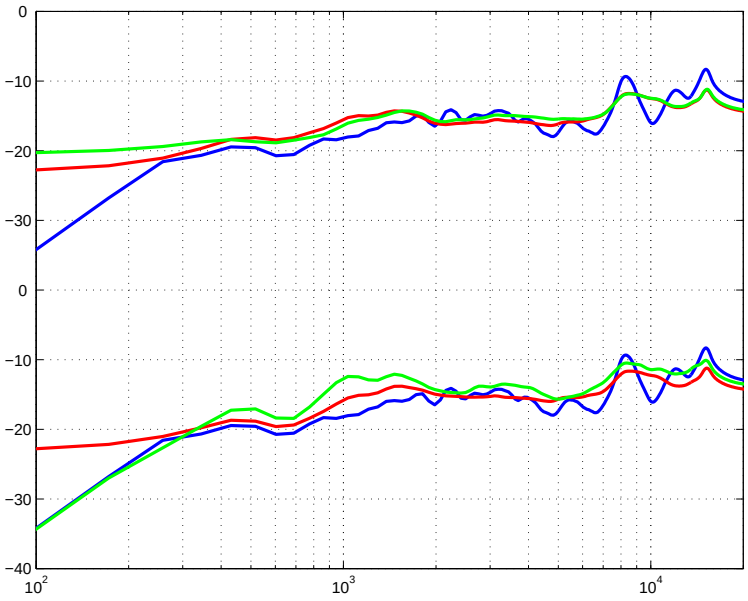


FIG. 3.22 – Erreur de reconstruction des HRTF à phase minimale dans le plan horizontal pour des fonctions spatiales avec compacité accentuée : ICA (en rouge) et subset selection (en vert). Courbe de référence : Binaural B (en bleu).

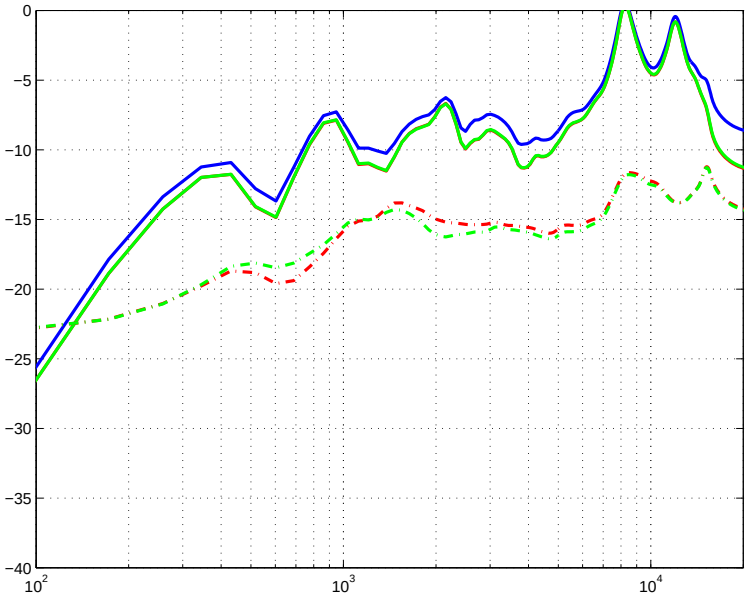


FIG. 3.23 – Distance entre têtes avant décomposition (trait continu bleu) et après décomposition ICA universelle et idéalement compacte (trait continu vert). Erreur de reconstruction pour une ICA sur données centrées (trait pointillé rouge), sur données centrées (trait pointillé vert). Seul le plan horizontal est considéré.

1. Analyses statistiques des données : Analyse en composantes principales (ACP) et en composantes indépendantes (ACI),
2. Projection des HRTF sur les harmoniques sphériques, conduisant au format Binaural B,
3. Méthode de subset selection.

ACP et ACI minimisent la reconstruction aux moindres carrés des HRTF. Pour toutes ces méthodes, on peut définir des fonctions spatiales universelles, et utiliser ainsi un encodeur binaural multicanal indépendant de la tête. Pour l'approche d'universalisation retenue, l'erreur de reconstruction se trouve altérée en hautes fréquences par rapport à une décomposition "individuelle".
ACI et subset selection conduisent à des fonctions spatiales discrètes, ou à support compact, pour une décomposition d'ordre assez élevé (>3). Cette propriété permet d'envisager une réduction du coût d'implantation de l'encodeur, dont seuls quelques canaux sont actifs simultanément pour recréer une position. Centrer les données avant la décomposition améliore les performances de reconstruction en basses fréquences, mais pénalise cette économie, du fait de la présence d'une fonction spatiale omnidirectionnelle. L'ACI permet seule de s'affranchir de ce dernier problème et de réaliser ainsi le meilleur compromis "qualité-coût".

Les seules mesures objectives montrent que ces méthodes de décomposition préservent une large proportion des différences inter-individuelles, information concentrée dans les filtres de reconstruction du décodeur.

Chapitre 4

Synthèse binaurale bicanale

4.1 Introduction

L'implantation de la synthèse binaurale a été examinée dans le cas standard d'une structure bicanale (chapitre 2), comme dans celui d'une structure multicanale (chapitre 3). Pour les deux approches, des mesures objectives ont permis de mesurer les pertes consenties. Dans ce chapitre, nous proposons une évaluation perceptive de la qualité de localisation offerte par une implantation bicanale. Un second test permet de confirmer que les formats multicanaux envisagés atteignent une qualité proche de l'implantation bicanale de référence.

4.2 Performances subjectives de l'implantation bicanale

Dans cette partie, nous présentons une étude perceptive visant à caractériser les performances d'un système de synthèse binaurale tel que le *Spat˜*. Ses caractéristiques sont les suivantes :

1. il repose sur un modèle (ITD+filtre à phase minimal) des HRTF,
2. la structure d'implantation est bicanale,
3. les HRTF utilisées ne sont pas celles de l'auditeur.

En outre, nous nous concentrons sur la reproduction d'un champ libre créé par une source sonore statique, donc sans suivi de la position de la tête de l'auditeur.

4.2.1 Objectifs du test

Notre objectif est double :

1. Evaluer l'influence de la tête choisie pour les mesures de HRTF sur la qualité de localisation. La comparaison d'un ensemble de têtes a fait l'objet de plusieurs investigations : comparaison de 2 têtes humaines pour la synthèse binaurale (Asano [ASS90]), comparaison de 30 têtes humaines (Moller et al. [MJHS96]) ou de 8 têtes artificielles ([MHJS99]) pour l'enregistrement binaural.
2. Evaluer les performances de systèmes utilisant la synthèse binaurale, qui serviront de référence pour l'analyse du test mené en chapitre 3. On pourra rapporter les résultats aux étude menées sur la synthèse binaurale individuelle (Wightman et Kistler [WK89b], Chateau [Cha96], Huopaniemi et Zacharov [HZK99], Bronkhorst [Bro95]), synthèse binaurale avec tête artificielle (Gardner [Gar97]), ou avec tête humaine (Wenzel et al. [WAKW93]).

Pour évaluer qualitativement la dégradation de la qualité localisation par rapport à une écoute naturelle, on se reportera également aux travaux de Blauert ([Bla97]), Makous et Middlebrooks ([MM90]) et Oldfield et Parker ([OP84]).

4.2.2 Mise en place du test

4.2.2.1 Formulation des questions posées aux sujets

Nous souhaitons connaitre l'incidence perçue par le sujet pour un ensemble de positions de la source et à ce titre mettons en oeuvre un test de localisation absolue de sources sonores statiques. Nous avons envisagé, sans les retenir, certaines alternatives :

- tests de comparaison A/B de sources sonores statiques, où le sujet doit exprimer la différence perçue entre deux stimuli, comme l'ont entrepris Huopaniemi et Zacharov. Cette approche est d'un intérêt immédiat pour notre objectif de comparaison des têtes. En effet, elle donne un accès immédiat à une distance subjective inter-tête, par opposition à un test de localisation absolue qui, nous le verrons, conduit à plusieurs distances, une pour chaque critère d'analyse. Toutefois, ses inconvénients sont doubles : la comparaison par paire conduit à une durée de test très longue ; la différence perçue peut être ambigue, et peut s'apparenter à une différence de position tout comme à une différence de timbre. Il est alors difficile d'isoler la modalité qui nous intéresse, la localisation.
- test d'identification de la source, dans lequel une représentation visuelle de l'ensemble des sources est installée physiquement autour du sujet. Celui-ci doit donc indiquer la source réelle coïncidant avec le son perçu, parmi un ensemble limité de possibles. C'est la procédure adoptée par Moller et al. ou par Hartmann et Rakerd ([RH85]).
- test de localisation dynamique, dans lequel le stimulus est une trajectoire. Cette trajectoire peut être par exemple un cercle d'iso-ITD (cône de confusion), ou une tranche d'élévation constante. Le sujet doit alors exprimer le degré de latéralisation ou l'élévation perçue. L'intérêt de cette approche, que nous n'avons pas rencontrée dans la littérature, consiste en l'obtention d'un jugement global, "intégrant toutes les positions". Cela nous soustrait à l'arbitraire du choix d'un nombre réduit de positions à tester.

4.2.2.2 Choix d'une interface de réponse

L'interface doit permettre au sujet d'expliciter sa perception. On souhaite notamment éviter tout biais introduit par une interface trop peu intuitive.
Plusieurs solutions ont été adoptées dans la littérature :

- un rapport verbal (Wightman et Kistler, Wenzel, Gardner, Begault [Beg92]),
- un geste en direction de l'incidence perçue : orientation de la tête, du torse, pointage avec le bras (Makous et Middlebrooks, Bronkhorst, Brungart [BDR99b], Oldfield et Parker, Asano). C'est la technique apparemment la plus précise ([GGE+95]).
- une interface un peu spéciale : une boule de 20cm, allégorie du repère dans lequel la source se déplace, que le sujet tient entre les mains et sur laquelle il pointe un stylet (Gilkey et al. [GGE+95]). Cette interface semble permettre une réponse beaucoup plus rapide et aussi précise que la précédente (de 2 à 8 fois).
- une interface informatique sur laquelle le sujet indique l'incidence perçue (Chateau).

C'est cette dernière solution que nous avons retenue, pour sa commodité de mise en place. Notre interface est représentée en Figure 4.1. Le sujet mentionne azimut et élévation perçus à l'aide de deux curseurs, de même qu'un paramètre de distance. Les réponses à ce dernier paramètre ne seront que succintement analysées, vu l'apparente difficulté d'utilisation exprimée par les sujets.

En outre, le sujet a la possibilité de spécifier l'occurence d'un son non-localisable à l'aide d'une bascule qu'il lui suffit de "cocher". Les sons non localisés désignent des sons dont l'incidence ne peut être décrite en termes d'azimut et d'élévation. Cette option, échappatoire à la tâche de localisation, ne fait pas partie des "classiques" des interfaces proposées dans la littérature. Selon nous, il était important d'expliciter un phénomène courant dans le cadre des simulations binaurales : celui d'un son perçu au centre de la tête. Ce phénomène est un cas particulier de la localisation intra-crânienne, ou "Inside-the-head locatedness", pouvant apparaître lors de la simulation d'une source située dans le plan médian. Il est caractéristique de la localisation en synthèse binaurale statique, apparaissant en conséquence de l'absence d'indices dynamiques ([Bla97], p. 382). La formulation "son non-localisable" plutôt que "son perçu au centre de

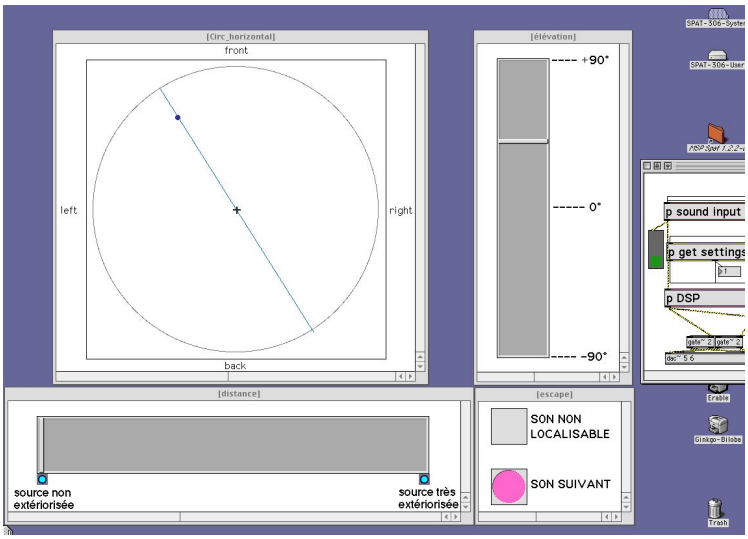


FIG. 4.1 – Interface de réponse au test perceptif.

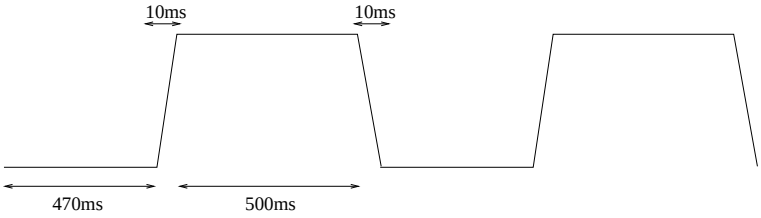


FIG. 4.2 – Enveloppe périodique du bruit blanc utilisé comme stimulus du test.

la tête" a été choisi pour son apparente simplicité, rendant son sens peut-être plus intuitif aux sujets non-expérimentés (n% des sujets du test). Les occurences de localisation au centre de la tête peuvent également s'exprimer par une distance perçue nulle. C'est la seule modalité pour laquelle nous utilisons les données de ce paramètre.

4.2.2.3 Stimuli

Le son utilisé est constitué de trains de bruit blanc, dont l'enveloppe est représentée en Figure 4.2. Ce son périodique est entendu sans limite. Il est filtré par les HRTF de 17 têtes, mesurées et mises à disposition par l'Université de UC Davis (voirle chapitre 1). Les filtres à phase minimale sont modélisés sous forme IIR à l'ordre 20. Les sons sont présentés sur casque à un niveau voisin de 75dB SPL. La fréquence d'échantillonnage est de 44.1kHz.
Pour faire face à des contraintes de temps, nous avons limité le nombre de positions à tester. Nous avons choisi 16 positions de l'hémisphère gauche, représentées en Figure 4.3. Les canaux droite/gauche pouvaient être inversés, selon une procédure aléatoire, afin de créer autant de sensations à droite qu'à gauche.
Chacun des sujets devait donc localiser $17 \times 16 = 272$ stimuli, présentés dans un ordre aléatoire. Le test durait en moyenne 1h15, et était précédé par un exercice d'échauffement de 22 stimuli.

4.2.2.4 Sujets

Le test a été passé par 22 sujets, qui se divisent en 3 groupes :

- 7 sujets non expérimentés (AF, AG, BL, CD, EH, MM, VR),
- 9 sujets habitués à l'écoute critique au casque (CG, DP, EC, EF, FD, GD, OL, PM, TH, SL),
- 5 sujets familiers avec la tâche de localisation (AL, ED, IS, OW, VL).

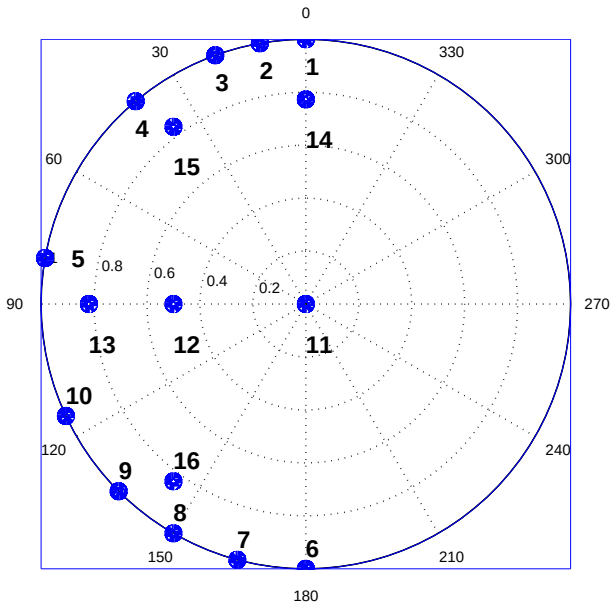


FIG. 4.3 – Positions choisies pour le test (vue de haut).

Il se déroule dans la pénombre, dans un studio silencieux. L'écran de l'ordinateur est calé dans un support incliné dégageant le champ de vision de l'auditeur.

4.2.3 Méthodes d'analyse statistique

L'analyse des résultats se concentre sur deux opérations élémentaires :

- déterminer le degré de significativité de la différence observée entre deux échantillons,
- étudier l'influence d'un facteur sur les réponses (typiquement : tel critère d'analyse dépend-il de la tête observée ?). Formellement, cela revient à comparer 3 échantillons ou plus, chacun d'entre eux correspondant à un niveau différent du facteur (tête 1, tête 2, tête 3, etc....).

Le choix des méthodes à utiliser dépend tout d'abord de la distribution de la variable observée. Si le critère d'analyse est issu d'une variable à distribution gaussienne, nous avons recours à la famille des tests paramétriques, par opposition aux tests dits non-paramétriques, qui permettent d'étudier des variables à distribution quelconque. Utiliser un test non-paramétrique alors que la variable a une distribution gaussienne est sous-optimal dans la mesure où des différences significatives risquent de demeurer masquées.

Les test paramétriques, qui sont les plus fréquemment employés, font l'hypothèse supplémentaire de l'égalité des variances des échantillons comparés. Or, comme nous le verrons, la différence de variance est au contraire un des éléments caractéristique de nos données. Par exemple : certaines positions sont perçues avec un plus grand écart-type car notre système de localisation est moins précis dans cette région de l'espace. Nous utilisons donc principalement des tests non-paramétriques. Ce qui les distingue fondamentalement des précédents, c'est que la comparaison porte non sur les valeurs des échantillons mais sur les rangs issus du classement de ces valeurs ([Bea96]). Ils ne s'intéressent donc qu'aux relations ordinales entre les échantillons.

Notre choix doit également tenir compte de la relation entre les échantillons comparés. Ces échantillons peuvent être appareillés (corrélés) ou indépendents. Si l'on prend notre exemple, des échantillons seront appareillés si nous comparons les réponses d'une même population dans plusieurs conditions (comparaison du taux de confusion obtenu à deux positions par chaque tête). Nous utiliserons les tests pour échantillons indépendants dans les cas où cette corrélation n'est pas totale, par exemple lorsque nous n'avons pas le même nombre d'éléments dans chaque échantillon (comparaison de l'azimut perçu à deux positions, qui ne totalisent pas le même nombre de sons non localisés).

Dans les deux sections suivantes, nous esquissons une description des principes des tests non-paramétriques utilisés, et mentionnons les tests paramétriques "équivalents".

4.2.3.1 Comparaison de 2 échantillons

type de test :	paramétrique	non-paramétrique
échantillons indépendants	t-test	Mann-Whitney
échantillons appareillés	t-test (repeated measures)	Wilcoxon

Le test de Mann-Whitney consiste à former l'ensemble des valeurs des deux échantillons (A et B), à les classer de la plus petite (rang 1) à la plus grande (rang N). Si A est significativement différent de B, ses éléments auront une tendance à être systématiquement supérieurs aux éléments de B (par exemple). Pour détecter cette occurence, on comptabilise le nombre total U de couples où un élément de A est supérieurs à un élément de B. Ce nombre U est asymptotiquement gaussien, approximation excellente dès que les échantillons ont plus de 8 éléments chacun ([Sap90]). Pour deux échantillons issus de la même population, on sait estimer la moyenne et la variance de U. Le test de Mann-Withney consiste donc à évaluer l'écart entre la valeur observée pour U et la moyenne théorique. Si cet écart est grand, au niveau de risque d'erreur consenti, les deux échantillons sont significativement différents.

Le test de Wilcoxon (Wilcoxon Signed-rank test) utilise l'hypothèse de corrélation entre les échantillons pour effectuer le classement non sur les valeurs des échantillons, mais sur la différence de ces valeurs (A-B), ou, plus exactement sur la valeur absolue de la différence. Toutefois, on doit tenir compte du signe de (A-B) : si cette valeur prend la valeur 5 puis -5, ces deux occurrences vont obtenir le même rang alors qu'elles ne traduisent pas une "systématicité" de la supériorité des échantillons de A par rapport à ceux de B (ou inversement). On affecte donc à chaque rang le signe de la différence (A-B) (d'où le nom de "Signed-rank"). La somme de ces rangs signés, W, suit asymptotiquement une loi gaussienne si les échantillons ne sont pas significativement différents. La décision est prise par comparaison à cette distribution théorique.

4.2.3.2 Comparaison de 3 échantillons ou plus

type de test :	paramétrique	non-paramétrique
échantillons indépendants	ANOVA	Kruskal-Wallis
échantillons appareillés	ANOVA (repeated measures)	Friedman

Le test de Krukal-Wallis suit la même approche que le test de Mann-Withney. Toutefois, le rang moyen obtenu pour chaque échantillon est également utilisé pour évaluer la variance entre les échantillons SS_{bg} (between groups sum of squared deviates). On sait évaluer la valeur théorique de SS_{bg} dans le cas d'échantillons issus de la même population :

$$SS_{bg}^{th} = df.\frac{N.(N+1)}{12}$$

où df (degrees of freedom) vaut le nombre d'échantillons comparés moins un, et où N désigne la taille de chaque échantillons. Le rapport $SS_{bg}/SS_{bg}^{th} \times df$ suit asymptotiquement une distribution du χ_2(chi deux). Le test consiste à comparer la valeur obtenue pour ce rapport à la valeur critique du χ_2, fixée par le degré de liberté de nos données et le risque consenti.

Le test de Friedman se distingue du précédent sur le mode de classement utilisé. Ici, il est effectué pour chaque observation, et fournit une estimation du rang moyen, moyenne sur l'ensemble des observations, pour chaque échantillon. L'influence du facteur observé est détectée selon les mêmes modalités que pour le test de Kruskal-Wallis. Pour les représentations graphiques, nous affichons le rang moyen de chaque tête ainsi que le rang théorique dans le cas d'échantillons issus de la même famille (différence non-significative entre les échantillons).

4.2.4 Resultat 1 : Sons non localisés

Le sujet mentionnait les stimuli non-localisables à l'aide d'un élément prèvu à cet effet sur l'interface de réponse (clic sur une bascule). Ce phénomène est étudié avec soin à plusieurs titres :

- il constitue un premier critère d'estimation de la qualité de localisation obtenue, et éventuellement un premier critère pour départager les têtes.
- il conditionne l'approfondissement de cette première analyse par l'étude de la localisation en azimut et en élévation : si une proportion trop importante des sons proposés a été non-localisée, l'exploration des réponses ne peut aller plus loin.

4.2.4.1 Origine du phénomène

Par "son non-localisable", nous désignons un son qui ne peut être décrit en termes d'azimut et d'élévation : un son perçu au centre de la tête. Comme le précise Blauert (pp. 132 à 137 [Bla97]), les occurences de localisation intra-crânienne trouvent leur origine dans ***"l'altération de la contribution de l'oreille externe par rapport à une situation d'écoute naturelle"*** : mauvaise équalisation du casque d'écoute, utilisation de HRTF non individuelles, et, nous le rajoutons, compromis liés à l'implantation.

4.2.4.2 Dépendance vis à vis de la tête

Un test de Friedman mené sur la répartition des occurences montre que la tête n'est pas un facteur de variation significative (χ_2=11.4, df=16). On choisit donc d'observer globalement l'ensemble des réponse. Pour la fréquence de localisation en revanche, la tête est un facteur significatif de variation (risque 5%), et nous présentons en Tableau 4.1 les performances de la meilleure tête, la tête 11, ainsi que celles d'une tête hybride minimisant le nombre d'occurences pour chaque position.

4.2.4.3 Distribution spatiale des occurences de non-localisation

Comme on peut l'observer sur la Figure 4.4, les occurences de non localisation concernent pour une forte majorité le plan médian (60%). Près d'une fois sur 5 (17%), il s'agit de la position 14 (0°,34°). Ce score est suivi par celui de la position 1 (0°,0°), qui représente 15% des occurences, de la position du zénith (13%), et enfin de la position 6 (11%). Les occurences sont en revanche quasi inexistantes pour les autres positions, mis à part les voisines du plan médian. C'est également pour la région du plan médian que Begault ([Beg92]) et Gardner ([Gar97]) observent la plus faible distance perçue avec synthèse binaurale.

4.2.4.4 Fréquence des occurences de non-localisation

La fréquence des non-localisation désigne la proportion de sons non localisés parmi l'ensemble des stimuli. On constate que sur 3 jugements émis pour les positions 1, 11 ou 14, l'un d'entre eux fait part d'une non-localisation, proportion descend à 1 jugement sur 5 pour les positions du plan horizontal concernées (2 et 6). Pour les autres positions, l'occurence d'une non-localisation survient dans moins de 10% des cas.

4.2.5 Resultat 2 : Phénomène de confusion en azimut

La confusion désigne le phénomène par lequel une source sonore située à une position frontale est localisée à l'arrière (confusion avant→arrière), et réciproquement. Pour une implantation pratique ne faisant pas cas particulier des positions latérales, nous suivons la définition de [BDR99b] : il y a confusion dès que l'écart entre l'azimut redressé et la référence est diminuée de plus de 10° par rapport à l'écart initial. Nous écartons nénamoins les positions du plan frontal (12 et 13).

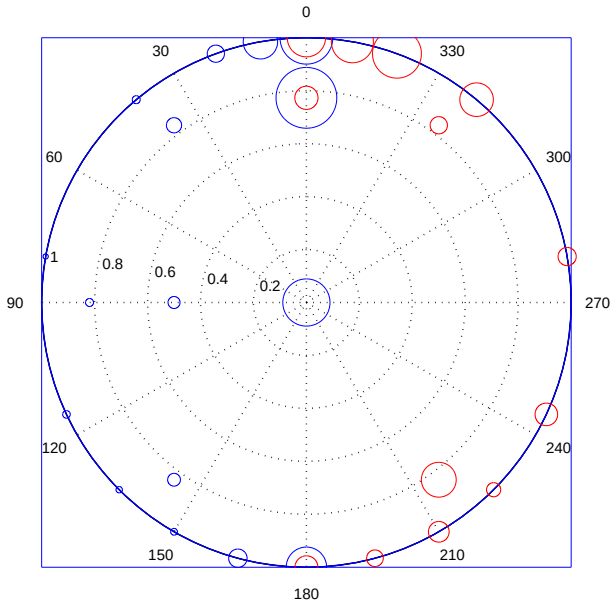


FIG. 4.4 – Distribution des artefacts de localisation pour les 15 positions testées : demi-plan de gauche˜ : occurences de non localisation (bleu) ; demi-plan de droite : occurences de confusion (rouge). Les calculs sont effectués en sommant les occurences observées pour les 22 sujets et les 17 têtes.

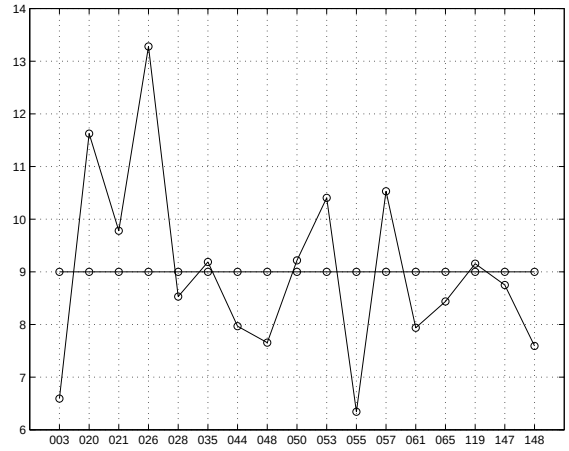


FIG. 4.5 – Influence de la tête sur les occurences de non-localisation (test de Friedman) : rang théorique si la tête n'était pas un facteur déterminant, et rang moyen observé (bleu).

n°	position	non-localisation		
		tête 055	taux min	tête
1	(0°,0°)	32%	23%	044
2	(10°,0°)	23%	9%	061
3	(20°,0°)	14%	5%	020
4	(40°,0°)	0%	0%	003
5	(80°,0°)	0%	0%	003
10	(115°,0°)	9%	0%	044
9	(135°,0°)	0%	0%	003
8	(150°,0°)	0%	0%	044
7	(165°,0°)	5%	0%	003
6	(180°,0°)	5%	18%	055
11	(-,90°)	32%	0%	028
12	(90°,60°)	9%	0%	021
13	(90°,35°)	9%	0%	003
14	(0°,34°)	32%	9%	028
15	(37°,33°)	9%	0%	119
16	(143°,33°)	0%	0%	055

TAB. 4.1 – Fréquence d'occurrence des "non-localisation".

n°	position	confusions en azimut				confusions en élévation
		taux moyen	tête "028"	taux min	tête	
1	(0°,0°)	–	32%	23%	020	–
2	(10°,0°)	–	27%	14%	065	–
3	(20°,0°)	–	27%	18%	148	–
4	(40°,0°)	–	9%	9%	028	–
5	(80°,0°)	–	23%	5%	057	–
10	(115°,0°)	–	5%	5%	028	–
9	(135°,0°)	–	0%	0%	003	–
8	(150°,0°)	–	23%	5%	020	–
7	(165°,0°)	–	0%	0%	028	–
6	(180°,0°)	–	9%	9%	028	–
11	(-,90°)	–	–	–		11%
12	(90°,60°)	–	–	–		9%
13	(90°,35°)	–	–	–		7%
14	(0°,34°)	20%	–	–		8%
15	(37°,33°)	15%	–	–		5%
16	(143°,33°)	30%	–	–		20%

TAB. 4.2 – Fréquence d'occurrence des confusions.

4.2.5.1 Origine du phénomène

Le phénomène de confusion est lié à l'existence de positions générant des différences interaurales identiques. Le lieu de ces positions est couramment dénommé "cône de confusion". En ces points, seuls les indices de localisation spectraux diffèrent.

Pour la localisation de sources réelles, les confusions sont quasi-inexistantes car de petits mouvements de la tête permettent au système auditif de s'appuyer sur la variation des indices interauraux pour recouvrir l'incidence de la source ([Wal40], [Bro95], [WK99]). L'efficacité de cette stratégie n'est efficace que lorsque le stimulus a une durée suffisante, que Makous et Middlebrooks évaluent à 650ms±320ms. Lorsque les sujets localisent des sources réelles avec tête immobile (ou en présence d'un stimulus trop court), les indices spectraux seuls ne suffisent pas à résoudre l'ambiguité, puisque le taux de confusion augmente de façon significative ([MHJS99]). Makous et Middlebrooks observent alors un taux oscillant entre 2 à 10% suivant les sujets, résultats semblables à ceux de Wightman et Kistler (de 3 à 12%). D'après Asano et al., ces scores témoignent du manque d'habitude des auditeurs à localiser en l'absence d'indices dynamiques, et pourraient en partie être réduits grâce à un plus grand entraînement des sujets à la tâche.

Pour une synthèse binaurale statique et individuelle, le taux de confusion augmente jusqu'à doubler, observent Wightman et Kistler. Cela peut s'expliquer par le fait qu'en général, les sujets ont les yeux bandés ou fermés lors des expériences. Ils n'ont donc pas accès aux stimuli visuels, qui, dans le cas de sources frontales, contribuent au jugement de localisation, voire le détermine. D'ailleurs, dans les expériences de Moller et al., les sujets ont les yeux ouverts et peuvent voir les haut-parleurs ayant servi lors de l'enregistrement. Les résultats obtenus ne montrent pas d'augmentation significative du taux de confusion par rapport à la localisation statique de sources réelles.

Asano et al. montrent en outre que les confusions sont d'autant plus nombreuses que le modélisation des HRTF à phase minimale est grossière. Pour un modèle IIR d'ordre 20 tel que le notre, le taux de confusion oscille entre 25.9 et 35.2% selon le sujet.
Enfin, on peut s'attendre à ce que ce taux d'occurences soit encore augmenté par l'utilisation d'indices spectraux non-individuels, diffèrant de ceux auxquels l'auditeur se réfère habituellement pour résoudre l'ambiguité (modèle d'association, [The86]).

4.2.5.2 Dépendance vis à vis de la tête

Le choix de la tête n'influe pas sur la répartition des occurences de confusions (test de Friedman : χ_2=20.4, df=16), qui sera donc analysée toutes têtes confondues. En revanche, la fréquence des confusions dans le plan horizontal varie de façon significative en fonction de la tête (Friedman à 10% de confiance). Comme on l'observe en Figure 4.6, c'est la tête "028" qui montre le plus faible taux. Ce sont donc les performances de cette tête que l'on présente dans le Tableau 4.1 pour le plan horizontal.
Une alternative au choix de la tête favorisant le moins de confusions "en moyenne" sur toutes les positions, on peut chercher à minimiser ce taux pour chaque position, et définir une tête hybride constituée des HRTF de têtes éventuellement différentes pour chaque position. Les performances de cette tête sont mentionnés dans le tableau 4.2.

Pour les positions hors du plan horizontal, en revanche, la tête n'est pas un facteur influençant la fréquence des confusions, et l'on peut alors se permettre de mélanger les têtes pour obtenir une etimation plus robuste.

4.2.5.3 Distribution spatiale des confusions en azimut

Les résultats illustrés en figure 4.4 montrent tout d'abord que si les inversions se sont produites sur les 15 positions observées, elles apparaissent de façon privilégiées autour du plan médian. Les positions les moins sujettes aux confusions sont les positions les plus latérales, ce qu'appuient les résultats de Chateau,

pour lequel les confusions du plan horizontal sont confinées aux azimut inférieurs à 60°.

Dans le plan horizontal, les inversions apparaissent de façon prédominante pour les positions frontales : dans 65% des cas, ce sont des sons situés vers l'avant qui sont reportés comme provenant de l'arrière. La prédominance des confusions avant→arrière a été fréquemment observée dans la littérature et trouve en partie son origine en l'absence de stimuli visuels. Wenzel observe que 60% des confusions sont des confusions avant→arrière, prédominance que l'on retrouve chez Gardner et Chateau pour lesquels les confusions arrière→avant sont quasi-inexistantes.
On note néanmoins que ce déséquilibre s'estompe, voir même s'inverse pour les positions en élévation : pour les positions 15 et 16, c'est le son provenant de l'arrière qui a majoritairement été inversé, le taux de confusion passant du simple au double (5% de confusion avant→arrière contre 10% de confusion arrière→avant). Les résultats détaillés de Chateau permettent de constater que, comme nous, cette prédominance est inversée pour une élévation intermédiaire (40°), pour des positions semi-latérales (45°et 135°) : c'est la position arrière qui est perçue devant dans 70% des occurences de confusion.

Nous essayons d'interpréter ces résultats au regard des indices monauraux en charge de la discrimination avant-arrière. Le rapport entre les HRTF ipsilatérales des azimuts 0° et 180° du plan horizontal d'une part, des positions 15 et 16 d'autre part, sont présentés en Figure 4.7. Plusieurs éléments apparaissent sur nos données :

1. pour les positions horizontales, les indices de l'avant sont très faiblement saillants, voire inexistants pour la zone autour de 500Hz. Au contraire, l'indice arrière autour de 1kHz est net, puisqu'il atteint un contraste moyen de 7dB. Nous observons les mêmes caractéristiques pour les rapports formés avec HRTF voisines du plan médian (positions 2 et 3) et la position arrière à 180°.
2. pour les positions en élévation, c'est au contraire la bande avant autour de 4kHz qui apparait comme l'indice de discrimination avant-arrière prépondérant.

Ces rapides observations permettent d'ajouter quelques explications à la prépondérance des confusions avant→arrière observées pour les positions du plan horizontal proches du plan médian. Outre l'absence d'indices visuels, il apparait que les indices spectraux susceptibles d'une perception frontale sont très faibles. Au contraire, pour l'arrière, les indices sont peu dépendant de la tête et représentent nettement les zones du spectre les plus riches en énergie.
Pour les positions 15 et 16, en revanche, les indices avant sont saillant et peuvent induire une perception en faveur de la zone frontale. Pour la position arrière en revanche, la HRTF ne présente pas de différence marquante entre les bandes.

4.2.5.4 Fréquence des confusions

Le taux de confusion des positions frontales vaut en moyenne 23% (1 son sur 4 est confus), alors que pour les positions symétriques à l'arrière, le taux moyen n'est que de 9% (cf Tableau 4.2). Ces chiffres sont inférieurs à ceux observés par Gardner avec la tête artificielle KEMAR (resp. 70.5% et 3.4%), du même ordre que ceux de Wenzel obtenus avec des HRTF humaines sans modélisation des HRTF (31% en moyenne), et compatibles avec les résultats d'Asano et al. observés avec des HRTF humaines modélisées à l'ordre 20.

4.2.6 Résultat 3 : Localisation en azimut

Pour l'analyse de la localisation en azimut, nous écartons la position du zénith, ainsi que les réponses formalisées au zénith. Nous nous appuyons sur deux estimateurs, réalisant un consensus dans la littérature :

– l'erreur signée, qui permet de détecter tout biais systématique par rapport à la position de la source. On peut attribuer deux origines à ce biais perceptif, reflet d'une distorsion de la localisation en écoute réelle, et erreur propre au système, liée par exemple à la non-adaptation de la tête aux sujet, à l'absence d'indices visuels ou d'indices dynamiques. Cette erreur ne constitue pas à notre avis un obstacle majeur

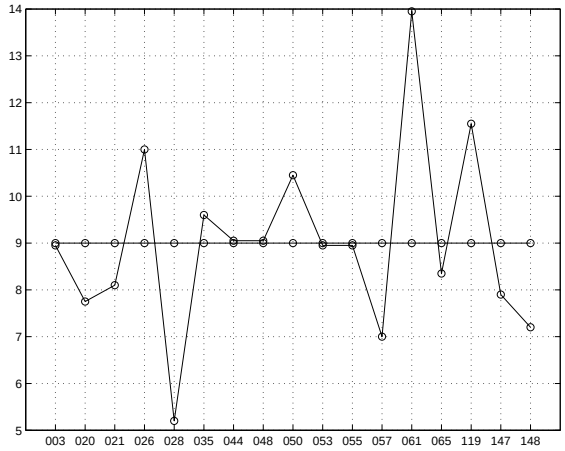


FIG. 4.6 – Influence de la tête sur le taux de confusion (test de Friedman) : rang théorique si la tête n'était pas un facteur déterminant (rouge), et rang moyen observé (bleu).

