Individualisation d'indices acoustiques pour la synthèse binaurale

Sylvain Busson

Académie d'Aix-Marseille
Université de la mediterranée Aix-Marseille II

THÈSE DE DOCTORAT

École Doctorale de
MECANIQUE, PHYSIQUE ET MODÉLISATION

Présentée
par
Sylvain Busson

pour obtenir le grade de
DOCTEUR DE L'UNIVERSITÉ DE LA MEDITERRANÉE
AIX-MARSEILLE II

INDIVIDUALISATION D'INDICES ACOUSTIQUES POUR LA SYNTHÈSE BINAURALE

version provisoire

Rapporteurs et membres du jury

GEORGES CANÉVET	Examinateur
V. RALPH ALGAZI	Rapporteur
CLAUDE DÉPOLLIER	Examinateur
MATHIEU PAQUIER	Examinateur
PIERRE-OLIVIER MATTEI	Directeur de Thèse
DORTE HAMMERSHØÏ	Rapporteur
OLIVIER WARUSFEL	Examinateur

THÈSE

INDIVIDUALISATION D'INDICES ACOUSTIQUES POUR LA SYNTHÈSE BINAURALE

Sylvain Busson

version provisoire

Résumé

La synthèse binaurale est la technique de spatialisation sonore la plus proche de l'écoute naturelle. Elle permet un rendu spatialisé d'une source monophonique à une position donnée avec seulement deux filtres qui correspondent aux oreilles gauche et droite : les HRTF (Head Related Transfer Function). L'inconvénient majeur de la technique binaurale repose sur le fait que les HRTF, liées à la morphologie de l'auditeur, sont propres à chaque utilisateur. Une écoute avec des HRTF non-individuelles comporte des artefacts audibles. Il faut donc acquérir des HRTF individuelles. Cette thèse aborde le problème de l'individualisation de la synthèse binaurale dans le cadre de son implémentation en un retard pur, la différence interaurale de temps (ITD), et un filtre à phase minimale déterminé par le module de la HRTF. Le travail sur l'ITD permet de valider l'implémentation choisie même pour les positions où les HRTF sont mal décrites par des filtres à phase minimale et permet de déterminer, parmi les méthodes classiques de calcul de l'ITD, celles qui estiment une ITD proche de la perception. Une étude expérimentale est aussi menée pour établir la résolution de l'ITD avec l'angle d'élévation. Les résultats indiquent la nécessité perceptive de reproduire les variations de l'ITD en élévation. Une nouvelle formule d'estimation de l'ITD créée sur la base d'un modèle de tête sphérique, la formule de déplacement des oreilles (FDO), est développée pour rendre compte de ces variations. L'optimisation des paramètres de cette formule aux ITD de toute une base de données de HRTF permet d'entrevoir une formulation moyenne convenant pour un grand nombre de personne et pour de nombreuses applications. L'étude s'est ensuite focalisée sur la modélisation du module spectral (filtre à phase minimale). Le travail réalisé sur l'application des méthodes de calcul par éléments de frontière (BEM pour Boundary Element Method) pour l'acquisition de HRTF, indique que cette méthode, peut notamment être utilisée en complément des mesures pour l'acquisition de la partie basse fréquence des HRTF. Une approche originale, qui applique des techniques d'apprentissage statistique, est proposée et étudiée pour la modélisation de HRTF. Un réseau de neurones artificiels (RNA) est entraîné pour calculer des HRTF d'un individu à partir de la connaissance des HRTF mesurées en un nombre réduit de positions. Les premiers résultats sont encourageants : le modèle permet d'atteindre un degré assez fin d'individualisation, ce qui suggère un protocole simplifié d'acquisition de HRTF. Un faible nombre de mesures est acquis et les autres sont prédites par le modèle.

Mots clés : Synthèse binaurale, son 3D, HRTF, filtre à phase minimale, psychoacoustique, ITD, JND, BEM, réseaux de neurones artificiels.

Table des matières

Table des figures

Liste des tableaux

Notations et conventions de langage

- **HRTF** : fonction de transfert binaurale (de l'anglais Head Related Transfer Function)
- **HRIR** : réponse impulsionnelle binaurale (de l'anglais Head Related Impulse Response)
- **ITD** : différence interaurale de temps (de l'anglais Interaural Time Difference)
- **ILD** : différence interaurale de niveau (de l'anglais Interaural Level Difference)
- **JND** : plus petite différence audible (de l'anglais Just Noticeable Difference)
- **MAA** : plus petit écart angulaire audible (de l'anglais Minimum Audible Angle)
- **RNA** : Réseau de Neurones Artificiels
- **CHA** : Classification Hiérarchique Ascendante
- **RIF** : filtre à Réponse Impulsionnelle Finie
- **RII** : filtre à Réponse Impulsionnelle Infinie
- **BEM** : méthode de calcul par éléments de frontière (de l'anglais Boundary Element Method)
- **Hilbert** : transformée de Hilbert
- **ANOVA** : méthode d'analyse de la variance (de l'anglais ANalyse Of VAriance)
- $HRTF_{mixte}$: HRTF non décomposée (par opposition aux composantes à phase minimale et à excès de phase)
- $HRTF_{min}$: composante à phase minimale d'une fonction de transfert binaurale
- $HRTF_{excess}$: composante à excès de phase d'une fonction de transfert binaurale
- $\{ITD \bigoplus HRTF_{min}\}$: implémentation de la synthèse binaurale sous la forme d'un retard pur et d'un filtre à phase minimale
- θ : angle d'azimut
- ϕ : angle d'élévation
- $\mathbb{C}_{p,f}$: espace vectoriel des matrices de dimension $p \star f$ à coefficients complexes
- $\mathbb{R}_{p,f}$: espace vectoriel des matrices de dimension $p \star f$ à coefficients réels

Introduction

Les systèmes de rendu sonore spatialisé connaissent un essor considérable. Ces systèmes sont aujourd'hui utilisés dans de nombreux domaines et depuis peu dans des applications grand public. La spatialisation sonore apporte une autre dimension à la diffusion d'informations visuelles. Dans les systèmes de visio-conférence, les voix de locuteurs distants peuvent être spatialisées. Dans les systèmes de réalité virtuelle, la spatialisation sonore est complémentaire de la vision et l'auditeur est immergé dans un espace sensoriel virtuel. Les jeux multi-média gagnent en réalisme grâce à la spatialisation sonore et l'intérêt porté par le joueur est augmenté. Enfin, la dimension supplémentaire apportée par un rendu sonore spatialisé sert autant pour la sonification d'interface homme-machine que de message d'alerte : par exemple un pilote d'avion de chasse peut suivre la trajectoire d'une cible venant de l'arrière.

La synthèse binaurale est la technique de spatialisation sonore la plus proche de l'écoute naturelle. Le principe de cette technique repose sur la reproduction au niveau des oreilles d'un auditeur de toutes les informations nécessaires pour la construction d'une image sonore extra-crânienne. La synthèse binaurale permet une localisation précise des sources sonores en trois dimensions ainsi qu'un rendu fidèle de l'environnement sonore (effet de salle). Contrairement aux autres systèmes de spatialisation sonore qui nécessitent parfois un nombre important de haut-parleurs, la diffusion des informations binaurales est réalisée grâce à un casque stéréophonique classique. L'utilisation d'un casque, et d'un système de suivi de mouvement, permet une immersion totale de l'auditeur dans une scène sonore sans interaction avec le monde extérieur et la synthèse n'est pas réduite à une zone de reconstruction optimale. De plus, la synthèse binaurale est réalisée grâce à des filtres de spatialisation et les indices perceptifs qui gouvernent la localisation sonore peuvent en être extraits. Cette spécificité autorise l'étude psychoacoustique des indices de localisation.

Le principe de la synthèse binaurale repose sur la connaissance de filtres de spatialisation. Ces filtres contiennent les informations liées aux phénomènes de diffraction, diffusion et réflexion que subit une onde sonore lors de son trajet entre son point source et l'entrée du canal auditif de l'auditeur. Ces filtres sont communément appelés **Head Related Transfer Function** ou **HRTF** et leurs équivalents temporels sont nommés **Head Related Impulse Response** ou **HRIR**. Toutes les informations de spatialisations sont consignées dans les HRTF. C'est un des avantages de la synthèse binaurale par rapport aux autres systèmes de spatialisations sonores qui se basent sur des modèles de reconstruction du champ sonore. La perception d'une source sonore à une position donnée est alors obtenue par la convolution du signal monophonique avec les HRIR droite et gauche

correspondant à la position souhaitée. Les HRTF sont propres à l'individu et à la position de l'espace simulée. Pour assurer un rendu optimal de la synthèse binaurale, il faut utiliser un grand nombre (environ 1000 par oreille) de HRTF individuelles décrivant des positions de sources autour de l'auditeur. Le principe de la mesure de HRTF est de placer des microphones dans les oreilles et d'enregistrer les signaux qui correspondent à différentes positions de source. Les HRTF sont les fonctions de transfert entre les signaux sources et les signaux au niveau des oreilles.

Si le principe de la mesure est simple, sa mise en oeuvre n'en reste pas moins délicate, coûteuse et longue. C'est pourquoi un des axes de recherche sur la synthèse binaurale est dédié à la simplification des procédés d'acquisition des HRTF. Un autre axe de recherche concerne la réduction du coût d'implémentation des filtres binauraux. Cette réduction s'accompagne souvent d'artefacts audibles qu'il convient d'estimer et de corriger. Il est aussi proposé d'adapter la synthèse binaurale plutôt que de mesurer les filtres de spatialisation pour chaque utilisateur : de nombreuses études se sont consacrées à l'individualisation de la synthèse binaurale, c'est-à-dire à l'adaptation des indices binauraux à un individu.

L'ensemble des travaux de thèse présentés ici est dédié au problème de l'individualisation de la synthèse binaurale. Son implémentation $\{ITD \bigoplus HRTF_{min}\}$ est la plus commune : elle est utilisée comme base de travail. Les efforts d'individualisation se sont alors portés sur l'ITD (Interaural Time Difference ou Différence Interaurale de Temps) d'une part et sur les $HRTF_{min}$ (HRTF à phase minimale) d'autre part. Le premier chapitre de la thèse présente la synthèse binaurale dans le contexte de la spatialisation sonore et met en valeur ses spécificités. Les hypothèses simplificatrices ainsi que l'implémentation choisie sont décrites. Les voies d'analyses empruntées pour le travail de thèse sont alors détaillées.

L'ITD (l'indice perceptif de spatialisation le plus important) a été très étudiée dans le passé comme indice de localisation isolé et peu d'études l'ont abordée dans le cadre de l'implémentation choisie. Afin de déterminer la meilleure technique d'estimation de l'ITD, il est proposé de l'estimer du point de vue psychoacoustique. La sensibilité du système auditif à l'ITD est ensuite évaluée. Cette étude fait l'objet du second chapitre. L'individualisation de l'ITD est alors abordée à la lumière de la sensibilité et de la variation inter-individuelle de l'ITD. La validité de l'implémentation choisie est aussi examinée.

Le troisième chapitre décrit les techniques d'acquisition des HRTF par la mesure et par le calcul par éléments de frontières. L'apport de modélisations géométriques simples de la morphologie de l'auditeur pour le calcul de HRTF est évalué.

La mesure d'un grand nombre de HRTF ne peut être réalisée dans le cadre d'application grand public. Le quatrième chapitre décrit une solution proposée pour réduire la complexité de l'acquisition des HRTF individuelles. Cette solution est basée sur la mise en oeuvre d'un réseau de neurones artificiels (RNA). Une étude de faisabilité de la modélisation de HRTF par RNA est menée. Un des objectifs est de réduire le nombre de points de mesure et de prédire les autres points. Cette étude établit les performances de prédiction du modèle neuronal construit.

Enfin le dernier chapitre résume les principaux résultats de la thèse. Un début de réflexion y est proposé quant à l'interaction entre le trio perception/coût/application liée à l'utilisation de la synthèse binaurale.

I

Contexte de l'étude

1 CONTEXTE ET HISTORIQUE DE LA SPATIALISATION SONORE.

La synthèse binaurale est un procédé de reproduction sonore en trois dimensions. Le premier système de reproduction sonore est inventé en 1876 par Bell et n'est autre que le téléphone, suivi une année après par le phonographe de Cros et Edison. Avant ces deux inventions, un événement sonore était indissociable de l'endroit et de l'instant où il était émis. La problématique générale des systèmes de reproduction sonore est de *donner l'illusion d'une scène sonore réelle*. Pour ce faire, il faut restituer à l'auditeur toutes les informations qu'il percevrait en situation d'écoute réelle. Dans une scène sonore, notre système auditif capte principalement le message sonore. Ce message peut être la voix d'un interlocuteur, le phrasé d'un instrument de musique ou encore un signal d'alerte (klaxon). Mais un message sonore ne pourrait rendre compte de la globalité d'une scène sonore et notre système auditif est capable aussi d'identifier le **positionnement des sources sonores** et la nature de leur **environnement acoustique**. La perception de l'environnement acoustique, ou *effet de salle*, est un phénomène complexe lié principalement aux multiples réflexions, atténuations, diffractions et diffusion sur les éléments constitutifs de l'environnement physique autour de la source sonore que subit l'onde acoustique dans sa propagation, de la source jusqu'au tympan. Le schéma représenté en figure I.1 illustre ce propos avec l'exemple d'un musicien soliste écouté par un auditeur dans une salle de concert. Les systèmes de reproduction sonore doivent donc reproduire à la fois le positionnement et les mouvements des sources sonores par rapport à l'auditeur mais aussi l'effet de salle. Un des premiers systèmes prenant en compte ses deux aspects est

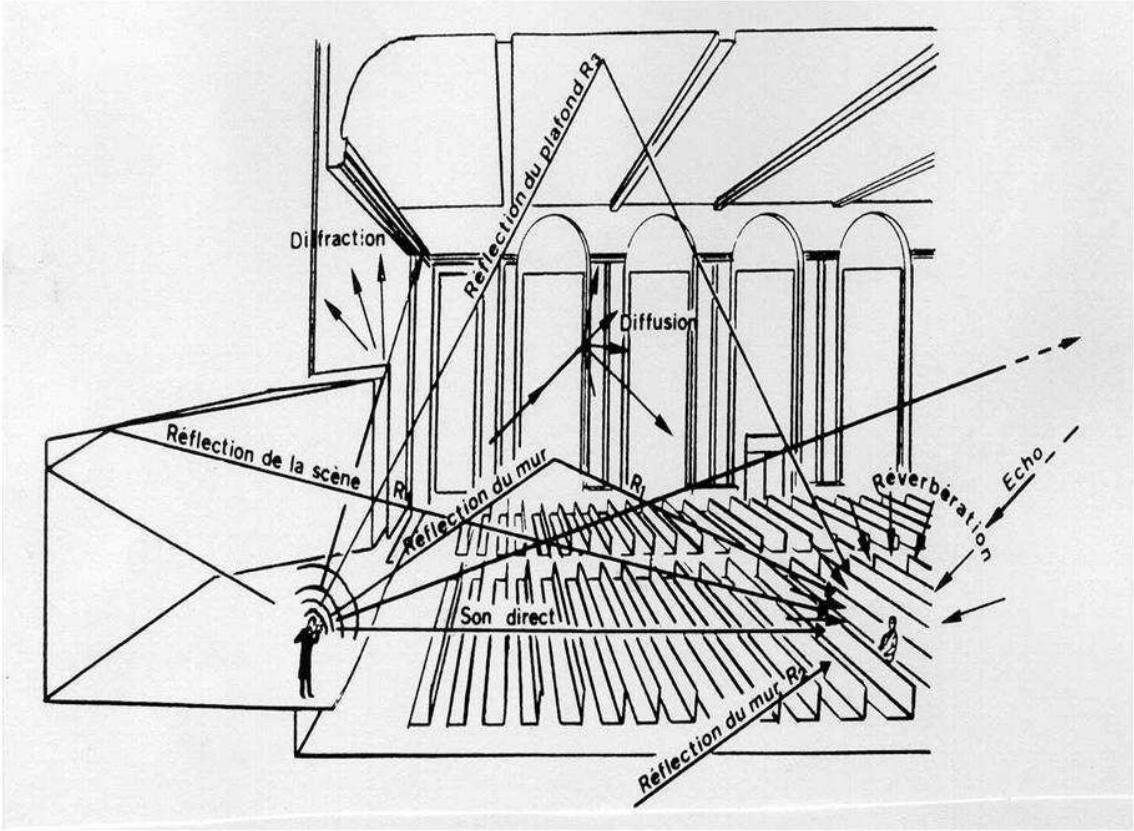


FIG. I.1 – L'effet de salle est la résultante de multiples transformations d'un événement acoustique.