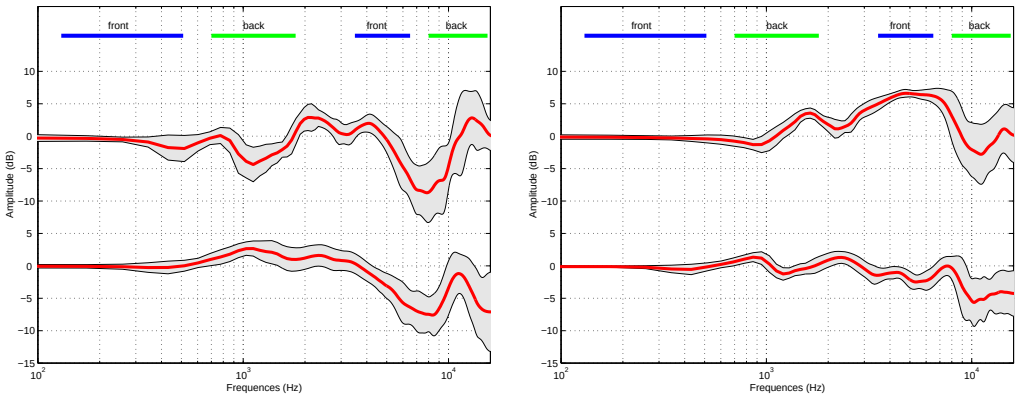


FIG. 4.7 – HRTF ipsilatérale avant (en haut) et arrière (en bas) : positions (0°,0°) et (180°,0°) (a), positions (37°,33°) et (143°,33°) (b). La moyenne géométrique des courbes sur l'ensemble des 17 têtes est représentée en rouge. La variance inter-tête est indiquée par la zone grisée. En haut sont mentionnées les bandes directionelles de Blauert pour les zones avant et les zones arrière.

pour la synthèse binaurale, dans la mesure où elle peut être en partie compensée par l'adaptation de la procédure de contrôle.
Pour chaque tête i, et chaque position k, le biais est obtenu par moyenne sur les N sujets n'ayant pas mentionné de "non-localisation" pour ce stimulus :

$$b(i,k) = \frac{1}{N(i,k)} \sum_{j=1}^{N} [az(j,i,k) - \overline{az_0}(k)]$$

– l'écart-type autour de ce biais, qui mesure la dispersion des réponses autour du biais. Nous l'interprétons comme un indicateur de la robustesse (ou reproductibilité) de la localisation. Ce critère tient à nos yeux une importance prépondérante, dans la mesure où une robustesse raisonnable conditionne seule l'analyse et l'éventuelle correction des défauts observés. Elle est estimée par :

$$\sigma(i,k) = \sqrt{\frac{1}{N(i,k)-1} \sum_{j=1}^{N} [az(j,i,k) - b(i,k)]^2}$$

4.2.6.1 Intégration des confusions

Les sons confus peuvent ou bien être écartés de l'analyse, ou bien y être inclus après un "redressement" de l'azimut perçu. Cette correction consiste à adopter l'azimut symétrique par rapport au plan frontal. Afin de faire notre choix, nous estimons biais et robustesse dans deux conditions :

– sons non confus,
– sons confus (après redressement) intégrés aux sons non confus.

Pour chacune des 15 positions considérées, nous comparons les deux critères à l'aide d'un test de Wilcoxon. Nous constatons que le biais est parfois significativement différent (risque : 1%, positions 2, 4, 5, 10, 11, 12). Toutefois, pour la plupart de ces positions, cette variation du biais liée à l'intégration des azimuts redressés va de paire avec une augmentation significative de la robustesse de la localisation. Cela semble indiquer que l'estimation du biais à partir des seuls sons non confus peut manquer de consistence du fait du faible nombre de réponses sur lesquelles elle s'appuie (17 maximum). L'écart observé entre les biais ne permet pas de conclure sur une différence structurelle entre l'azimut perçu avec ou sans confusion. Nous choisissons donc d'intégrer les confusions après redressement pour la suite de l'analyse. C'est également le choix réalisé par Wightman et Kistler ou Makous et Middlebrooks.

4.2.6.2 Positions et têtes atypiques

Afin de dégager des propriétés de la qualité de localisation permise avec chacune des têtes, nous souhaitons identifier et écarter les positions peu robustes. La Figure 4.8 présente ces valeurs pour les 17 têtes et les 15 positions. Plusieurs phénomènes saillants apparaissent :

– la position 12 (90°,60°) provoque une instabilité de la localisation supérieure aux autres positions, de façon quasi-systématique,
– la tête 10 est facteur d'un accident de localisation pour la position 1, sans que ce manque de reproductibilité ne soit observable sur les autres têtes. Voir si ça s'explique par leurs ITD/ITF (car localisation en azimut est prédominée par indices interauraux).

Nous retirons de l'analyse la position 12 ainsi que la tête 053. Sur ce sous-ensemble de réponses, l'analyse de Friedman montre une influence significative du choix de la tête pour le plan horizontal (risque de 5%). Comme on l'observe en Figure 4.9, c'est la tête 028 qui, comme pour l'analyse des confusions, présente les meilleures performances. Voir si ça s'explique par leurs ITD/HRTF.

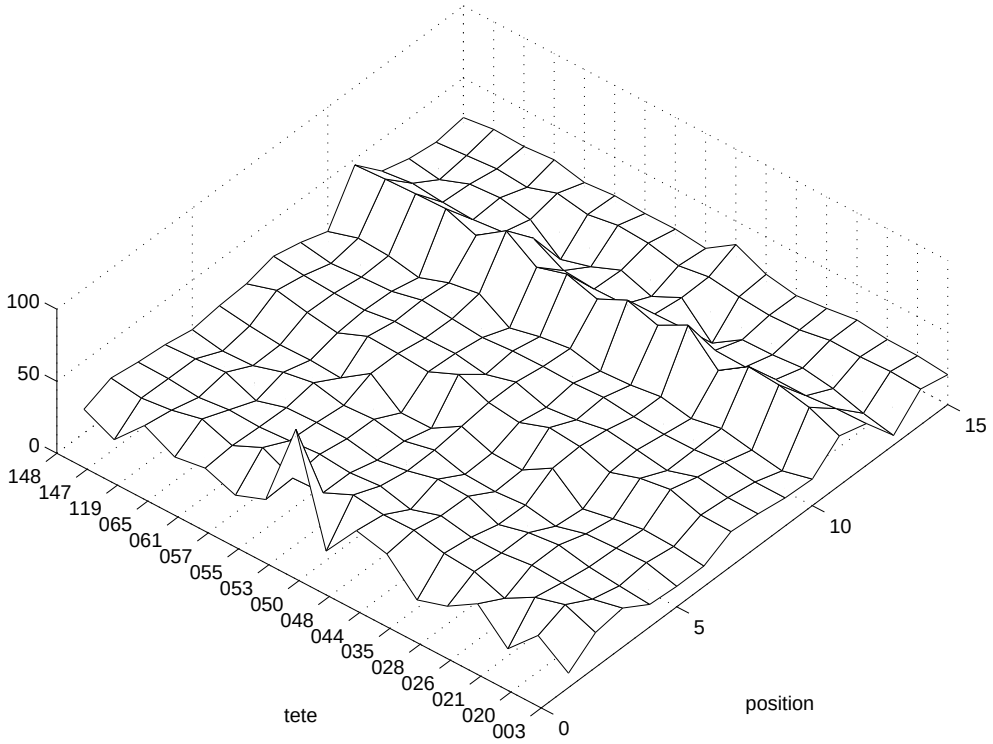


FIG. 4.8 – Robustesse de la localisation en azimut : valeur de l'écart-type de localisation en fonction de la tête et de la position.

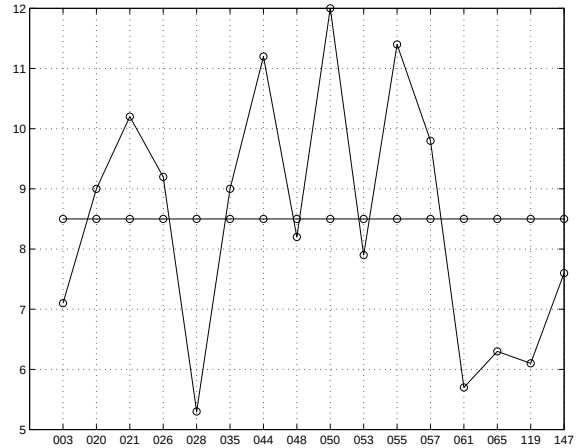


FIG. 4.9 – Influence de la tête sur la robustesse de la localisation en azimut (test de Friedman) : rang théorique si la tête n'était pas un facteur déterminant (rouge), et rang moyen observé (bleu).

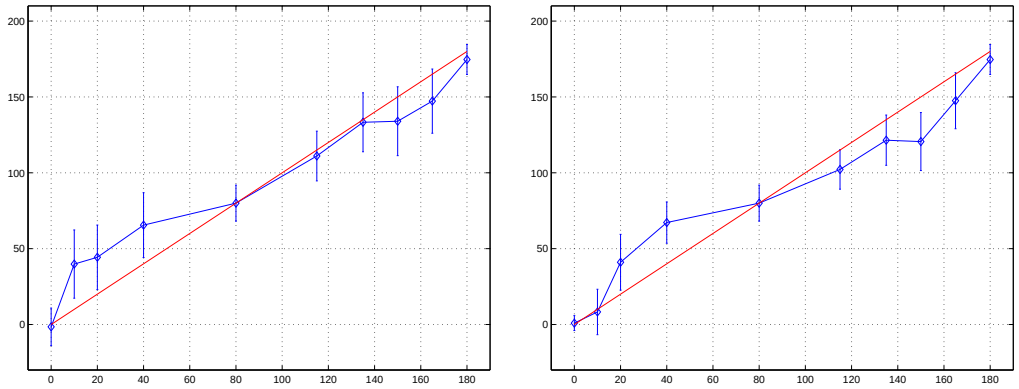


FIG. 4.10 – Biais de localisation et écart-type (en °) dans le plan horizontal. tête 028 à gauche et tête hybride maximisant la robustesse de la localisation (minimisant l'écart-type) à droite.

4.2.6.3 Biais et robustesse de localisation

Une fois les éléments atypiques écartés, nous nous intéressons à la valeur moyenne perçue pour chaque azimut. Nous pourrions nous attendre à ce que cette valeur moyenne soit l'azimut source. Il apparait toutefois que cette valeur s'en écarte pour constituer un biais perceptif systématique, amplifiant un phénomène déjà connu de la perception des sources réelles ([Bla97], p. 49). Nous présentons les résultats obtenus pour la tête la plus robuste (tête 028) ainsi que pour une tête hybride maximisant la robustesse de localisation pour chaque position. On a observé que biais et écart-types n'étaient pas corrélés ($r_{max} \approx 0.6$ sur l'ensemble des têtes). Par conséquent minimiser l'écart-type ne conduit pas à minimiser le biais.
Les résultats illustrés pour le plan horizontal en Figure **??** pour la tête 028, présentent une certaine symétrie par rapport à 90°, azimut vers lequel les perceptions semblent "attirées" : on constate une surlatéralisation systématique, mis à part pour les positions du plan médian. Autrement dit, les positions identifiées au plus près de la position source sont d'une part les positions à différences interaurales nulles (positions 1 et 6), et d'autre part les positions à différences interaurales maximum (positions 5 et 10). Chateau observe des résultats semblables pour la synthèse binaurale individuelle, ainsi que Makous et Middlebrooks pour l'écoute naturelle. En outre, Oldfield et Parker justifient cette attirance vers 90° comme un corrolaire d'une forte erreur en élévation (cf section 4.2.8).

L'azimut perçu présente néanmoins un fort écart-type, et nous avons pris soin de vérifier la significativité des écarts observés d'une position à l'autre, à l'aide d'un test de Mann-Whitney. Le test permet de s'assurer que les écarts observés en Figure **??** sont significatifs mis à part pour les positions 2 et 3, situées resp. à 10° et 20° et perçues autour de 42°, et les positions 8 et 9, situées resp. à 135° et 150° et perçues toutes deux autour de 133°. Si l'on étend l'analyse aux positions en élévations, on peut observer que :

- les positions 1 et 14, situées au même azimut, sont perçues avec le même biais, à 3° environ.
- les positions 5 et 13, écartées de 10° en azimut, sont perçues toutes deux à 76°.
- les positions 3 et 15, écartées de 17° en azimut, sont perçues toutes deux à 43°.
- la position 16, située à 143°, ne se distingue pas significativement des positions 7 (165°) et 8 (150°).

La robustesse de la localisation autour de ce biais est représentée pour le plan horizontal en Figure 4.11. Nous observons que les positions favorisant la plus forte robustesse de localisation sont les positions du plan médian, également affectées du plus faible biais (mais aussi du plus fort taux de confusion). Les localisations les plus instables sont obtenues pour les positions justes voisines (positions 2 et 8). On peut ainsi supposer que le système auditif détecte aisément des sons identiques aux deux oreilles. Cette caractéristique des positions du plan médian est indépendante de la tête écoutée, et est confortée par les performances du système auditif pour séparer deux sources sonores voisines, maximales pour les positions frontales ([Bla97] p. 41). Les positions latérales présentes une plus faible robustesse que les précédentes, mais néanmoins supérieure que celle des positions à "faibles différences interaurales".

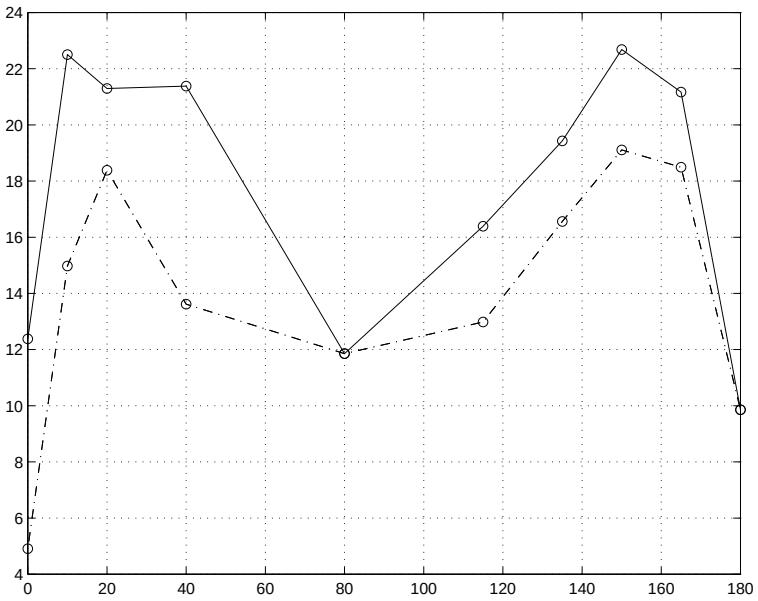


FIG. 4.11 – Robustesse de la localisation dans le plan horizontal : variation de l'écart-type de l'erreur en azimut en fonction de la position. tête 028 (trait continu) et tête hybride maximisant la robustesse de la localisation (trait pointillé).

Pour la tête 028, l'erreur moyenne de localisation dans le plan horizontal est de 17.3°, score supérieur à celui de Gardner (14.3°). Si l'on étend l'analyse aux positions en élévation (cf Tableau 4.3), on constate que même pour la position élevée à différences interaurales nulles (position 14), la robustesse de localisation est la meilleure, du même ordre que pour les autres positions du plan médian. En outre, les positions symétriques 15 et 16 sont localisées avec une reproductibilité comparable, ce qui ne laisse entrevoir aucun privilège ni pour l'avant ni pour l'arrière. Enfin, les performances à la position 13 sont les plus médiocres.

4.2.7 Résultat 4 : Phénomène de confusion en élévation

Tout comme Wenzel et al., nous constatons des confusions "haut-bas" : certains sons en élévation sont perçus à une position symétrique par rapport au plan horizontal. Nous les détectons suivant le même principe que les confusions en azimut, en constatant un écart par rapport à la position de la source diminuée de plus de 10° lorsque l'on considère l'élévation miroir de l'élévation perçue. Les confusions haut-bas ne sont pas étudiées pour les positions du plan horizontal.

4.2.7.1 Origine du phénomène

Comme les confusions avant-arrière, ce phénomène est lié à l'existence de lieux d'iso-ITD ou iso-ILD, qualifiés de "cône de confusion", conduisant à des ambiguités de position que seuls les indices spectraux permettent de résoudre.
Wenzel et al. observent des confusions haut-bas avec une fréquence moyenne de 6% pour la localisation de sources sonores réelles. Ce taux augmente jusqu'à 18% dans le cas de la localisation avec une synthèse binaurale. Aucune distorsion n'est observée entre les hémisphères supérieurs et inférieurs. Dans notre cas, puisque toutes les sources ont une élévation positive, nous n'observerons que des confusions haut→bas.

4.2.7.2 Dépendance vis à vis de la tête

Un test de Friedman permet d'observer que ni la fréquence des occurence ni leur répartition spatiale ne dépend de la tête. L'analyse sera donc effectuée globalement sur toutes les têtes.

n°	position	azimut perçu		robustesse		
		tête 028	Min	tête 028	Min	n°
1	(0°,0°)	-2°	1°	12°	5°	021
2	(10°,0°)	40°	8°	23°	15°	026
3	(20°,0°)	44°	41°	21°	18°	048
4	(40°,0°)	66°	67°	21°	14°	061
5	(80°,0°)	80°	70°	12°	12°	028
10	(115°,0°)	111°	102°	16°	13°	053
9	(130°,0°)	133°	122°	19°	17°	119
8	(150°,0°)	134°	121°	23°	19°	048
7	(165°,0°)	147°	148°	21°	18°	061
6	(180°,0°)	175°	175°	10°	10°	028
13	(90°,35°)	73°	71°	24°	23°	050
14	(0°,34°)	7°	0°	19°	5°	044
15	(37°,33°)	41°	41°	22°	19°	147
16	(143°,33°)	142°	144°	23°	20°	050

TAB. 4.3 – Biais perceptif et robustesse de localisation, pour la tête réalisant le meilleur consensus (tête 028) et pour une tête hybride maximisant la robustesse. Pour chaque position, le biais de la tête hybride correspond au biais de la tête sélectionnée sur le critère de robustesse.

4.2.7.3 Distribution spatiale des confusions en azimut

Comme on le constate en Figure 4.12, les confusions haut-bas affectent toutes les positions en élévation. Toutefois, elles sont uniformément réparties sur chacune d'entre elles, mise à part sur la positions 16 qui représente un tiers des occurences.
Pour trois des positions (1, 11, 15), les confusions haut-bas coïncident de façon prépondérante avec des confusions avant-arrière. Ce phénomène représente plus de 62% des cas, score à rapprocher des 55% de confusions "combinées" chez Wenzel et al. Bien qu'ils n'aient pas étudié les confusions haut-bas, Makous et Middlebrooks constatent que les erreurs en élévation sont supérieures dans le cas de sons confus en azimut. Cela semble souligner la complexité du phénomène de confusion et peut remettre en question les approches simplificatrices de redressement pratiquées dans la littérature, que par souci de comparaison, nous avons adoptées. En ce sens, les critères d'analyse définis par Moller et al., consistant à comptabiliser 4 types d'erreur dont "within cone error" respectent l'intégrité du phénomène de confusion.

4.2.7.4 Fréquence des confusions

Comme chez Wenzel et al., ce phénomène est moins fréquent que les confusions avant-arrière. Il atteint un score moyen de 10% des stimuli, à rapprocher des 18% mentionnés dans [WAKW93].

4.2.8 Résultat 5 : Localisation en élévation

Pour l'analyse de la localisation en élévation, nous observons les 16 positions initiales. Nous nous appuyons sur les deux estimateurs, biais et robustesse, utilisés pour l'analyse de la localisation en azimut.

4.2.8.1 Intégration des confusions

Le problème des confusions se pose ici à double titre. Pouvons-nous conserver les confusions haut-bas après redressement ? Devons-nous intégrer à l'analyse les sons ayant donné lieu à une confusion avant-arrière ?
Nous avons choisi, comme Wenzel, de redresser les confusions en élévation. Le faible nombre de valeurs pour les estimations ne nous a pas permis de vérifier l'effet de l'intégration de ces confusions.

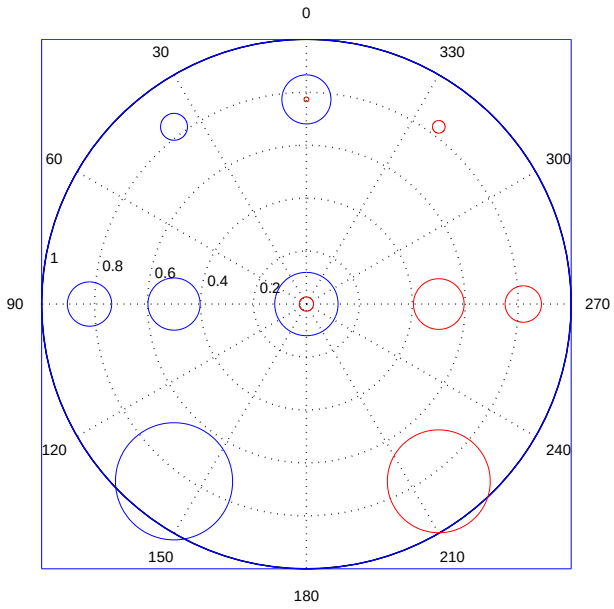


FIG. 4.12 – Distribution des confusions haut-bas pour les 16 positions testées : confusions avant-arrière incluses (en bleu), exclues (en rouge). Les calculs sont effectués en sommant les occurences observées pour les 22 sujets et les 17 têtes.

En revanche, comme il est illustré en Figure 4.13, les confusions en azimut contribuent à augmenter sensiblement l'écart-type des réponses, essentiellement pour les positions frontales. On peut penser qu'il s'agit de l'effet des confusions azimut/élévation combinées observées au paragraphe précédent. En outre, Makous et Middlebrooks ont constaté que l'inclusion des confusions en azimut conduisait à une surestimation l'élévation perçue. Nous choisissons d'écarter les sons confus en azimut pour l'analyse de l'erreur en élévation.

4.2.8.2 Dépendance vis à vis de la tête

Le test de Friedman montre que la tête n'est pas un facteur influençant la robustesse de localisation. En revanche, le critère de biais, significatif quant à lui, nous permet de classer les têtes, et, comme on le voit en Figure 4.14, la tête 035 est celle qui montre le biais moyen le plus faible.

4.2.8.3 Biais et robustesse de localisation

Pour toutes les positions, la robustesse de localisation en élévation apparait plus faible que pour la localisation en azimut. Les forts écart-types observés sont cohérents avec les résultats de la littérature. On peut rapprocher nos résultats des $\pm 10°$ estimés par Makous et Middlebrooks en localisation champ libre et aux $\pm 40°$ observés par Gardner pour une synthèse avec tête artificielle.

Dans le plan horizontal, cette robustesse apparait plus faible à l'arrière qu'à l'avant, puisque l'écart-type moyen passe de $\pm 25°$ à l'avant à $\pm 33°$ à l'arrière (cf Figure 4.15). La position la plus robuste est la position latérale avant. La robustesse ne diminue pas avec l'élévation. La meilleure reproductibilité est même obtenue pour la position 15 ($\pm 21°$). Ces meilleures performances obtenues pour les positions frontales coincident avec une meilleure exposition aux indices fournis par l'oreille externe, interprétation apportée par Oldfield et Parker, que rappelle Chateau : ***"les stimuli provenant de l'arrière transmettent moins d'énergie et moins d'indices spectraux que ceux provenant de l'avant et sont ainsi moins bien localisés"***.

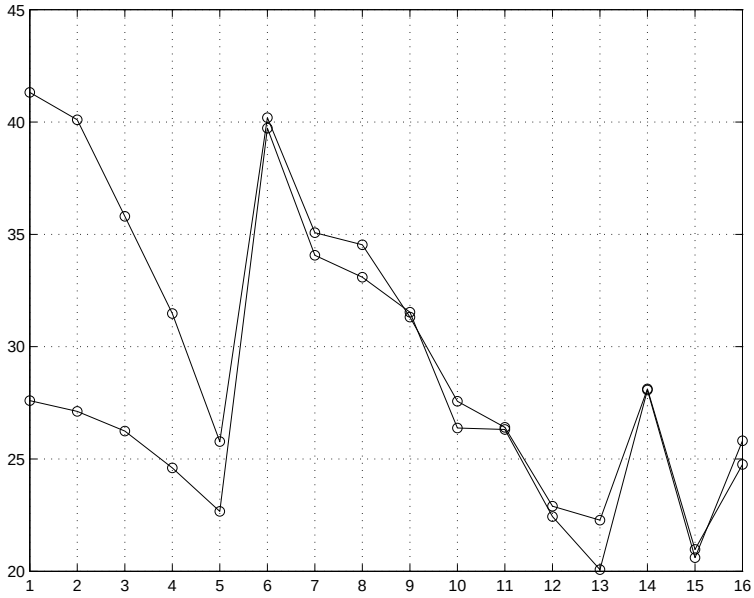


FIG. 4.13 – Robustesse de la localisation en élévation pour chacune des 16 positions : confusions en azimut incluses (bleu) et exclues (rouge).

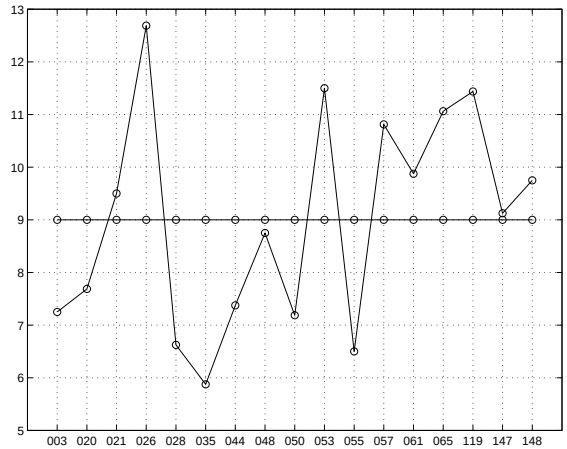


FIG. 4.14 – Influence de la tête sur le biais de localisation en élévation (test de Friedman) : rang théorique si la tête n'était pas un facteur déterminant, et rang moyen observé (bleu).

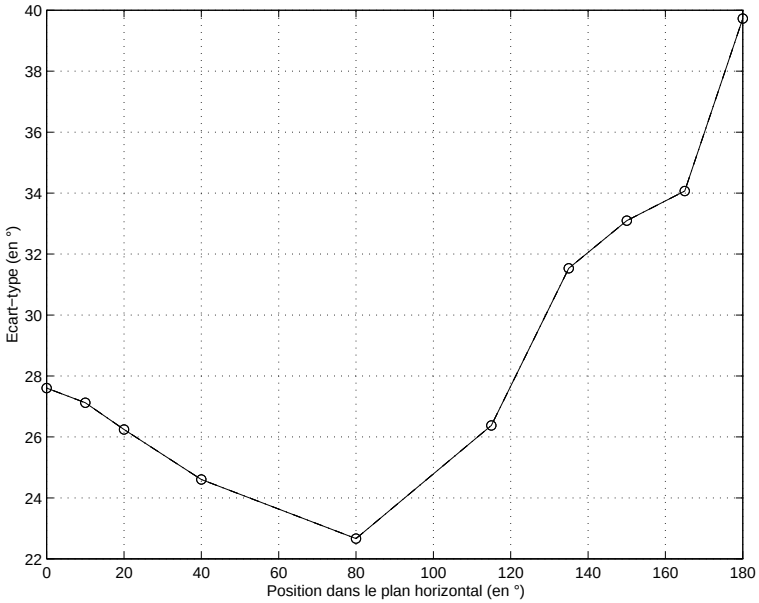


FIG. 4.15 – Robustesse de la localisation en élévation pour les positions du plan horizontal.

Dans le plan vertical médian, la robustesse est significativement inférieure. C'est une zone où le halo de localisation est très large en écoute naturelle.

Dans le plan horizontal, les positions avant ont tendance à être perçues avec une élévation positive, alors que les positions arrière sont plutôt perçues au dessous du plan horizontal. Un test de Wilcoxon nous permet de vérifier la significativité de l'écart observé entre les positions 1 et 6. On notera que ces résultats s'écartent de ceux de Makous et Middlebrooks qui observent un biais nul devant et un biais positif à l'arrière pour une synthèse binaurale individuelle.
C'est pour les azimuts frontaux intermédiaires (de 10° à 50° environ) que l'on obtient les plus forts biais en élévation. On constate que c'est également la région pour laquelle l'erreur en azimut était maximum. Le modèle du cône de confusion donne une première justification de cette interprétation (cf [OP84]). Si le système auditif a identifié le cône de confusion auquel appartient le stimulus, le choix de la position sur ce cône détermine conjointement azimut et élévation perçue. Notamment, une surestimation (en valeur absolue) de l'élévation conduit à une sur-latéralisation en azimut. Cette attraction vers 90° observée pour l'erreur en azimut peut ainsi être directement reliée à la surestimation de l'élévation pour les positions du plan horizontal. Ce phénomène est spécifique aux positions hors du plan médian, pour lesquelles l'élévation n'est pas déterminée par les seuls indices spectraux, mais également par les indices interauraux.

Comme on l'observe sur les Figures 4.17 et 4.18, notre synthèse binaurale a du mal à suggérer des positions fortement élevées. Dans le plan médian, l'élévation maximale se situe autour de 40° pour la tête 035, et ne monte que jusqu'à 60° pour la tête minimisant le biais. En écoute naturelle, un évènement sonore au zénith, est perçu à une élévation inférieure proche de 74° (cf [Bla97]p. 44). Ce biais naturel est renforcé dans notre cas du fait de l'inadéquation des indices spectraux aux sujets. Dans le plan frontal, le biais observé sur la tête 6 ne présente aucune différence significative entre les élévations. Tous les sons ont été perçus à une élévation voisine de 40°. Pour la tête hybride, en revanche, cette différence est significative pour les positions 12 et 11, qui sont toutes deux perçues à une élévation supérieure, voisine de 60°.

4.3 Performances subjectives de l'implantation multicanale

Nous souhaitons comparer à l'aide d'un test perceptif la qualité de localisation offerte par différentes techniques d'implantation muticanales de la synthèse binaurale. L'implantation bicanale, dont nous venons de caractériser les performances subjectives, est prise comme référence. Les techniques de décomposition

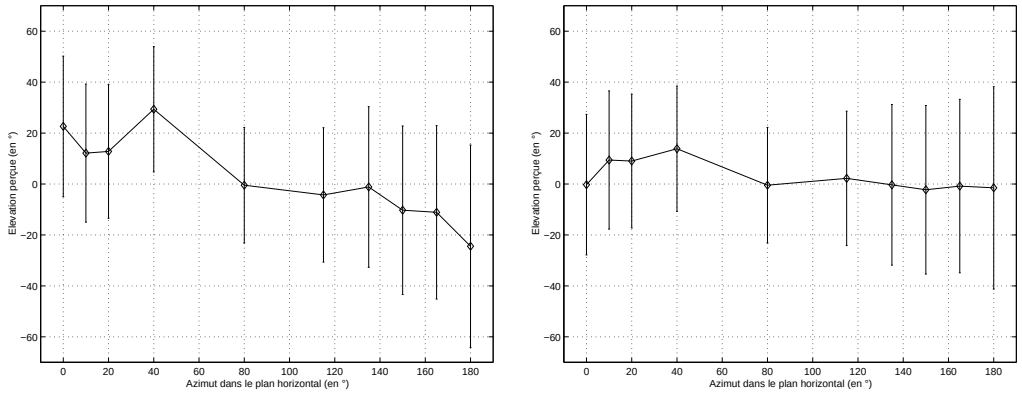


FIG. 4.16 – Biais de localisation en élévation et écart-type (en °) dans le plan horizontal. tête 035 à gauche et tête hybride maximisant la robustesse de la localisation (minimisant le biais) à droite.

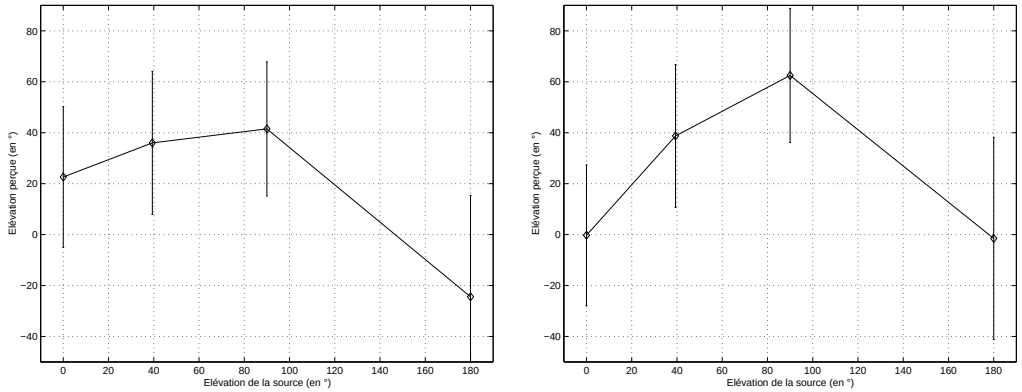


FIG. 4.17 – Biais de localisation en élévation et écart-type (en °) dans le plan plan médian.

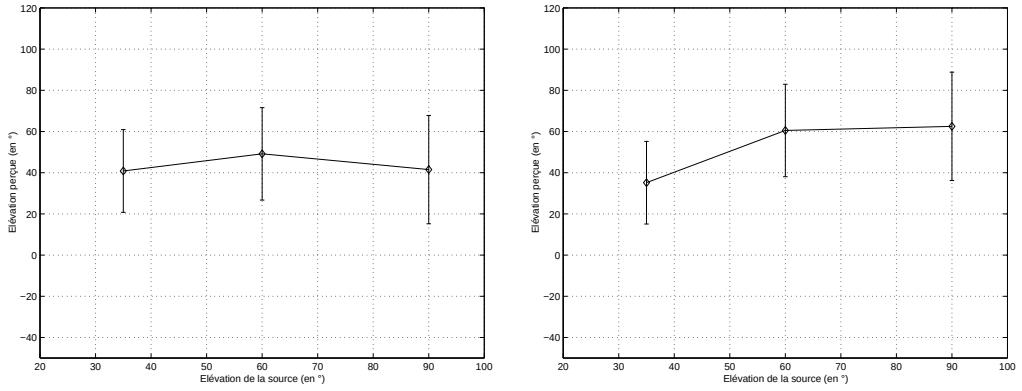


FIG. 4.18 – Biais de localisation en élévation et écart-type (en °) dans le plan plan frontal.

n°	position	élévation perçue			robustesse
		tête 035	Min	tête	
1	(0°,0°)	23°	0°	028	28°
2	(10°,0°)	12°	9°	050	27°
3	(20°,0°)	13°	9°	020	26°
4	(40°,0°)	29°	14°	003	25°
5	(80°,0°)	0°	0°	035	23°
10	(115°,0°)	-4°	2°	057	26°
9	(130°,0°)	-1°	0°	147	32°
8	(150°,0°)	-10°	-2°	048	33°
7	(165°,0°)	-11°	-1°	048	34°
6	(180°,0°)	-24°	-1.5°	065	40°
11	(-,90°)	41.5°	62.5°	119	26°
12	(90°, 60°)	49°	60.5°	119	22°
13	(90°,35°)	41°	35°	053	20°
14	(0°, 34°)	36°	39°	048	28°
15	(37°,33°)	54°	36°	021	21°
16	(143°,33°)	39°	34.5°	048	26°

TAB. 4.4 – Biais perceptif et robustesse de localisation en élévation, pour la tête réalisant le meilleur consensus (tête 6) et pour une tête hybride maximisant la robustesse. Pour chaque position, le biais de la tête hybride correspond au biais de la tête sélectionnée sur le critère de robustesse.

TAB. 4.5 – Caractéristiques des méthodes de décomposition linéaire des HRTF comparées perceptivement.

Méthode	ITD explicite	Encodeur universel	Nombre de canaux actifs de l'encodeur (plan hor.)	Nombre de filtres du décodeur (plan hor.)
Binaural B ordre 1	oui	oui	2×3	2×3
Binaural B ordre 3	oui	oui	2×7	2×7
ACI spatiale	oui	oui	2×2	2×7
ACI temporelle	oui	non	2×4	2×3
HP virtuels	non	oui	1×2	2×7
HP virtuels à phase minimale	oui	oui	2×2	2×7

linéaire des HRTF sont choisies parmi celles abordées au chapitre 3, ou en dérivent directement, et réalisent différents compromis sur les critères de performance objectifs que résume le Tableau 4.5.
Pour ce test, le choix de la tête n'est pas un facteur, puisque les filtres utilisés sont ceux d'une seule tête, ayant montré de bonnes performances pour l'implantation bicanale, la tête 028.

4.3.1 Mise en place du test

Les conditions expérimentales sont très proches de celles du test de localisation du chapitre 2. Nous ne précisons ici que les éléments s'en écartant.

4.3.1.1 Interface de réponse

Afin de mettre à profit les critiques formulées par les sujets lors du premier test, l'interface de réponse à été modifiée sur trois aspects :

1. le jugement sur la distance perçue est supprimé, du fait de la difficulté apparente des sujets à l'utiliser.

2. sur un conseil d'un des sujets, la valeur par défaut des curseurs de réponse est aléatoire, alors que pour le test du chapitre 2, ces paramètres étaient remis à la position (0°,0°) à chaque nouveau stimulus.
3. le jugement d'élévation est rapporté sur un cercle, vue symbolique d'un profil de tête, et non plus sur une règle.
4. le train de bruit blanc est limité à trois périodes, durée au delà de laquelle le son s'arrête. Le sujet peut rejouer la série des trois bruits blancs autant de fois qu'il le désire. Ce choix permet de diminuer la fatigue auditive des sujets et d'accéler la réponse. En effet, on a pu observé que notre jugement de localisation pouvait être formalisé très rapidement.
5. Le niveau sonore moyen a été diminué de 5dB environ.

4.3.1.2 Stimuli

Afin de diminuer la durée du test, nous n'avons testé qu'un sous-ensemble des 16 positions utilisées au chapitre 2. Nous avons conservé toutes les positions du plan horizontal, soit un total de 10 positions (pour la numérotion des positions, on peut se reporter au tableau 4.6 ou a la Figure 4.21). Les méthodes de décomposition linéaire des HRTF que nous comparons sont les suivantes :

- Binaural B d'ordre 1.
- Binaural B d'ordre 3 pour le plan horizontal, avec la fonction spatiale d'élévation de l'ordre 1.
- ACI spatiale avec encodeur universel et idéalement compact. L'ACP initiale a été pratiquée sur des données centrées.
- ACI temporelle avec décodeur universel. Cette technique est calquée sur la description donnée par Dudouet dans ([DM98]), et consiste à appliquer l'ACI non aux fonctions spatiales issues de l'ACP mais aux réponses impulsionnelles des filtres reconstruction. Avec cette technique, les HRTF peuvent être reconstruites à partir de réponses impulsionnelles à support limité et disjoint. Comme dans l'étude [DM98], nous choisissons une représentation sur 3 canaux par oreille, aucun cas particulier n'est fait pour le plan horizontal. En revanche, nous étendons cette approche à un décodeur universel en appliquant l'approche aux données concaténées, suivant la dimension spatiale cette fois. Un quatrième canal est ajouté pour la fonction spatiale moyenne.
- haut-parleurs virtuels, situés aux positions 40°, 0°, -40°, -80°, -150°, 150° et 80° dans le plan horizontal, ainsi qu'au zénith. Pour cette méthode, les positions coïncidant avec celles des haut-parleurs virtuels ne sont pas testées, soit un total de 6 positions testées : positions 2, 3, 6, 7, 9 et 10.

En outre, on teste également l'implantation bicanale de référence sur les 8 positions retenues pour la méthode des haut-parleurs virtuels. Les erreurs objectives de reconstruction sont présentées en Figure 4.20.
Trois têtes ont été testées, parmi lesquelles la tête 028, qui a obtenu les meilleurs scores au test du chapitre 2. Cette tête est la seule commune à tous les sujets et nous concentrons l'analyse sur les stimuli qui y sont attachés, soit $5 \times 10 + 2 \times 6 = 62$ stimuli par sujet.

4.3.1.3 Sujets

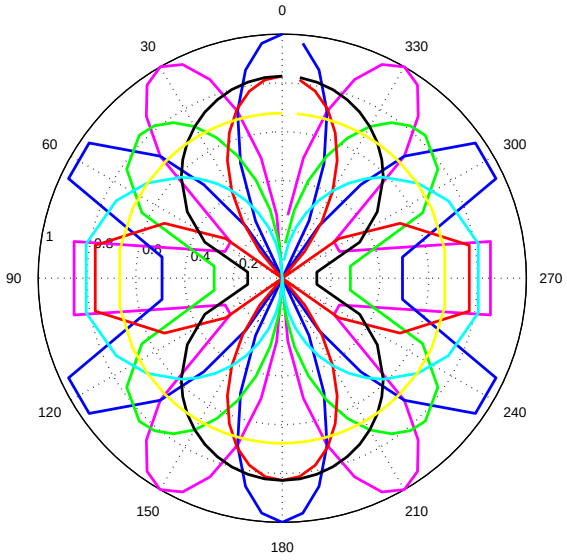
Dix-sept des 22 sujets de l'expérience du chapitre 1 ont passé ce test.