réalisé en 1881 pour l'exposition Universelle de Paris par Ader avec l'expérience du *théâtrophone*. Cette expérience avait pour but de retransmettre un spectacle donné à l'Opéra Garnier jusqu'au Palais de l'Industrie (2 km). La prise de son était assurée par une dizaine de microphones répartis autour de la scène et la reproduction était assurée par des récepteurs téléphoniques. Ader a ainsi mis au point le premier système stéréophonique car les auditeurs pouvaient entendre avec deux écouteurs téléphoniques, les informations venant de la droite et celles de la gauche (auditeur placé au centre de la figure I.3).

Cette expérience publique eut un succès considérable et chaque soir d'Opéra durant l'exposition Universelle, la salle de restitution du théâtrophone était remplie. Les auditeurs utilisant les deux écouteurs décrivaient une sensation d'espace et de relief et un très bon rendu des mouvements des acteurs. Le théâtrophone permet d'illustrer toutes les composantes des systèmes de spatialisation sonore :

Création du contenu Le contenu peut être une scène sonore réelle ou artificielle. Dans le cas du théâtrophone, l'événement sonore est l'information captée par les microphones. Dans les systèmes actuels, il peut s'agir du contenu pré-enregistré ou créé en temps réel.

Encodage L'encodage est lié au contenu. Pour des scènes sonores réelles, l'encodage dépend du système microphonique de captation, mais peut être modifié pour un

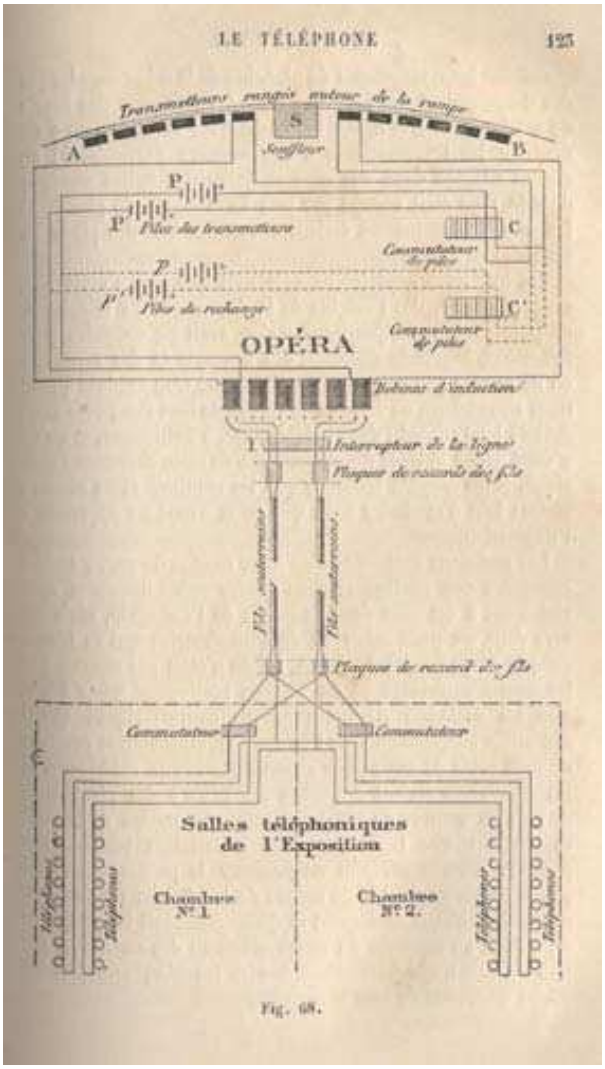
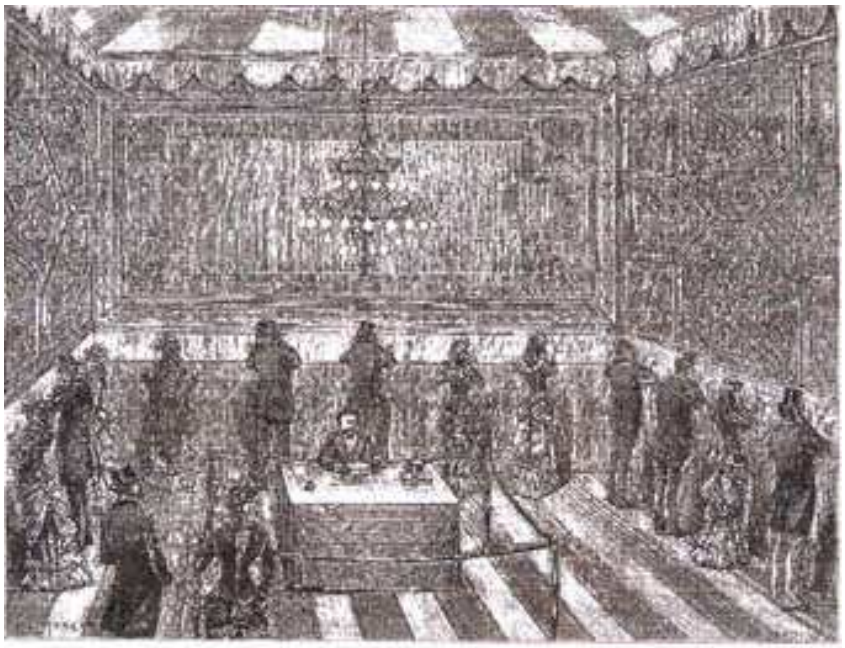


FIG. I.2 – Schéma du dispositif du théâtrophone.

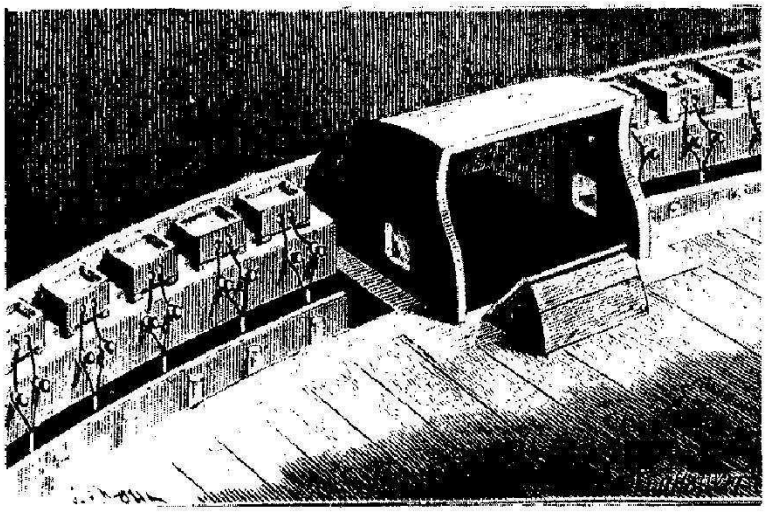
format particulier plus adapté au transport. L'encodage du théâtrophone correspond au nombre et à la disposition des microphones autour de la scène (cf. fig. I.3). Dans le cas de scène artificielle, l'événement peut être directement encodé à sa création.

Transport Selon le canal de transport de l'information, un type d'encodage peut être privilégié. C'est l'étape la plus contraignante et nombre de recherches son axées sur la réduction du débit d'information (voir les travaux sur le BCC (Binaural Cue Coding) [Faller and Baumgarte (2002); Baumgarte and Faller (2003); Faller and Baumgarte (2003)]. Cette étape est encore plus délicate quand il s'agit d'une application de diffusion en temps réél. Autant de canaux de transmissions téléphoniques que de capteurs ont été utilisés pour le théâtrophone. Ce genre d'architecture est bien sur inapplicable à grande échelle.

Décodage et restitution Le format de décodage dépend du système de restitution et sera différent selon qu'est utilisé un système à deux, quatre ou N haut-parleurs. Un premier décodage est réalisé pour l'expérience de Ader : un mixage des voies



a) Récepteurs téléphoniques du théâtrophone.



b) Captation du son près de la scène du théâtrophone.

Fig. I.3 – Le théâtrophone : diffusion (en haut) et captation du spectacle (en bas).

de droite et des voies de gauche est effectué pour l'écoute sur deux écouteurs.

L'exemple du théâtrophone peut paraître pittoresque, mais peu de systèmes actuels sont capables de transmettre en temps réel une prise de son multi-canal. Plus tard en 1930, l'ingénieur anglais Blumlein étudie la stéréophonie et le procédé fut exploité de manière commerciale par Disney pour le film *Fantasia* en 1939. Ce fut le premier film disposant d'une bande sonore diffusée en stéréophonie. Ce procédé est encore très largement utilisé dans le cinéma, la télévision, la radio et la prise de son musicale. Bien plus tard, les avancées technologiques et l'avénement de l'ordinateur, ont permis l'étude et la création de procédés innovants tels que les systèmes multi-canaux, Ambisonic et la synthèse binaurale. Ces procédés améliorent les effets de spatialisation par rapport à la stéréophonie et permettent notamment un meilleur rendu de l'effet de salle. Ces techniques se sont longtemps heurtées à la complexité de leur installation et l'augmentation du nombre de procédés a conduit à des problèmes de compatibilité de formats. Ces procédés sont désormais de plus en plus utilisés grâce aux progrès techniques et scientifiques de ces 20 dernières années. De nouvelles applications grand public sont apparues et les systèmes de spatialisation sonore connaissent aujourd'hui un essor considérable (jeux sur ordinateur, cinéma, home-cinéma, réalité virtuelle, etc...). Les systèmes de télécommunication utilisent encore peu les informations de spatialisation principalement du fait d'une bande passante réduite. Une partie des recherches dans le domaine du son spatialisé est consacrée à la définition de codage et d'encodage optimaux pour les télécommunications et les récentes avancées ont permis des applications telles que le *mur de téléprésence* ou la diffusion sur téléphone mobile d'une bande d'annonce spatialisée du film *Stars War* via des écouteurs stéréophoniques.

La réalisation de procédé de spatialisation sonore doit prendre en compte les performances du système auditif. La première partie de ce chapitre présente les résultats fondamentaux issus de la littérature consacrée à l'étude des performances de localisation sonore et des indices perceptifs qui la gouvernent. Deux types de système de spatialisation sonore peuvent être distingués : ceux qui visent la reproduction d'un champ sonore et ceux qui utilisent des indices perceptifs pour rendre compte d'effets de spatialisation. La troisième partie de ce chapitre décrit brièvement les différents procédés les plus répandus. La quatrième partie est consacrée à la synthèse binaurale et expose ses spécificités. La cinquième partie présente les problématiques globales liées à l'utilisation de l'écoute binaurale pour une large audience et introduit le concept de *binaumaton*. Enfin, la dernière partie est consacrée au plan détaillé de l'ensemble des travaux réalisé pour la thèse ainsi qu'aux problématiques abordées.

2 LA LOCALISATION SONORE

La localisation sonore désigne la capacité du système auditif à déterminer la position relative d'une source sonore. Cette capacité est à différencier de la perception de l'espace sonore qui correspond à la perception de l'*effet de salle*. Dans la suite du paragraphe, et pour des raisons de clarté de l'exposé, la description des mécanismes de localisation sonore se restreint au cas d'une seule source en écoute en champ libre. Cette description ne prend pas en compte la perception de la distance, car d'une part, les performances d'estimation de la distance absolue d'une source sonore en écoute anéchoïque sont faibles et d'autre part, les bases de données de HRTF sont généralement mesurées à rayon

constant. De plus, il semble qu'une loi classique de décroissance du niveau sonore de 6 dBSPL par doublement de la distance est satisfaisante du point de vue perceptif.

Une fois présentés les systèmes de coordonnées servant à repérer une source sonore, les performances de localisation du système auditif sont décrites succinctement. Les mécanismes de localisation sont ensuite présentés.

2.1 Le référentiel auditeur

Pour repérer la position d'une source sonore par rapport à un auditeur, il est pratique d'utiliser un système de coordonnées dont le centre est confondu avec le centre de la tête. Le référentiel est celui de l'auditeur (cf. fig. I.5) : on dira qu'une source sonore a une trajectoire et non que l'auditeur décrit un mouvement autour de la source. Deux principaux systèmes de coordonnées sphériques employés sont représentés en figure I.4 : le système polaire-vertical dont l'axe de référence est l'axe (Oz) et le système polaire-interaural dont l'axe de référence est l'axe interaural (Ox). Le système polaire-vertical est le plus intuitif, mais il n'est pas adapté pour la description des plans sagittaux répartis le long de l'axe interaural. De manière usuelle, une source sonore sera repérée par ses

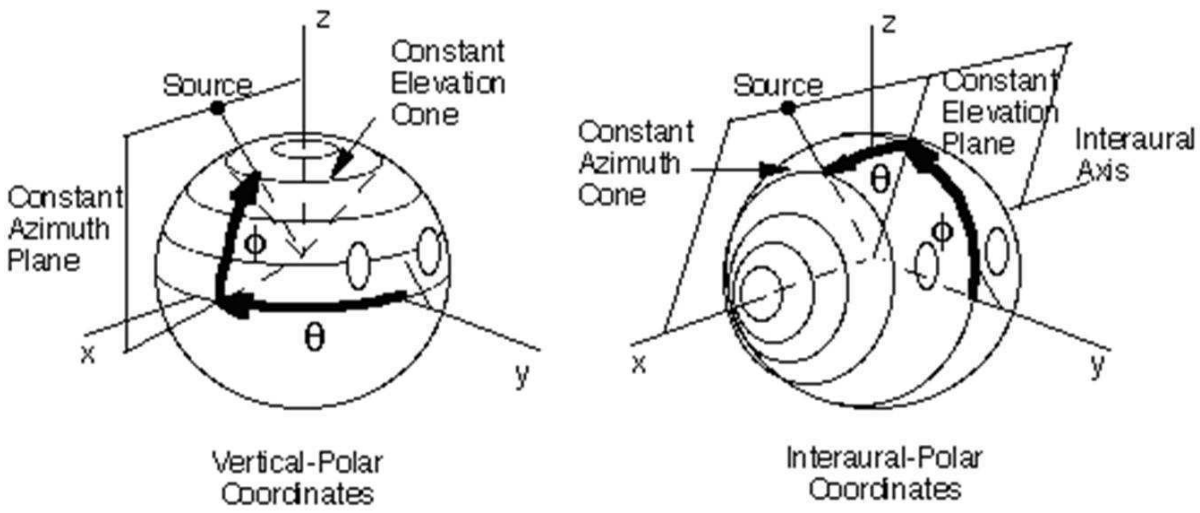


FIG. I.4 – Systèmes de coordonnées sphériques utilisés en spatialisation sonore. A gauche, système polaire-vertical, à droite, système polaire-interaural.

coordonnées (r, θ, ϕ). Trois plans sont couramment employés pour décrire l'analyse de la localisation sonore : le plan médian qui est défini par $\theta = 0°$, le plan horizontal décrit par $\phi = 0°$ et le plan frontal par $\theta = 90°$ en coordonnées polaires-verticales (cf. fig. I.5). La position frontale désigne la position à $\theta = 0°$ et $\phi = 0°$.