4.3.2 Résultat 1 : Sons non localisés

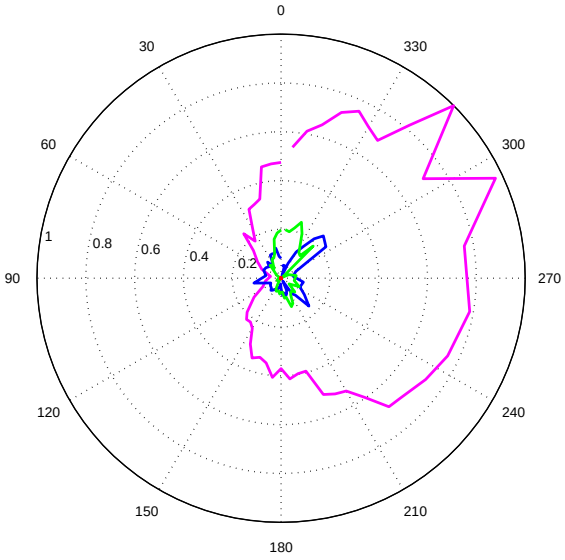
Le phénomène de non-localisation n'est pas un paramètre de discrimination des méthodes, ni pour la répartition de ses occurences, ni pour leur fréquence. Les résultats sont présentés sur la Figure 4.21 et dans le tableau 4.6. Comme pour le binaural bicanal étudié en chapitre 2, les occurences se concentrent autour du plan médian. Les fréquences observées se démarquent en revanche du premier test. En effet, les occurences de non localisation apparaissent ici de façon marginales. Deux éléments peuvent expliquer ce résultat :

1. l'entrainement acquis par les sujets,

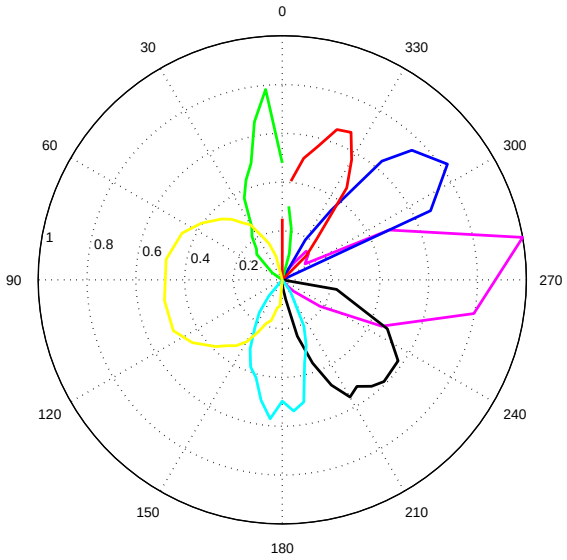
FIG. 4.19 – Fonctions spatiales considérées pour le test. Sauf mention contraires, elles sont universelles.



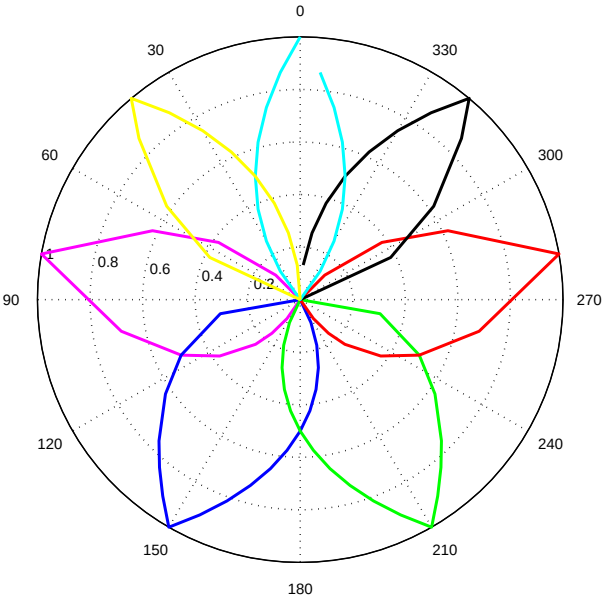
Binaural B d'ordre 1 à 3



ACI temporelle tête "028"



ACI spatiale



HP virtuels

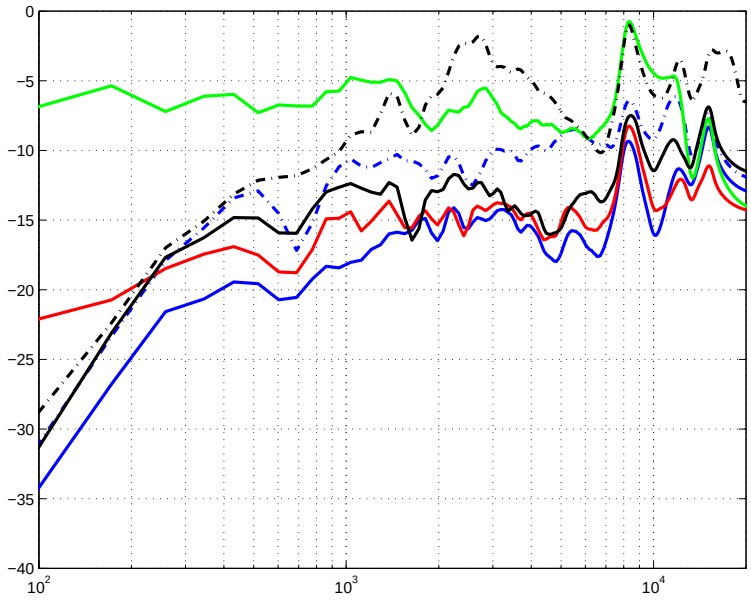


FIG. 4.20 – Erreur de reconstruction dans le plan horizontal : Binaural B d'ordre 1 (trait interrompu bleu), Binaural B d'ordre 3 (bleu), ACI spatiale sur 7 canaux (rouge), ACI temporelle sur 3 canaux (vert), haut-parleurs virtuels sur 7 canaux (trait interrompu noir), haut-parleurs virtuels à phase minimale sur 7 canaux (noir).

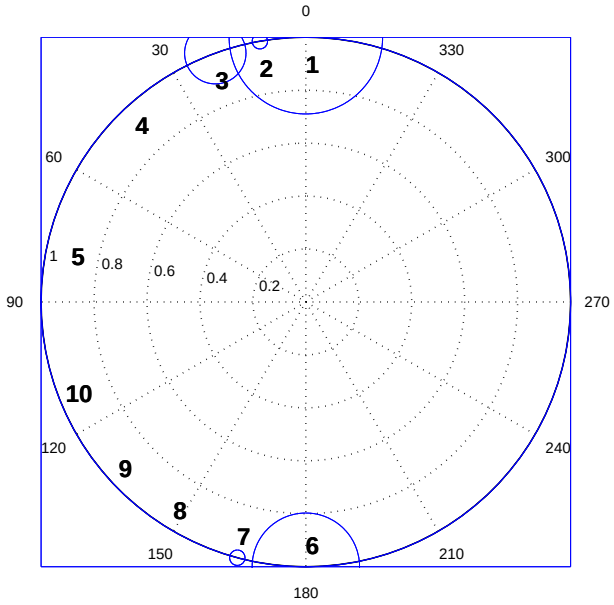


FIG. 4.21 – Répartition des occurences de non localisation dans le plan horizontal.

n°	position	fréquence			
		moyenne multicanale	HP virtuels	bino2	*bino2 (section 4.2)*
1	0°	12%	-	-	*23%*
2	10°	1%	12%	6%	*9%*
3	20°	5%	0%	0%	*5%*
4	40°	0%	-	-	*0%*
5	80°	0%	-	-	*0%*
10	115°	0%	0%	0%	*0%*
9	135°	0%	0%	0%	*0%*
8	150°	0%	-	-	*0%*
7	165°	1%	0%	0%	*0%*
6	180°	8%	35%	12%	*18%*

TAB. 4.6 – Fréquence d'occurence des "non-localisation".

n°	position	répartition	fréquence				
			HP phase min	ACI temporelle	HP virtuels	bino2	*bino2 (chap2)*
1	0°	20%	35%	59%	-	-	*32%*
2	10°	18%	47%	50%	53%	29%	*27%*
3	20°	16%	35%	41%	53%	24%	*27%*
4	40°	10%	12%	53%	-	-	*9%*
5	80°	12%	24%	35%	-	-	*23%*
10	115°	10%	24%	6%	12%	25%	*5%*
9	135°	6%	18%	0%	6%	12%	*0%*
8	150°	3%	6%	0%	-	-	*23%*
7	165°	3%	12%	6%	7%	12%	*0%*
6	180°	4%	18%	12%	0%	12%	*9%*

TAB. 4.7 – Répartition et fréquence des confusions "avant-arrière".

2. le paramètre de distance, qui, réglé à 0, était utilisé dans la première expérience pour faire part d'un son "non-localisé". Ce paramètre a été retiré pour la seconde expérience car jugé redondant avec la bascule "son non-localisable" de l'interface. La distance était peut-être un élément plus intuitif à manipuler.

4.3.3 Résultat 2 : Phénomène de confusion en azimut

La répartition des confusions ne diffère pas d'une méthode à l'autre : comme pour le test du chapitre 2, si toutes les positions sont affectées, les confusions avant→arrière représentent la grande majorité des occurences.

La fréquence des confusions avant→arrière révèle quant à elle une différence significative entre les méthodes (test de Friedman à 5%). Celle qui présente le taux le plus faible est la méthode des haut-parleurs virtuels à phase minimale, le taux le plus important étant obtenu par l'ACI temporelle (cf Tableau 4.7). Si l'on concentre la comparaison sur les deux positions avant testées pour les haut-parleurs virtuels et l'implantation bicanale, on observe en outre que les résultats de l'implantation bicanale sont meilleurs que ceux des autres méthodes, et que les plus forts taux de confusion sont obtenus par l'ACI temporelle et par la méthode des haut-parleurs virtuels. Ces différences significatives ne se retrouvent pas pour les positions arrières.

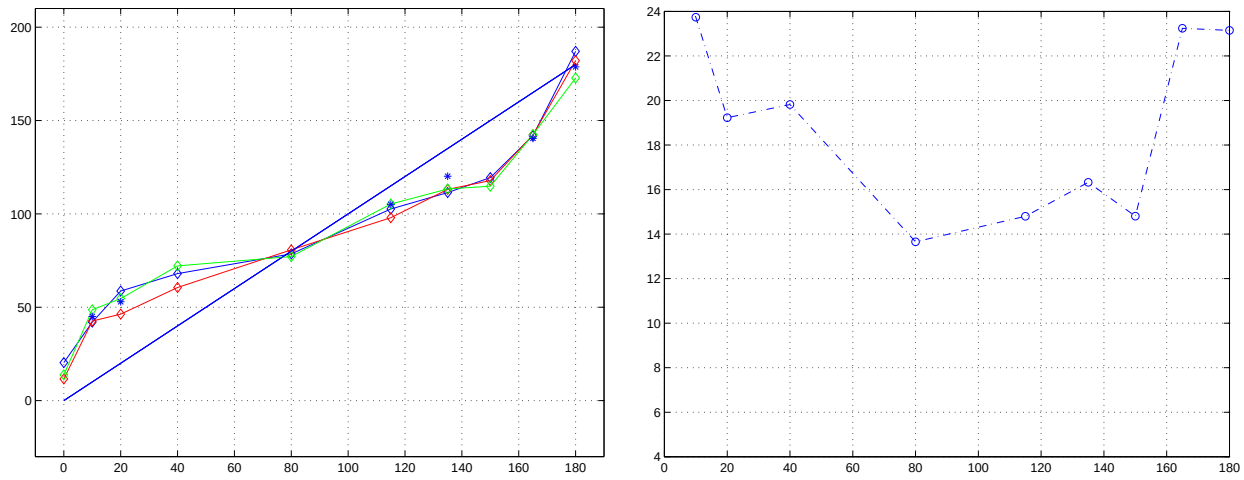


FIG. 4.22 – Biais et robustesse de localisation en azimut, dans le plan horizontal : Binaural B d'ordre 3 (bleu), Binaural B d'ordre 1 (rouge), ACI spatiale (vert), Binaural bicanal (étoiles bleues). Robustesse de localisation moyenne pour l'ensemble des méthodes.

4.4 Résultat 3 : Localisation en azimut

La localisation en azimut permet de mettre en évidence la différence entre les deux décomposition Binaural B. En effet, le biais est significativement plus important dans le cas d'une décomposition au premier ordre (test de Wilcoxon : z=1.9552), l'écart avec l'ordre deux pouvant aller jusqu'à 12° (cf Figure 4.22). L'ACI spatiale présente également un biais signitificativement plus important que celui du Binaural B d'ordre 3, pour les positions avant. Pour cette dernière méthode, le biais présente un écart oscillant entre 10° et 30° par rapport à l'azimut cible. La Figure 4.22 peut être rapprochée de celle obtenue au chapitre 2 : l'azimut perçu est "attiré" par la position latérale à 90°, le biais étant minimum pour cette position ainsi que pour les positions du plan médian.

La robustesse de localisation est comparable pour toutes les méthodes, et est représentée en Figure 4.22. De valeur moyenne 20°, elle est comparable à celle obtenue pour l'implantation bicanale au chapitre 2 mis à part pour les positions 0° et 180°. En effet, dans ce dernier cas, un grand nombre des réponses pour ces positions a été catégorisé comme son non localisé, tandis que pour ce test, les sujets ont fait plus d'efforts pour les localiser, au prix d'une indécision supérieure.

4.4.1 Résultat 4 : Localisation en élévation

Nous ne testons que des positions du plan horizontal, pour lesquelles la notion de confusion en élévation n'a pas de sens.

L'élévation perçue pour les positions du plan horizontal ne permettent pas de discriminer les méthodes. Comme on l'observe sur la Figure 4.23, les positions avant sont perçues en élévation positive, alors que les positions arrière sont plutôt perçues avec une élévation négative, d'amplitude plus réduite. Comme mentionné en section 4.2, ce phénomène explique en partie l'"attraction vers 90°" observée pour l'azimut.

4.5 Conclusion

Notre étude perceptive a principalement porté sur la qualité de localisation d'un système de synthèse binaurale bicanal, statique, et non-individuel. Les HRTF étaient modélisées sous forme d'un retard pur et d'un filtre à phase minimal d'ordre 20. La localisation dans le plan médian s'est caractérisée par un

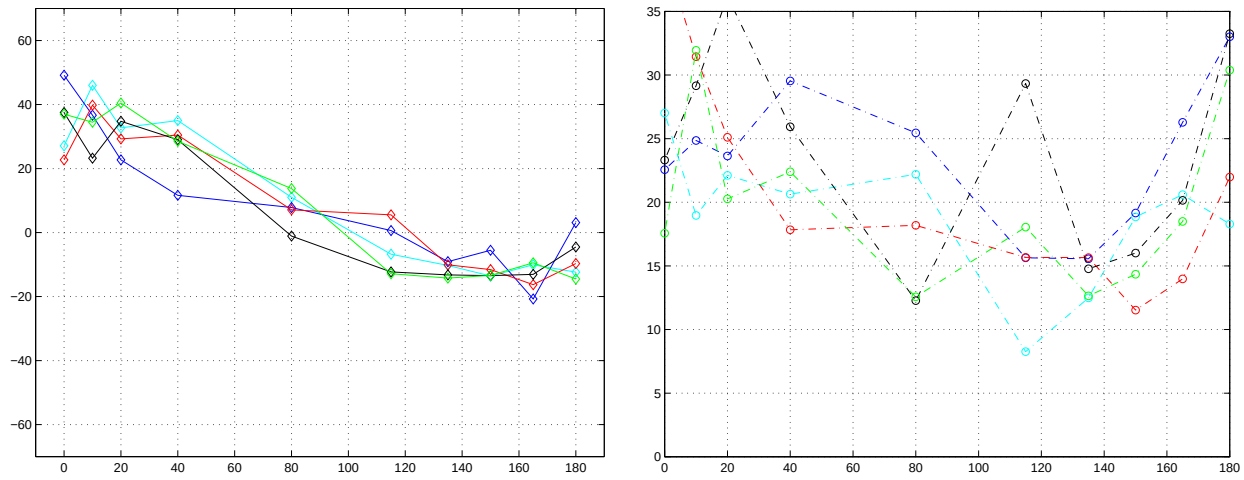


FIG. 4.23 – Biais et robustesse de localisation en élévation, plan horizontal.

fort taux de sons perçus au centre de la tête (33%), ainsi que par un fort taux de confusion avant-arrière (23%). L'azimut est perçu avec un biais, minimum pour les positions du plan médian et extra-latérales, mais qui atteint 20° pour les positions intermédiaires, surtout devant. Ce phénomène s'accompagne d'une forte erreur en élévation, conformément à ce que prédit l'analyse du cône de confusion. Nous observons également des confusions haut-bas, qui représentent 10% des stimuli (hors du plan horizontal). Enfin, l'élévation est perçue avec un fort écart-type (±28° en moyenne), surtout pour les positions arrières. Nous constatons la difficulté pour les têtes étudiées de suggérer une position en forte élévation (zénith), puisque l'élévation perçue ne dépasse pas 60°. Souvent, le stimulus au zénith aura plutôt été perçu au centre de la tête.
Un second test perceptif a permis de comparer la qualité de localisation pour le plan horizontal obtenues avec plusieurs formats d'encodages multicanal. Il a permis de montrer que :

1. le biais en azimut provoqué par un Binaural B au premier ordre est supérieur à celui obtenue pour un ordre 3,
2. l'ACI "spatiale", méthode que nous avons proposé, provoque un biais en azimut supérieur à celui obtenu pour un Binaural B d'ordre 3.

Ce test n'a pas permis de mettre en évidence d'autres différences significatives par rapport à l'implantation bicanale. Ces premiers résultats sont poursuivis au chapitre 5 pour évaluer l'influence du choix de la base de données de HRTF pour la qualité de localisation.

Chapitre 5

Adaptation individuelle de la synthèse binaurale

5.1 Introduction

Les HRTF, filtres à la base de la synthèse binaurale, varient en fonction de l'incidence, mais également en fonction de la tête sur laquelle ils ont été mesurés. A l'origine de ce phénomène : la morphologie de cette tête, qui, faisant obstacle au son incident, modèle les caractéristiques temporelles et fréquentielles des HRTF. Une écoute binaurale non individuelle, i. e. proposant à l'auditeur des HRTF différentes des siennes, entraine une augmentation des artefacts de localisation, notamment les confusions avant-arrière et les occurences de sons perçus à l'intérieur de la tête ([Bla97], [WAKW93]). Domaine morphologique, domaine du signal et domaine perceptif, constituent donc trois espaces d'observation des différences inter-individuelles. Différentes approches peuvent être envisagées pour en atténuer les effets :

1. l'approche "sans ajustement" consiste à proposer un jeu de filtres identique pour tous les auditeurs, choisi au mieux pour réaliser le meilleur compromis perceptif. Elle suppose la définition de HRTF "universelles". Certaines études résolvent le problème dans le domaine morphologique, par la conception d'une tête artificielle aux caractéristiques morphologiques moyennes, ou "mannequin anthropomorphique" ([BS75], [BPA+91]). D'autres approches se concentrent sur le domaine du signal et s'appliquent à moyenner les HRTF ou un modèle structurel de ces dernières ([MM77], [Pla79]). Ces modèles structurels paramétrisent ou bien les HRTF elles-mêmes (positions des pics et des vallées, facteur de qualité) ou bien la transformation permettant de passer d'une tête à une autre. Enfin,la solution proposée par Wenzel, s'appuie sur des jugements exprimés dans le domaine perceptif, et désigne les HRTF d'un "bon localisateur" comme HRTF universelles ([WW91]).
2. l'approche "avec ajustement discret" consiste à proposer plusieurs jeux de filtres à l'auditeur qui doit choisir et utiliser celui qui "lui va le mieux". Pour que cette approche soit efficace, il faut que les têtes présentées segmentent de façon homogène l'ensemble de la population. Des tentatives de segmentation de la population dans le domaine du signal ont été proposées par Middlebrooks et Green, par Wightman et Kistler et par Shimada et al. ([MG92], [WK93], [SHH94]). Le choix parmi ces têtes candidates peut être fait par l'auditeur en écoutant les différentes têtes pour quelques positions révélatrices. On peut également envisager un choix automatique réalisé par proximité de certaines caractéristiques morphologiques de l'auditeur et des têtes. Cela nécessite bien sûr la détermination des caractéristiques prépondérantes pour la localisation. Enfin, une autre possibilité "automatique" consisterait à mesurer la proximité en s'appuyant cette fois sur quelques HRTF mesurées.
3. l'approche "avec ajustement continu" consiste à synthétiser le jeu de filtres qui auraient été obtenus par mesure sur la tête de l'auditeur. Elle peut s'appuyer sur les modèles structurels introduits pour l'approche 1., et consiste alors à régler les paramètres décrivant un jeu de HRTF générique ou bien à en ajuster les paramètres de transformation ([Mid99a]). Une alternative consiste à partir des paramètres morphologiques pertinents, introduits dans l'approche 2., et à modéliser le filtrage réalisé

par chacun d'entre eux ([Gen84], [LPM96], [AAD99]). La connaissance des paramètres morphologiques de l'auditeur permet ainsi d'estimer les paramètres du modèle. On parle de modélisation physique des HRTF. Enfin, les HRTF "individuelles" peuvent être approchées de façon perceptive, par un ajustement piloté par l'auditeur couplé à un algorithme d'optimisation ([RW96]). Cet ajustement peut conduire à l'exagération de certains traits des HRTF, rendant l'effet plus robuste, mais définissant peut-être des HRTF qui ne pourraient être mesurées sur aucun individu.

Dans ce chapitre, nous commencons par quantifier les différences inter-individuelles en définissant trois espaces de représentation des têtes :

- un espace "signal", qui divise en un sous-espace s'appuyant sur l'ITD des têtes et un sous-espace s'appuyant sur leurs HRTF à phase minimale,
- un espace morphologique, témoignant de l'écart des caractéristiques corporelles des têtes.
- un espace perceptif, qui, à l'aide des réponses au test perceptif du chapitre 4, traduit les différences audibles entre les têtes.

La définition de ces espaces permet ensuite d'envisager l'adaptation discrète de la synthèse binaurale, le principal enjeux consistant à trouver une méthode pour insérer le nouvel auditeur dans l'un de ces espaces afin d'élire la tête la plus proche. Nous proposons ainsi une solution pratique permettant un tel appairage à partir de la mesure d'un nombre réduit de paramètres morphologiques.

Enfin, nous étudions un mode d'adaption continue de la synthèse binaurale, introduit par Middlebrooks dans [Mid99a] et validée perceptivement dans [Mid99b]. Cette approche à base de déformation spectrale, que nous désignerons par *scaling fréquentiel*, cherche à transformer une tête en une autre par une homothétie de l'axe des fréquence portant les HRTF. Nous envisageons plusieurs extensions au travail de Middlebrooks, comprenant notamment la définition d'un "scaling multiple", impliquant un facteur de transformation spécifique pour différentes bande de fréquences, et pour différentes régions spatiales.

5.2 Mesure des différences inter-individuelles

Cette section a pour objectif de mettre en évidence et de quantifier les différences inter-individuelles, afin de montrer que toutes les têtes "ne se valent pas" : qu'elles ont des différences mesurables (HRTF, ITD, morphologie), et que ces différences sont perceptibles (jugements perceptifs). Cette étude est appliquée aux mesures de 17 têtes mises à notre disposition par R. Algazi de UC Davis, pour lesquelles nous disposons de HRTF mesurées sur 825 positions (calotte supérieure), de caractéristiques morphologiques, et pour lesquelles nous disposons des jugements perceptifs du test décrit au chapitre 4.
Ces observations justifieront alors la recherche de méthodes pour l'adaptation individuelle de la synthèse binaurale, que nous abordons aux sections suivantes.

5.2.1 Dépendance inter-individuelle des HRTF

5.2.1.1 Définition d'une distance inter-spectre

Dans les chapitres précédents, nous avons utilisé deux mesures de distance inter-spectre :

- une mesure d'écart entre deux spectres complexes, permettant d'évaluer l'erreur de modélisation de spectres complexes. L'information de phase y est prise en compte tout comme l'information d'amplitude :

$$d_{cx}(f) = \sqrt{\frac{\sum_{\theta_i} w(\theta_i).\left|H_2(\theta_i,f) - H_1(\theta_i,f)\right|^2}{\sum_{\theta_i} w(\theta_i).\langle H_1|H_2\rangle}}$$

  où $w(\theta)$ est une fonction de pondération des différentes positions vérifiant $\sum w(\theta) = 1$, et les $H_k(\theta, f)$ désignent les spectres complexes. On rappelle que, si cette distance est évaluée sur des spectres à symétrie hermitienne autour de la fréquence de Nyquist, $\langle H_1|H_2\rangle$ est réel.
  Une mesure indépendante de la fréquence requierrait le lissage "perceptif" de $d_{cx}(f)$. Or, les méthodes traditionnelles de lissage, telles que celles que nous avons utilisées aux chapitres précédents, s'appliquent

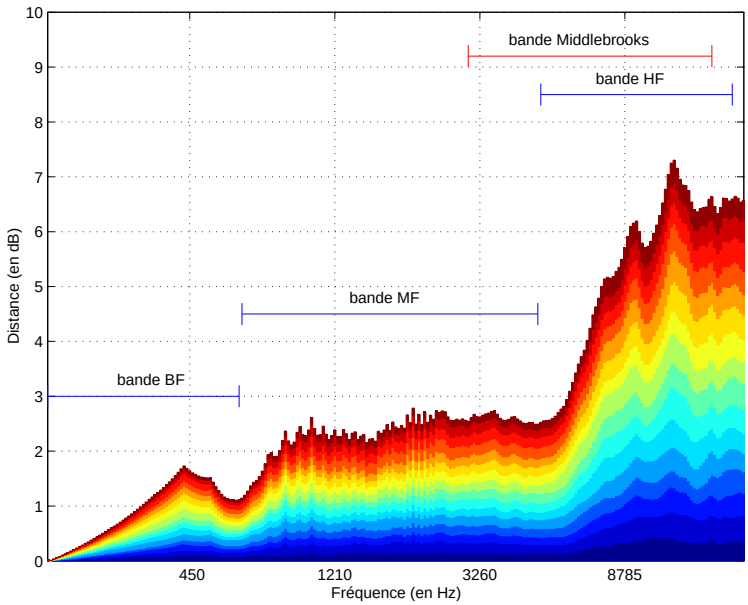


FIG. 5.1 – Distance moyenne entre les spectres d'amplitudes de la tête 148 et ceux des 16 autres. Les strates correspondent à la contribution de chacune des 16 têtes dans la distance moyenne. On a repéré les intervalles que nous proposons pour l'étude (BF, MF, HF) ainsi que l'intervalle retenu par Middlebrooks pour le scaling fréquentiel (cf section 3).

à des spectres de puissance, et perdent donc l'information de phase. L'étude du lissage des spectres complexes a été décrit par Hatziantoniou et Mourjopoulos, mais n'a pas été implantée dans le cadre de notre étude (cf [HM00]).

– une mesure d'écart entre deux spectres d'amplitude, pratiquée sur les log-magnitudes :

$$d_{mag}(f) = \sqrt{\sum_{\theta_i} w(\theta_i).\left|\log(\frac{mag_2(\theta_i,f)}{mag_1(\theta_i,f)})\right|^2}$$

Pour obtenir une mesure indépendante de la fréquence de d_{mag}, il suffit de lisser les spectres de puissance et de les re-échantillonner sur une échelle de fréquence perceptivement pertinente, afin de pouvoir sommer les valeurs obtenues pour chaque fréquence.

Nous choisissons de nous concentrer sur cette dernière expression, et nous re-échantillonnons les spectres d'amplitude avec une résolution de un 35ème d'octave. Cet échantillonnage est celui retenu par Middlebrooks aux résultats duquel nous souhaitons rapporter les notres.

5.2.1.2 Intervalle fréquentiel d'étude

Sur la Figure 5.1, nous présentons un exemple de la distance entre les spectres d'amplitude de deux têtes, d_{mag}. D'une manière générale, trois zones fréquentielles se distinguent : une zone basses-fréquence sur laquelle la distance ne dépasse par 2dB, une zone moyennes fréquences, sur laquelle elles valent autour de 3dB, et une zones hautes fréquences sur laquelle elle peut atteindre 9dB. La dynamique des HRTF est effectivement plus importante en hautes fréquences, ce qui explique qu'une mauvaise superposition des spectres se traduise par une plus grande distance sur cette zone. Ces écarts dépassent le seuil d'audibilité d'un écart d'intensité, et à ce titre doivent être audibles. En outre, ils dépassent l'écart observé pour une même tête entre positions voisines. En effet, sur la Figure 5.2, on présente la distance moyenne entre deux positions d'un même cône de confusion écartées d'environ 17°. C'est l'écartement angulaire que nous avons retenu pour atteindre une distance inter-spectre moyenne proche de celle présentée avec la même tête en Figure 5.1. Bien que les données manquent pour les cônes hors du plan médian, on peut penser que ces 17° excèdent le hâlot sonore du système auditif et que les deux sources sont discriminées.

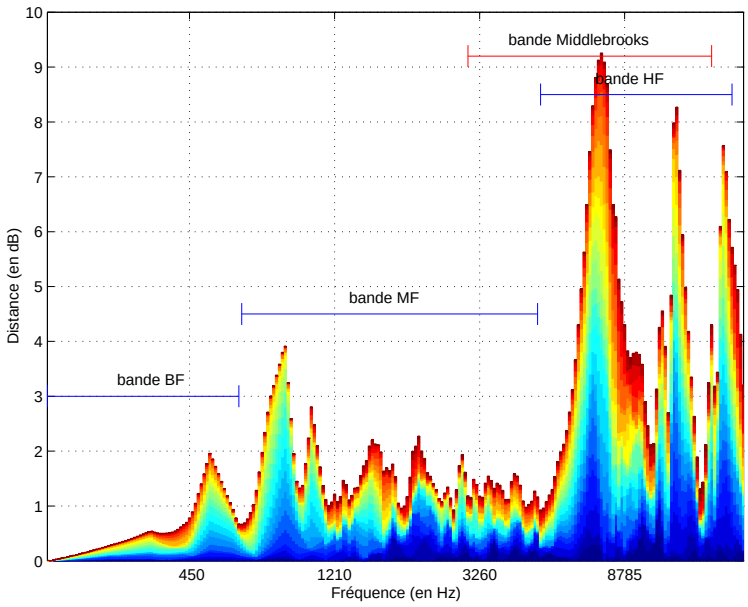


FIG. 5.2 – Distance moyenne entre les spectres d'amplitudes de la tête 148 pour différentes positions. La position de référence est la position frontale du plan horizontal d'un cône de confusion. La distance est calculée par rapport à une position élevée de 16.9° sur le même cône. La moyenne est calculée sur 25 cônes (de -80° à 80° selon les conventions Algazzi).

Par conséquent, on peut penser que non seulement la distance "signal" entre deux têtes est perceptible (seuil d'audibilité d'un ΔI), mais qu'en outre elle est d'ampleur à altérer la position perçue de la source. Elles sont donc non négligeables.

En outre, on peut considérer que les trois intervalles fréquentiels BF, MF et HF délimitent les plages de variation "locales" de certaines caractéristiques fréquentielles. D'après cette hypothèse, par exemple, les HRTF possèdent des caractéristiques marquées en moyennes fréquences, dont l'emplacement en fonction de la tête tout en restant confiné dans l'intervalle MF. Cette plage de variation est plus faible en basses et en hautes fréquences. Nous ferons l'hypothèse que ces intervalles traduisent des différences inter-individuelles a priori indépendantes, et la distance inter-spectres sera évaluée pour les trois intervalles de fréquences :

1. d_{mag}^{BF} sur l'intervalle basses fréquences [170Hz-640Hz] (noté BF), soit 67 points fréquentiels,
2. d_{mag}^{MF} sur l'intervalle moyennes-fréquences [641Hz-4800Hz] (noté MF), soit 103 points fréquentiels,
3. d_{mag}^{HF} sur l'intervalle hautes-fréquences [5000Hz-18000kHz] (noté HF), soit 67 points fréquentiels.

5.2.1.3 Technique d'analyse multidimensionnelle

Nous recherchons une représentation de l'espace des 17 têtes à partir de la connaissance de la distance les séparant deux à deux. Ce résultat peut être obtenu à l'aide d'une analyse muldimensionnelle, ou MDS (Multidimensional scaling) dont nous rappelons brièvement les mécanismes.
Comme le décrit Saporta ([Sap90]), l'analyse mutidimensionnelle a pour objectif de *"trouver une configuration de n individus dans un espace de faible dimension"*, objectif identique à celui de l'analyse en composantes principales (ACP). Les deux méthodes d'analyse se distinguent toutefois :

– les données de départ sont des distances entre individus pour l'analyse MDS alors qu'il s'agit des coordonnées des individus pour l'ACP. La distance utilisée par le mécanisme de minimisation de l'ACP est "imposée" (du type $\sqrt{\sum (x_i - y_i)^2}$).
– la propriété "d'emboitement" des représentation de l'ACP (*la meilleure représentation à p dimensions se déduit de la meilleure représentation à p−1 dimensions en rajoutant un p^{eme} axe*), n'est pas vérifiée pour les techniques de MDS. Cela n'empêche par un classement des axes par variance expliquée croissante.

La première étape des MDS consiste à dériver une distance euclidienne δ de la mesure d fournie en entrée. Celle-ci doit vérifier :

- $d_{ij} \geq 0$
- $d_{ij} = d_{ji}$
- $d_{ij} \leq d_{ik} + d_{kj}$

L'inégalité triangulaire est la propriété la moins "naturellement" vérifiée par la mesure de distance définie entre nos données. La distance δ peut être obtenue par la méthode de la constante additive, mettant en évidence la constante c telle que $\delta_{ij}^2 = d_{ij}^2 + c^2$ avec $\delta_{ii} = 0$, soit euclidienne. On a coutume de choisir $c = \max_{jkl} (\delta_{jl} - \delta_{jk} - \delta_{kl})$ ([Van01]).

L'espace multi-dimensionnel recherché doit être tel que les distances δ respectent au mieux l'ordre défini par d. On peut chercher une relation monotone entre les deux distances, telle que :

$$d_{ij} < d_{kl} \Rightarrow \delta_{ij} < \delta_{kl}$$

L'analyse est alors seulement "ordinale". Nous avons choisi une contrainte plus forte, imposant la linéarité de cette transformation. Nous cherchons donc M telle que :

$$M(d_{ij}) = \alpha.d_{ij} + \beta \quad \forall\, i, j$$

Nous procédons ainsi à une analyse "métrique". L'algorithme MDSCAL de J.B. Kruskal, que nous avons utilisé, cherche alors à minimiser un indicateur de l'adéquation du modèle, nommé *stress* :

$$\min_M \frac{\sum_{i,j} \left(\delta_{ij} - M(d_{ij})\right)^2}{\sum_{i,j} \delta_{ij}^2}$$

Le stress prend ses valeurs entre 0 et 1. Enfin, notons que les axes donnés par l'analyse sont arbitraires : le modèle est invariant par rotation. Une présentation en détails de l'analyse multidimentionnelle et de son application est donnée par G. Vandernoot dans [Van01]. On peut également se reporter à [SRY81]. Nous avons utilisé l'implantation de MDSCAL du logiciel SYSTAT ([WL89]).

5.2.1.4 Application à un espace de représentation des distances inter-HRTF

Nous appliquons l'analyse MDS aux distance inter-têtes $d_{mag}^{\cdot\cdot}$ évaluées pour nos 17 têtes sur les 825 positions de la calotte supérieure. Pour choisir le nombre de dimensions de l'espace de représentation, nous observons le stress et la variance expliquée qui sont des sorties standard de SYSTAT (cf Figure 5.3). En essayant de repérer un coude dans ces deux courbes, indiquant une moins grande efficacité marginale de l'analyse, on aboutit à un espace de représentation des têtes de dimension 4, pour lequel le stress est inférieur à 0.06 et la variance expliquée dépasse 95%.

La projection de cet espace sur les deux premiers axes est présentée en Figure 5.4. La distribution des têtes apparaît spécifique à la plage de fréquence. Pour chacune apparait au moins une tête "atypique" : la tête 028 pour BF, les têtes 061 et 028 pour MF, et, moins nette, la tête 061 pour HF. Afin de vérifier la stabilité de la configuration obtenue, et notamment la singularité de ces têtes, nous avons utilisé une application freeware de visualisation de données interactive, XGvis ([BSLD98])[1]. Nous vérifions tout d'abord que nos distances suivent une distribution quasi-normale. Pour BF, la tête 028 maintient son écart aux autres têtes en dépit d'un "jitter" introduit dans les données de départ. On note également une grande robustesse de l'écart entre les têtes 028 et 050. De même pour MF, on vérifie que les têtes 061 et 028 "se repoussent", ce qui confirme l'information fournie par notre analyse MDS. Toutefois, on observe également que la singularité de la tête 061 est plus patente que celle de la tête 028. Pour HF, enfin, la tête 061 sort de façon consistante du lot des autres têtes.
La singularité de certains éléments est susceptible d'accaparer beaucoup de l'effort réalisé par l'analyse MDS, puisqu'ils génèrent une forte erreur en cas de mauvais positionnement. Par conséquent, une distribution plus fidèle des autres têtes peut être obtenue en pratiquant l'analyse MDS sur ce sous-ensemble. C'est ce que nous présentons en partie (b) de la Figure 5.4. On peut observer que la distribution des têtes

[1] Merci à Suzanne Winsberg, chercheur à l'Ircam, pour avoir porté à notre attention cet outil très précieux.

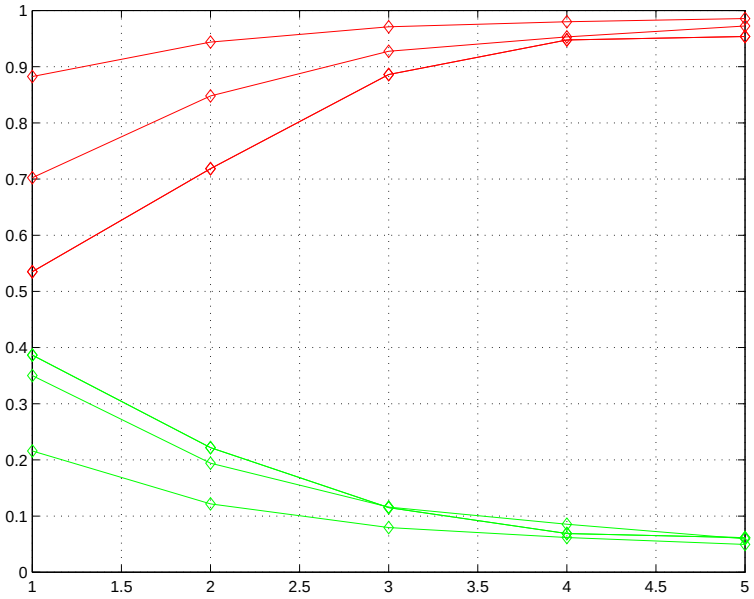


FIG. 5.3 – stress (vert) et variance expliquée (rouge) en sortie de SYSTAT pour les 3 bandes de fréquence, en fonction du nombre de dimensions retenues.

est faiblement modifiée par cette opération, mis à part, par exemple, pour les têtes 148 (BF et MF) et 119 (MF) qui, initialement parmi le groupe, migrent vers la périphérie après le retrait des éléments singuliers. Pour l'ensemble, la modification se comporte essentiellement comme une combinaison de dilatations et de rotations.
Parmi les têtes s'opposant, on peut noter la paire (050,061), qui pour les trois intervalles fréquentiels, présente une des plus grandes distances. Les autres paires "distantes" sont plus spécifiques à chaque zone de fréquence :

– (003,15), (003,061) en BF,
– (065,119) et (057,119) en MF,
– (055,119), (055,147) et (050,119) en HF.