2.2 Performances de localisation du système auditif

De nombreux travaux psychoacoustiques ont étudié la capacité du système auditif à localiser une source sonore. Ces travaux ont commencé avec [Lord Rayleigh (1876)] et se continuent de nos jours [Blauert (1983)]. Les performances de localisation sont ici

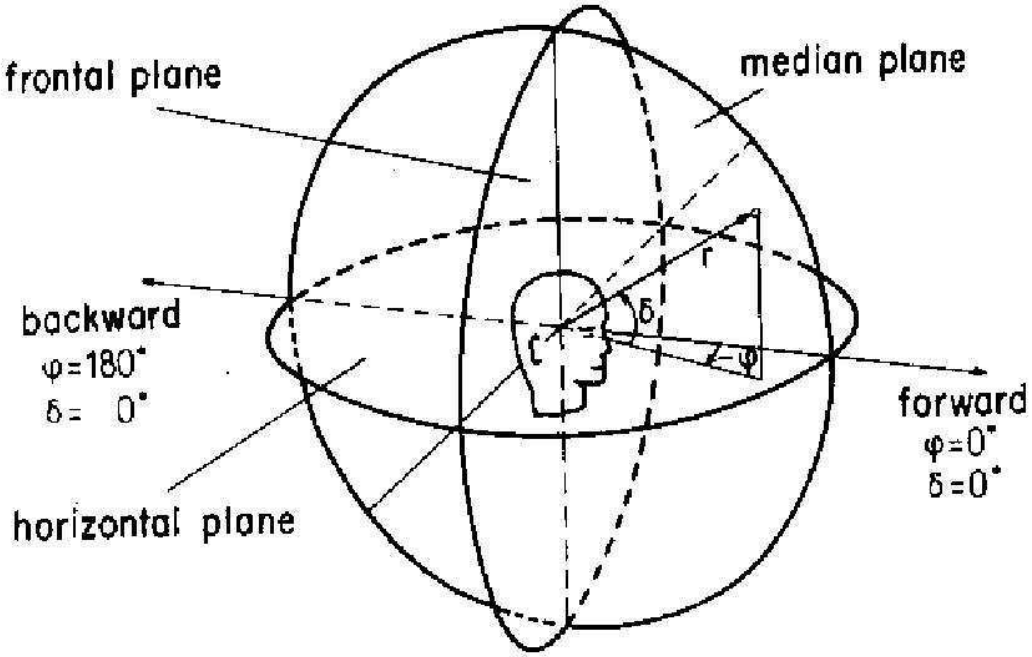


FIG. I.5 – Le référentiel auditeur.

présentées en deux parties : la localisation en azimut puis en élévation. Cette dichotomie ne présume en rien d'un traitement similaire du système auditif et cérébral, mais présente un avantage pratique. Les mécanismes et les performances sur les plans diagonaux sont encore soumis à des hypothèses multiples [Grantham et al. (2003)].

La localisation auditive est une capacité que le cerveau améliore au cours du plus jeune âge. Il s'agit d'un processus complexe d'intégration d'informations auditives, visuelles et cognitives. Le nourrisson *règle* sa perception sonore sur sa perception visuelle et la première tâche qu'il effectue est de repérer la provenance de la voix de sa mère. C'est peut-être alors parce que la localisation auditive est étalonnée par la localisation visuelle[1], et ceci tout au long de la vie[2], que les performances de localisation auditive sont très proches des performances de discrimination visuelle.

La figure I.6 indique de manière synthétique les performances de localisation sur le plan horizontal. Les sources sont mieux discriminées devant que derrière et que sur les côtés. La précision de localisation est de l'ordre de 3° devant et atteint 10° sur les côtés. La précision moyenne en azimut en fonction de l'élévation est quasi-constante et reste inférieure à 10° (cf. fig. I.7).

[1]L'étalonnage de la localisation auditive met en oeuvre d'autres processus qui peuvent notamment être mis en évidence dans le cas des aveugles.

[2]Des expériences ont montré que la localisation sonore pouvait être modifiée par des stimuli visuels.

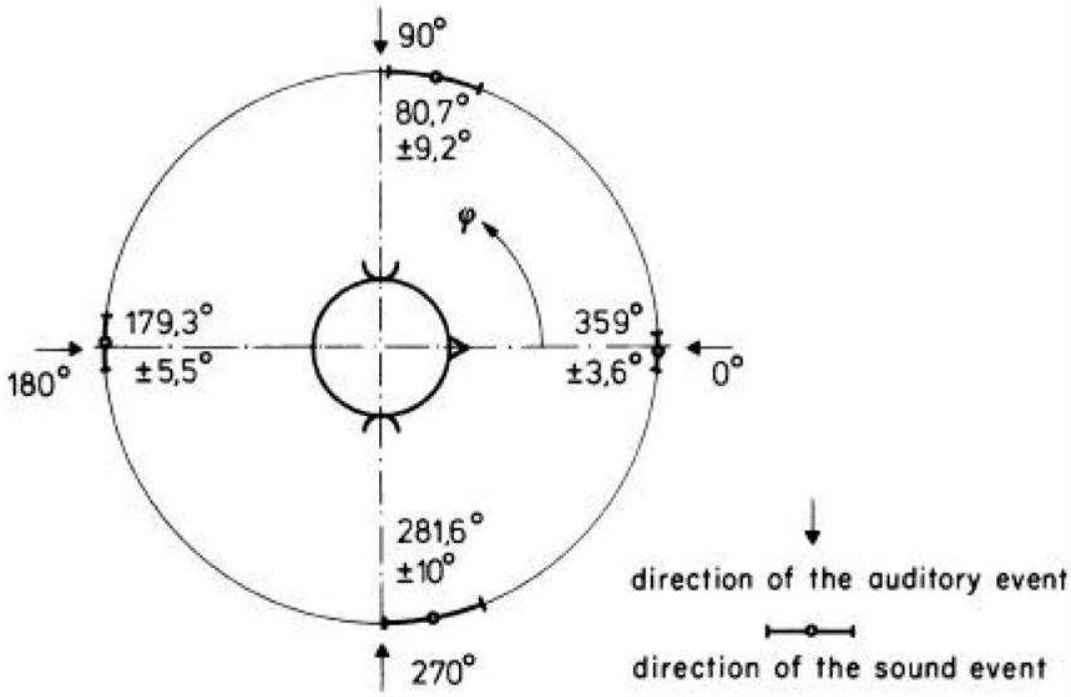
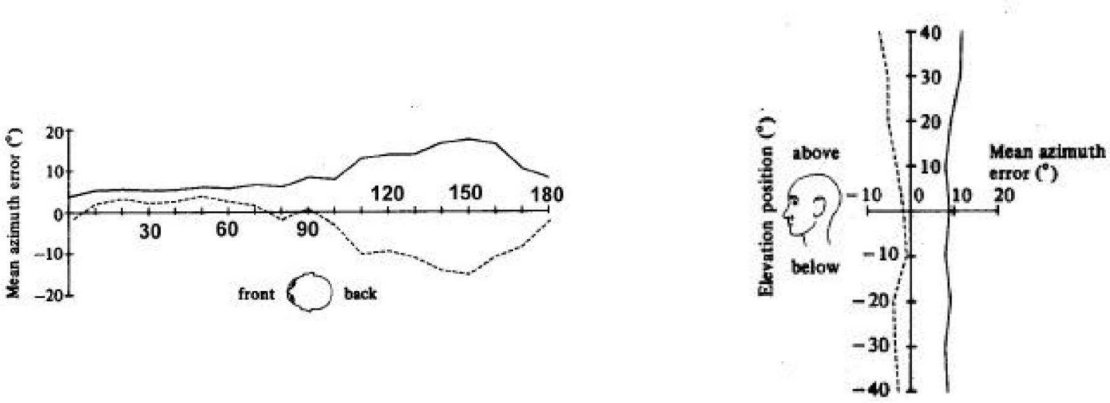


FIG. I.6 – Précision de localisation en azimut [Blauert (1983)].