Si l'on se place dans le cadre d'une synthèse binaurale sans adaptation individuelle, le problème se pose de choisir la tête qui sera proposée à tous les auditeurs. Cette sélection peut se porter sur la tête aux HRTF "moyennes", i.e. sur le barycentre de notre nuage de têtes, représenté sur la Figure 5.4 (calcul effectué sans les têtes atypiques avec l'ensemble des coordonnées). Le plus proche représentant est la tête 048 en BF, 035 en MF, et la tête 065 en HF.

5.2.1.5 Conclusion

Les HRTF de différentes têtes présentent ainsi de fortes dissimilarités, parfois spécifiques à la zone fréquentielle considérée. Ces écarts sont supérieurs à celui que l'on observe entre deux position discriminables et justifient à ce titre que l'on cherche à les réduire. Une analyse statistique nous a permis d'obtenir une représentation des têtes dans un espace de dimension réduite, d'observer la singularité de certaines têtes (028 en BF,061 et 028 en MF, 061 en HF), et de déterminer une tête "moyenne", barycentre de l'ensemble, pour chaque intervalle de fréquences.

5.2.2 Dépendance inter-individuelle du retard interaural

Nous nous sommes précédemment intéressés aux différences inter-individuelles des HRTF à phase minimales. Dans cette section, c'est la composante complémentaire, le retard interaural, qui est considérée.

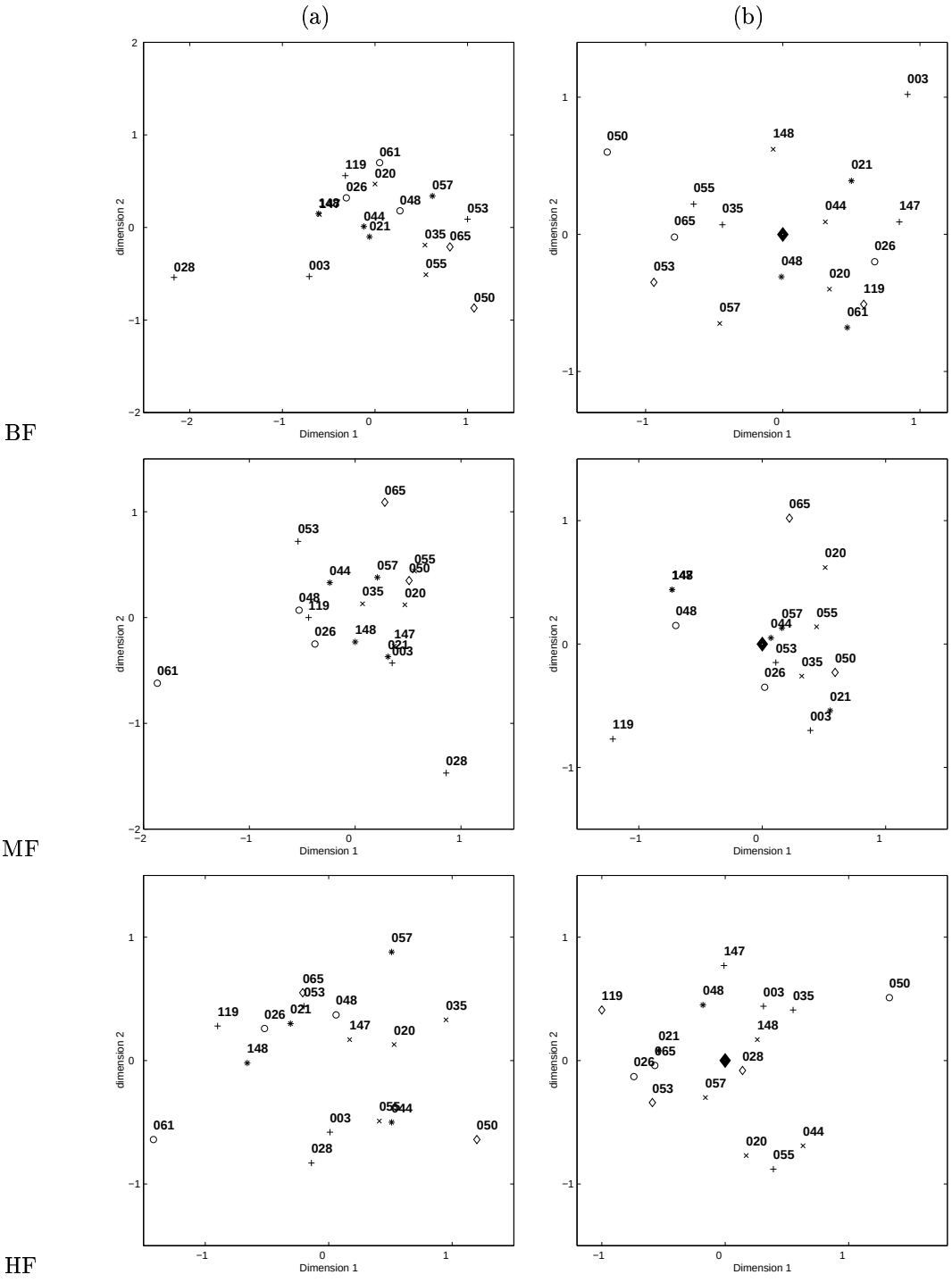


FIG. 5.4 – Espaces de représentation des têtes à partir d'une mesure de distance spectrale évaluée sur chaque zone fréquentielle : (a) les 2 premiers axes donnée par l'analyse MDS, (b) même chose qu'en (a) après avoir retiré les têtes "atypiques" (tête 028 pour BF, têtes 028 et 061 pour MF, 028 pour HF). Le carré noir représente le barycentre du nuage de points.

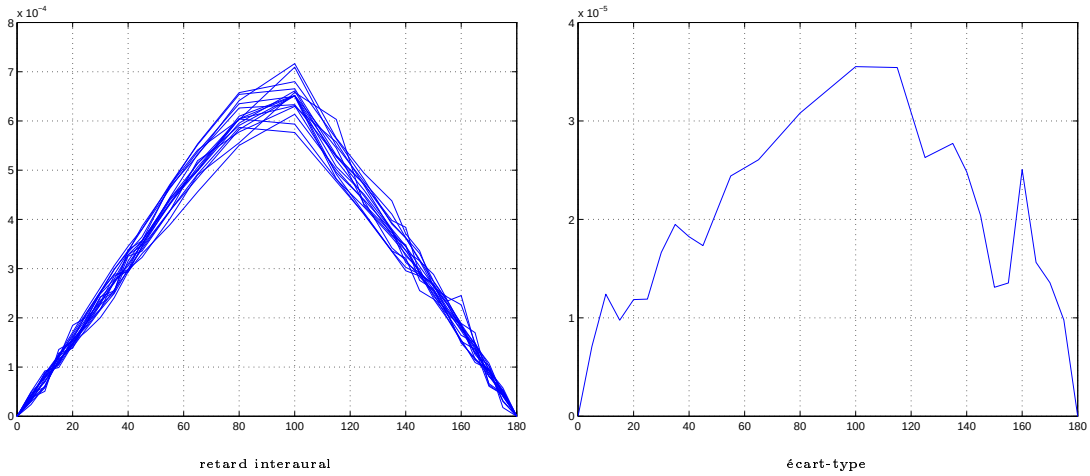


FIG. 5.5 – Différence inter-individuelle du retard interaural pour les 17 têtes de UC Davis

5.2.2.1 Erreur de localisation induite par un retard interaural non individuel

Nous présentons en Figure 5.5 le retard interaural pour plusieurs têtes. On constate un écart entre les têtes, caractérisé par la différence de pente de leur ITD. Le modèle "sphérique" présenté en chapitre 1 décrivait en effet l'ITD par la loi :

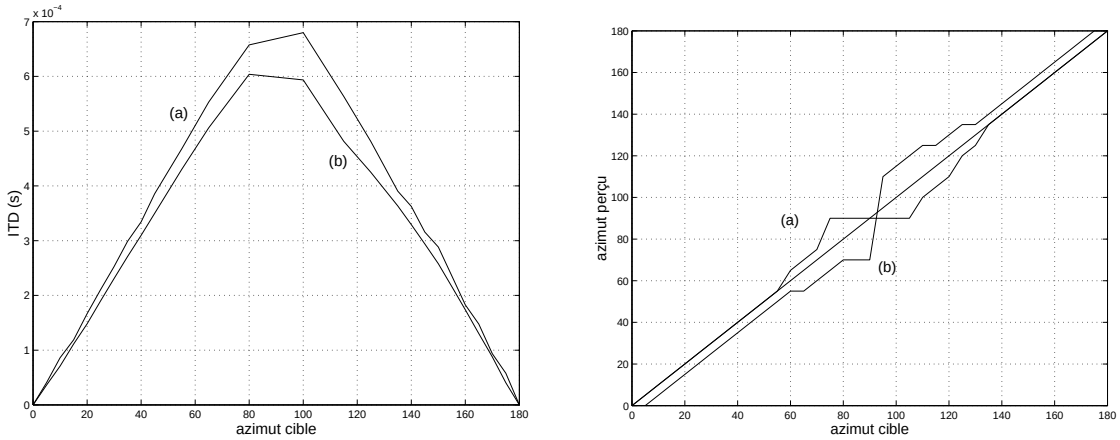
$$ITD(\theta) = \frac{r}{c}.f(\theta)$$

où $f(\theta)$ est une fonction de l'azimut, croissante de 0° à 90° et décroissante de 90° à 180°. Dans cette expression, la différence entre les individus s'exprime par le "rayon sphérique équivalent", ou radius, r, et on constate que $\frac{\partial ITD}{\partial r}$est proportionnel à $f(\theta)$. On peut donc s'attendre à ce que la différence inter-individuelle d'ITD suive les mêmes variations que l'ITD lui-même et par conséquent qu'elles augmente quand l'azimut se rapproche des positions latérales.
Pour en donner un ordre de grandeur, nous étudions les distorsions d'azimut perçu dans le cas où un individu à "petite tête" écouterait au casque une stéréophonie en $\triangle T$ utilisant comme loi de panpot l'ITD d'une "grosse tête", et réciproquement. Comme on l'observe en Figure 5.6, la distorsion peut atteindre 20° d'erreur sur l'azimut. Dans le cas d'une loi d'ITD trop "grande", l'azimut perçu balaie tout l'intervalle utile, mais entre 70° et 110° (pour notre exemple), l'ITD perçu reste constantamment à 90°. Des écoutes informelles laissent néanmoins penser que l'augmentation du retard sur cet intervalle est audible, et se traduit par une sensation d'extériorisation (distance perçue de plus en plus lointaine). Utiliser une "grande" tête pour la simulation de l'ITD produit ainsi une distorsion moins pénalisante que le cas contraire, où la simulation est faite avec un ITD trop "petit". Dans ce cas en effet, l'azimut perçu ne dépasse pas une valeur seuil (70° dans notre exemple), et suit une loi discontinue, sautant subitemment de 70° à 110°. L'adaptation de l'ITD à l'auditeur apparait donc comme un facteur important pour la qualité de localisation pour une stéréophonie en $\triangle T$. En outre, on peut penser que l'adaptation seule des HRTF à phase minimale ne peut suffire pour pallier les distorsions d'un ITD non adapté. En effet, comme le rapportent Wightman et Kistler [WK92], l'ITD constitue un indice de localisation prépondérant sur les indices spectraux fournis par les HRTF. Il est donc nécessaire d'envisager une adaptation individuelle de l'ITD pour préserver la qualité de localisation de la synthèse binaurale.

5.2.2.2 Espace de représentation des distances inter-ITD

Une approche pour l'adaptation individuelle du retard interaural passe par la paramétrisation de ce dernier, par exemple à l'aide d'un modèle sphérique équivalent de la tête. Les paramètres de ce modèle peuvent ensuite être reliés aux dimensions morphologiques de la tête, et l'adaptation est obtenue en utilisant les dimensions appropriées de l'auditeur. C'est l'approche adoptée par Algazi et al. ([AAD01]) :

1. l'ITD est modélisé à l'aide de la formule de woodworth dans le plan horizontal (azimut repéré par



ITD de deux têtes de taille différente : "grosse" tête (a), "petite" tête (b)

distorsion induite

FIG. 5.6 – Effet de l'écoute d'un ITD non-individuel sur l'azimut perçu : (a) "petite" tête utilisant l'ITD de la "grosse" tête, (b) "grosse" tête utilisant l'ITD de la "petite" tête.

θ) :

$$ITD(\theta) = \frac{r}{c}.(\theta + \sin(\theta)) \tag{5.1}$$

en obtenant r par projection orthogonale de l'ITD individuel mesuré sur la fonction spatiale $\theta + \sin(\theta)$,

2. le rayon équivalent r est corrélé avec les dimensions de la tête, et obtiennent :

$$r \simeq 0.48.w_1 - 0.003.w_2 + 0.22.w_3 + 3[cm]$$

avec :

$$\begin{array}{lll} w_1 : & 1/2 & largeur\ \ de\ \ tête \\ w_2 : & 1/2 & hauteur\ \ de\ \ tête \\ w_3 : & 1/2 & profondeur\ \ de\ \ tête \end{array}$$

Les rayons équivalents que nous obtenons à partir de 17 têtes fournissent des résultats comparables à ceux d'Algazi et al. :

1. le rayon sphérique équivalent varie entre 7.91cm et 9.22cm (7.95cm et 9.5cm pour Algazi et al.), avec une valeur moyenne de 8.55cm (contre 8.7cm).
2. l'erreur quadratique moyenne est comprise entre 10 et 32μs (22 et 47μs pour Algazi et al.), et atteint en moyenne 18.5μs (contre 32μs), soit une erreur angulaire inférieure à 3° dans le plan horizontal. L'erreur est donc en général plus faible en utilisant le modèle sphérique avec le rayon approprié qu'en utilisant un ITD mesuré non individuel pris au hasard.

Nous proposons d'étendre le modèle sphérique de Woodworth aux positions hors du plan horizontal par l'expression ([LJ97]) :

$$ITD(az, el) = \frac{r}{c}.(f(az, el) + g(az, el)) \tag{5.2}$$

avec :

$$f(az, el) = \cos(el).\sin(az)$$

et :

$$g(az, el) = \arcsin(\cos(el).\sin(az))$$

Cette expression est donnée par des considérations géométriques simples. Les valeurs précédentes sont différentes si l'optimisation est étendue à toutes les positions de la calotte supérieure. Le rayon obtenu est systématiquement supérieur, d'environ 5%, mais reste très corrélé au "rayon sphérique horizontal" ($r = 0.88$). L'erreur moyenne globale augmente légèrement jusqu'à 22μs, et atteint cette fois 26μs pour les positions horizontales. Comme dans l'étude d'Algazi et al., on constate que la largeur de la tête est le

TAB. 5.1 – Rayons du modèle d'ITD à deux paramètres, pour une optimisation basée sur les positions horizontales seules ou sur l'ensemble des positions.

	optimisation 3D		optimisation 2D	
	r_1	r_2	r_1	r_2
min	2.4cm	10.7cm	0	11.3cm
max	8.4cm	14.7cm	5.3cm	15.0cm
mean	4.9cm	12.3cm	3.1cm	13.1cm

paramètre morphologique mesuré le plus fortement corrélé avec le rayon sphérique équivalent ($r = 0.84$). Reprenant la corrélation multiple qu'il a proposée, nous obtenons des paramètres légèrement différents :

$$r_{2D} \simeq 0.73.w_1 - 0.06.w_2 + 0.17.w_3 + 2.07$$

$$r_{3D} \simeq 0.66.w_1 - 0.04.w_2 + 0.11.w_3 + 3.33$$

Comme Algazi, nous pouvons conclure quant à la faible contribution du paramètre de la hauteur de la tête. L'erreur quadratique moyenne qu'il obtient pour l'estimation du radius dans le plan horizontal (r_{2D}) est de l'ordre de 0.12cm, que nous pouvons rapprocher des 0.18cm que nous obtenons, contre 0.17cm pour r_{3D}.

Pour améliorer l'ajustement entre ITD mesuré et modèle paramétrique, nous proposons d'introduire un degré de liberté supplémentaire, en décomposant l'ITD mesuré sur chacune des deux fonctions spatiales permettant de l'approximer. Dans le modèle proposé dans l'expression 5.2, ces deux composantes sont pondérées par un gain commun, mais une optimisation par projection orthogonale conduit à une approximation plus précise, suivant l'expression :

$$ITD(az, el) = \frac{1}{c}.(r_1.f(az, el) + r_2.g(az, el)) \tag{5.3}$$

Avec ce nouveau modèle, l'ITD est, comme on peut s'y attendre, modélisé avec une meilleure précision, : l'optimisation dans le plan horizontal conduit à une erreur quadratique de $15.5\mu s$, tandis que pour une optimisation 3D, elle approche $21\mu s$, valeurs systématiquement inférieures à celles observées dans le cas d'une approximation "mono-rayon".
Les observations faites sur l'ensemble des têtes montrent que les rayons r_1 et r_2 sont décorrélés avec le rayon sphérique équivalent, et qu'ils sont fortement anti-corrélés entre eux. Les valeurs caractéristiques de ces deux rayons sont recensées dans le Tableau 5.1. La corrélation avec les paramètres morphologiques ne donne pas de résultat saillant : le coefficient de corrélation ne dépasse pas 0.61.

A l'aide de ces modèles, on peut représenter les têtes dans un espace de dimension 1 (modèle de tête sphérique), ou de dimension 2 (modèle (r_1, r_2)). Ces espaces sont donnés en Figures 5.7 et 5.8. Afin de mener des comparaisons avec les représentations obtenus à l'aide des distance inter-spectres, nous nous focalisons sur les optimations 3D. Notons qu'il n'y a aucune raison a priori pour que les espaces ITD et HRTF à phase minimale se superposent : ils représentent deux composantes complémentaires de la distance signal entre les HRTF à phase mixte.

Analyse de l'espace donné par le modèle à un seul rayon (radius de l'éq. 5.3)
Pour ce qui est de l'espace obtenu avec le radius, on remarque que la tête 061, "hors-norme" pour l'espace HRTF, sort également du groupe en se plaçant parmi les têtes à plus faible radius. La tête 028 en revanche, se place au milieu de la distribution. Parmi les oppositions les plus fortes observées dans l'espace HRTF, on retrouve ici la forte distance entre les têtes 055 et 119 (HF), la tête 003 et les têtes 119 et 061 (BF). Suite à la réflexion menée sur les artefacts d'un ITD non individuel, le choix d'un ITD optimal pour une synthèse binaurale non adaptée à l'auditeur nous conduirait à privilégier la tête à plus grand radius. Les

têtes 020, 055, 003 et 044 constituent à ce titre les meilleures candidates : on garantit ainsi que tous les azimuts sont balayés pour toutes les têtes (Figure 5.6). Les têtes moyennes "préférées" pour l'espace HRTF figurent quant à elles parmi les têtes à plus faible radius, mis à part la tête 035, retenue pour l'intervalle de fréquences MF.

Analyse de l'espace donné par le modèle à deux rayons (r_1 et r_2 de l'éq. 5.2)
L'espace de représentation fourni par le modèle d'ITD à deux variables confirme la singularité de la tête 028, qui présente un très fort rayon r_2. Les têtes à ITD maximum se retrouvent aux extrémités des axes r_1 (020) ou r_2 (055), ou bien à une position éloignée sur la première diagonale (003). L'écart entre les tête 050 et 061, présentent pour tous les intervalles fréquentiels dans l'espace HRTF, constitue ici aussi l'une des plus fortes distances inter-têtes.

5.2.2.3 Conclusion

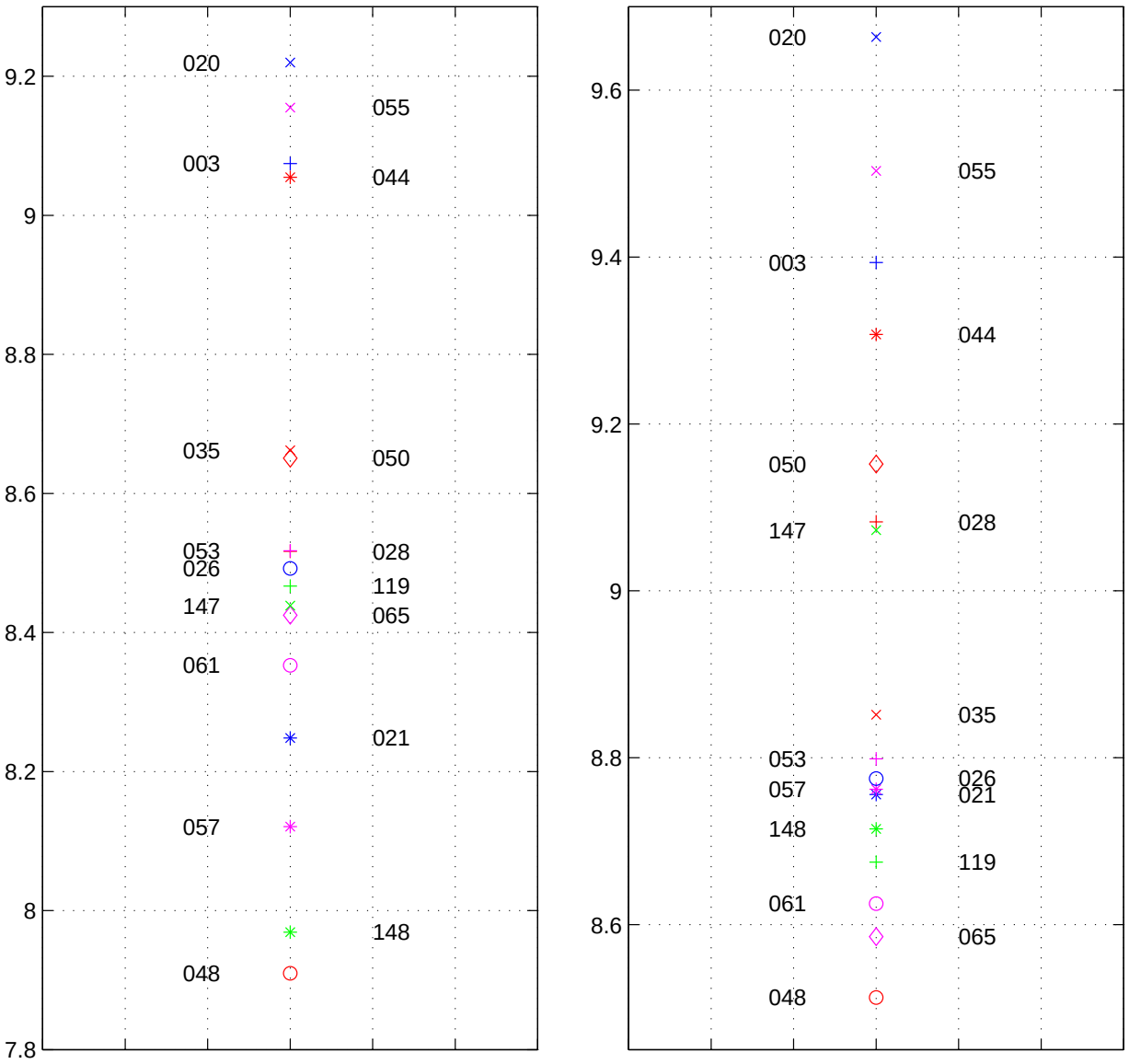
L'absence d'adaptation individuelle de l'ITD est susceptible de créer des artefacts de localisation audibles, qui peuvent être néanmoins limités dans le cas de l'utilisation d'un ITD surestimé.
La loi de variation de l'ITD peut être paramétrisée à l'aide d'un modèle standard faisant l'hypothèse d'une tête sphérique, dont nous avons étendu l'expression aux positions en élévation. L'adaptation de l'ITD est alors possible par le réglage d'un seul paramètre, fortement corrélé avec la largeur de la tête. Nous proposons un modèle à deux degrés de libertés découlant du précédant, et permettant une meilleure approximation de l'ITD. Les paramètres du modèle n'offrent pas, néanmoins, l'avantage pratique d'être corrélés à un paramètre morphologique.
Les espace de représentation des têtes qui se déduisent de ces deux modèles constituent tous deux un espace inter-ITD. Il ne se superposent pas avec l'espace inter-HRTF. Certains écarts ou têtes atypiques se retrouvent néanmoins, et notamment la singularité des têtes 028 (espace r_1/r_2) et 061 (espace radius), ou l'opposition entre les têtes 050 et 061. Enfin, les têtes 028, 055 et 119 constituent de bons candidats pour baliser l'espace signal, puisqu'elles apparaissent balayer "les possibles" tant dans l'espace ITD que dans l'espace des HRTF à phase minimale.

5.2.3 Dépendance inter-individuelle des caractéristiques morphologiques

Dans cette section, nous mesurons les différences inter-individuelles des caractéristiques morphologiques, à l'origine des différences de type "signal" observées dans les section précédentes.

5.2.3.1 Définition d'un protocole de mesure

La conception de têtes artificielles a suscité de nombreuses campagnes de relevés morphologiques, que l'objectif visé soit l'enregistrement ou la synthèse binaurale, ou le design industriel (casques de protection, combinés téléphoniques , ...). Nous nous inscrivons dans le prolongement des études concernant le premier de ces points, parmi lesquels figurent les travaux de Burkhard et Sachs ([BS75]) pour la conception de la tête KEMAR, et ceux de Genuit ([Gen84]), appliqués à la conception des têtes artificielles Head Acoustics. Les travaux de Genuit ont suscité l'intérêt pour la modélisation physique des HRTF, avec pour objectif de reconstruire les HRTF par l'association du modèle de filtre équivalent de chaque caractéristique morphologique. C'est dans ce but qu'Algazi et al. ont pratiqué des relevés morphologiques sur un grand nombre de têtes, en complément des mesures de HRTF sur les mêmes têtes ([AAD99], [AAD01]). D'autres mesures morphologiques ont été réalisées par Middlebrooks, pour l'interprétation et le réglage automatique de paramètres pilotant le "scaling fréquentiel" des HRTF, approche décrite en section 5.4 ([Mid99a]). Les relévés morphologiques que nous avons menés avaient pour objectif l'étude de l'adaptation individuelle des HRTF ([Auz99]).
La multiplicité de ces campagnes s'accompagne d'une certaine hétérogénéité dans les protocoles de mesure et dans les facteurs relevés. Toutefois, ce "désordre" apparent ne doit pas cacher les efforts de normalisation. Le standard de l'ANSI par exemple ([Ano85]) a fourni en 1985 un ensemble de dimensions moyennes de la tête, dont on a désormais convenu qu'elles sont sous-estimées. De même, la communauté internationale a défini un protocole de mesure pour 60 points de la tête ([Glo73]), mais les caractéristiques retenues sont visiblement mal adaptées à la définition d'une tête artificielle ([BPA+91]). Nous présentons dans le



optimisation “2D” optimisation “3D”

FIG. 5.7 – Espace de représentation des têtes à l'aide des écart de rayons obtenus avec un modèle sphérique de la tête à un seul degré de liberté (cf éq. 5.2) : le radius est obtenu à partir des mesures dans le plan horizontal (optimisation “2D”) ou à partir des mesures de la calotte supérieure (optimisation “3D”).

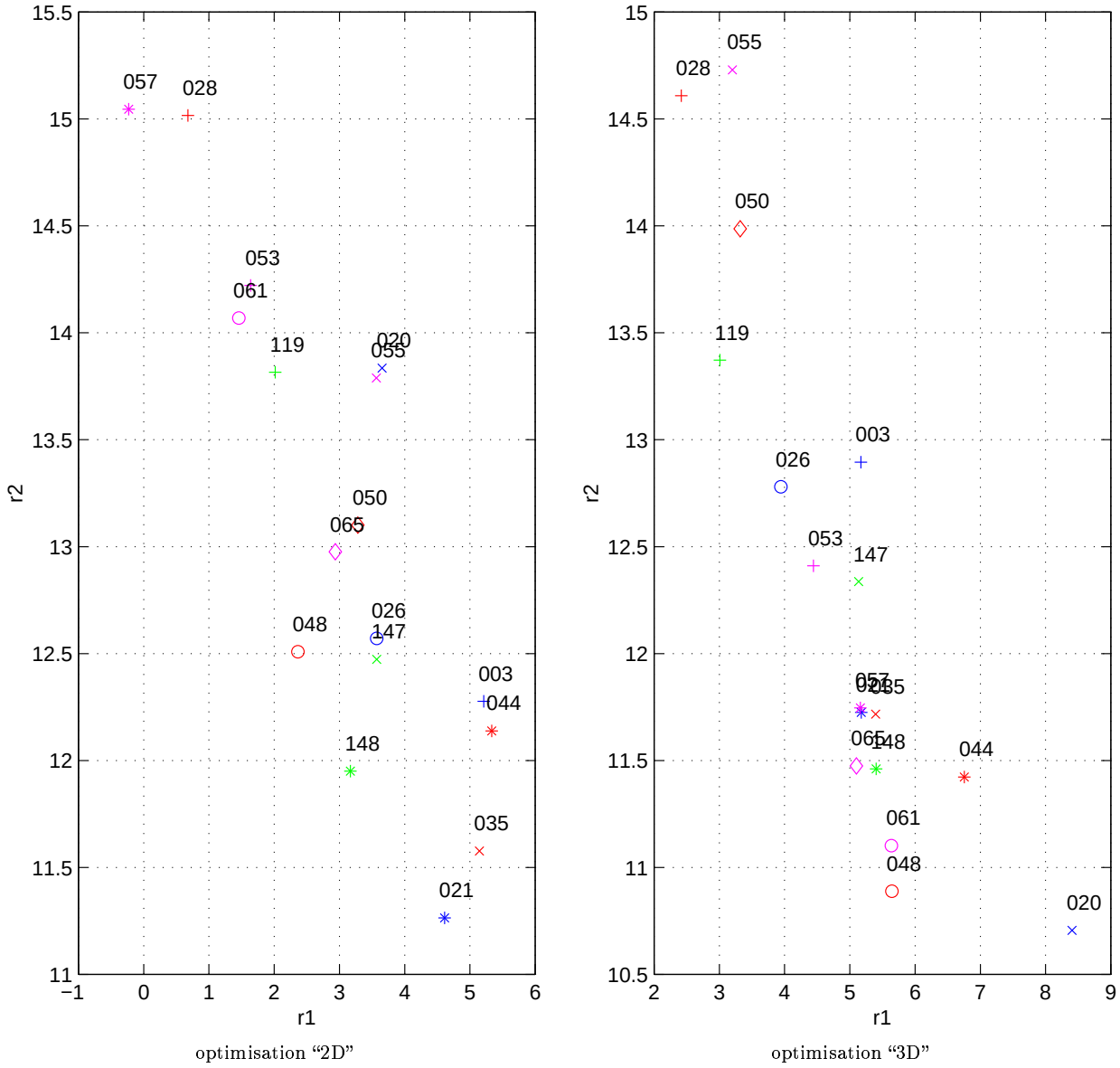


FIG. 5.8 – Espace de représentation des têtes à l'aide des écart de rayons obtenus avec un modèle sphérique à deux degrés de liberté (cf éq. 5.3) : r_1 et r_2 sont obtenus à partir des mesures dans le plan horizontal (optimisation “2D”) ou à partir des mesures de la calotte supérieure (optimisation “3D”).

Burkhard75		
10 paramètres tête+torse	données moyennes extraites de [CT57] et [Dre67]	
13 paramètres pour l'oreille	12 hommes et 12 femmes	photo, pied à coulisse, moulage
Genuit 86		
28 paramètres	6 hommes	photo, règle
Burandt 91		
5 paramètres	77 hommes âgés de 18 à 56 ans	pied à coulisse
11 paramètres	id.	*Aachen Ear Anthropometer*
Larcher 99		
17 paramètres	14 hommes et 6 femmes	photo, pied à coulisse, cordelette
Middlebrooks 99		
8 paramètres	33 sujets	règle, pied à coulisse
Algazi 99		
23 paramètres	80 sujets	photo, stylet optique
Larcher 00		
9 paramètres	11 hommes et 4 femmes	photo, pied à coulisse

TAB. 5.2 – Vue d'ensemble des campagnes de mesures morphologiques comparées dans ce chapitre.

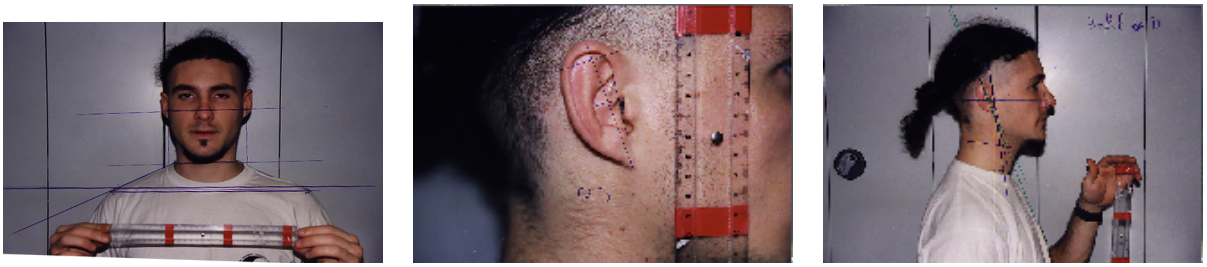


FIG. 5.9 – Exemple de prises de vue pour la session 99

Tableau 5.2 une vue d'ensemble des protocoles utilisés dans les études que nous comparerons par la suite.

1. Outils de mesure des caractéristiques morphologiques
   On distingue deux grands types de technique de mesure :

   (a) Certains paramètres sont mesurés sur le "relief naturel", le plus souvent à l'aide d'un pied à coulisse. Celui utilisé par Burandt et al., le *Aachen Ear Anthropometer*, a été construit ad hoc et peut tourner autour de l'axe interaural. Nous avons également utilisé un pied à coulisse "amélioré", construit avec de longues machoires pour pouvoir mesurer les dimensions de la tête.
   D'autres outils peuvent être adjoints au pied à coulisse : Burkhard et Sachs ont également recours au moulage de la conque pour en mesurer le volume et les dimensions ; Algazi et al. utilisent un stylet optique pour mesurer avec le plus de précision les dimensions du pavillon et de la conque ; nous avons eu besoin d'une cordelette afin d'estimer le tour de tête à l'altitude des oreilles prenant en compte le contour des surfaces.

   (b) D'autres dimensions sont estimées à partir de projections 2D du relief naturel de la tête, obtenues par photographie. Les Figures 5.9 et 5.10 présentent les photos que nous avons réalisées lors de nos deux campagnes de mesures. Pour pouvoir comparer les dimensions observées sur plusieurs photos, il est nécessaire qu'elles partagent un étalon de référence. Pour la session de 1999, cet échelon est constitué par la règle graduée que tiennent les sujets. Ce système est très fragile puisque la règle doit alors être dans le plan de projection, ce qui est rarement rigoureusement vérifié. Pour la seconde session, nous avons utilisé un matériel de photographie semi-professionnel[2] , et le cadrage est réalisé dans un rapport de proportion constant avec la réalité pour les plans rapprochés. Pour la photo de profil, un échelon de 10cm est fixé sur le

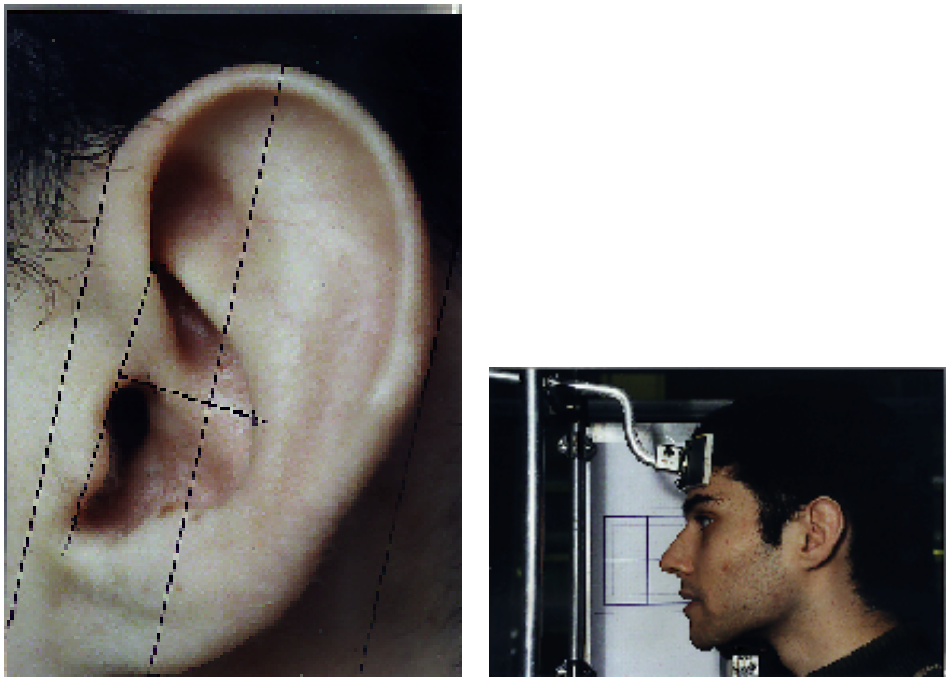


FIG. 5.10 – Exemple de prises de vue pour la session 2000

dispositif utilisé pour immobiliser la tête. Nous avons estimé les dimensions morphologiques à partir de relevés manuels sur les tirages photographiques.
Une précision de mesure supérieure aurait pu être atteinte grâce à la numérisation de ces tirages (ou bien grâce à une prise de vue numérique) en s'aidant pour les relevés d'un utilitaire informatique pour le traitement des images. Les avantages à tirer d'une numérisation des relevés morphologiques vont au delà d'une meilleure précision. En effet, rien n'oblige à fixer les points de mesure lors du relevé, contrairement à la mesure directe. Et si le choix de la prise de vue contraint l'ensemble des points candidats, cette limite disparait si l'on utilise une prise de vue 3D, comme il est décrit par exemple dans [MFBBG00]. Dans cette étude, les données morphologiques de 40 têtes sont saisies à l'aide d'un scanner ([Lab90]), puis modélisées en 3D. Les auteurs explorent alors ces modèles, afin d'en tirer des dimensions moyennes pour un standard sur les combinés de radio-téléphones[3] .

2. Paramètres morphologiques mesurés
   Si les paramètres mesurés sont souvent très semblables d'une étude à l'autre, les estimations obtenues diffèrent parfois, du fait des points de références choisis pour la mesure. Pour faire ce choix, nous nous sommes inspirés des préconisations de Genuit (tableau 3.4-1 de [Gen84]). Les paramètres que nous avons recensés lors de nos deux sessions de mesure sont schématisés dans les Figures 5.11, 5.12, 5.13 et 5.14.

Il s'agit plus précisement de :

[2]Guillaume Vandernoot, qui a réalisé les prises de vue, avec son matériel, est chaleureusement remercié.
[3]Un résumé de l'étude est accessible à http ://www.tsi.enst.fr/~marquez/ABS_BEMS2000/text_BEMS2000.html

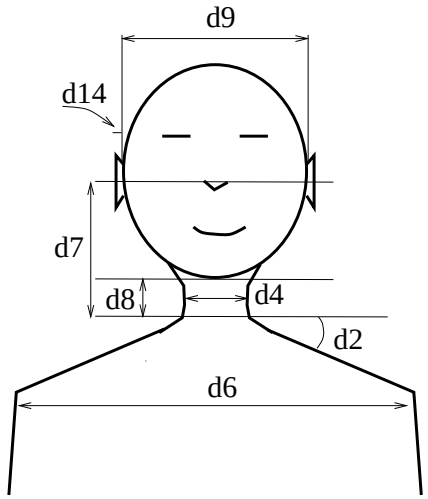


FIG. 5.11 – Paramètres mesurés à partir de photos (sauf d14 mesuré avec un pied à coulisse).

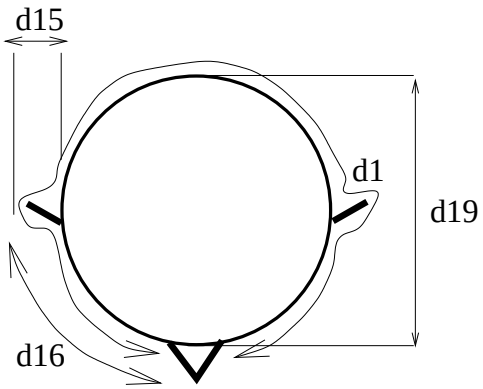


FIG. 5.12 – paramètres mesurés avec cordellette ou pied à coulisse.

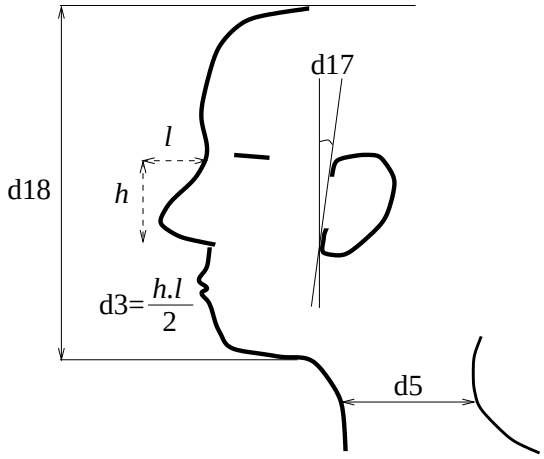


FIG. 5.13 – Paramètres mesurés à partir de photos de profil.

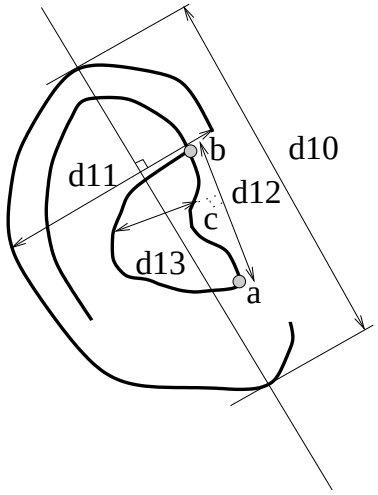


FIG. 5.14 – Paramètres du pavillon et de la conque mesurés avec un pied à coulisse (session 99), à partir de photos (session 2000).

d1	tour de tête sans nez	d11	largeur du pavillon*
d2	pente des épaules	d12	longueur de la conque*
d3	surface du nez	d13	largeur de la conque*
d4	largeur du cou*	d14	profondeur de la conque
d5	profondeur du cou	d15	décollement du pavillon
d6	largeur du torse	d16	quart de contour facial
d7	"altitude" des oreilles	d17	inclinaison du pavillon*
d8	"altitude" du menton	d18	hauteur de la tête$^\diamond$
d9	diamètre minimal de la tête*	d19	profondeur de la tête$^\diamond$
d10	longueur du pavillon*		

La définition des paramètres mesurés devrait être indépendante du moyen technique choisi pour la mesure. Nous essayons ci-après de décrire nos points de mesure aussi généralement que possible. Toutefois, nous parlerons parfois de dimension "projetée" pour préciser que le relevé a été fait à partir d'une photo, par opposition à une mesure suivant le volume du corps (pour les mesures de l'oreille d10 à d13, nous ne considérons que des dimensions projetées) :

- Le tour de tête sans nez mesure la circonférence de la tête au niveau des protubérences cartilagineuses des pavillons droite et gauche (points c de la Figure 5.14). Ce tour de tête s'arrête de part et d'autre du nez.
- La pente des épaules est définie comme l'angle entre l'horizontale et la ligne tangente au corps entre la base du cou et le décrochement des bras. La largeur du torse est mesurée comme la distance séparant ces points de rupture de pente à droite et à gauche.
- La largeur du cou est donnée par la plus petite distance horizontale entre les parois droite et gauche du cou. Pour la profondeur du cou, nous commençons par inscrire la vue de profil du cou au sein de deux droites parallèles. Ces droites sont rarement verticales. La profondeur est donnée par la distance entre ces deux droites.
- La mesure des "altitudes" requiert la définition préalable de la base du cou. Celle-ci est repérée par une ligne horizontale marquant la rupture de pente entre le cou et la chute des épaules. L'altitude des oreilles est alors définie comme la distance verticale entre la base du cou et l'entrée du conduit auditif. L'altitude du menton est donnée par la distance verticale entre la base du cou et la base du menton.
- Nous définissons le diamètre minimal de la tête comme la distance projetée entre les protubérances cartilagineuses des pavillons droite et gauche (points c de la Figure 5.14).
- Nous proposons tout d'abord d'inscrire le pavillon dans un rectangle. On commence par déterminer l'axe de plus grande longueur joignant deux points du pavillon. La longueur "extérieure" est parallèle à cet axe et tangente au point le plus extérieur du pavillon. La longueur "intérieure" lui est parallèle et passe par le point d'accolement de l'oreille à la tête. Notons qu'Algazi et al. ont choisi une convention

		session 1999				session 2000	Algazi 99		
		$\overline{m}$		σ		$\overline{m}$	$\overline{m}$		σ
d1	tour de tête sans nez	50.3 cm		2%		-	-		-
d2	pente des épaules	19.5°		4%		-	-		-
d3	surface du nez	3.6 cm^2		9%		-	-		-
d4	largeur du cou	9.6 cm		4%		9.5 cm	10.5 cm		6%
d5	profondeur du cou	8.7 cm	8.7 cm	11%	10%	-	9.7 cm		-
d6	largeur du torse	32.7 cm		11%		-	-		-
d7	"altitude" des oreilles	8.9 cm		7%		-	-		-
d8	"altitude" du menton	1.4 cm		80%		-	-		-
d9	diamètre minimal de la tête	13.4 cm		0%		14.1 cm	13.9 cm		3%
d10	longueur du pavillon	6.1 cm	6.0 cm	0%	1%	6.1 cm	5.9 cm	6.1 cm	1%
d11	largeur du pavillon	3.3 cm	3.2 cm	1%	0%	3.6 cm	2.9 cm	2.9 cm	9%
d12	longueur de la conque	2.5 cm	2.5 cm	4%	8%	2.5 cm	2.4 cm	2.5 cm	1%
d13	largeur de la conque	1.6 cm	1.6 cm	9%	0%	1.7 cm	1.6 cm	1.9 cm	5%
d14	profondeur de la conque	1.2 cm	1.2 cm	3%	0%	-	1.0 cm	0.8 cm	20%
d15	décollement du pavillon	1.9 cm	2.1 cm	4%	0%	-	-	-	-
d16	quart de contour facial	12.7 cm	12.3 cm	6%	5%	-	-	-	-
d17	inclinaison du pavillon	26.5°	25.2°	3%	10%	22°	31.4°	27.1°	-
d18	hauteur de la tête	-	-	-	-	20.8 cm	18.9 cm		7%
d19	profondeur de la tête	-	-	-	-	19.0 cm	19.3 cm		1%

TAB. 5.3 – Paramètres morphologiques mesurés sur le même sujet lors de trois sessions de mesure . Le cas échéant, les paramètres mesurés sur chacun des deux profils sont reportés.

différente, puisqu'il font passer ce deuxième axe par un point du pli intérieur du pavillon. Les largeurs sont définies par les perpendiculaires aux longueurs tangentes au pavillon.
- Pour la conque, on définit la longueur comme la ligne joignant les points a et b de la Figure 5.14. Le support de la largeur s'en déduit en prenant la perpendiculaire à la longueur passant par le point supérieur du cartillage c. L'autre extrémité du segment est donnée par l'extrémité de la conque. La profondeur de la conque est mesurée par une tige comme le segment entre le point le plus profond de la conque et la surface rasante du pavillon.
- Le quart de contour facial est défini comme la longueur du contour joignant l'entrée du conduit auditif et le centre de la base du nez.
- L'inclinaison du pavillon est donnée par l'angle entre la verticale et l'axe de la longueur du pavillon.
- La hauteur de la tête est donnée par la distance entre la base du menton (tête "droite") et le sommet du crâne, ce dernier pouvant être difficile à repérer du fait de la présence des cheveux.
- Nous définissons la profondeur de la tête comme la distance entre le sommet du nez et l'arrière de la tête, les deux points formant une ligne horizontale.

1. Validation de notre protocole de mesures
   Pour la session de 1999, notre première expérience de relevé morphologique, nous avons procédé à un test de la reproductibilité de la mesure, afin d'éliminer les paramètres mesurés avec trop d'incertitude. Ce premier test a été réalisé sur l'un des sujets, mesuré à deux reprises. L'écart-type en proportion de la valeur moyenne est présenté dans le Tableau 5.3, et nous constatons le manque de fiabilité de notre protocole pour :

   (a) la profondeur du cou et l'"altitude du menton" : les photos à partir desquelles sont estimés ces paramètres montrent que la recommandation faite au sujet de "tenir sa tête droite" est interprétée de façon équivoque. Ce résultat nous a conduit à construire un dispositif "stabilisateur de tête"[4] pour la session de mesure 2000.

   (b) la largeur du torse : le flou introduit par les vêtements nuit à l'estimation précise du décrochement des épaules.

   (c) l'inclinaison du pavillon : l'axe de la longueur de la conque est déterminée de façon assez imprécise sur les photos contenant le profil entier.

		Larcher 99		Larcher 00		Burandt 91	Tilley 93	
		hommes	*femmes*	*hommes*	*femmes*	*hommes*	*hommes*	*femmes*
d1	tour de tête sans nez	52.1 cm	51.5 cm	-	-	-	-	-
d2	pente des épaules	24°$^{\diamond}$	19°$^{\diamond}$	-	-	-	-	-
d3	surface du nez	5.6 cm$^{2\diamond}$	3.4 cm$^{2\diamond}$	-	-	-	-	-
d4	largeur du cou	10.4 cm	9.8 cm	10.8 cm	9.8 cm	11.7 cm	11.7 cm	10.9 cm
d7	"altitude" des oreilles	10.3 cm	9.5 cm	-	-	-	9.6 cm	9.6 cm
d9	diamètre minimal de la tête	13.3 cm	13.5 cm	14.3 cm	13.2 cm	15.5 cm	14 cm	13 cm
d10	longueur du pavillon	6.5 cm$^{\diamond}$	5.8 cm$^{\diamond}$	6.4 cm	6.1 cm	-	6.4 cm	5.8 cm
d11	largeur du pavillon	3.8 cm$^{\diamond}$	3.4 cm$^{\diamond}$	3.4 cm	3.7 cm	-	3.6 cm	3.3 cm
d12	longueur de la conque	2.8 cm$^{\diamond}$	2.5 cm$^{\diamond}$	2.7 cm	2.8 cm	-	-	-
d13	largeur de la conque	2.0 cm	1.8 cm	1.5 cm	1.7 cm	-	-	-
d14	profondeur de la conque	1.3 cm$^{\diamond}$	1.1 cm$^{\diamond}$	-	-	-	-	-
d15	décollement du pavillon	2.1 cm	1.9 cm	-	-	-	-	-
d16	quart de contour facial	12.2 cm	12.1 cm	-	-	-	-	-
d18	hauteur de la tête	-	-	20.7 cm$^{\diamond}$	20.2 cm$^{\diamond}$	23.1 cm	22.1 cm	21.8 cm
d19	profondeur de la tête	-	-	19.7 cm$^{\diamond}$	18.5 cm$^{\diamond}$	20.4 cm	19.6 cm	18 cm

TAB. 5.4 – Ventilation par sexe des paramètres morphologiques moyen .

La meilleure fiabilité est obtenue pour la mesure des dimensions du pavillon et pour le diamètre minimale de la tête. Pour tous les paramètres dont la mesure est jugée robuste, nous observons que les différences droite-gauche sont du même ordre de grandeur ou inférieure à l'erreur de mesure. Par conséquent, pour la suite de l'analyse, nous retirons les paramètres d_5, d_6, d_8 et d_{17}. En outre, pour les paramètres robustes, nous considérons leur moyenne sur les deux séances de mesure et le cas échéant sur les deux profils.

Le même sujet a également été mesuré lors de notre session 2000, ainsi que par Algazi et al.. La comparaison des différentes valeurs obtenues pour les mêmes paramètres montrent une bonne cohérence globale, illustrée par le faible écart-type (Tableau 5.2). C'est le cas de :

(a) la largeur du cou,

(b) le diamètre minimal de la tête,

(c) la longueur du pavillon,

(d) la longueur et la largeur de la conque,

(e) la profondeur de la tête.

Ce sont les paramètres pour lesquels nous pourrons "mélanger" les sessions de mesure dans la section 5.3. En revanche, l'écart-type est non négligeable pour la mesure de la hauteur de la tête, ce qui traduit sans doute la difficulté d'estimation de l'épaisseur des cheveux à partir de photos (protocole Algazi). Le mauvais résultat obtenu pour la mesure de la profondeur de la conque et la largeur du pavillon s'explique par la différence entre les points de mesure choisis entre les deux équipes, mais aussi entre nos deux séances de mesure.

5.2.3.2 Analyse des différences inter-individuelles observées

Connaissant la consistence de nos données, nous essayons de dégager quelques caractéristiques de la population observée.

1. Différences hommes-femmes observées
   Burandt et al. prétendent qu'il est illusoire de vouloir spécifier une tête unique pour représenter l'ensemble de la population, et qu'il est au contraire nécessaire de diviser celles-ci en clusters afin

[4]Ce dispositif a été conçu par Alain Terrier, de l'atelier mécanique de l'Ircam, et réduit également les riques de défaut d'orientation des sujets.

		Larcher 99		Larcher 00		Algazi 99 (17 têtes)		Algazi 99 (80 têtes)	
		$\overline{m}$	σ	$\overline{m}$	σ	$\overline{m}$	σ	$\overline{m}$	σ
d1	tour de tête sans nez	52.0 cm	6%	-	-	-	-	-	-
d2	pente des épaules	22.5°	24%	-	-	-	-	-	-
d3	surface du nez	5 cm^2	20%	-	-	-	-	-	-
d4	largeur du cou	10.3 cm	10%	10.4 cm	8%	11.7 cm	10%	11.6 cm	10%
d7	"altitude" des oreilles	10.1 cm	11%	-	-	-	-	-	-
d9	diamètre minimal de la tête	13.4 cm	5%	14.0 cm	5%	14.8 cm	6%	14.4 cm	7%
d10	longueur du pavillon	6.3 cm	8%	6.3 cm	9%	6.7 cm	8%	6.3 cm	9%
d11	largeur du pavillon	3.7 cm	12%	3.5 cm	11%	3.0 cm	8%	2.9 cm	9%
d12	longueur de la conque	2.7 cm	9%	2.7 cm	13%	2.6 cm	6%	2.6 cm	8%
d13	largeur de la conque	1.9 cm	18%	1.6 cm	15%	1.6 cm	17%	1.5 cm	19%
d14	profondeur de la conque	1.2 cm	14%	-	-	1.1 cm	16%	1.0 cm	16%
d15	décollement du pavillon	2.0 cm	17%	-	-	-	-	-	-
d16	quart de contour facial	12.1 cm	6%	-	-	-	-	-	-
d18	hauteur de la tête	-	-	20.6 cm	5%	21.4 cm	7%	21.5 cm	6%
d19	profondeur de la tête	-	-	19.3 cm	4%	20.1 cm	6%	20.0 cm	6%

TAB. 5.5 – Comparaison des paramètres morphologiques moyens : pour nos deux séances de mesure, pour celle d'Algazi et al avec toutes les têtes ou seulement les 17 têtes utilisée dans le test du chapitre 2.

d'atteindre une représentativité suffisante de la tête "barycentre". Nous nous interrongeons donc sur la pertinence de réaliser une distinction hommes-femmes à partir des mesures dont nous disposons. La significativité des différences observées est analysée à l'aide d'un test de Wilcoxon, tel que nous l'avons décrit en chapitre 4, et les paramètres significativement différents (confiance à 5%) sont indiqués par les losanges ◇ dans le Tableau 5.4. On peut remarquer que pour aucune de nos deux sessions, le diamètre minimal de la tête (ou "largeur" de la tête), la largeur du cou, le tour de tête et l'altitude des oreilles, ne permettent de distinguer hommes et femmes. A contrario, on peut penser que si les différences observées pour les dimensions du pavillon sont non significatives pour la session 2000, alors qu'elles le sont pour la session 1999, c'est avant tout lié au faible nombre de sujet féminins de la deuxième session. On peut donc penser que les paramètres de l'oreille externe, tout comme hauteur et profondeur de tête sont significativement différents entre les hommes et les femmes. Par conséquent, construire une tête artificielle spécifique pour les hommes d'une part et pour les femmes de l'autre semble faire sens.
Les valeurs moyennes obtenues peuvent être comparées à celles de l'étude de Tilley menée sur plusieurs milliers de personnes ([Til93]) ainsi que de celle de Burandt et al. menée sur une cinquantaine de sujets ([BPA+91]). On peut remarquer que nous semblons sous-évaluer la hauteur de la tête ainsi que la largeur du cou, phénomène "de groupe" que nous n'observions pas sur un sujet isolé (cf Tableau 5.3). Nous attribuons cet écart à un manque de représentativité de l'échantillon que nous observons, qui en moyenne semble avoir une tête moins haute et un cou moins large.

2. Représentativité de notre échantillon de population
Afin d'affiner la conclusion tirée au paragraphe précédent, nous nous interrogeons sur la variance observée sur les paramètres dans les campagnes de mesure de grande ampleur. Si notre échantillon de population est représentatif, la variance que nous obtenons doit être comparable. Nous dressons dans le Tableau 5.5 une synthèse des valeurs moyennes et des variances des paramètres communs à plusieurs campagnes de mesure : celle d'Algazi et al. ainsi que nos deux sessions de mesure.
D'une manière générale, on observe que les paramètres varient plus entre individus (données du Tableau 5.5) qu'entre deux mesures du même individu (données du Tableau 5.3) : les écart-type passent le plus souvent du simple au double. Par conséquent, nos données morphologiques contiennent effectivement des informations sur les différences interindividuelles.
Nos écart-types sont plutôt inférieurs à ceux obtenus par Algazi, ce qui souligne une représentativité moins complète de la population que nous avons mesurée. Cela tient peut-être au nombre réduit de sujets mesurés (resp. 20 et 15 contre 80 pour Algazi). Cette sous-représentativité est moins nette pour les petites dimensions (celles du pavillon).

3. Corrélation entre les paramètres morphologiques mesurés.
   Nous souhaitons pouvoir décrire une tête à l'aide d'un nombre minimal de paramètres. Aussi est-il important de mesurer la corrélation entre les caractéristiques mesurées, afin d'éliminer celles qui n'apportent pas d'information.
   L'analyse des 23 paramètres mesurés par Algazi et al. sur 80 sujets montre une très forte corrélation entre :
   (a) largeur du cou et largeur des épaules ($r = 0.9$),
   (b) largeur du torse et largeur des épaules ($r = 0.86$),
   (c) largeur du cou et profondeur du cou ($r = 0.84$),
   (d) orientation du pavillon et décollement du pavillon ($r = 0.8$).

   Ces résultats permettent de penser que largeur du torse, des épaules et profondeur du cou peuvent être représentées par la seule mesure de la largeur du cou, mesure qui des 4 semble pouvoir être mesurée avec la plus grande fiabilité.
   En ce qui concerne les sessions de mesure Ircam, nous observons en général de plus faibles corrélations que pour les paramètres précédents, ce qui s'explique par le simple fait que nous avons mesuré moins de paramètres qu'Algazi et al. et avons ainsi réduit le risque de redondance au prix d'une description moins précise de la morphologie des sujets. Toutefois, deux corrélations non négligeables sont observées :
   (a) largeur du pavillon et largeur de la conque ($r = 0.79$ pour la session 99 et $r = 0.6$ pour la session 2000),
   (b) longueur du pavillon et longueur de la conque ($r = 0.6$ pour la session 99 et $r = 0.73$ pour la session 2000).

5.2.3.3 Conclusion

La mesure des paramètres morphologiques est un élément encore mal maîtrisé. Pour nos deux sessions de mesures, nous avons identifié les paramètres réellement porteurs d'information et cohérents avec les relevés pratiqués sur une tête commune par Algazi et al. Il s'agit de :

- la largeur du cou,
- la largeur de la tête,
- la profondeur de la tête,
- la longueur du pavillon,
- la longueur de la conque,
- la largeur de la conque.

Pour ces paramètres, il nous sera possible de "mélanger" les données mesurées lors de notre session 2000 et celles d'Algazi et al, pour l'adaptation discrète de la section 5.3. Toutefois, ces paramètres ne sont pas suffisant pour décrire une tête. Partant de l'étude exhaustive d'Algazi et al., nous retenons 19 paramètres, que, par souci de compatibilité avec les documents fournis par celui-ci, nous désignerons par la suite par :

x_1	largeur de la tête	x_2	hauteur de la tête
x_3	profondeur de la tête	x_4	décalage de l'oreille vers le bas
x_5	décalage de l'oreille vers le bas	x_6	largeur du cou
x_7	hauteur du cou	x_{10}	altitude du sommet du torse
x_{11}	épaisseur du torse	x_{13}	décalage de la tête vers l'avant
$d_1 + d_2$	longueur de la conque	d_3	largeur de la conque
d_4	hauteur de la fossa	d_5	longueur du pavillon
d_6	largeur du pavillon	d_7	*"intertragal incisure width"*
d_8	profondeur de la conque		
θ_1	inclinaison du pavillon	θ_2	décollement du pavillon

Sur ces différents paramètres, on constate de fortes disparités entre individus : les grandes dimensions de la tête présentent une variance d'environ 10% de leur valeur moyenne, tandis que la conque, ces écarts

sont plus près de 20%. Les plus fortes variances sont obtenues pour les paramètres de décalage (jusqu'à 154% pour le paramètre x_5).
Ces données permettent de définir un espace de représentation révélateur de la morphologie des têtes. Seules des opérations de corrélation avec les données perceptives ou signal peuvent nous renseigner de la pertinence de cet espace du point de vue de la localisation.

5.2.4 Dépendance inter-individuelle des jugements perceptifs

Dans cette section, nous mettons en évidence l'importance perceptive des différences inter-têtes observées objectivement dans les sections précédentes. Cela permet de montrer que le choix d'une base de données de HRTF est un facteur de qualité significatif pour la localisation en synthèse binaurale. En outre, nous observons également que le choix de la base de données optimale est différente pour chaque sujet, ou du moins pour chaque famille de sujets. Nous obtenons donc un espace de représentation des jugements perceptifs des sujets. Ces résultats sont obtenus à partir des données du test perceptif présenté au chapitre 4, pour lequel 22 sujets ont répondu à une tâche de localisation menée avec les HRTF de 17 têtes. Nous ne chercherons pas à superposer ces espaces perceptifs avec les domaines signal et morphogique.

5.2.4.1 Représentation des têtes dans un espace subjectif

Nous recherchons tout d'abord un espace de représentation des têtes "partagé par tous les sujets", ou plus exactement obtenu en moyennant les réponses des sujets. Celles-ci font apparaître des sensations significativement différentes pour 5 critères de localisation :

1. taux de non-localisation,
2. taux de confusions avant-arrière,
3. robustesse de localisation en azimut,
4. biais de localisation en azimut (plan horizontal),
5. biais de localisation en élévation.

Deux autres critères, le taux de confusion haut-bas et la robustesse de localisation en élévation, n'ont pas permis de départager ces têtes.

Puisqu'ils sont discriminants, ces 5 critères mettent en évidence le caractère audible des différences entre les têtes. Pour chacun d'entre eux, on peut déterminer la tête optimale et la tête critique, obtenant respectivement le meilleur ou le plus mauvais score (cf Figure 5.15) :

- la tête 055 offre le meilleur taux de localisation. C'est également l'une des têtes à plus fort ITD.
- les têtes 028 et 061 s'opposent sur le critère du taux de confusions avant-arrière, tout comme elles s'opposaient sur leur distance inter-HRTF pour l'intervalle de fréquence BF, et sur leur distance inter-ITD. La tête 061 est également l'une des têtes à plus petit ITD.
- la tête 028 est aussi celle qui fournit la meilleure localisation en azimut , i.e. le plus faible biais et le plus faible écart-type.
- la tête 035, qui est la tête barycentre du nuage de têtes pour l'espace inter-HRTF HF, est celle qui engendre le plus faible biais en élévation.

Ces écarts montrent qu'il existe une distance perceptive entre les têtes, dont nous n'apercevons qu'une projection sur chaque critère. Il nous serait pourtant utile de définir un espace de représentation global, les combinant : pour une implantation de la synthèse binaurale n'utilisant qu'une seule base de données de HRTF, il serait en effet sous optimal de choisir une tête au hasard, et il convient plutôt de prendre celle réalisant le meilleur compromis global. Dans cet objectif, nous souhaiterions déterminer une loi de combinaison de nos critères. Connaissant les coordonnées des têtes sur les 5 axes définis par nos critères, ce résultat pourrait être obtenu à l'aide d'une analyse en composantes principales. Toutefois, cette combinaison serait alors linéaire, alors que de par nos définitions, les critères interagissent de façon multiplicative : un très fort taux de non localisation ne doit être compensé par une bonne localisation en azimut.
Sans proposer de solution pour cette combinaison, nous pouvons supposer qu'en fonction de l'application, la "meilleure tête" peut être choisie comme celle obtenant le meilleur score sur le critère prépondérant.

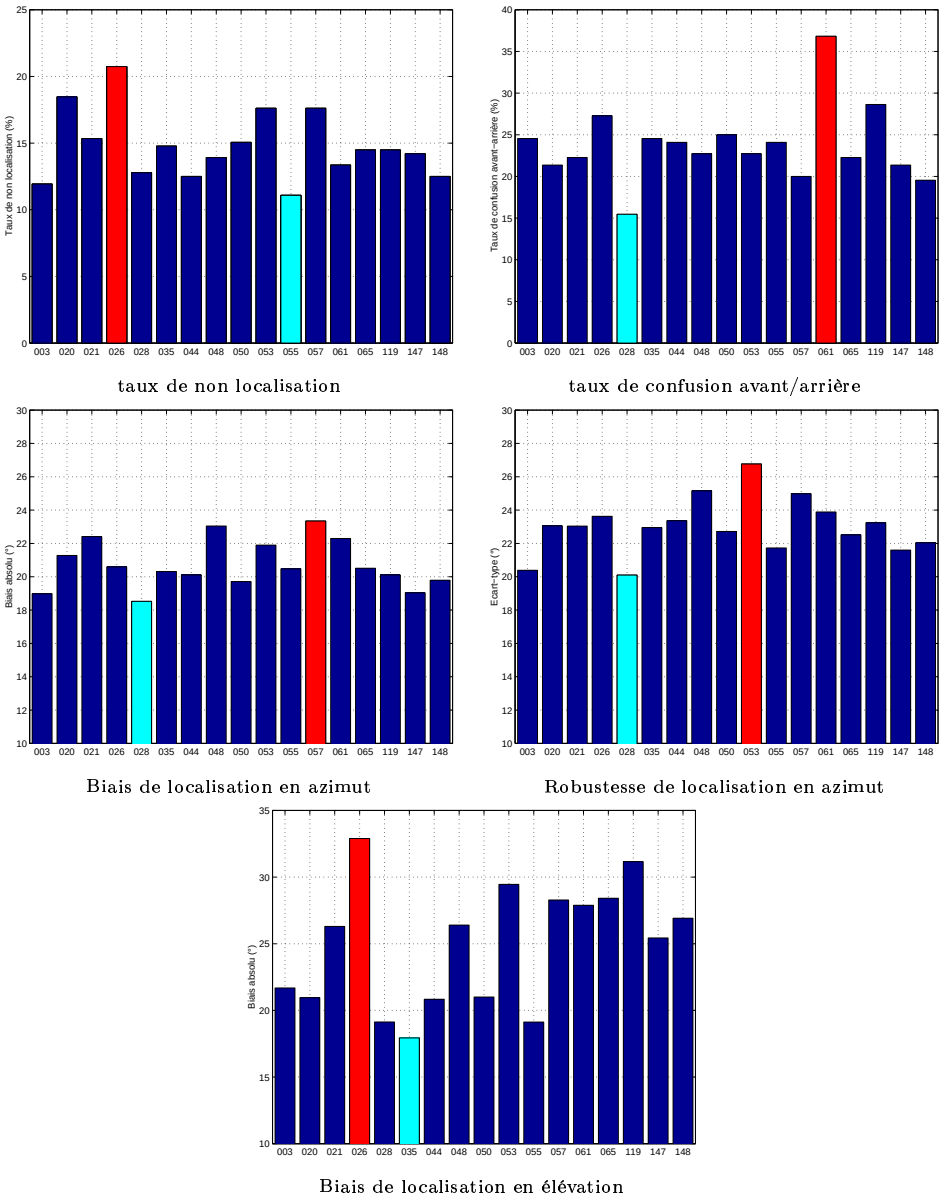


FIG. 5.15 – Critères subjectifs présentant une différence significative entre nos têtes : tête la plus performante (cyan) et la moins performante (rouge).

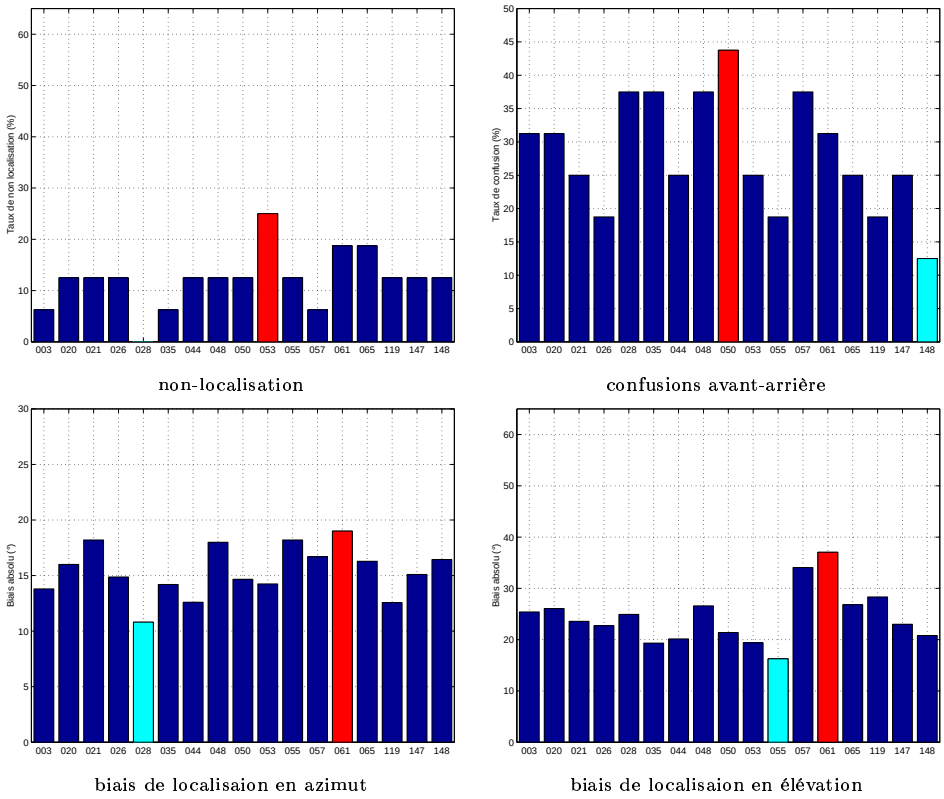


FIG. 5.16 – Performances de localisation obtenues par le sujet VL, dont la tête appartient à l'échantillon testé (tête 148).

Ainsi, pour une oeuvre musicale enregistrée en binaural, par exemple, on peut souhaiter que tous les auditeurs partagent la même expérience de l'oeuvre. Ainsi, c'est une tête présentant la plus forte robustesse de localisation que l'on privilégiera. En revanche, pour un simulateur de vol, où les signaux d'alarme doivent permettre d'éveiller l'attention du pilote vers un danger, les confusions avant-arrière constituent un artefact majeur. On choisira donc la tête garantissant le plus faible taux.

5.2.4.2 Représentation des sujets dans un espace perceptif

Quand bien même il est possible de dégager une tête obtenant les meilleurs suffrages en moyenne sur l'ensemble des sujets, ce n'est forcément celle qui convient le mieux à chaque sujet pris indépendemment. En effet, plusieurs auteurs ont montré que les performances de localisation pouvaient tirer profit du choix individuel de la "meilleure" tête. Cette meilleure tête est le plus souvent la tête du sujet, mais peut également être celle d'un "bon localisateur", avec lesquelles les indices de localisation sont accentués.

Nous pouvons tout d'abord observer les préférences exprimées par un sujet, le sujet VL, dont la tête appartient au groupe de têtes comparées (tête 148). Comme on le constate en Figure 5.16, c'est avec sa tête que le sujet VL commet le moins de confusions avant-arrière. En outre, le sujet fait figurer sa tête dans la bonne moyenne d'ensemble pour le biais en élévation et le taux de non-localisation. Sur ce dernier critère, il lui préfère notamment la tête 028, dont on a déjà noté les performances globales à l'ensemble des sujets. C'est également la tête avec laquelle le sujet VL localise dans le plan horizontal avec le plus faible biais, peut-être du fait du fait du fort ITD de la tête 028, supérieur à celui de la tête 148. Pour être généralisés, ces résultats nécessiteraient bien sûr d'être confrontés à d'autres cas d'écoute individuelle, conditions qu'il n'a pas été possible de réunir dans le cadre du test perceptif décrit au chapitre 4.

Il est également possible de construire un espace de représentation des sujets : chacun d'entre eux est représenté par ses coordonnées sur les 17 têtes ; ces coordonnées sont successivement le taux de non localisation, le taux de confusions avant-arrière, le biais de localisation en azimut et le biais de localisation

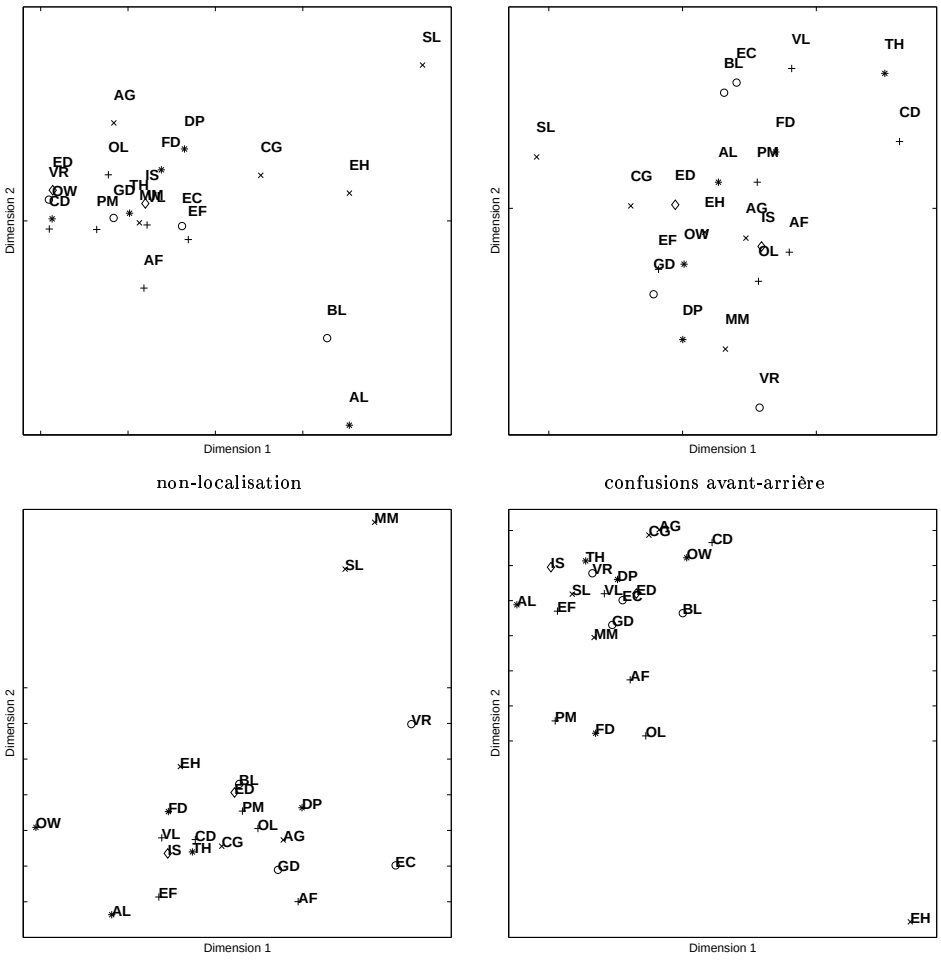


FIG. 5.17 – Espace perceptif de représentation des sujets, pour 4 critères de localisation.

en élévation. Les critères de robustesse, utilisés à la section précédente, évaluaient la variabilité des sensations entre sujets, et n'ont donc plus de fondement dans le cadre d'une analyse individuelle.
A l'aide d'une décomposition en valeur singulière[5], nous réduisons la dimension de l'espace de représentation des sujets de 17 à 2. Ces deux axes permettent de reproduire 91% de la variance expliquée pour les critères de non localisation et de confusion, et plus de 96% pour les autres. Puisque nous travaillons sur des données non centrées, le premier axe des 4 espaces, présenté en abscisse, note les performances des sujets en moyenne sur l'ensemble des têtes. Par opposition, l'axe des ordonnées explique toujours les performances relatives obtenues sur certaines têtes par rapport à d'autres. Nous avons représenté en Figure 5.17 la distribution des préférences individuelles obtenue à partir des jugements sur chaque critère.

D'une manière générale, les performances moyennes des sujets varient d'un critère à l'autre. Seuls trois sujets, EH, SL, et OW se distinguent assez systématiquement :

- SL totalise un fort taux de non localisation, et un fort biais en azimut, mais il obtient également un faible taux de confusions avant-arrière (9%) : c'est l'un des sujets ayant le mieux localisé les sons frontaux, notamment avec les têtes 020 et 119 pour lesquelles tous les sons localisés le sont dans le bon cadrant.
- EH présente un fort taux de non localisation et très fort biais en élévation.
- OW apparait comme l'un des sujets les plus performants, puisqu'il obtient parmi les meilleurs scores pour le taux de non localisation (1%) et le biais en azimut (8°). C'est aussi le sujet le plus expérimenté du groupe, habitué aux tâches de localisation, contrairement aux deux sujets précédents. Néanmoins,

[5] Nous ne parlons pas d"'Analyse en Composantes Principales" car nous ne centrons pas les données avant de décomposer l'ensemble en valeurs singulières.

nous n'avons pu dégager de résultats plus généraux quant à la ségrégation sujets expérimentés/sujets non expérimentés.

Nous pouvons poursuivre l'analyse en expliquant les axes principaux, non représentés ici. Si l'on se concentre sur l'information du second axe de l'analyse, on constate que pour le taux de non localisation, l'analyse oppose sur l'axe des ordonnées les sujets obtenant de meilleurs résultats avec les têtes 028 et 053 (sujets SL et AG) qu'avec la tête 050 (sujet AL). Pour les confusions avant-arrière, on oppose les bonnes performances obtenues sur les têtes 028 et 035 (sujets VL, TH, EC) aux moins bons résultats obtenues avec la tête 119, et réciproquement pour les sujets à l'autre extrémité de l'axe (VR, MM, DP). Pour le biais en azimut, ce sont les performances obtenues sur la tête 065, bonnes pour le sujet MM et faibles pour le sujet AF, que l'on oppose à celles obtenues avec les têtes 035 et 053. Pour le biais en élévation, enfin, l'axe des ordonnées accorde un poids prépondérant au sujet EH, et exprime le bon score obtenu par ce dernier avec la tête 119, par opposition avec celui qu'il atteint avec la tête 057.
Ces observations montrent l'écart entre les jugements individuels de localisation, et permettent d'apprécier l'apport d'une adaptation de la synthèse binaurale à l'auditeur. C'est sur les méthodes pour réaliser cette adaptation que se concentrent les sections 5.3 et 5.4.