A) Erreur angulaire en fonction de l'azimut B) Erreur angulaire en fonction de l'élévation

FIG. I.7 – Précision de localisation en azimut [Oldfield and Parker (1984)]. Les courbes en trait plein indiquent les erreurs absolues et les courbes en pointillé, les erreurs relatives.

La figure I.12 et I.9 indiquent les performances de localisation dans le plan médian. Les sources frontales sont les mieux discriminées. Les positions zénithales semblent être les moins bien discriminées avec une précision la localisation proche de 20°. Oldfield et Parker indiquent quant à eux une précision de localisation quasi-constante en élévation de 8.3° en moyenne [Oldfield and Parker (1984)]. La précision en élévation en fonction de l'azimut est quasi-constante et reste inférieure à 10° (cf. fig.I.7). Les positions situées dans l'hémisphère arrière pour des élévations autour de 40° sont les moins bien discriminées.

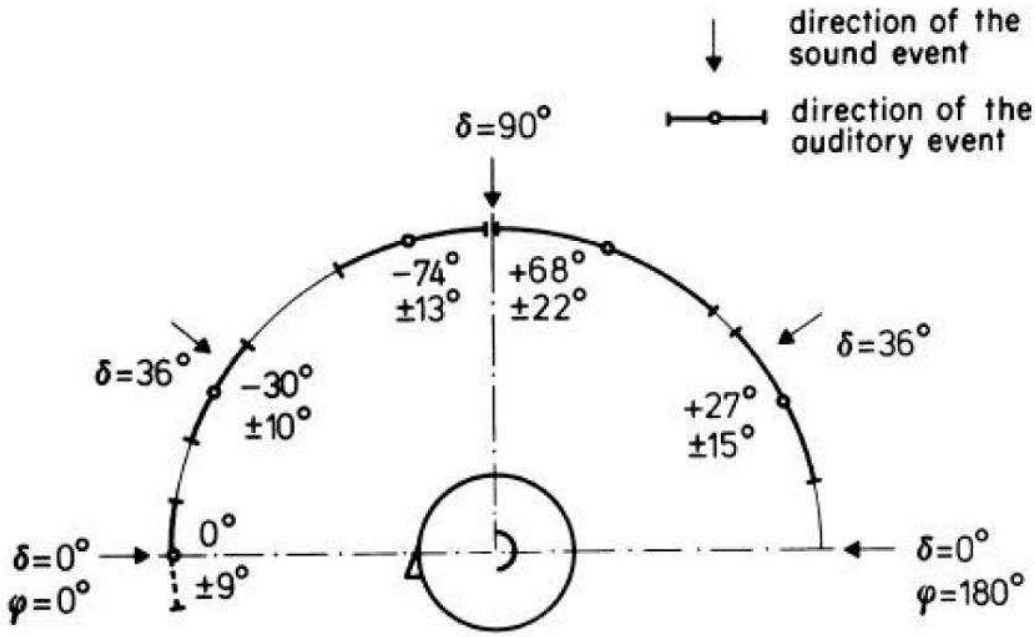


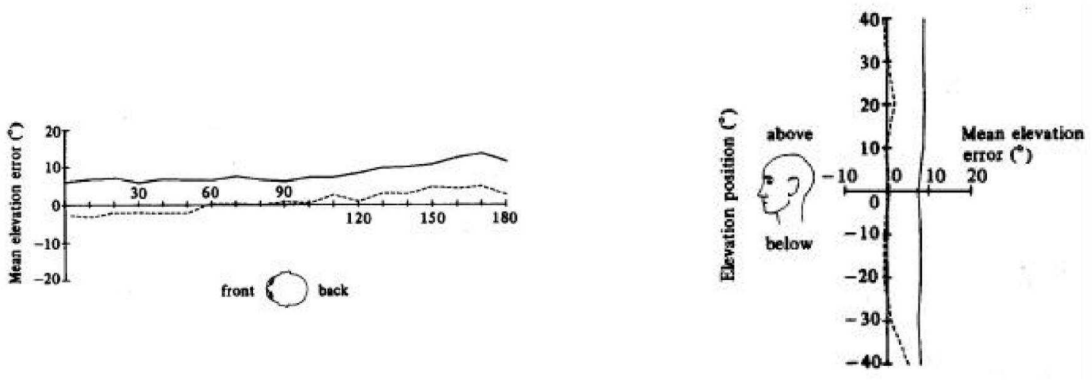
Fig. I.8 – Précision de localisation en élévation [Blauert (1983)].

2.3 Indices perceptifs de la localisation sonore

2.3.1 La théorie duplex

En 1907, Rayleigh [Lord Rayleigh (1907)] publie la *Duplex Theory* qui permet une description simple des mécanismes de perception en azimut (dimension gauche-droite). Cette théorie se base sur la définition de deux indices acoustiques la différence interaurale (cf. fig. I.10) de temps et de niveaux (cf. fig. I.11) :

- **ITD** La différence interaurale de temps d'arrivée de l'onde sonore entre les deux oreilles, ou ITD pour Interaural Time Difference, est l'indice de localisation pré-



A) Erreur angulaire en fonction de l'azimut B) Erreur angulaire en fonction de l'élévation

FIG. I.9 – Précision de localisation en élévation [Oldfield and Parker (1984)]. Les courbes en trait plein indiquent les erreurs absolues et les courbes en pointillé, les erreurs relatives.

dominant[3]. Sa contribution majeure se situe en basses fréquences (f < 1500 Hz). La localisation en azimut est une fonction quasi-linéaire de l'ITD. Une source est généralement perçue plus proche de l'oreille qui est atteinte en première par le front d'onde. Plus l'ITD est grande, plus la source paraît latéralisée. En première approximation, les variations de l'ITD en fonction de l'azimut θ sont données par la formule de Woodworth [Woodworth and Schloesberg (1962)] :

$$ITD_{sphere}(\theta) = \frac{a}{c}(\sin(\theta) + \theta) \tag{I.1}$$

Cette formule correspond à une estimation de l'ITD haute fréquence d'un modèle de tête sphérique rigide de rayon a.

- ILD La différence interaurale d'amplitude de l'onde sonore entre les deux oreilles, ou ILD pour Interaural Level Difference, est l'indice de localisation complémentaire de l'ITD. Sa contribution majeure se situe en hautes fréquences (f > 1500 Hz). La fonction entre l'angle d'azimut perçu et ILD est aussi quasi-linéaire pour une onde monochromatique. Mais elle est de nature moins triviale pour un son complexe. Une source est généralement perçue plus proche de l'oreille recevant le plus d'énergie acoustique. Plus l'ILD est importante, plus la source paraît latéralisée.

Malgré la simplicité de la théorie duplex, elle représente une très bonne approximation de la perception dans le plan horizontal, et l'importance relative des indices binauraux ainsi que la séparation effectuée à 1500 Hz entre ITD et ILD a été utilisée pendant très longtemps. Ce n'est que récemment que cette théorie a été ré-éxaminée à la lumière de la synthèse binaurale [Macpherson and Middlebrooks (2002)]. Cette théorie ne peut toutefois pas rendre compte de la perception des sources avec l'élévation. En effet, pour un même angle d'azimut, il existe un nombre infini de positions ayant même ILD et même ITD (cf. fig.I.13.B). Ces positions sont appelées *cônes de confusion* et selon la théorie duplex, elles ne peuvent pas être distinguées les unes des autres.