5.2.4.3 Conclusion

Nous avons pu observer des différences inter-individuelles, mesurables objectivement à partir des différence entre HRTF, ITD et paramètres morphologiques de différentes têtes. Nous avons montré dans cette section, à l'aide du test perceptif du chapitre 4, que les différence inter-individuelles influençaient également les performances subjectives de localisation. Bien que ayons dérivé plusieurs mesures de qualité subjective de chaque tête (5 critères de localisation), il semble que la tête 028 présente un compromis global satisfaisant : dans le cas d'une synthèse binaurale sans adaptation individuelle, c'est la tête la plus consensuelle. Cette tête est également sortie du lot avec les distances de type "signal" : tête atypique en BF et MF, tête moyenne en HF. Il est difficile de savoir si les bons résultats perceptifs de cette tête sont expliqués par le caractère atypique ou au contraire consensuel de ses HRTF.
Il est apparu clairement que les sujets expriment des préférences individuelles, et que ce compromis "sans adaptation" n'autorise pas les performances optimales de localisation. Il convient alors de proposer des méthodes pour l'adaptation individuelle de la synthèse binaurale.

5.3 Adaptation discrète de la synthèse binaurale

L'adaptation discrète de la synthèse binaurale consiste à appairer chaque nouvel auditeur avec l'une des têtes dont on possède les HRTF. La simulation n'est alors pas faite avec les HRTF propres de l'auditeur, mais avec des HRTF qui lui conviennent le mieux possible, choisies parmi une base de données limitée. Pour choisir la tête la mieux adaptée, on peut s'appuyer sur un espace de représentation des têtes, dans lequel on place chaque nouvel auditeur. On peut envisager plusieurs espaces de représentation des têtes équivalents : un espace "signal", où les distances entre têtes sont calculées à partir de distance entre HRTF ou entre ITD, un espace perceptif, où les distances entre têtes sont calculées à partir de réponses à un test d'écoute. Nous pourrions également définir un espace morphologique, où les distances entre têtes sont calculées à partir de dimensions corporelles révélatrices de la localisation. Cet espace de représentation présente l'avantage pratique de pouvoir être construit à partir de relevés plus rapides que ceux requis pour les deux espaces précédents. L'adaptation discrète serait ainsi permise par la mesure, éventuellement automatique, de quelques dimensions morphologiques de chaque nouvel auditeur, permettant de le placer dans l'espace de représentation des têtes de la base de données.
L'objectif de cette section est de définir la combinaison linéaire des paramètres morphologiques permettant de représenter les écarts de type signal entre les têtes (relations 1 de la Figure 5.18) afin de dégager un protocole simple et rapide pour l'adaptation discrète de la synthèse binaurale.

5.3.1 Méthode pour superposer deux espaces de représentation

L'analyse MDS menée sur les distances "signal" entre les têtes a mis en évidence 4 axes. De même, la modélisation de l'ITD donne d'autres "coordonnées des têtes", un axe pour le modèle sphérique pur, deux

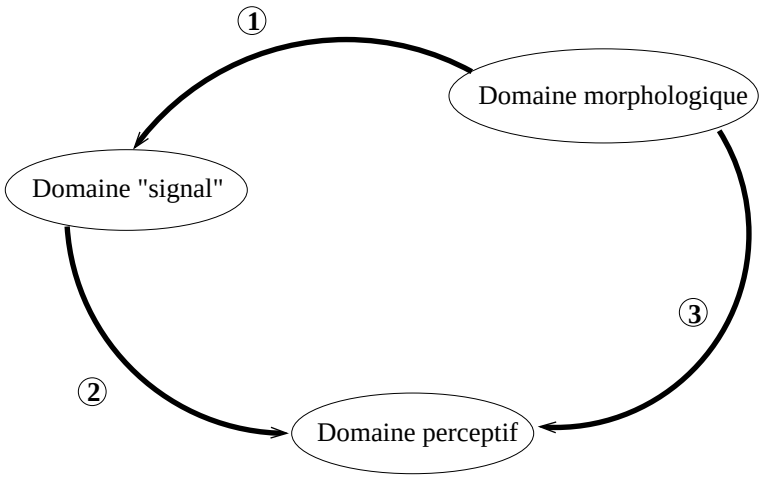


FIG. 5.18 – Relations à établir entre les domaines de représentation des têtes.

axes pour le modèle étendu (r_1,r_2). Nous cherchons à approximer la distribution des têtes sur chacun de ces axes par une combinaison linéaire des paramètres morphologiques. La combinaison minimisant l'erreur aux moindres carrés peut être obtenue par projection orthogonale de l'espace engendré par les variables "signal" sur l'espace engendré par les variables morphologiques. Cette méthode simple que nous décrivons peut être rapprochée d'un cas particulier d'analyse canonique[6] décrit par Saporta : l'analyse en composantes principales de variables instrumentales ([Sap90]).

D'un point de vue formel, nous cherchons donc la meilleure approximation des variables y_i par une combinaison linéaire des caractéristiques morphologiques x_j (+ constante), au sens des moindres carrées. Cela nous conduit à chercher la projection orthogonale des y_i sur l'espace engendré par les x_j , vérifiant :

$$\forall i \quad y_i = \alpha_i^0 + \sum_{j=1}^{p} \alpha_i^j . x_j + \epsilon_i$$

avec :

$$[\epsilon_i | x_j] = 0$$

Cela nous conduit à une estimation aux moindres carrés des α_i :

$$[\alpha_i^0 .....\alpha_i^p]^t = \left(X^t.X\right)^{-1} .X^t.y_i$$

où X est la matrice des paramètres morphologiques (et une colonne constante) $[1\, x_1 ... x_p]$.
La projection des y_i sur X, $\widetilde{y_i}$est donnée par :

$$\widetilde{y_i} = X.\left(X^t.X\right)^{-1} .X^t.y_i$$

Ce calcul requiert bien sûr que $X^t.X$ soit inversible, donc que X soit de rang plein, ce qui est le cas avec les données que nous manipulons. Dans le cas où les variables seraient liées, il conviendrait d'utiliser une pseudo-inversion.
On peut mesurer l'efficacité de cette approximation en mesurant la distance euclidienne moyenne entre y_i et $\widetilde{y_i}$:

$$s = \frac{1}{N}.\sum_{i=1}^{N} (y_i - \widetilde{y_i})^2$$

[6]L'analyse canonique est une technique permettant d'examiner les liens entre deux ensembles de variables décrivant les mêmes n individus. Si les espaces engendrés par chacun des deux ensembles sont confondus, alors les deux groupes de variables mesurent les mêmes propriétés. Si au contraire, les espaces sont orthogonaux, les deux groupes de variables appréhendent des phénomènes totalement différents. Avec l'analyse canonique, il est possible de mettre à jours l'intersection de ces ensembles. Une forme qui en découle, l'analyse en composantes principales de variables instrumentales donne les combinaisons linéaires d'un des ensembles, présentant une redondance maximale avec le second.

où N est le nombre de tête considéré (nombre d'éléments du vecteur y_i). On peut également observer le coefficient de corrélation entre y_i et $\widetilde{y_i}$.
Le vecteur de combinaisons linéaires α_i associé à la coordonnée de rang i des têtes, a été optimisé à partir d'un ensemble de têtes que l'on souhaite le plus représentatif possible. Il est alors possible de placer une nouvelle tête k dans cet espace si l'on connait ses caractéristiques morphologiques, contenues dans le vecteur x_k. L'approximation de sa coordonnée de rang i est donnée par :

$$\widetilde{y_i} = x_k^t.\alpha_i$$

5.3.2 Application à l'insertion de nouvelles têtes dans un espace de représentation "prédéfini"

5.3.2.1 superposition des espaces morphologie et "signal"

Nous appliquons la technique d'analyse précédente pour relier les paramètres morphologiques aux paramètres de représentation des têtes fournis par :

1. l'analyse MDS appliquée aux distance inter-HRTF, dont on ne représente que le premier axe qui explique la plus grande part de variance,
2. la modélisation de l'ITD faisant l'approximation d'une tête sphérique, qui fournit un ou plusieurs "rayon équivalents" pour chaque tête.

La technique précédente nous permet d'avoir une superposition parfaite de deux espaces de représentation dès lors que l'on possède davantage de paramètres "expliquants" libres (ici les paramètres morphologiques) que de têtes. Dans ces conditions en effet, le système linéaire à résoudre contient plus d'inconnues (les α_i) que de contraintes (une équation pour chaque tête), et la résolution que nous appliquons, à base de pseudo-inverse, conduit à l'une des solutions du système. Dans le cas qui nous concerne, où l'on dispose de moins d'inconnues que de contraintes, la superposition est optimisée aux moindres carrées. Nous considérons cette étape comme une étape d'analyse puisqu'elle nous permet d'accéder aux relations entre la distance séparant les têtes et les paramètres morphologiques.

Pour illustrer notre approche, nous nous concentrons sur les 6 paramètres communs aux deux sessions de mesure et dont on a vérifié la robustesse en section 5.2.3 :

- largeur de la tête x_1,
- profondeur de la tête x_3,
- largeur du cou x_6,
- longueur du pavillon d_5,
- longueur de la conque $d_1 + d_2$,
- largeur de la conque d_3.

On vérifie que le rang de la matrice X contenant ces données est de rang 6. Nous aboutissons donc à combinaison linéaire des paramètres morphologiques permettant de construire un espace de représentation des têtes approximativement superposable à celui donné par les distances inter-HRTF ou inter-ITD. Les coefficients de pondération associés à chaque paramètre morphologique sont présentés sur les Figures 5.19 et 5.20. Pour pouvoir en comparer le poids respectifs, nous avons standardisé les paramètres morphologiques avant l'analyse.
Les combinaisons varient fortement d'un espace à l'autre. On peut toutefois noter la forte prépondérance de la largeur de la tête pour la reconstruction de l'espace ITD à un radius, ou encore de la longueur de la conque pour l'espace HF. Les autres espaces s'appuient sur davantages de paramètres, notamment l'espace BF pour lequel tous les paramètres ont une contribution non négligeable.

5.3.2.2 Insertion des sujets du test dans l'espace des têtes

Nous souhaitons insérer de nouvelles têtes, celles des sujets du test perceptif du chapitre 4, au sein des têtes représentées à l'aide de la distance inter-HRTF ou à l'aide de la distance inter-ITD. L'objectif est de

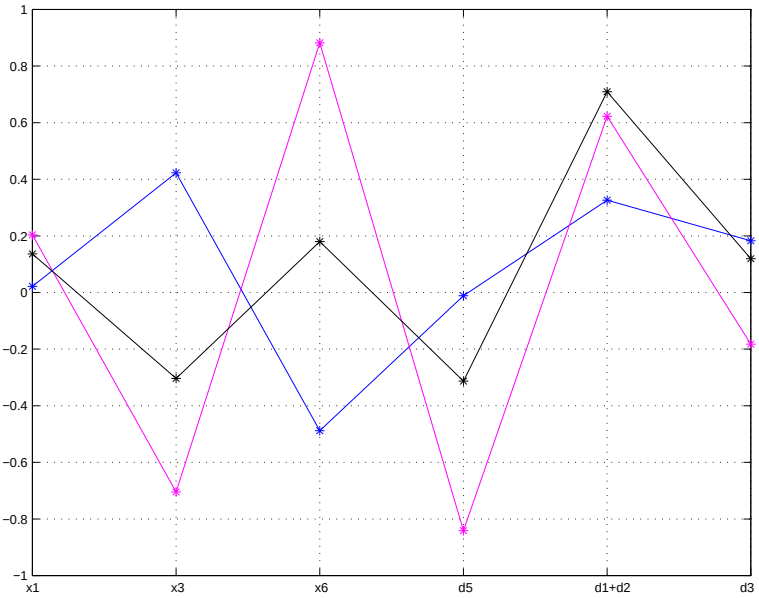


FIG. 5.19 – Poids de chaque paramètre morphologique pour la combinaison linéaire permettant d'approximer l'espace de représentation HRTF des têtes : intervalle BF (magenta), intervalle MF (bleu), intervalle HF (noir).

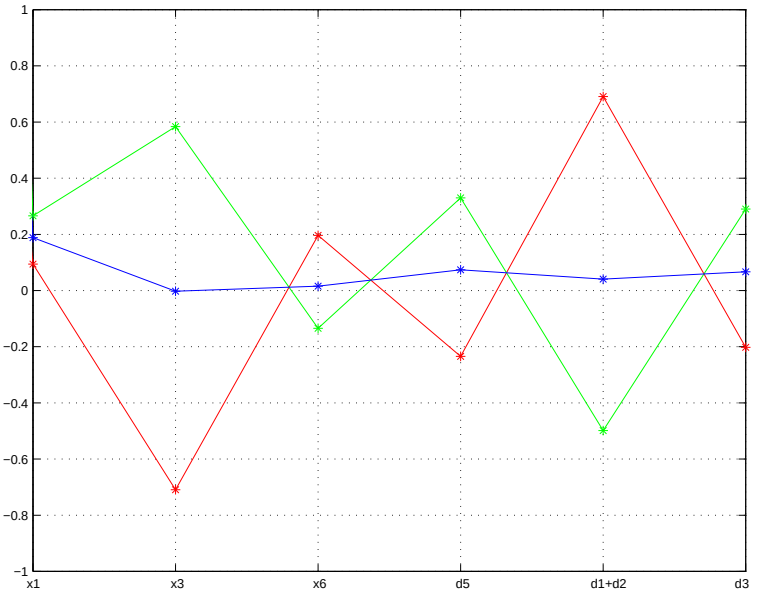


FIG. 5.20 – Poids de chaque paramètre morphologique pour la combinaison linéaire permettant d'approximer l'espace de représentation ITD des têtes : rayon r_1 (rouge), rayon r_2 (vert), rayon sphérique équivalent (bleu).

pouvoir ainsi déterminer la tête la plus "proche" de chaque sujet afin de pouvoir lui attribuer des HRTF ou un ITD appropriés.
Nous faisons l'hypothèse que les lois de combinaison des paramètres morphologiques obtenues précédemment sont "universelles", i.e. que les têtes desquelles elles sont issues sont suffisamment représentatives pour que l'ajout d'une tête supplémentaire ne change pas la combinaison qui en découle. Dans ces conditions, nous pouvons placer toute nouvelle tête par la simple donnée de ses caractéristiques morphologiques, auxquelles nous appliquons la combinaison précédente.

La cohérence de la méthode est soulignée par le fait que :

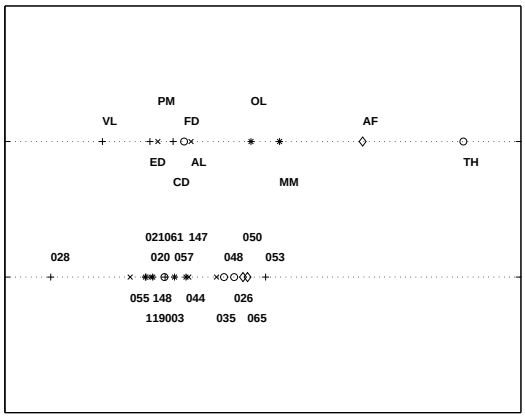
1. Les deux ensembles de têtes sont mélangés, bien que l'observation de l'espace inter-ITD laisse penser que nos têtes sont plutôt plus petites que celles étudiées par Algazi et al. En outre, si les nuages resp. de têtes et de sujets peuvent sembler disjoints pour l'espace inter-HRTF (5.21, (a)), on constate en (b) que les têtes extrêmes y sont pour beaucoup.
2. Ce sont des sujets de sexe féminin ('AF', 'MM', 'CD') qui présentent un "rayon équivalent" le plus faible,
3. Le sujet 'VL' est proche de sa tête, la tête 148 : l'écart entre les deux rayons sphériques équivalents est de l'ordre de 2.5mm, la superposition est aussi satisfaisante avec l'espace de représentation r_1/r_2, de même que pour les espaces inter-HRTF pour les bandes fréquentielles MF et HF. En BF, les têtes les plus proches sont plutôt les têtes 028 et 055, dont on a vu qu'elles lui permettait la meilleure localisation en azimut et en élévation.

5.3.3 Conclusion sur l'adaptation discrète

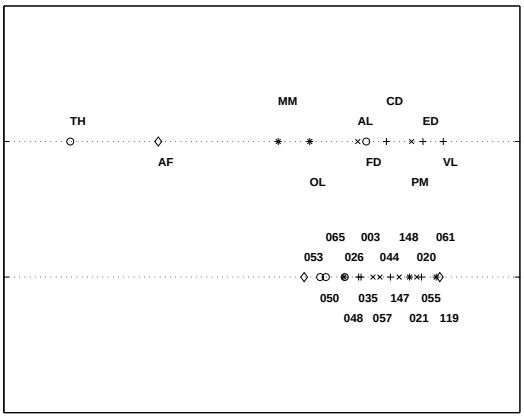
Nous avons proposé une méthode pour insérer un nouvel auditeur dans un espace de représentation de têtes, dans l'objectif de déterminer la tête qui lui convient le mieux. L'approche adoptée permet de s'affranchir de la mesure (longue) des paramètres de l'espace des têtes (HRTF, ITD) et s'appuie plutôt sur la mesure de paramètres rapidement accessible, 6 paramètres morphologiques dans notre étude. Il serait toutefois utile de réfléchir au choix de ces paramètres, éventuellement d'en augmenter le nombre, afin d'obtenir une reproduction fidèle de l'espace de représentation de départ.
Pour son application, notre méthode de superposition des espaces de représentation des têtes s'est appuyée sur des outils mathématiques traditionnels tels que la projection orthogonale. Elle bénéficierait néanmoins de l'utilisation d'outils plus spécifiques : la projection minimise l'erreur au moindres carrés sur la représentation des distances inter-têtes, alors qu'une erreur conservant les relations d'ordre serait plus appropriée.
D'autre part, il reste à appliquer la méthode à l'espace perceptif, qui doit être relié aux domaines morphologiques et au domaine signal, comme on le représente en Figure 5.18. Ces liaisons permettraient d'une part de définir un protocole de sélection de la "tête-soeur" à partir réponses à un test de localisation, qu'il faudrait, pour des raisons pratiques, rendre plus simple que celui que nous avons mené au chapitre 4. D'autre part, au delà de l'objectif d'adaptation individuelle, la liaison des espaces perceptifs et morphologiques permettrait de mettre en évidence les caractéristiques morphologiques prépondérantes pour la localisation. Cette sélection pourrait alors se greffer sur un modèle physique des HRTF, assemblant la contribution fréquentielle de ces caractéristiques pour définir des HRTF contenant un minimum de détail pour une localisation satisfaisante. Ce type de modèle a été réalisé par Genuit, qui a limité son investigation à quelques caractéristiques sans en motiver le choix par un critère objectif ([Gen84]).

5.4 Adaptation continue de la synthèse binaurale

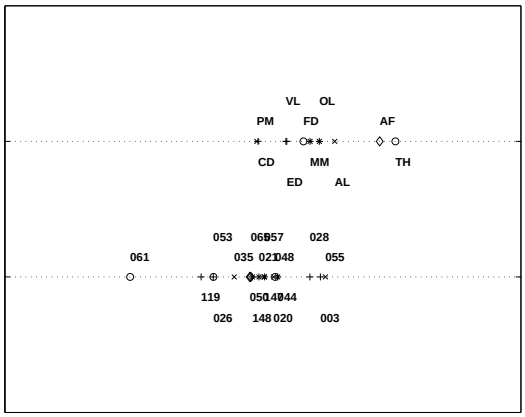
L'objectif de l'adaptation continue est de synthétiser les HRTF de l'auditeur final sans avoir à les mesurer. Cet objectif est plus satisfaisant qu'une adaptation discrète car il doit théoriquement conduire à une plus grande fidélité. En outre, l'adaptation continue permet de s'affranchir du temps consacré la constitution des bases de données et d'économiser la place en mémoire qui seraient requis pour l'implantation d'une adaptation discrète.
La technique de scaling fréquentiel que nous étudions est une possibilité pour l'adaptation continue. Elle



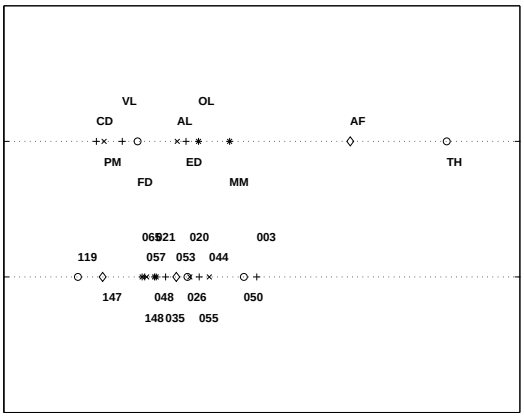
distance BF



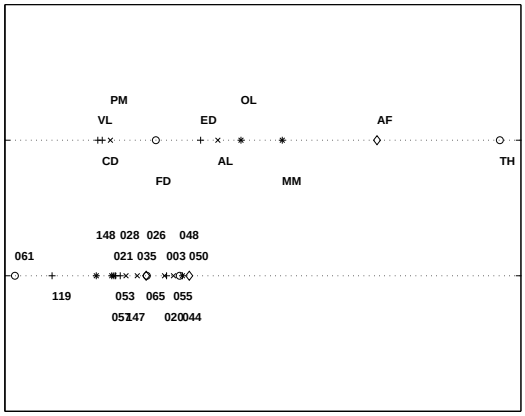
distance BF sans tête 028



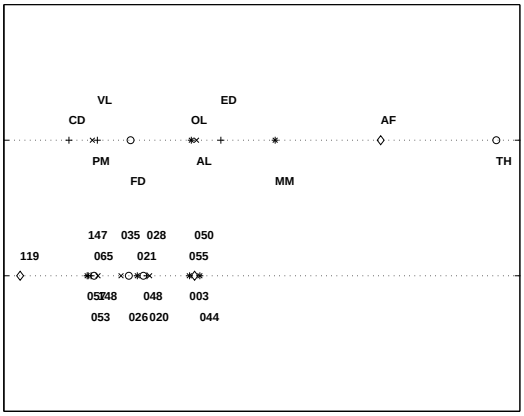
distance MF



distance MF sans tête 028 ni 061



distance HF



distance HF sans tête 061

FIG. 5.21 – Positionnement des sujets du test perceptif du chapitre 4 dans l'espace des têtes défini par leur distance inter-HRTF.

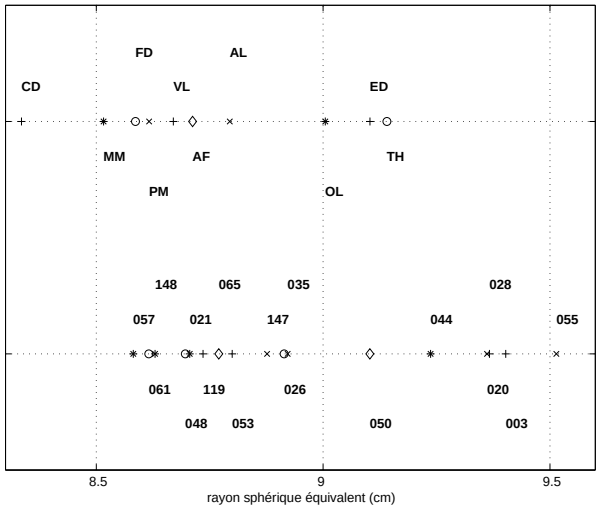


FIG. 5.22 – Positionnement des sujets du test perceptif du chapitre 2 dans l'espace de représentation "ITD" des têtes : rayon sphérique équivalent.

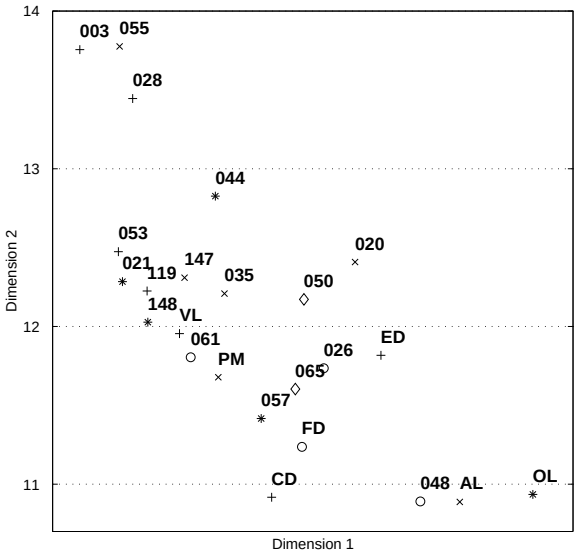


FIG. 5.23 – Positionnement des sujets du test perceptif du chapitre 2 dans l'espace de représentation "ITD" des têtes : modèle de l'ITD à deux paramètres.

n° de mode	*Mode 1*	*Mode 2*	*Mode 3*	*Mode 4*	*Mode 5*	*Mode 6*
fréquence centrale du pic	4.2kHz	7.1kHz	9.6kHz	12.1kHz	14.4kHz	16.7kHz
incidence privilégiée d'excitation	omni	el. 68°	el. 73°	el. n -6°	el. 8°	el. 7°

TAB. 5.6 – Modes de résonance de la conque, d'après Shaw ([Sha80], [Sha82], [Sha97b], [Sha97a]).

s'appuie sur les relations "physiques" entre paramètres morphologiques et caractéristiques des HRTF développées par exemple par E.A.G. Shaw. Middlebrooks en a proposé une mise en place simple ([Mid99a]), et a constaté l'efficacité de l'adaptation réalisée à l'aide d'un test perceptif ([Mid99b], [MZ00]).
Dans cette section, nous reprenons l'approche de Middlebrooks en l'appliquant à nos données, puis nous en proposons quelques extensions.

5.4.1 Principe de la méthode de scaling fréquentiel

5.4.1.1 Interprétation physique des HRTF

En 1885, Mach présentait déjà l'oreille comme un résonateur acoustique ([ref59]). Approfondissant cette approche, Shaw a identifié six modes de résonances de la conque, élément de l'oreille dont les dimensions permettraient d'expliquer les caractéristiques spectrales des HRTF au delà de 5kHz ([Sha80], [Sha82], [Sha97b], [Sha97a]). Selon l'hypothèse de Shaw, l'oreille serait ainsi un réseau de résonateurs en parallèle, dont les supports fréquentiels ne sont pas nettement disjoints, mais entrent en action pour des incidences spécifiques. Chacun de ces modes est représenté par une onde stationnaire au sein des cavités de la conque, qui crée une surpression à l'entrée du conduit auditif, et engendre à ce titre un maximum dans le spectre. Cette zone d'excitation privilégiée ainsi que la fréquence centrale du pic qu'ils engendrent sont rappelées en tableau 5.6. Le premier mode, créant un maximum à 4.2kHz, est actif pour toutes les incidences. Les modes 2 et 3 correspondent à des zones fortement élevées, et laissent penser, comme les bandes directionnelles de Blauert, que la sensation d'élévation s'accompagne d'une forte énergie dans le spectre autour de 7 à 10kHz. Les modes 4, 5 et 6 quant à eux sont excités pour des incidences frontales.

On peut remarquer qu'alors que l'approche de Shaw cherche à interpréter les "bosses" présentes dans les HRTF, celle de Batteau s'attache à en expliquer les vallées, comme effet d'interférence destructive entre les ondes réfléchies sur les parois de l'oreille externe ([Bat67]).

Le modèle par résonateurs de l'oreille permet de concevoir le lien entre les dimensions des paramètres morphologiques et la fréquence des pics des HRTF : si les dimensions de la conque, ou plus généralement de l'oreille, sont augmentées de 10%, les caractéristiques de résonances de toutes les cavités vont être modifiées dans la même proportion, et se traduisent par une translation des pics des HRTF vers les basses fréquences, transformation correspondant à une homothétie de 10% de l'échelle fréquentielle linéaire. Selon ce modèle physique simple des HRTF, leur adaptation individuelle se ramène donc à rechercher le coefficient d'homothétie approprié, ou facteur de scaling optimal. Sur une échelle de fréquence logarithmique, cette transformation se ramène à une translation.

5.4.1.2 Scaling fréquentiel de Middlebrooks

Middlebrooks applique la technique de scaling en trois étapes :

1. Re-échantillonnage des log-magnitudes afin de donner un poids perceptif équivalent à chaque échantillon fréquentiel. Le re-échantillonnage est réalisé par 35ème d'octave (0.0286).
2. Choix d'un intervalle fréquentiel d'étude.
3. Calcul d'une distance entre deux log-magnitudes pour une position donnée. L'un de ces deux spectres a été translaté sur l'axe des log-fréquences d'un certain nombre d'échantillons. La distance est obtenue par sommation sur des échantillons fréquentiels considérés sur l'intervalle [3700Hz-12900Hz], après centrage de chaque log-magnitude.
4. Calcul d'une distance globale, par moyennage sur l'ensemble des positions.

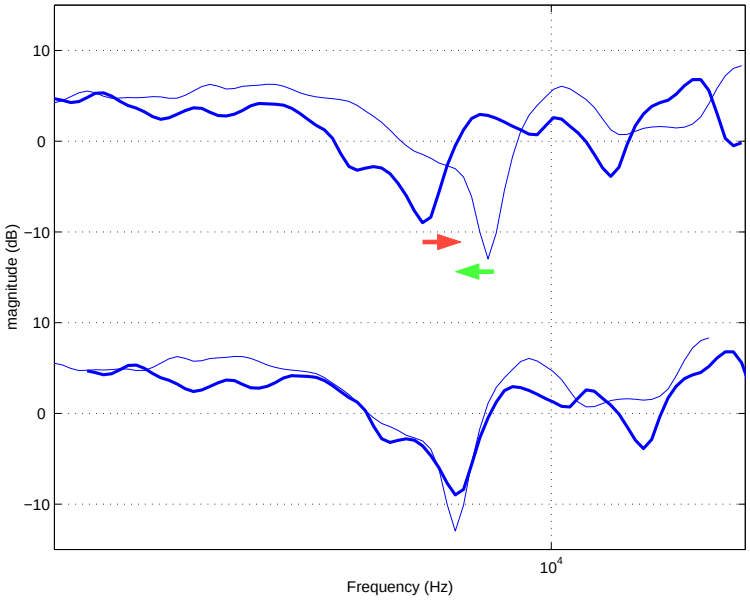


FIG. 5.24 – Principe de la technique de scaling fréquentiel.

Le facteur de scaling optimal doit permettre la superposition optimale des deux jeux de spectres, comme le schématise la Figure 5.24. Il est obtenu comme minimum de la distance inter-spectre (Figure 5.25).

Le facteur de scaling optimal est exprimé sur une échelle linéaire, et représente alors le coefficient d'homothétie qu'il faut appliquer à l'échelle de fréquence sur laquelle est représenté l'une des deux famille de spectres. Un facteur de scaling de 1 correspond donc à une superposition optimale des deux familles sans besoin de translation relative. Nous avons appliqué l'approche de Middlebrooks à nos données, et, comme l'illustre la figure 5.25, le facteur de scaling optimal entre les têtes 026 et 044 vaut 0.8366 , soit une translation de 9 échantillons entre les spectres.

On aboutit donc à un facteur de scaling "global", s'appuyant sur son calcul sur un intervalle fréquentiel unique, et sur toutes les positions mesurées.

5.4.1.3 Extensions proposées de la méthode

L'approche de Middlebrooks a l'avantage d'être simple, puisqu'elle ne requiert qu'un seul paramètre d'ajustement. Toutefois, il semble abusif de réduire la différence entre deux têtes à une transformation unique. Dans le prolongement des travaux de Middlebrooks, nous développons ainsi quelques raffinements permettrant de mieux rendre compte de la transformation réelle entre deux têtes. Nos extensions conduisent à un "scaling multiple", où plusieurs facteurs de scaling "locaux" se sustituent au facteur de scaling "global" de Middlebrooks :

– Les facteurs de scaling sont recherchés suivant plusieurs régions fréquentielles, à savoir sur les deux intervalles étudiés en section 5.2.1 (intervalle basses-fréquences [100Hz-1600Hz] et intervalle hautes-fréquences [800Hz, 13000Hz]). Le lien établi par Shaw entre caractéristiques morphologiques et caractéristiques spectrales des HRTF laisse penser que Middlebrooks concentre son analyse sur les paramètres de petite taille tels le pavillon, et écarte par exemple les dimensions de la tête, susceptibles d'être prises en compte par notre analyse BF.
La méthode de scaling présente par construction un risque de minima locaux dans le calcul de la distance : ils correspondent à la superposition optimale de deux caractéristiques hétérogènes mais de forme voisine, situées dans des plages de fréquences différentes. Middlebrooks propose de résoudre ce problème en utilisant un petit intervalle de translation localisé sur l'intervalle témoin des caractéristiques du pavillon. Nous proposons de renforcer la robustesse aux minima locaux en élargissant l'intervalle d'étude. Les intervalles BF et HF sont donc plus grands que celui retenu par Middlebrooks.

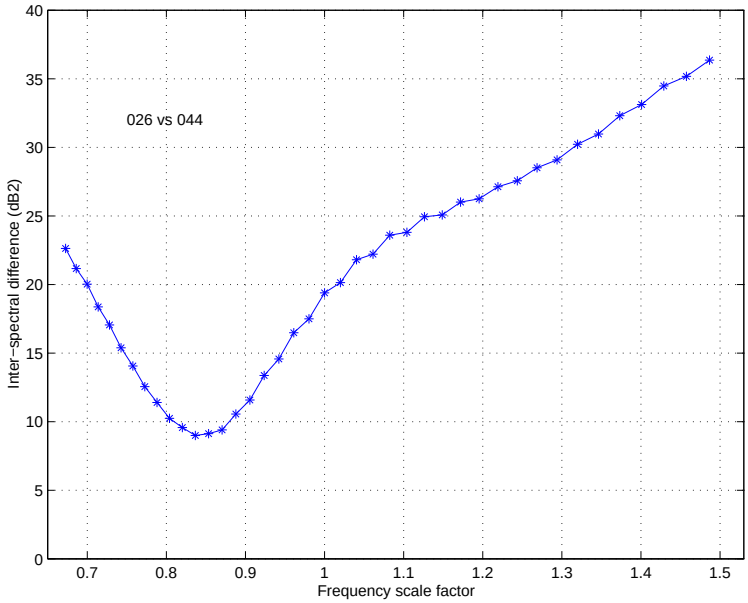


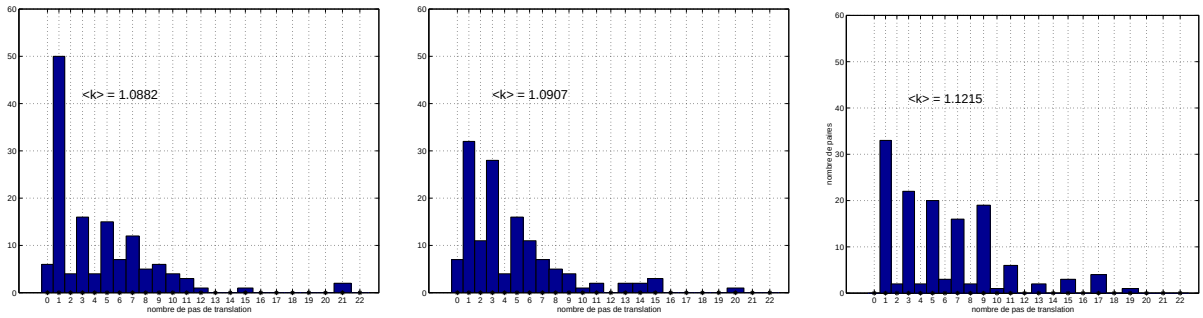
FIG. 5.25 – Variation de la distance inter-spectre de Middlebrooks en fonction du facteur de scaling.

- Le facteur de scaling global de Middlebrooks peut être obtenu en s'appuyant sur un sous-ensemble de positions. Cette solution un intérêt pratique puis qu'elle réduit le nombre de mesures nécessaires pour réaliser l'adaptation.
- Les facteurs de scaling sont recherchés suivant plusieurs régions spatiales, par opposition à une recherche globale s'appuyant sur l'ensemble des positions mesurées. En effet, on peut penser que certaines caractéristiques morphologiques interviennent davantage en fonction de la zone d'incidence, par exemple, la longueur de la tête pour les positions latérales, la largeur de la tête pour des positions frontales ou arrières. Les déformations spectrales à compenser dépendraient ainsi de la position. De plus, l'analyse statistique des données menée en chapitre 3 a mis en évidence des régions de l'espace "privilégiées", sur lesquelles s'appuyent plusieurs techniques d'implantation multicanale (haut-parleurs virtuels, Analyse en Composantes (spatiales) Indépendantes). Un scaling adapté à chaque région spatiale conduirait alors à transformer de façon idéale chaque filtre de reconstruction du décodeur de la synthèse binaurale .

5.4.2 Apport d'un scaling par bandes fréquentielles

5.4.2.1 Relation entre les facteurs de scaling "locaux"

Les facteurs de scaling "locaux", i.e. pour chaque bande de fréquence, sont obtenus en minimisant la distance entre spectres d'amplitude sur chaque intervalle. On parlera indifféremment du facteur de scaling linéaire, dont les valeurs varient autour de 1, ou de sa version logarithmique en base 2.
Comme dans l'étude de Middlebrooks, les valeurs de k sont quantifiées "logarithmiquement", espacées d'un pas de 0.0286 qui correspond à une translation d'un échantillon des spectres que nous avons préalablement re-échantillonnés. Bien que nos intervalles d'étude soient de taille différente (45 points fréquentiels pour BF et HF, 81 points pour MF), nous considérons le même nombre de valeurs possibles pour k, 45. Ces 45 valeurs correspondent à 22 pas de translation possibles dans un sens, 22 pas dans l'autre, plus le cas d'une absence de translation.
La latitude que nous avons retenue est identique à celle choisie par Middlebrooks, et, correspond à une homothétie d'un coefficient dépassant 46% ($k_{max} = 1.5467$, $k_{min} = 0.6465$). Selon le raisonnement de Shaw, le coefficient de scaling optimal doit être rapproché de la variance inter-individuelle des caractéristiques de la conque, qui, d'après la section 5.2.3, ne dépasse pas 20%. Par conséquent, la plage de variation que nous étudions pour k semble être suffisant. La Figure 5.26 présente les valeurs obtenues pour le scaling de nos 136 paires. Pour étudier les valeurs moyennes, nous ne considérons que la valeur absolue des facteurs, qui correspondent ainsi à des pas variant entre 0 et 22.



Intervalle BF | Intervalle MF | Intervalle HF

FIG. 5.26 – Histogramme des facteurs de scaling obtenus pour chaque bande fréquentielle, pour 136 paires.

	BF	MF	HF	Mi
BF	1	0.25	0.01	0.02
MF	0.25	1	0.63	0.62
HF	0.01	0.63	1	0.99
Mi	0.02	0.62	0.99	1

TAB. 5.7 – Corrélation entre les 136 facteurs de scaling obtenus pour toutes les bandes fréquentielles étudiées.

Plusieurs résultats peuvent être mentionnés :

1. Les facteurs de scaling obtenus dans chaque bande sont différents. Pour les trois intervalles, il existe des paires requierant un très fort scaling, pouvant même atteindre 21 pas de translation pour l'intervalle BF. Les moyennes moyennes en revanche, varient fortement d'un intervalle à l'autre. Pour BF et MF, une translation de 4 pas ou moins suffit pour plus de la moitié des paires, tandis que 6 pas au moins sont nécessaires dans le cas HF. Cet écart traduit également la présence de différences inter-individuelles plus importantes en HF que pour les autres intervalles.
2. Les facteurs de scaling obtenus pour les bandes BF, MF et HF sont peu corrélés. C'est ce qu'illustre le Tableau 5.7. La corrélation entre les facteurs MF et HF atteignent 0.63, ce qui est insuffisant pour que l'on puisse assimiler les deux coefficients. On peut donc penser que ces facteurs de scaling traduisent des phénomènes différents. Ainsi, appliquer un facteur de scaling unique sur toute la bande de fréquence conduit à un résultat sous-optimal, ce qui plaide en faveur d'un scaling différent pour chaque bande de fréquence. En outre,si l'on souhaite néanmoins appliquer un scaling "global", le choix de la bande pour déterminer le facteur de scaling n'est pas anodin, et l'étude de l'efficacité de la méthode est requise pour départager les trois facteurs candidats.
3. Les facteurs de scaling obtenus pour notre bande HF sont très corrélés à ceux de Middlebrooks, sans toutefois être identiques.