[3]L'acronyme anglais est utilisé dans tout le document car il est d'usage courant

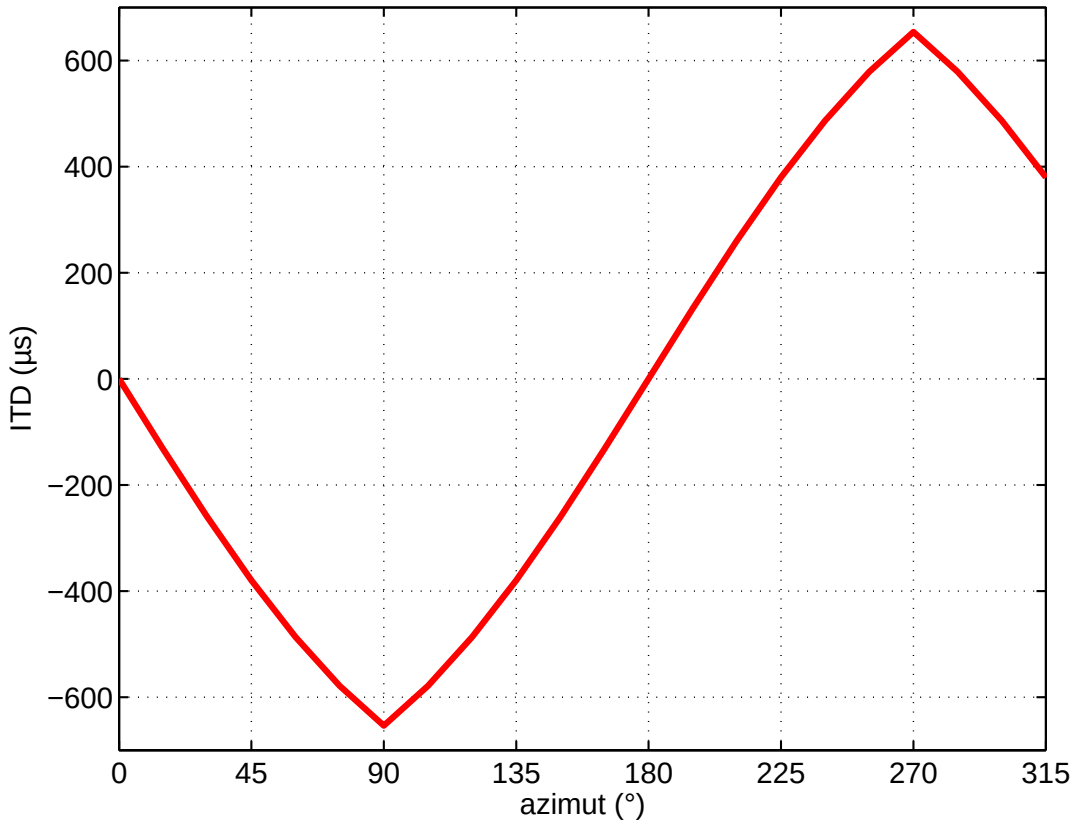


FIG. I.10 – ITD en fonction de l'angle d'azimut selon la formule de Woodworth (cf. équation II.1).

2.3.2 Les cônes de confusion

Un cône de confusion est défini comme le lieu des points pour lesquels la distance entre les deux oreilles est la même, pour le modèle simplifié de tête composé seulement de deux oreilles (cf. fig.I.13.A). Ce lieu est décrit par une hyperboloïde dont le centre est situé sur l'axe interaural et qui peut être approchée par la surface extérieure d'un cône en champ lointain [Blauert (1983); Shin-Cunningham et al. (2000)]. L'intersection des cônes de confusion avec une sphère centrée sur la tête définit des cercles dont les centres sont répartis sur l'axe interaural (cf. fig.I.13.B). Dans le système de coordonnées polaires-interaurales, ces cercles sont inclus dans des plans d'azimut constant. L'ITD de la formule de Woodworth (cf. equation .II.1) ne dépend pas de l'élévation (en coordonnées polaires-interaurales). Une trajectoire à azimut et rayon constant décrit une ligne iso-ITD. **Dans la suite du document, le terme *cône de confusion* désigne les lignes iso-ITD du modèle de tête sphérique, c'est-à-dire des trajectoires à azimut et rayon constant en coordonnées polaires-interaurales**. Les principaux artefacts audibles liés aux cônes de confusion sont *les confusions avant/arrière* : le sujet indique qu'une source située dans l'hémisphère avant est perçue dans l'hémisphère arrière. Ce type de confusion est généralement résolu par des petits mouvements de tête [Wightman and Kistler (1999)].

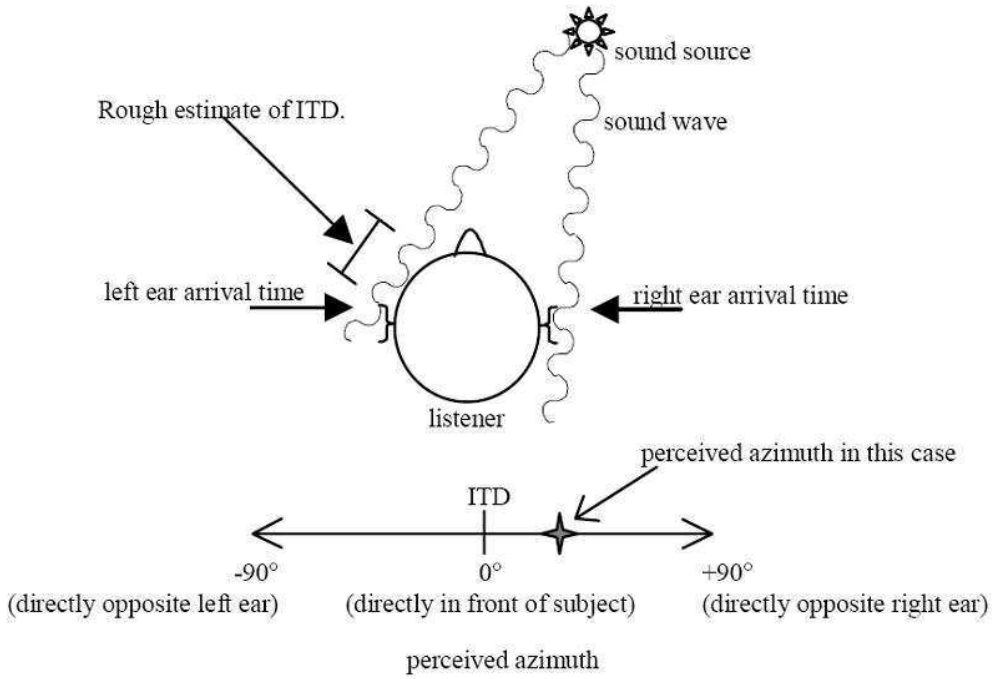


FIG. I.11 – Utilisation de l'ITD pour la localisation en azimut. Une source est généralement perçue plus proche de l'oreille atteinte en premier par le front d'onde. Plus l'ITD est grand, plus la source paraît latéralisée [Cheng and Wakefield (2001)].

La théorie duplex ne décrit pas la perception des sons en élévation et ne peut expliquer que les positions situées sur le plan médian, par exemple, sont discriminées. Ce pouvoir de discrimination fut la preuve pour de nombreux auteurs de mécanismes de localisation monaurale qui reposent sur les colorations spectrales induites par les multiples réflexions, diffraction et diffusion de l'onde sur notre corps avant d'atteindre les tympans.