5.4.2.2 Réduction des distances inter-spectre

La Figure 5.27 présente un exemple de scaling global, utilisant le coefficient HF.

Pour mesurer l'efficacité du scaling, plusieurs critères sont évalués. Deux d'entre eux sont proposés par Middlebrooks : le pourcentage de réduction de la distance entre log-magnitudes (Figure 5.28), et la corrélation entre la distance entre log-magnitude après scaling et le facteur de scaling (Figure 5.29). Nous observons également la distance entre les spectres complexes avant et après scaling (Figure) .

Les histogrammes de la Figure 5.28 sont beaucoup plus étalés pour les HF et la bande Mi qu'en MF et (encore plus) qu'en BF, premier élément traduisant l'efficacité du scaling en hautes fréquences (bandes HF et Mi) : on observe une réduction de la distance inter-spectres de plus de 15% pour la moitié des paires (15.5% pour Middlebrooks), et de plus de 30% pour le tiers d'entre elles. Comme Middlebrooks,

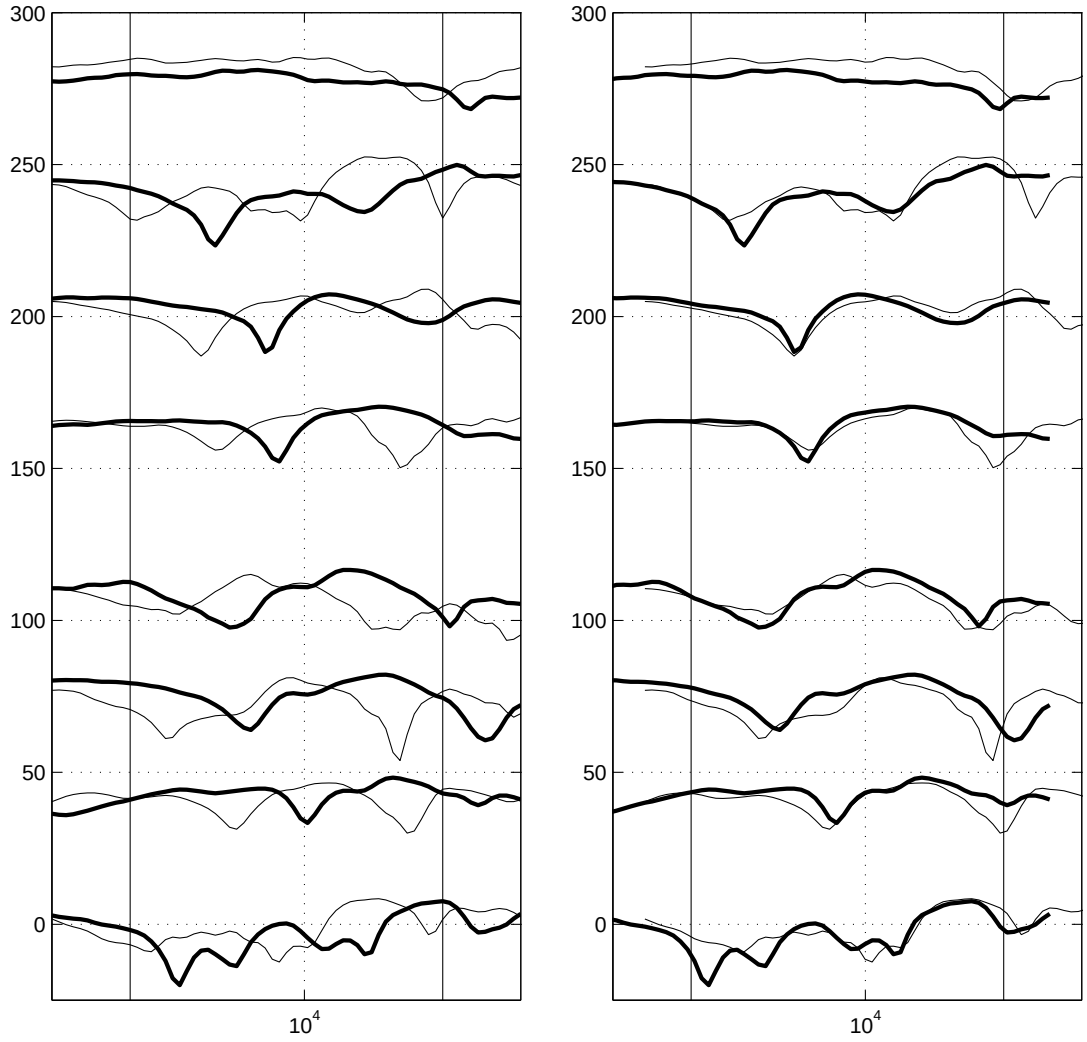


FIG. 5.27 – Résultat d'un scaling utilisant le facteur de scaling HF, pour adapter la tête 4 (trait gras) à la tête 7 (trait fin).

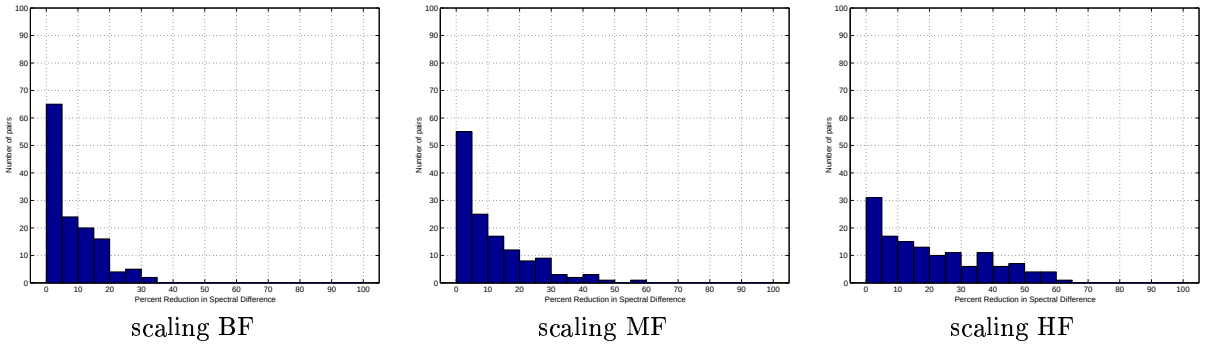


FIG. 5.28 – Efficacité du scaling par la réduction de la distance entre log-magnitudes.

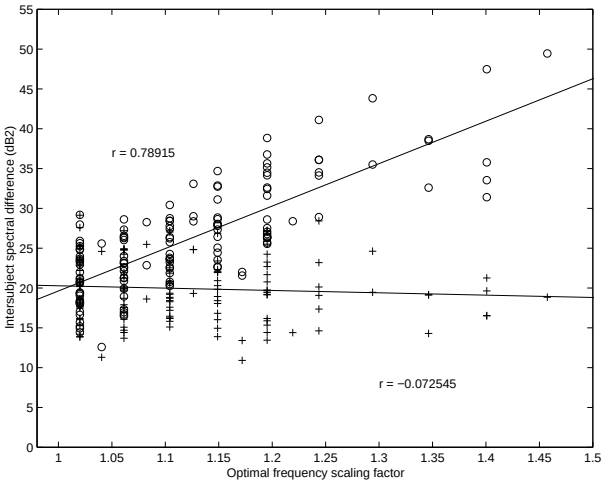


FIG. 5.29 – Corrélation entre le facteur de scaling optimal en HF et la distance entre log-magnitude sur ce même intervalle : distance avant scaling ('o'), distance après scaling ('+').

nous observons une forte corrélation entre la distance initiale et le coefficient de scaling, pour les hautes fréquences (Figure 5.29) : plus la distance est grande, et plus la translation à réaliser est importante. En revanche, la distance résiduelle est fortement décorrélée de ce même coefficient, et apparait même presque constante. Nous pouvons ainsi considéré que le scaling a réalisé tous les efforts possibles, et que l'écart demeurant entre les têtes représente des différences qui sortent des hypothèses d'homothétie formulées par Shaw.

Sur les spectres complexes également, on constate que l'amélioration en BF est très faible. Sur l'exemple donné en Figure 5.30, elle se traduit par une diminution inférieure à 0.1dB. Elle est non négligeable en MF, où elle vaut 2dB en moyenne, et va jusqu'à 8dB pour les HF. Ces valeurs correspondent à une diminution de la distance inter-spectre respectivement de 3.5%, de 31% et de 58%.

Ces résultats plaident en faveur d'un scaling multiple, qui s'appliquerait avec des coefficients éventuellement différents sur MF et sur HF. En revanche, il apparait inutile de faire subir cette transformation aux basses fréquences. Ainsi, un double scaling semble nécessaire et suffisant.

5.4.2.3 Sens physique des facteurs de scaling

Donner un sens physique au facteur de scaling, pourrait nous conduire à une relation simple avec un paramètre morphologique. Ces derniers étant facilement mesurables, cette relation fournirait une méthode simple pour accéder au facteur de scaling. Cette recherche est justifiée par l'hypothèse sous-jacente au scaling, selon laquelle 10% d'écarts entre les paramètres morphologiques de deux têtes doit se traduire par un décalage homothétique de 10% des caractéristiques fréquentielles des HRTF.
Tous les param morpho ne se prêtent pas à cette hypothèse : on retire de nos paramètres d'étude les caractérisques repérant des déplacements et des angles.Pour ceux qui restent (17 paramètre), on forme les rapports $\frac{m_k(i)}{m_k(j)}$ de la valeur du paramètre k mesuré sur les têtes i et j. Ce sont ces rapports que l'on tente de corréler avec le facteur de scaling permettant de transformer la tête i en tête j (exprimé en linéaire ou en log).

On observe une faible corrélation des facteurs de scaling avec chaque paramètre pris individuellement. La plus forte relation a été obtenue entre le facteur de scaling HF et la hauteur de la conque, pour lesquels le coefficient de corrélation atteint 0.78 en linéaire et en log (cf Figure 5.31). Cette valeur est légèrement supérieure à celle que nous obtenons pour ces deux paramètres sur l'intervalle de fréquence de Middlebrooks (r=0.75). Middlebrooks observe également une corrélation de 0.72 entre facteurs de scaling et longueur de la conque (c'est la corrélation maximale). Il semble logique d'observer une relation avec un paramètre de la conque, dont les modes justifient l'approche de scaling, et qui plus est, sur les HF

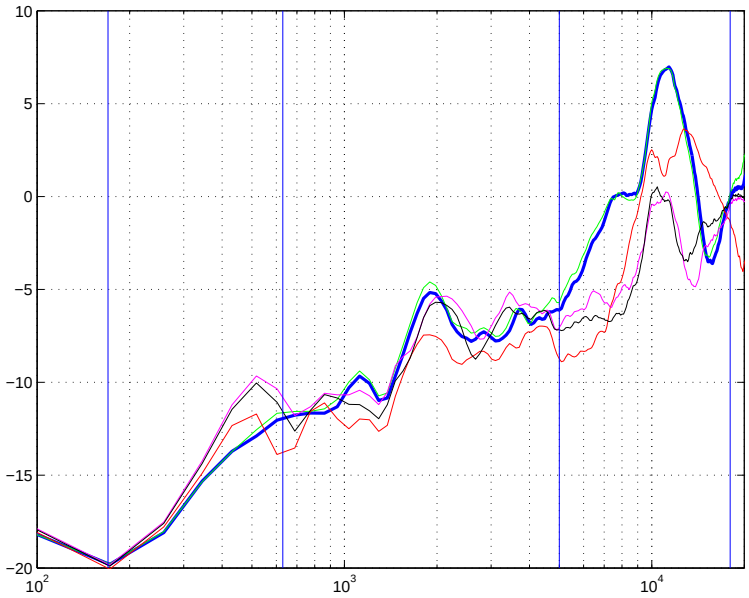


FIG. 5.30 – Efficacité du scaling pour la réduction de la distance complexe entre les têtes 6 et 13 : distance avant scaling (bleu), scaling BF (vert), scaling MF (rouge), scaling HF (magenta), scaling Middlebrooks (noir).

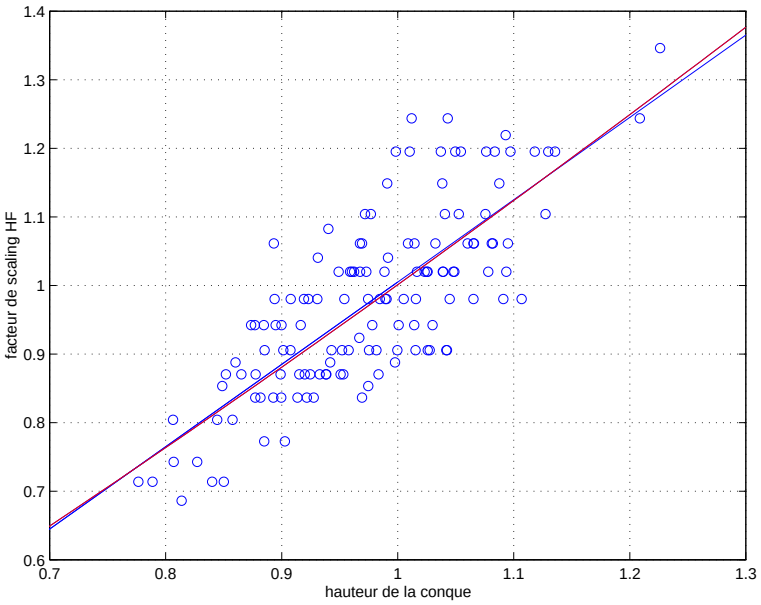


FIG. 5.31 – Corrélation entre le facteur de scaling HF et la hauteur de la conque : meilleure approximation linéaire (en bleu), et logarithmique (en rouge).

(pour BF : r=0.07 et pour MF : r=0.57). On aboutit à deux relations (cf Figure 5.31) :

$$k_{j\to i} \approx \left(\frac{r_j}{r_i}\right)^{1.21}$$

$$k_{j\to i} \approx 1.20 * \frac{r_j}{r_i} - 0.19$$

Une meilleure approximation du facteur de scaling par les paramètres morphologiques peut être trouvée par une technique de corrélation multiple, telle que nous l'utilisions en partie 5.3. Nous notons toutefois qu'aucune relation "triviale" n'a pu être établie pour les bandes BF et MF.

Middlebrooks a également cherché une relation entre le facteur de scaling et le rayon du modèle sphérique équivalent, ou radius. Trouver une forte similitude entre les deux serait effectivement très satisfaisant, puis que les paramètres d'adaptation individuel, le radius pour le réglage de l'ITD, et le facteur de scaling, seraient confondus. Etant données les dimensions de la tête, auxquelles le radius est fortement corrélées, on pourrait ainsi s'attendre à voir un fort lien avec le facteur de scaling BF. Or, cette hypothèse n'est pas vérifiée, puisque le coefficient de corrélation obtenu est de l'ordre de 0.22. Nous interprétons alors ce résultat par le fait que le scaling n'est pas approprié pour les BF : même si les paramètres de la têtes gouvernent effectivement les différences inter-individuelles en BF, celles-ci ne s'expriment pas par une homothétie de l'axe des fréquences.
La plus forte corrélation est obtenue pour le facteur de scaling HF (=0.64), à nouveau légèrement supérieur à celui que l'on obtient pour la bande Midd (r=0.62). Comparaison avec le coeff observé par Midd. On peut d'ailleurs comparer l'expression donnée par Midd :

$$\frac{r_j}{r_i} \approx (k_{j\to i})^{0.26}$$

à celle que nous obtenons :

$$\frac{r_j}{r_i} \approx (k_{j\to i})^{0.23}$$

Nous précisons que nos résultats ont été obtenus en utilisant le radius estimé à partir de toutes les positions. Pour le radius "horizontal" en effet, la corrélation ne dépasse en aucun cas 0.5. Dans tous les cas, la relation entre facteur de scaling et radius semble fragile. On peut néanmoins penser que ces corrélations sont pénalisées par la quantification des facteurs de scaling, qui masque les variations fines de ces derniers éventuellement requises.

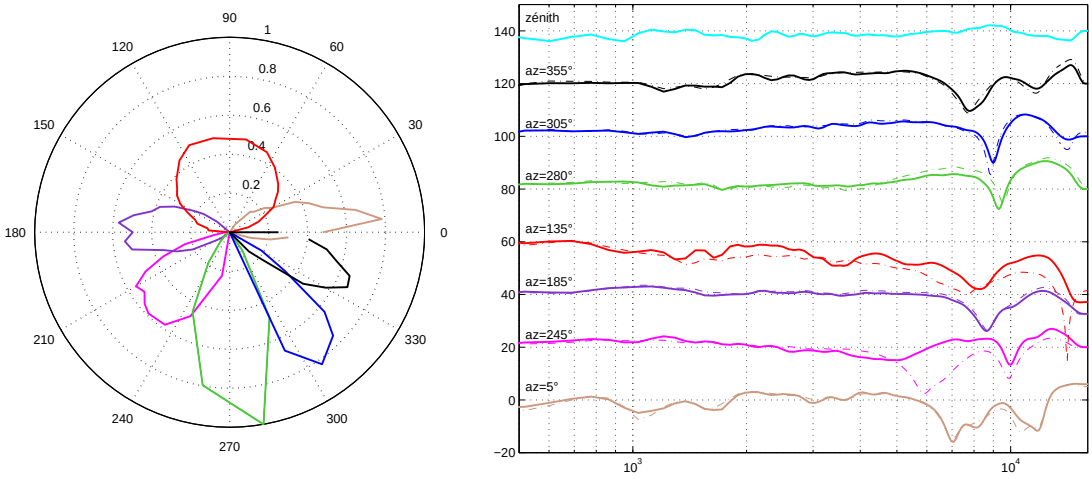
5.4.3 Choix de régions spatiales déterminantes pour le scaling

Dans cette section, nous nous concentrons sur le scaling de la bande HF, et nous qualifierons de "local" un facteur de scaling optimisant la transformation pour une seule position. Il sera donc local spatialement.

5.4.3.1 Facteur de scaling global abtenu avec un nombre réduit de positions

Pour être optimal, le scaling devrait être appliqué à chaque HRTF, avec un facteur propre à chaque position. On peut penser que les facteurs de scaling de positions voisines seraient fortement corrélés et qu'ainsi un coefficient par région spatiale pourrait suffire. Toutefois, le réglage manuel des paramètres du scaling par l'utilisateur deviendrait rapidement mal-pratique, de par la multiplicité des paramètres. L'approche de Middlebrooks, au contraire, ne requiert l'ajustement que d'un seul paramètre, mais puisqu'il s'appuie sur toutes les positions pour dériver un facteur de scaling "moyen",[7] il atteint une adaptation sous-optimale pour chacune.
Entre ces deux solutions extrêmes, on envisage tout d'abord un scaling global, dont le facteur est déterminé à partir d'un sous-ensemble de positions. En effet, comme nous le décrivons en chapitre 3, l'Analyse en

[7] Bien sûr, on affecte toutefois à chaque position un poids proportionnel à l'angle solide occupé.



Fonctions spatiales universelles issues de l'ACI

Filtres de reconstruction associés

FIG. 5.32 – Comparaison des filtres de reconstruction obtenus par ICA et des HRTF des directions "pointées".

Composantes Indépendantes met en évidence certaines positions plus "importantes" que d'autres. Comme nous l'illustrons en Figure 5.32, les fonctions spatiales fournies par l'analyse statistique des HRTF du plan horizontal présentent des lobes pointant dans les directions -5°, 245°, 185°, 135°, 80° 55° et 25°. Comme on l'observe, les filtres de reconstruction associés à ces fonctions spatiales sont très proches des HRTF pointant dans ces directions, mis à part peut-être la position contralatérale, pour laquelle la fonction spatiale a un lobe large. Les HRTF mesurées à ces positions sont celles qui permettent donc de reconstruire au mieux toutes les autres par combinaison linéaire. Nous choisissons comme filtre d'élévation la HRTF du zénith.

Nous nous intéressons au facteur de scaling obtenu en ne s'appuyant que sur ces positions spécifiques. Dans le cas d'une implantation multicanale, ces 8 filtres de reconstructions concentrent en effet toute l'individualité du sujet et c'est eux qu'il convient d'adapter. Dans une implantation bicanale, il peut également paraître intéressant de pouvoir baser l'estimation du facteur de scaling sur un nombre réduit de positions, puisque seules quelques HRTF sont alors à mesurer. En outre, si l'on considère une implantation "hybride" où coexistent les deux implantations, alors un intérêt supplémentaire consiste à pouvoir utiliser un re-échantillonneur commun en sortie du décodeur binaural. Il convient donc de s'assurer de la corrélation entre le facteur global obtenu à partir de toutes les positions, tel que celui de Middlebrooks, et le facteur obtenu selon cette nouvelle approche.

Comme l'illustrent les Figures 5.33 et 5.34, on obtient effectivement une forte corrélation entre ces deux familles de facteurs de scaling (136 paires) : sur la bande HF, le coefficient de corrélation atteint 0.86, et on propose la relation suivante pour les relier :

$$\log_2\left(k_{global}\right) \approx 0.89.\log_2\left(k_{global\,ICA}\right) + 0.0249$$

En revanche, la corrélation est plus faible pour les autres bandes de fréquences (0.58 pour les BF et 0.6 pour les MF). Cette variante du scaling global ne facitera donc l'estimation facteur que pour les HF. En outre, comme on l'observe sur la Figure 5.34, certaines positions demeurent "sacrifiées" : dans les deux cas d'adaptation globale, le creux autour de 8kHz, qui se retrouve aux HRTF 3 et 4 bénéficient du scaling, au détriment de la superposition des HRTF de la position 2, pour laquelle nous choisirions "à l'oeil" une translation dans le sens opposé. Si la différence entre têtes peut s'exprimer par un facteur d'homothétie, on constate ainsi que ce facteur dépend de la position.

5.4.3.2 Détermination de facteurs de scaling "locaux"

Les facteurs de scaling "locaux" sont ainsi recherchés, en appliquant l'algorithme de minimisation de la distance entre spectres d'amplitude à chaque HRTF séparemment.

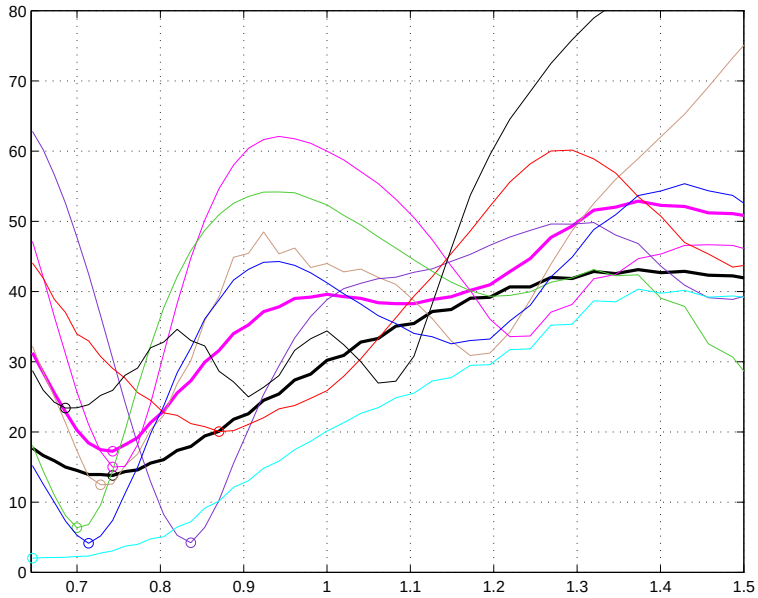


FIG. 5.33 – Distance entre les spectres d'amplitude des têtes 6 et 13, en fonction du facteur de scaling HF : facteur de scaling global (en noir), facteur de scaling global s'appuyant sur 8 position (en magenta), facteur de scaling local.

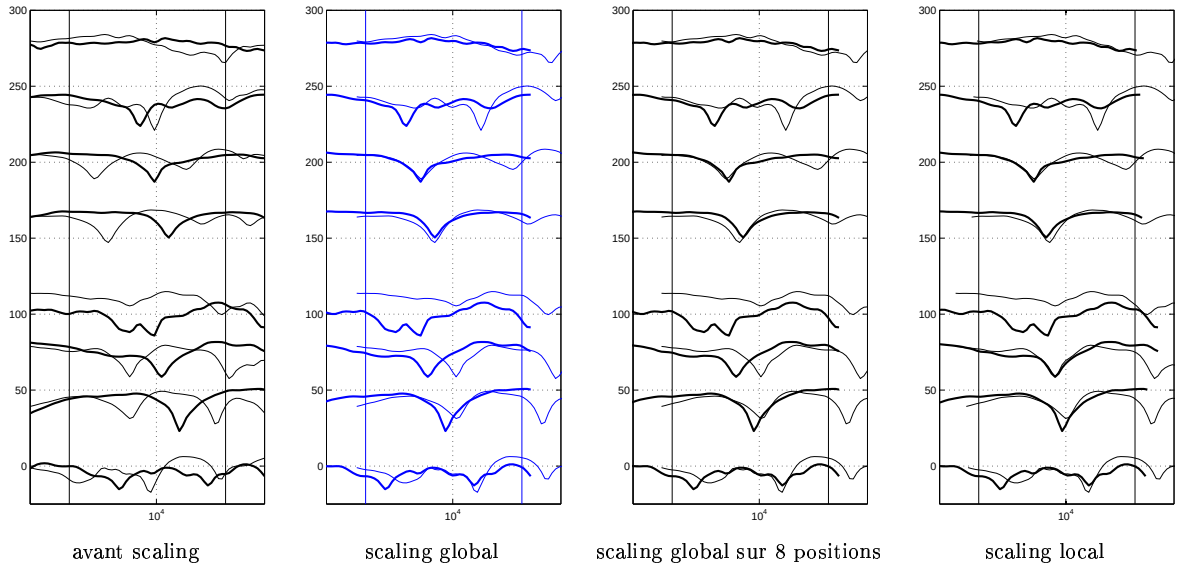


FIG. 5.34 – Spectres d'amplitude des 8 HRTF "cardinales" avant et après scaling, pour les têtes 6 et 13.

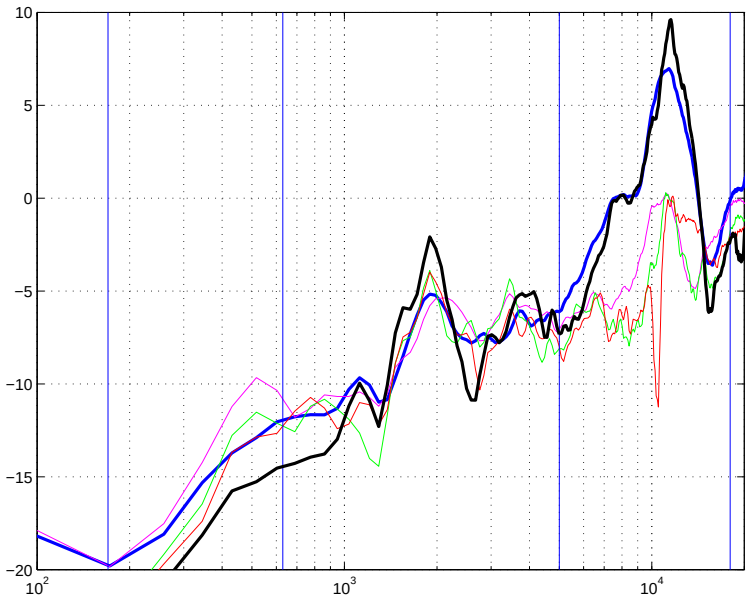


FIG. 5.35 – Réduction de la distance complexe inter-spectre pour les têtes 6 et 13 : distance avant scaling et sans décomposition ICA (bleu), distance avant scaling et après décomposition ICA (noir), scaling global HF sur toutes les positions (magenta), scaling global "réduit" HF des filtres de reconstruction ICA (vert), scaling local HF des filtres de reconstruction ICA (rouge).

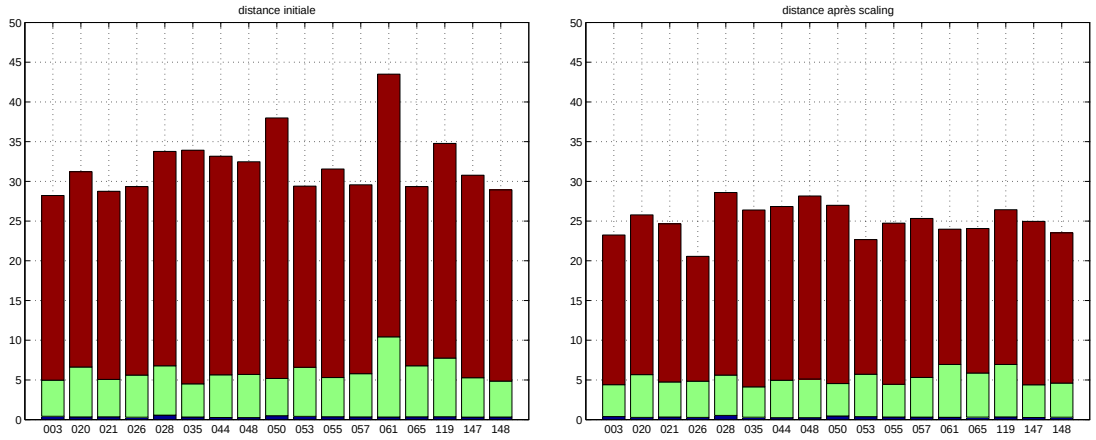
Comme on l'observe en Figure 5.33, les facteurs de scaling locaux offrent une dispersion non négligeable, ce qui traduit la légitimité de la démarche. Elle fait néanmoins apparaître plusieurs limites :

- la fonction de distance peut présenter plusieurs minima locaux, ce que l'on constate en Figure 5.33 pour les courbes de scaling local. Chacun correspond à une solution pour la superposition des HRTF. Pour certaines positions (par exemple courbe noire de la Figure 5.33), le choix du minimum minimorum apparait presque arbitraire, tant les autres minima en sont proches. Plus généralement, il apparait que le facteur de scaling optimal est parfois celui qui permet de rejeter les grosses différences. Ainsi, sur la Figure 5.34, on constate que même un scaling local n'a pas entraîné la superposition des HRTF n°2, qui semble pourtant aisément discernable à l'oeil dans la situation avant scaling (cette position correspond justement à la courbe noire précédente). La solution donnée par l'algorithme s'explique du fait de l'existence de différences inter-spectres non réductibles à une simple homothétie des fréquences. En l'occurence, la HRTF n°2 de la tête 6 présente une forte résonance autour de 12kHz, absente chez la tête 13. Le poids de cet écart dans le calcul de distance, encore amplifié par le centrage des données, entraîne un déplacement des courbes opposé à celui qu'intuitivement nous souhaiterions : pour obtenir une superposition optimale, la méthode conduit à écarter de l'intervalle d'étude "ce qui dérange".
- la fonction de distance peut être dépourvue de minimal local, phénomène que l'on observe pour la courbe cyan de la Figure 5.33, associée à la position du zénit. Il semble abusif de mettre en cause la trop faible dimension de l'intervalle de recherche pour le facteur de scaling. Ces artefacts se présentent plutôt pour les positions auxquelles la différence entre tête ne se réduisent pas à une homothétie des fréquences.

Ces deux limites sont inhérente à la méthode de scaling, mais sont en partie gommées par le "moyennage" réalisé dans l'approche s'appuyant sur plusieurs positions. Elles constituent un défaut important quand la méthode est appliquée indépendemment à chaque position.

5.4.3.3 Réduction des distances inter-spectre

La Figure 5.35 permet de comparer l'efficacité des scalings global et local pour une paire de têtes, représentative. La modélisation ICA, ou plus exactement la décomposition des HRTF sur les HRTF à phase minimale des directions pointées par les fonctions spatiales de l'ICA, augmente l'écart initial entre tête



Avant scaling Après scaling

FIG. 5.36 – Distance entre têtes d_{Midd} cumulée pour les différentes bandes de fréquence : BF (bleu), MF (vert) et HF (rouge). Chaque tête est tour à tour prise comme tête de référence.

au niveau des résonances, sans doute parce que leur position diffère entre les 2 têtes, mais elle diminue cet écart partout ailleurs. Le scaling pratiqué sur les filtres de reconstruction permet de réduire cet écart jusqu'à un niveau inférieur à celui obtenu par un scaling global appliqué à l'ensemble des HRTF. Le scaling des filtres de reconstruction apparait donc comme une méthode très efficace. Sur notre exemple, la distance entre têtes est réduite de 5dB en hautes fréquences. En revanche, comme nous le présentions par l'analyse des limites de l'algorithme, l'apport d'un scaling local des filtres de reconstruction ne semble pas significativement différent de celui du scaling global appliqué sur ces mêmes filtres.

5.4.4 Implantation de l'adaptation individuelle

5.4.4.1 Recherche de la tête la plus "adaptable"

L'implantation de la technique de scaling requiert tout d'abord le choix d'une tête qui sera utilisée comme point de départ de la transformation et à partir de laquelle il s'agit donc de déduire toutes les autres.

Un critère pour départager nos 17 têtes est la distance inter-spectre d_{Midd} après scaling, que nous représentons en Figure 5.36. Sur cette Figure, d_{Midd} a été calculée pour chaque bande de fréquence, et un score "global" est présenté pour chaque tête, comme somme des scores de chaque bande. On observe qu'avant scaling, les têtes 9 et surtout 13 présentent de très fortes distances. Pour la première, ce sont les BF et MF qui la font sortir du lot, tandis qu'il s'agit plutôt des HF pour la tête 13. En outre, bien que son score global ne soit pas significativement élevé, la tête 5 présente une forte singularité en BF. Cette tête apparaît en outre mal adaptée au scaling, puisque la distance après scaling est parmi les plus importantes.
D'une manière générale, les apports du scaling se situent en HF, comme nous l'avons observé sur les spectres complexes des exemples précédents. La tête présentant la plus faible distance après scaling est la tête 4. Ce bon score s'explique par :

1. une distance initiale faible,
2. une forte réduction grâce au scaling.

Pour d'autres têtes, comme les têtes 8 et 12, le scaling est peu efficace, comme en témoigne le taux de réduction de la distance d_{Midd} initiale (Figure 5.37), pénalisée par de mauvais résultats en BF et MF.

Une autre approche consisterait à coupler adaptation discrète et scaling fréquentiel : le sujet est tout d'abord placé au sein des têtes de référence, par exemple suivant le protocole de superposition des espaces que nous avons présenté, et le représentant le plus proche du sujet est alors utilisé pour le scaling. La tête "de départ" serait ainsi éventuellement différente pour chaque sujet, mais la distance à compenser par le scaling serait a priori plus faible donc plus accessible.

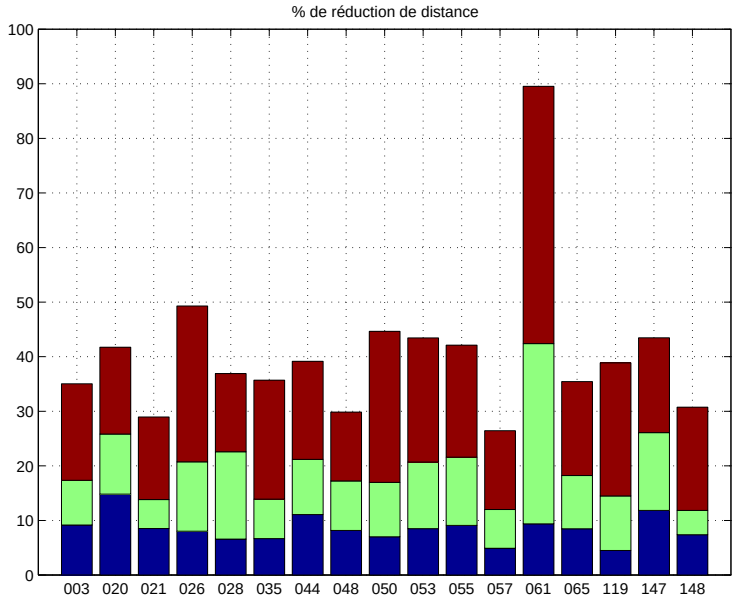


FIG. 5.37 – Taux de réduction de la distance initiale entre têtes d_{Midd}, cumulé pour les différences de fréquence : BF (bleu), MF (vert) et HF (rouge). Chaque tête est tour à tour prise comme tête de référence.

5.4.4.2 Synthèse des HRTF "adaptées"

Pour que le scaling puisse être réglé de façon interactive par l'auditeur, il est nécessaire d'implanter la transformation en temps réel.

Une première méthode, utilisée par Middlebrooks, consiste à changer la fréquence d'échantillonnage du signal de sortie : si elle est multipliée d'un facteur β, l'axe des fréquence subit une homothétie de facteur β et le retard interaural est réduit d'un facteur β. Cette approche a deux principaux inconvénients :

- c'est le même facteur qui transforme spectres et ITD , alors que, comme nous l'avons vu, la corrélation entre leur facteur de scaling optimal n'est pas très forte.
- elle ne permet pas d'envisager un scaling local par bandes fréquentielles, mis à part en dédoublant chaque signal et en appliquant à chaque réplique un reéchantillonnage approprié. Cela double le coût d'implantation par rapport au scaling simple.

Une alternative pourrait consister à utiliser une transformation bilinéaire, telle que nous la décrivions pour l'optimisation de la modélisation paramétrique de filtres ([Smi83]). La technique de "warping" consiste à substituer, dans la structure de filtre, le retard z^{-1} par un filtre passe-tout de degré un $\frac{a-z^{-1}}{1-a^*.z^{-1}}$. Le scaling de l'ITD peut alors être réalisé par changement de fréquence d'échantillonnage, tandis que le paramètre a contrôle le scaling des spectres à phase minimale. Pour étendre cette solution à un scaling multiple, on peut penser à une transformation conforme d'ordre 2, équivalant à la mise en série de deux warping. Toutefois, on introduit alors un filtre passe-tout d'ordre 2 à la place du z^{-1}, ce qui, comme pour la première solution, double le coût de la structure.

5.4.5 Conclusion sur le scaling fréquentiel

Le scaling fréquentiel est une technique de morphing permettant de transformer au mieu une tête en une autre, utile pour réduire l'écart entre ces deux têtes pour des fréquences supérieures à 1kHz. Son application peut tirer profit d'une transformation indépendente des MF et des HF. Il semble qu'elle doive en revanche s'abstenir de modifer les BF, intervalle fréquentiel sur lequel les différences interindividuelles ne répondent pas aux hypothèse du scaling : les têtes possèdent le mêmes résonances structurelles, qui ne se distinguent que par une translation constante sur l'axe des fréquence.

Le facteur de scaling obtenu en HF est fortement corrélé aux dimensions de la conque, et plus spécialement avec sa longueur, ce qui laisse ouvert la perspective d'une adaptation des HRTF en HF à partir de la simple mesure de ce paramètre morphologique. En outre, autre approche pour faciliter la mise en oeuvre du scaling, le facteur de scaling HF peut être estimé à partir d'un sous-ensemble de positions, celles qu'indiquent l'analyse statistique des HRTF avec l'ICA.
En outre, et en première approximation, on peut utiliser le même facteur de scaling pour les structure bicanales et pour les structures multicanales, élément utile dans le cas d'une implantation hybride.

L'application d'un scaling spécifique pour chaque HRTF fait apparaître certaines limites de la méthode et ne semble donc pouvoir être mise à profit. Ces limites semblent en partie liées au choix de la distance inter-spectre. Nous avons en effet justifié l'approche de Middlebrooks par les théories de Shaw, accordant une importance prédominante aux résonances. Or la distance qu'il utilise, et que nous avons reprise, évalue des écarts de log-magnitude, qui auront précisemment effet de donner plus d'importance aux anti-résonance, par comparaison à un écart calculé sur des amplitudes linéaires.

Enfin, la capacité d'adaptation des têtes par scaling varie. Il apparaît que la tête 5, dont on a vu qu'elle constituait une bonne candidate pour une synthèse binaurale sans adaptation, est difficilement adaptable aux autres têtes. Ce résultat conforte la thèse du "bon localisateur" de Wenzel, selon laquelle une tête aux traits "exagérés" pourrait favoriser la localisation. En outre, cela nous indique qu'il est important de ne pas choisir au hasard la tête "générique" à partir de laquelle toutes les autres doivent être déduites par scaling.

5.5 Conclusion

Dans ce chapitre, nous avons d'abord caractérisé la distance entre têtes. Cette distance peut être quantifiée à l'aide de paramètres objectifs impliqués dans le processus de localisation : différences entre HRTF, entre ITD ou entre paramètres morphologiques. En outre, on a pu vérifier que ces différences mesurables sont également audibles, en tirant profit du test perceptif décrit au chapitre 4.
Pour atténuer les effets d'une écoute non indivuelle, l'approche "a minima" consiste à sélectionner une tête moyenne, réalisant le meilleur compromis pour l'ensemble des auditeurs. On ne pratique donc pas d'adaptation, mais la tête partagée par tous n'est pas choisie au hasard. Les caractéristiques de la tête 028, que nous avons exhibées pour les différents espaces d'observation, nous conduisent à la proposer comme une tête "moyenne" satisfaisante.
Un premier effort d'adaptation individuelle peut être réalisé à l'aide d'une adaptation discrète, appairant un auditeur avec l'une des têtes constituant la base de données. Nous avons proposé une méthode pour évaluer la distance entre ces têtes et tout nouvel auditeur, ne s'appuyant que sur les relevés morphologiques de ce dernier. Cette procédure tirerait grandement profit de la mise en place d'un dispositif de mesure plus robustes et flexibles que ceux que nous avons étudiés. L'utilisation d'un scanner par exemple, offrirait l'avantage d'une capture numérique 3D de la morphologie des auditeurs, et autoriserait le choix a posteriori des paramètres utiles à relever pour l'adaptation.
C'est l'adaptation continue, enfin, qui permet de réduire les plus significativement les différence inter-individuelle. Nous avons repris et développé une approche initialement exposée par Middlebrooks : le scaling fréquentiel. Une tête est alors transformée en une autre par un morphisme des caractéristiques spectres de leurs HRTF. L'opération de base est une homothétie de l'axe des fréquences, qu'il est utile, comme nous l'avons montré, de pratiqué indépendemment sur deux bandes de fréquences, au delà de 700Hz. La tête 028, à nouveau, se présente comme un bonne tête générique, facilement déformable par scaling.
Plusieurs amélioration peuvent être envisagées pour l'adaptation continue, et notamment la prise en compte de transformations complémentaires au scaling fréquentiel. C'est dans cette direction que s'est engagée l'équipe de recherche dirigée par Sibbald à Sensaura Ltd, qui, s'inspirant également des travaux de Shaw, propose plusieurs paramètres d'implantation pour sa technologie "virtual ear" [Sib99] :

- l'ITD est adapté à l'aide de la *taille de la tête*.
- Comme pour Shaw, la *taille de l'oreille* contrôle un facteur d'homothétie de l'axe fréquentiel. Cette homothétie a pour effet de déplacer la fréquence des pics des HRTF et d'en modifier la largeur.

- La *profondeur de la conque* semble être le facteur principal des variations inter-individuelles de la fréquence du premier mode, autour de 5kHz. L'ajustement de cette fréquence est réalisé à l'aide d'une translation sur l'échelle linéaire appliquée après l'homothétie de l'étape précédente.
- L'*"ouverture" de la conque* est fortement influencée par la forme des plis extérieurs du pavillon. D'après Shaw, ces caractéristiques ont un effet amplificateur/atténuateur des premiers pics des HRTF. L'adaptation individuelle intègre donc un contrôle de gain des HRTF.

Parmi les transformations à envisager, on note donc : le gain en amplitude de certaines anti/résonances, et la translation de zones du spectre plus étroites que celle nous avons envisagé. Il y aurait ainsi un scaling macroscopique sur de larges intervalles, auxquel devraient s'ajouter des ajustements "microscopiques" sur des anti/résonances à comportement plus autonome.

Conclusion

Nous avons décrit et proposé plusieurs axes de perfectionnement pour la synthèse binaurale :

1. Post-traitement des fonctions de transfert binaurales (HRTF) mesurées, visant à simplifier les données à modéliser en éliminant les caractéristiques inutiles pour la synthèse binaurale.
2. Optimisation de son implantation en termes de coût de calcul et d'encombrement mémoire, grâce à l'approximation des HRTF par des filtres numériques d'ordre réduit, éventuellement obtenus par interpolation spatiale.
3. Optimisation supplémentaire pour la spatialisation de sources multiples, grâce à l'implantation multicanale de la synthèse binaurale.
4. Amélioration de la qualité du rendu sonore grâce à une simulation adaptée "sur mesure" à l'auditeur.

Sur ces différents domaines, on peut souligne les principales contributions de cette thèse :

- Définition et mise en oeuvre d'une nouvelle méthode expérimentale pour l'égalisation par rapport au champ diffus,
- Implantation multicanale par décomposition des HRTF sur une base de fonctions indépendantes de l'individu, présentant un recouvrement spatial minimal, et mettant en évidence un jeu réduit de "directions principales".
- Adaptation individuelle de la synthèse binaurale par déformation spectrale des HRTF, traitant séparément basses et hautes fréquences. Cette technique d'adaptation se marie idéalement à une implantation multicanale de la synthèse binaurale utilisant des fonctions spatiales indépendantes de l'individu, comme dans la méthode précédente.

Ce dernier domaine, encore peu exploré, peut tirer profit des recherches sur les modèles physiques des HRTF ([Gen84], [Kat98], [AAD99]), qui offriraient pour l'adaptation individuelle continue une alternative à la technique de déformation spectale que nous avons étudiée.

La validation perceptive que nous avons menée a souligné la persistance de deux principaux artefacts : défauts d'extériorisation des sons *(Inside-the-head localization)* et confusions avant-arrière. Ils constituent des défits majeurs à relever pour le déploiement des techniques binaurales vers des applications qui, au delà d'une simple "plausibilité", cherchent la reproduction pointilliste d'une distribution de sources sonores. Il convient de rappeler que deux facteurs n'ont pas été introduits dans nos tests : le suivi de position de la tête, qui permet de prendre en compte les petits mouvements effectués naturellement par l'auditeur et de restituer des indices de localisation dynamiques, et l'effet de salle, qui apparaît comme facteur d'amélioration de l'extériorisation des sons ([Har83]). Ces deux axes d'étude prolongeraient avantageusement les résultats de cette thèse.

Nous avons concentré notre étude sur la synthèse binaurale, mais la plupart des résultats peuvent être étendus aux "techniques binaurales" en général, et notamment à l'enregistrement. C'est le cas de l'égalisation par rapport au champ diffus que nous avons abordée au chapitre 2. La définition de formats d'encodage directionnel multicanaux peut également être prolongée par la réalisation pratique de systèmes de prise de sons. Nous avons ainsi mis en oeuvre les principes du Binaural B décrits par Jot ([JWL98]) pour une évaluation subjective de la qualité de localisation offerte par une prise son équivalente, constituée de deux microphones soundfield non coïncidents ([Auz99], [Lar99]). De nouvelles voies d'amélioration ont également été ouvertes pour la simulation de microphones de directivité idéale (donnée par exemple par une technique de décomposition linéaire des HRTF) à partir de microphones réels, aux

garabits de phase et d'amplitude variant le plus souvent avec la fréquence ([Lab00]). Parmi les domaines encore peu explorés, on peut mentionner les modalités d'ajout d'un effet de salle synthétique aux canaux d'un format binaural multicanal. Dans le cas bicanal, une restitution satisfaisante est obtenue en adjoignant aux signaux binauraux deux canaux de réverbération décorrélés sur toute la bande audible sauf en basses fréquences ([Jot92]). Rien n'indique que ces règles sont directement transposables aux signaux parallèles des formats multicanaux.

Bibliographie

[AAD99] C. Avendano, V.R. Algazi, and R.O. Duda. A head-and-torso model for low frequency binaural elevation effects. *in Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 1999.

[AAD01] V.R. Algazi, C. Avendano, and R.O. Duda. Estimation of a spherical-head model from anthropometry. *to be published*, 2001.

[AAT99] V. R. Algazi, C. Avendano, and D. Thomson. Dependence of subject and measurement position in binaural signal acquisition. *J. Audio Eng. Soc.*, 47(11), 1999.

[AF97] J.S. Abel and S.H. Foster. Method and apparatus for efficient representation of high-quality three-dimensional audio. *United States Patent*, (5596644), January 1997.

[Ano85] Anon. Specification for a manikin for simulated in-situ airborne acoustic measurements. *American National Standard ANSI S3.36 1985 (ASA 58-1985)*, 1985. Standards secretariat, Acoustical Society of America.

[APS97] C.-Y. Ahn, H.-S. Pang, and K.-M. Sung. Model of hrtf based on complex-valued pca considering group delay. *Presented at the International Symposium on Simulation, Visualization, and Auralization for Acoustic Research and Education, in Tokyo (Japan)*, 1997.

[ASS90] F. Asano, Y. Suzuki, and T. Sone. Role of spectral cues in median plane localization. *J. Acoust. Soc. Am.*, 88(1), 1990.

[Auz99] Bruno Auzet. Recherche de caractéristiques morphologiques impliquées dans le processus de localisation auditive. Master's thesis, Ecole Nationale Supérieure Louis Lumière, 1999.

[Bat67] D.W. Batteau. The role of pinna in human localization. *B. Proc. Roy. Soc.*, 168, 1967.

[BDR99a] D.S. Brungart, N.I. Durlach, and W.M. Rabinowitz. Auditory localization of nearby sources. i. head-related transfer functions. *J. Acoust. Soc. Am.*, 106(3) :1465–1479, 1999.

[BDR99b] D.S. Brungart, N.I. Durlach, and W.M. Rabinowitz. Auditory localization of nearby sources. ii. localization of a broadband source. *J. Acoust. Soc. Am.*, 106(4) :1956–1968, 1999.

[Bea96] B. Beaufils. *Statistiques appliquées à la psychologie.* Ed. Breal, Collection LEXIFAC, 1996.

[Beg92] D.R. Begault. Perceptual effects of synthetic reverberation on three-dimensional audio systems. *J. Audio Eng. Soc.*, 40 :895–904, 1992.

[Beg94] D. Begault. *3D Sound for Virtual Reality and Multimedia.* MIT Press, 1994.

[Ber49] L.L. Beranek. *Acoustical measurements.* New-York, 1949.

[BKC92] B. Beliczynski, I. Kale, and G.D. Cain. Approximation of fir by iir digital filters : an algorithm based on balanced model reduction. *IEEE Transactions on Signal Processing*, 40(3), 1992.

[BL95] J. Blauert and H. Lehnert. Binaural technology and virtual reality. *CIARM*, pages 3–10, 1995.

[Bla97] J. Blauert. *Spatial Hearing, the Psychophysics of human sound localization.* MIT Press, first published in 1974, re-edited in 1997.

[Blo96] M.A. Blommer. *Pole-zero modeling and principal component analysis of head-related transfer functions.* PhD thesis, University of Michigan, 1996.

[BPA+91] U. Burandt, C. Posselt, S. Ambrozis, M. Hosenfeld, and V. Knauff. Anthropometric contribution to standardising mannequins for artificial-head microphones and to measuring headphones and ear protectors. *Applied Ergonomics*, pages 373–378, 1991.

[Bro95] A.W. Bronkhorst. Localization of real and virtual sound sources. *J. Acoust. Soc. Am.*, 98 :2542–2553, 1995.

[BS75] M.D. Burkhard and R.M. Sachs. Anthropomorphic manikin for acoustic research. *J. Acoust. Soc. Am.*, 58(1) :214–222, July 1975.

[BS95] A.J. Bell and T.J. Sejnowski. An information-maximisation approach to blind separation and blind deconvolution. *Neural Computation*, 1995.

[BSLD98] A. Buja, D.F. Swayne, M.L. Littman, and N. Dean. Xgvis : Interactive data visualization with multidimensional scaling. *http ://www.research.att.com/areas/stat/xgobi, ou ftp ://ftp.research.att.com/dist/xgobi*, 106, 1998.

[BW94] M.A. Blommer and G.H. Wakefield. On the design of pole-zero approximations using a logarithmic error measure. *IEEE Trans. Signal Process.*, 42(11) :3245–3248, 1994.

[BW97] M.A. Blommer and G.H. Wakefield. Pole-zero approximations for head-related transfer functions using a logarithmic error criterion. *IEEE Trans. on Speech and Audio Process.*, 5(3) :278–287, 1997.

[Car94] Jean-François Cardoso. On the performance of orthogonal source separation algorithms. In *Proc. EUSIPCO*, pages 776–779, Edinburgh, September 1994.

[Car98] J.-F. Cardoso. Blind signal separation : statistical principles. *Proceedings of IEEE, Special issue on blind identification and estimation*, 9(10) :2009–2025, 1998.

[Car99a] J.-F. Cardoso. *Adaptative unsupervised learning*, chapter Entropic contrasts for source separation. 1999.

[Car99b] J.-F. Cardoso. High-order contrasts for independent component analysis. *To appear in Neural Computation*, 1999.

[CB89] D.H. Cooper and J.L. Bauck. Prospects for transaural recording. *J. Audio Eng. Soc.*, 37(1/2), Janvier/Février 1989.

[Cha96] N. Chateau. *Localisation de sources sonores multiples dans l'hémisphère supérieur.* PhD thesis, Université de la Méditerranée - Aix-Marseille II., 1996.

[Col62] P.D. Coleman. Failure to localize the source distance of an unfamiliar sound. *J. Acoust. Soc. of Am.*, 34 :345–346, 1962.

[Com94] P. Comon. Independent component analysis, a new concept ? *Signal Processing*, 36(3), 1994. Special issue on High-Order Statistics.

[CS93] Jean-François Cardoso and Antoine Souloumiac. Blind beamforming for non Gaussian signals. *IEEE Proceedings-F*, 140(6) :362–370, December 1993.

[CT57] E. Churchill and B. Truett. Metrical relations among dimensions of the head and face. Technical report, WADC Tech. Rep. 56-621, ASTIA docum. n° AD 110629, 1957.

[CVVH92] J. Chen, B. Van Veen, and K. Hecox. External ear transfer function modeling : a beamforming approach. *J. Acoust. Soc. Am.*, 91 :1333–1344, 1992.

[CVVH96] J. Chen, B.P. Van Veen, and K.E. Hecox. Methods and apparatus for producing directional sound. *United States Patent*, (5500900), March 1996.

[Dan00] J. Daniel. *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia.* PhD thesis, Université de Paris VI, 2000.

[Dat88] J. Dattoro. The implementation of recursive digital filters for high-fidelity audio. *Journal of the Audio Engineering Society*, 36-11 :851–877, 1988.

[Del95] J.-P. Delmas. *Eléments de théorie du signal : les signaux déterministes.* Ellipses, 1995.

[Dem99] C. Demars. *Représentations bidimensionnelles d'un signal de parole - éléments de monographie.* LIMSI-CNRS, 1999.

[Der97] P. Derogis. *Analyse des vibrations et du rayonnement de la table d'harmonie d'un piano droit et conception d'un système de reproduction du champ acoustique.* PhD thesis, Université du Maine, 1997.

[DGR93a] Ph. Depalle, G. Garcia, and X. Rodet. Analysis of sound for additive synthesis : tracking of partials using hidden markov models. *Proceedings of ICMC*, pages 94–97, 1993.

[DGR93b] Ph. Depalle, G. Garcia, and X. Rodet. Tracking of partials for additive sound synthesis using hidden markov models. *Proceedings of IEEE*, 1993.

[DM98] E. Dudouet and J. Martin. Analyses multidimensionnelles des hrtf pour la spatialsation de sources sonores. Technical report, CSTB, 1998.

[DR95] Y. Ding and D. Rossum. Filter morphing of parametric equalizers and shelving filters for audio signal processing. *J. Audio Eng. Soc.*, 43(10), 1995.

[Dre67] H. Dreyfus. *The measure of man : human factors in design.* 1967.

[EAT98] M. Evans, J. Angus, and A. Tew. Analyzing head-related transfer function measurements using surface spherical harmonics. *J. Acoust. Soc. Am.*, 104(4), October 1998.

[EM95] M. Emerit and J. Martin. Head-related transfer functions and high-order statistics. *Proceedings of the 15th Int. Congress on Acoustics (Trondheim)*, 1995.

[Fil00] T. Fillon. Etude de l'interpolation des fonctions "head-related transfer functions. Master's thesis, Ecole Nationale Supérieure des Télécommunications, 2000.

[FP84] B. Friedlander and B. Porat. The modified yule-walker method of arma spectral estimation. *IEEE Trans. Aerospace Electronic Systems*, 20(2) :158–173, 1984.

[Gai93] W. Gaik. Combined evaluation of interaural time and intensity differences : psychoacoustic results and computer modeling. *J. Acoust. Soc. Am.*, 94(1) :98–110, 1993.

[Gar68] M.B. Gardner. Historical background of the haas and/or precedence effect. *J. Acoust. Soc. Am.*, 43 :1243–1248, 1968.

[Gar94] W.R. Gardner. *Modeling and Quantization techniques for speech compression systems.* PhD thesis, University of California in San Diego, 1994.

[Gar97] W.G. Gardner. *3-D Audio Using Loudspeakers.* PhD thesis, MIT, 1997.

[Gar99] W. G. Gardner. Reduced-rank modeling of head-related impulse responses using subset selection. *IEEE Workshop on Applications of signal Processing to Audio and Acoustics*, 1999.

[Gen84] K. Genuit. *A model for the description of outer-ear transmission characteristics.* PhD thesis, faculty of Electrical Engineering in Düsseldorf, 1984.

[GGE+95] R. Gilkey, M. Good, M. Ericson, J. Brinkman, and J. Stewart. A pointing technique for rapidly collecting localization responses in auditory research. *Behavior Research Methods, Instruments and Computers*, 27(1) :1–11, 1995.

[Gir96] F. Giron. *Investigations about the directivity of sound sources.* PhD thesis, University of Ruhr-Bochum, 1996.

[Glo73] G. Glowatzki. *Der vermessene Mensch*, chapter Wissenschaftliche Anthropometrie - Anthropologische Methoden und ihre Anwendung. 1973.

[GM94] B. Gardner and K. Martin. Hrtf measurements of a kemar dummy head microphone. *MIT Media Lab Technical report 280*, 1994. http ://sound.media.mit.edu/KEMAR.html.

[Gre84] Y. Grenier. *Modélisation de signaux non-stationnaires.* PhD thesis, Université d'Orsay, 1984.

[GVL96] G.H. Golub and C.F. Van Loan. *Matrix computation.* 1996.

[GVM96] H.G. Giuliano, A.G. Velis, and A.M. Mendez. The reverberation chamber at the laboratorio de acustica y luminotecnia of the comision de investigaciones cientificas. *Applied Acoustics*, 40(1), 1996.

[Han94] H.L. Han. Measuring a dummy head in search of pinna cues. *J. Audio Eng. Soc.*, 42 :15–36, 1994.

[Har83] W.M. Hartmann. Localization of sound in rooms. *J. Acoust. Soc. Am.*, 74 :1380–1391, 1983.

[HBS99] K. Hartung, J. Braasch, and S. Sterbing. Comparison of different methods for the interpolation of head-related transfer functions. *16th Conference of the Audio Eng. Soc.*, 1999.

[HK97] J. Huopaniemi and M. Karjalainen. Review of digital filter design and implementation methods for 3d-sound. *Presented at the 102nd Convention of the Audio Eng. Soc. in Munich (Germany)*, Preprint 4461(I4), 1997.

[HKKM73] J.R. Haskew, J.M. Kelly, R.M. Kelly, and T.H. McKinney. Results of a study of the linear prediction vocoder. *IEEE Transactions on communications*, COM-21(9), 1973.

[HM00] P.D. Hatziantoniou and J.N. Mourjopoulos. Generalized fractional-octave smoothing of audio and acoustic responses. *J. Audio Eng. Soc.*, 2000.

[HS97] J. Huopaniemi and J.O. Smith. Spectral and time-domain preprocessing and the choice of modeling error criteria for binaural digital filters. *Presented at the 16th Conference of the Audio Eng. Soc. in Rovaniemi (Finland)*, Preprint 4461(I4), 1997.

[Huo99] J. Huopaniemi. *Virtual Acoustics and 3D sound in multimedia signal processing.* PhD thesis, Helsinki University of Technology, 1999.

[HW74] J. Hebrank and D. Wright. Are two ears necessary for localization of sound sources on the median plane ? *J. Acoust. Soc. Am.*, 56 :935–938, 1974.

[HY83] Y. Hiranika and H. Yamasaki. Envelope representations of pinna impulse responses relating to three-dimensional localization of sound sources. *J. Acoust. Soc. Am.*, 73 :291–296, 1983.

[Hyv99] A. Hyvärinen. Survey on independent component analysis. *Neural Computing Surveys*, 2 :94–128, 1999.

[HZK99] J. Huopaniemi, N. Zacharov, and M. Karjalainen. Objective and subjective evaluation of head-related transfer function filter design. *J. Audio Eng. Soc.*, 47(4) :218–239, 1999.

[Ita75] F. Itakura. Line spectrum representation of linear predictive coefficients of speech signals. *J. of the Acoust. Soc. of Am.*, 1975.

[JCV96] J.-M. Jot, L. Cerveau, and G. Vandernoot. Analysis and synthesis of room reverberation based on a statistical time-frequency model. *131st Meeting of the Acoust. Soc. of Am.*, Mai 1996.

[JLP99] J.-M. Jot, V. Larcher, and J.-M. Pernaux. A comparative study of 3-d audio encoding and rendering techniques. *Presented at the 16^{th} conference of the Audio Eng. Soc. in Rovaniemi (Finland)*, 1999.

[JLW95] J.-M. Jot, V. Larcher, and O. Warusfel. Digital signal processing issues in the context of binaural and transaural stereophony. *Presented at the 98th convention of the Audio Eng. Soc. in Paris*, Preprint 3980(I6), February 1995.

[Jot92] J.-M. Jot. *Etude et réalisation d'un spatialisateur de sons par modèles physiques et perceptifs.* PhD thesis, Télécom Paris, Département Signal, 1992.

[JWL98] J.-M. Jot, S. Wardle, and V. Larcher. Approaches to binaural synthesis. *Presented at the 105^{th} convention of the Audio Eng. Soc. in San Francisco*, N°4861(K4), 1998.

[Kap81] W. Kaplan. *Advanded Mathematics for engineers.* 1981.

[Kat98] B.F.G. Katz. *Measurement and calculation of individual head-related functions using a boundary element model including the measurement and effect of skin and hair impedance.* PhD thesis, Pennsylvania State University, 1998.

[KC95a] A. Kulkarni and H.S. Colburn. Efficient finite impulse response models of the head-related transfer functions. *J. Acoust. Soc. of Am.*, 97, 1995.

[KC95b] A. Kulkarni and H.S. Colburn. Infinite impulse response models of the head-related transfer functions. *J. Acoust. Soc. of Am.*, 97, 1995.

[KIC95] A. Kulkarni, S.K. Isabelle, and H.S. Colburn. On the minimum-phase approximation of head-related transfer functions. *IEEE ASSP Workshop*, Octobre 1995.

[KIC99] A. Kulkarni, S.K. Isabelle, and H.S. Colburn. Sensitivity of human subjects to hrtf phase spectra. *J. Acoust. Soc. of Am.*, 105(5), May 1999.

[KL81] S.Y. Kung and D.W. Lin. Optimal hankel-norm model reductions : multivariable systems. *IEEE Transactions on Automatic Control*, AC-26(4), Août 1981.

[Kuh77] G.F. Kuhn. Model for the interaural time differences in the azimuthal plane. *J. Acoust. Soc. Am.*, 62(1), Juillet 1977.

[Kuh79] G.F. Kuhn. The pressure transformation from a diffuse sound field to the external ear and to the body and head surface. *J. Acoust. Soc. Am.*, 65(4), April 1979.

[Lab90] Inc. Cyberware Laboratory. 4020/rgb 3d scanner with color digitizer. Technical report, available information at http ://www.cyberware.com/pressReleases/index.html, 1990.

[Lab00] A. Laborie. Reconstruction du format b à partir des canaux enregitrés avec un micro soundfield. Master's thesis, ENST, département Traitement du Signal et des Images, 2000.

[Lar94a] V. Larcher. Interpolation de filtres audio-numériques appliquée à la reproduction spatiale des sons sur écouteurs. Master's thesis, Ecole Nationale Supérieure des Télécommunications, 1994.

[Lar94b] J. Laroche. *Traitement des signaux audio-fréquences.* Télécom Paris, département Signal, 1994.

[Lar95] V. Larcher. *Paramétrisation des fonctions de transfert binaurales.* Mémoire d'Ingénieur de l'Ecole Nationale Supérieure des Télécommunications, 1995.

[Lar96] V. Larcher. *Correction individuelle pour la reproduction binaurale d'enregistrements effectués dans l'habitacle d'un véhicule automobile.* Mémoire de DEA ATIAM, Université de Paris VI, 1996.

[Lar98] J. Laroche. Using resonant filters for the synthesis of time-varying sinusoids. *Presented at the 105^{th} convention of the Audio Eng. Soc. in San Francisco*, (Preprint 4782), 1998.

[Lar99] V. Larcher. Comparaison de deux techniques de reproduction sonore tridimentionnelle sur casque. Technical report, Rapport d'avancement, 1999.

[LJ97] V. Larcher and J.-M. Jot. Techniques d'interpolation de filtres audio-numériques. application à la reproduction spatiale des sons sur écouteurs. *Presented at the 4^{th} French congress on Acoustics*, 97-100, 1997.

[LJGW00] V. Larcher, J.-M. Jot, J. Guyard, and O. Warusfel. Study and comparison of efficient methods for 3d audio spatialization based on linear decomposition of hrtf data. *Presented at the 108^{th} convention of the Audio Eng. Soc. in Paris*, N°5097(E1), 2000.

[LJV98] V. Larcher, J.-M. Jot, and G. Vandernoot. Equalization methods in binaural technology. *Presented at the 105^{th} convention of the Audio Eng. Soc. in San Francisco*, N°4858(K4), 1998.

[LPM96] E.A. Lopez-Poveda and R. Meddis. A physical model of sound diffraction and reflections in the human concha. *J. Acoust. Soc. Am.*, 100(5), 1996.

[Mar87] W. Martens. Principal components analysis and resynthesis of spectral cues to perceived direction. *ICMC proceedings*, pages 274–281, 1987.

[Mar96a] M. Marin. *Etude de la localisation en restitution du son pour la téléconférence de haute Qualité.* PhD thesis, Université du Maine, 1996.

[Mar96b] M. Marolt. A new approach to hrtf audio spatialization. *Proceedings of ICMC*, pages 365–367, 1996.

[MCM+99] P. Minnaar, F. Christensen, H. Moller, S. K. Olesen, and J. Plogsties. Audibility of all-pass components in binaural synthesis. *Presented at the 106^{th} convention of the Audio Eng. Soc. in Munich*, preprint n° 4911L5, 1999.

[MFBBG00] J. Marquez-Flores, T. Bousquet, I. Bloch, and G. Grangeat. Laser-scan acquisition of head models for dosimetry of hand-held mobile phones. *projet COMOBIO (RNRT program) - ENST / Alcatel Corporate Research Center*, 2000.

[MG90] J.C. Middlebrooks and D.M. Green. Directional dependence of interaural envelope delays. *J. Acoust. Soc. Am.*, 87 :2149–2162, 1990.

[MG92] J. Middlebrooks and D. Green. Observations on a principal components analysis of head-related transfer functions. *J. Acoust. Soc. Am.*, 92(1) :597–599, 1992.

[MHJS99] H. Moller, D. Hammershoi, C. Jensen, and M. Sorensen. Evaluation of artificial heads in listening tests. *J. Audio Eng. Soc.*, 47(3) :83–99, Mars 1999.

[MHVK97] J. Mackenzie, J. Huopaniemi, V. Välimäki, and I. Kale. Low-order modeling of head-related transfer functions unsing balanced model truncation. *IEEE Signal Processing letters*, 4(2), 1997.

[Mid99a] J. Middlebrooks. Individual differences in external-ear transfer functions reduced by scaling in frequency. *J. Acoust. Soc. Am.*, 106(3) :1480–1492, 1999.

[Mid99b] J. Middlebrooks. Virtual localization improved by scaling non-individualized external-ear functions in frequency. *J. Acoust. Soc. Am.*, 106(3) :1493–1509, 1999.

[MJHS95] H. Moller, C.B. Jensen, D. Hammerhoi, and M.F. Sorensen. Design criteria for headphones. *J. Audio Eng. Soc.*, 43(4) :218–232, April 1995.

[MJHS96] H. Moller, C. Jensen, D. Hammershoi, and M. Sorensen. Using a typical human subject for binaural recording. *100^{th} Convention of the Audio Eng. Soc in Copenhagen*, Preprint n°4157($C10$), 1996.

[MKBG90] J.N. Mourjopoulos, E.D. Kyriakis-Bitzaros, and C.E. Goutis. Theory and real-time implementation of time-varying digital audio filters. *J. Audio Eng. Soc.*, 38(7/8), 1990.

[MM77] S. Mehrgardt and V. Mellert. Transformation characteristics of the external human ear. *J. Acoust. Soc. Am.*, 61(6) :1567–1576, 1977.

[MM90] J.C. Makous and J.C. Middlebrooks. Two-dimensional sound localization by human listeners. *J. Acoust. Soc. Am.*, 87(5), 1990.

[Moh97] M.J. Mohlenkamp. *A fast transform for spherical harmonics.* PhD thesis, Yale University, 1997.

[Mol92] H. Moller. Fundamentals of binaural technology. *Applied Acoustics*, 36 :171–218, 1992.

[MSHJ95] H. Moller, M.F. Sorensen, D. Hammerhoi, and C.B. Jensen. Head-related transfer functions of human subjects. *J. Audio Eng. Soc.*, 43(5) :300–321, Mai 1995.

[MZ00] E.A. Middlebrooks, J.and Macpherson and A.O. Zekiye. Psychophysical customization of directional transfer functions for virtual sound localization. *J. Acoust. Soc. Am.*, 108(6) :3088–3091, 2000.

[OP84] S.R. Oldfield and S.P.A. Parker. Acuity of sound localization : a topography of auditory space. i. normal hearing conditions. *Perception*, 13 :581–600, 1984.

[Opp74] Oppenheim. *Digital signal processing.* Number pp.337-345. Prentice Hall, 1974.

[PHJ93] H.N. Pollack, S.J. Hunter, and J.R. Johnson. Heat flow from the earth's interior : Analysis of the global data set. *Rev. Geophys.*, 31 :267–280, 1993.

[PKH99] V. Pulkki, M. Karjalainen, and J. Huopaniemi. Analysing virtual sound sources using a binaural auditory model. *J. Audio Eng. Soc.*, 47(4), 1999.

[Pla79] H.J. Platte. *Zur Bedeutung der Aussenohrübertragungseigenschaften für den Narichtenempfänger menschliches Gehör.* PhD thesis, Aachen Technishe Hochschule, 1979.

[PM81] D.R. Perrott and A.D. Musicant. Dynamic minimum audible angle : binaural spatial acuity with moving sound sources. *The Journal of auditory research*, 21 :287–295, 1981.

[Pon92] F. Poncet. Simulation de la localisation de sources sonores dans l'espace. Master's thesis, Télécom Paris, département Signal, 1992.

[PW72] B.B. Platt and D.H. Warren. Auditory localization : the importance of eye movements and a textured visual environment. *Percept. Psychophys.*, 12 :245–248, 1972.

[Ray07] Lord Rayleigh. On our perception of sound direction. *Philos. mag.*, 13 :214–232, 1907.

[RC85] R. Rabenstein and R. Czanach. Stability of recursive time-varying digital filters by state vector transformation. *Signal Processing*, pages 75–92, 1985.

[ref59] *The analysis of sensations.* Dover Publications, New York, 1959.

[RH85] B. Rakerd and W.M. Hartmann. Localization of sound in rooms, ii : the effects of a single reflecting surface. *J. Audio Eng. Soc.*, 78(2) :524–533, 1985.

[Ros91] D. Rossum. The armadillo coefficients encoding scheme for digital audio filters. *in Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 1991.

[RV89] D.D. Rife and J. Vanderkooy. Tranfer-function measurement with maximum-length sequences. *J. Audio Eng. Soc.*, 37(6) :419–444, 1989.

[RW96] P. Runkle and G. Wakefield. On the perceptual optimization of synthetic acoustical systems. *ICMC Proceedings*, pages 210–211, 1996.

[Sap90] G. Saporta. *Probabilités. Analyse des données et statistique.* Editions Technip, 1990.

[SB00] A. Schmitz and H. Bietz. Free-field diffuse-field transformation of artificial heads. *107th conference of the Audio Eng. Soc. in New-York*, (Preprint n°4801), 2000.

[Sch76] H.W. Schlüssler. A stability theorem for discrete systems. *IEEE transactions on acoustics, speech and signal processing*, ASSP-24(1), 1976.

[SH96] J. Sanvad and D. Hammershoi. Binaural auralization. comparison of fir and iir filter representation of hirs. *Proceedings of the 96^{th} convention of the Audio Eng. Soc. of America*, page n°3862(P4.5), 1996.

[Sha80] E.A.G. Shaw. *Acoustical factors affecting hearing aid performance*, chapter The acoustics of the external ear. University Park Press, 1980.

[Sha82] E.A.G. Shaw. *Localization of sound : theory and applications*, chapter External ear Response and Sound localization. Amphora press, 1982.

[Sha88] E.A.G. Shaw. Diffuse field response, receiver impedance, and the acoustical reciprocity principle. *J. Acoust. Soc. of Amer.*, 84 :2284–2287, 1988.

[Sha97a] E.A.G. Shaw. *Encyclopedia of acoustics*, chapter Acoustical characteristics of the outer ear. John Wiley and sons, Inc., 1997.

[Sha97b] E.A.G. Shaw. *Real and Virtual environments*, chapter Binaural and Spatial Hearing. Lawrence Erlbaum Associates, 1997.

[SHH94] S. Shimada, N. Hayashi, and S. Hayashi. A clustering method for sound localization transfer functions. *J. Audio Eng. Soc.*, 42(7/8) :577–583, 1994.

[Sib99] A. Sibbald. Virtual ear technology. Technical report, Sensaura, www.sensaura.uk, 1999.

[SJ84] F.K. Soong and B.-H. Juang. Line spectrum pair (lsp) and speech data compression. *Proceedings of ICASSP*, 1984.

[SK62] M.R. Schroeder and H. Kutruff. On frequency response curves in rooms. comparison of experimental, theoretical and monte carlo results for average frequency spacing between maxima. *J. Acoust. Soc. of Am.*, 34, 1962.

[SMB65] K. Steiglitz and L.E. Mc Bride. A technique for the identification of linear systems. *IEEE Trans. Automatic Control*, AC-10 :461–464, 1965.

[Smi83] J.O. Smith. *Techniques for digital filtering design and system identification with the violin.* PhD thesis, CCRMA, département of Music, 1983.

[SRY81] S.S. Schiffman, M.L. Reynolds, and F.W. Young. Introduction to multidimensional scaling. *Academic Press, London*, 1981.

[SS80] B.R. Shelton and C.L. Searle. The influence of vision on the absolute identification of sound-source position. *Percept. Psychophys.*, 28 :589–596, 1980.

[ST68] E.G. Shaw and R. Teranishi. Sound pressure generated in an external ear replica and human ears by a nearby source. *J. Acoust. Soc. of Amer.*, pages 240–249, 1968.

[The80] G. Theile. *Localization in the superposed sound field.* PhD thesis, Technical university of Berlin, 1980.

[The86] G. Theile. On the standardization of the frequency response of high-quality studio headphones. *J. Audio Eng. Soc.*, 34(12), December 1986.

[Til93] A. R. Tilley. *The measure of man and woman : human factors in design.* 1993.

[Tou96] C. Touzé. Modélisation paramétrique de phénomènes acoustiques simples. Technical report, E.N.S.T.A., 1996.

[Val95] V. Valimaki. *Discrete-Time Modeling of Acoustics Tubes Using Fractional Delay Filters.* PhD thesis, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, 1995.

[Van95] G. Vandernoot. Analyse temps-fréquence de réponses impulsionnelles de salles. application pour la simulation sonore sur écouteurs. Technical report, E.N.S.T., 1995.

[Van01] G. Vandernoot. *Evaluation et optimisation de l'écoute binaurale.* PhD thesis, Université de Paris VI, 2001.

[Vin97] T. Vingtrinier. De la commutation des filtres numériques. Master's thesis, Ecole Nationale Supérieure des Télécommunications, 1997.

[VLM95] V. Valimaki, T.I. Laakso, and J. Mackensie. Elimination of transients in time-varying allpass fractinal delay filters with application to digital waveguide madeling. *Proceedings of ICMC*, 1995.

[VM75] R. Viswanathan and J. Makhoul. Quantization properties in transmission parameters in linear predictive coding. *IEEE Trans. Acoust., Speech, and Signal Process.*, ASSP-23 :309–321, 1975.

[VN86] W. Verhelst and P. Nilens. A modified-superposition speech synthesizer and its applications. *Proceedings of ICASSP*, 1986.

[VS87] I. Veit and H. Sander. Production of spatially limited diffuse sound field in an anechoïc room. *J. Audio Eng. Soc.*, 35(3), March 1987.

[WAKW93] E. Wenzel, M. Arruda, D. Kistler, and F. Wightman. Localization using non-individualized head-related transfer functions. *J. Acoust. Soc. Amer.*, 1993.

[Wal40] H. Wallach. The role of head movements and vestibular and visual cues in sound localization. *J. Exp. Psychol.*, 27 :339–368, 1940.

[WK89a] F. Wightman and D. Kistler. Headphone simulation of free-field listening, i : Stimulus synthesis. *J. Acoust. Soc. Am.*, 85(2) :858–867, February 1989.

[WK89b] F. Wightman and D. Kistler. Headphone simulation of free-field listening, ii : Psychophysical validation. *J. Acoust. Soc. Am.*, 85(2) :868–878, February 1989.

[WK92] F. Wightman and D. Kistler. The dominant role of low-frequency interaural time differences in sound localization. *J. Acoust. Soc. Am.*, 91 :1648–1661, 1992.

[WK93] F. Wightman and D. Kistler. Multidimensional scaling analysis of head-related transfer functions. *IEEE Digital Audio Workshop*, 1993.

[WK99] F. Wightman and D. Kistler. Resolution of front-back ambiguity in spatial hearing by listener and source movement. *J. Acoust. Soc. Am.*, 105(5) :2841–2853, 1999.

[WL89] Wilkinson and Leland. Systat : the system for statistics. *Evanston, IL*, 1989.

[WW91] E. Wenzel and F. Wightman. Localization with non-individualized virtual acoustics display cues. *Human Factors in computing systems 8th annual conference in New-Orleans*, 1991.

[ZZ88] L.H. Zetterberg and Q. Zhang. Elimination of transients in adaptative filters with application to speech coding. *Signal Processing*, 15, 1988.