2.3.3 Les indices spectraux

Le terme *indices spectraux* désigne, en localisation auditive, les modifications fréquentielles apportées au spectre d'une source sonore par le filtrage opéré par les réflexions sur les différentes parties du corps de l'auditeur. Le couplage source sonore / pavillon de l'oreille étant un système interférentiel, le moindre changement de position de la source sonore modifie les indices spectraux et donc il est attendu que les indices spectraux jouent un rôle dans la localisation auditive. Blauert [Blauert (1983)] a ainsi montré que, pour que des sujets localisent correctement des sources en élévation, les stimuli doivent être large bande et contenir des informations au-delà de 7 kHz. La localisation auditive à partir des indices spectraux est basée sur l'hypothèse que le système auditif effectue la

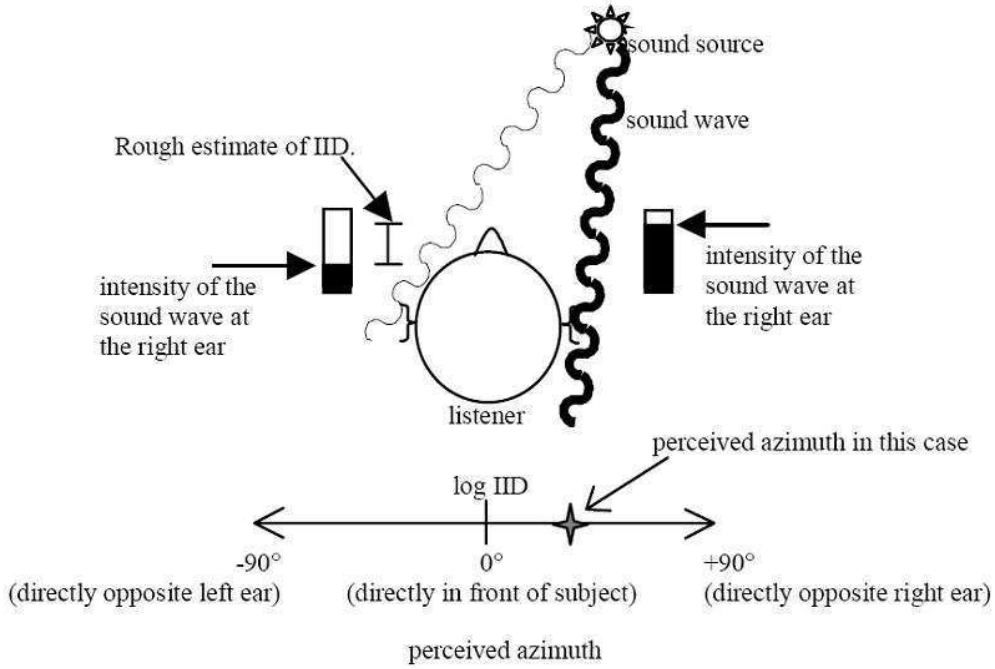
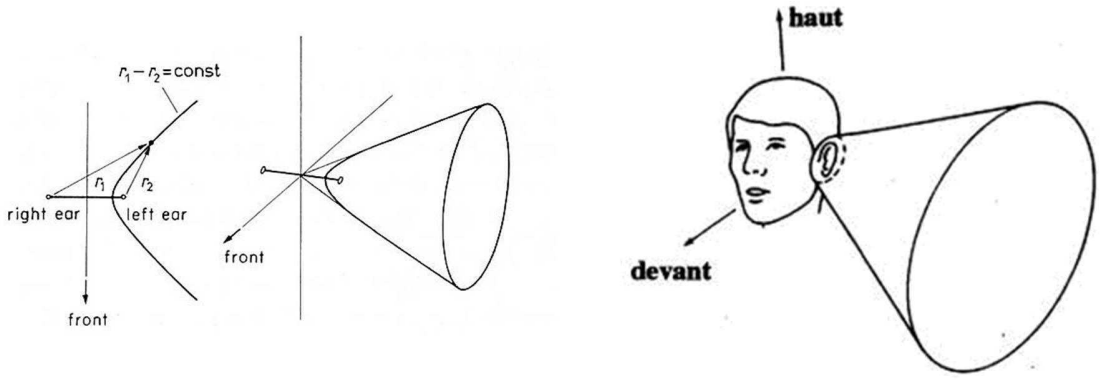


FIG. I.12 – Utilisation de l'ILD pour la localisation en azimut. Une source est généralement perçue plus proche de l'oreille recevant le plus d'énergie acoustique. Plus l'ILD est important, plus la source paraît latéralisée [Cheng and Wakefield (2001)].

déconvolution du spectre de la source et des indices spectraux liés à la morphologie et à la position de la source. Il est même montré que les indices spectraux seraient stockés en mémoire et que la détermination de la position d'une source serait effectuée par identification après l'étape de déconvolution. [Middlebrooks (1992)].

La relation entre la position perçue et les indices spectraux est complexe. La littérature rapporte généralement que les modifications spectrales d'une onde sonore sont responsables de la localisation en élévation. Cette hypothèse est confirmée notamment par les expériences reportées dans [Blauert (1983)] qui montrent que la position apparente de stimuli à bande étroite est premièrement gouvernée par leur fréquence centrale et non par leur position réelle (cf. fig.I.14). Plus récemment [Langendijk and Bronkhorst (2002)], il a été montré que les informations nécessaires à la perception des sources sur l'axe haut-bas se situent dans la bande [6 - 12] kHz et celles indiquant la dimension avant / arrière se situent dans la bande [8 - 16] kHz. Des études ont montré que la localisation auditive était sensible à la position des anti-résonances des indices spectraux [Shaw (1982)], c'est-à-dire la fréquences des minima locaux, d'autres montrent que c'est la pente des anti-résonances qui gouverne la localisation des sources sonores [Han (1994)] tandis que d'autres encore montrent la pertinence des pics, c'est-à-dire la fréquence des



A) Modèle simplifiée de la tête B) Tête humaine

FIG. I.13 – Définition des cônes de confusion. A) Modèle de tête ne comportant que deux oreilles [Blauert (1983)]. B) Représentation d'un cône de confusion au sens de la théorie duplex pour une tête humaine [Chateau (1996)].

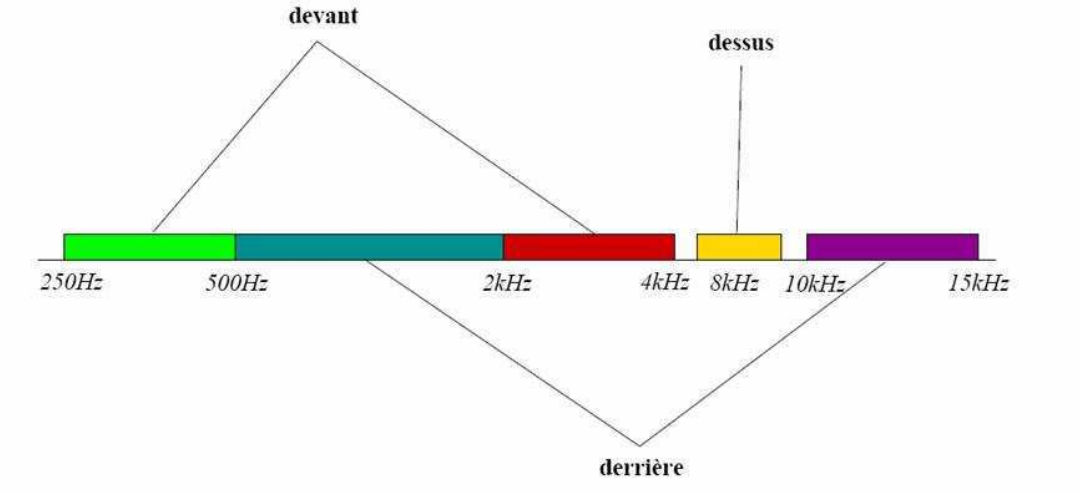


FIG. I.14 – Bandes directionnelles proposées par [Blauert (1983)].

maxima locaux [Blauert (1983); Middlebrooks et al. (1989)]. La forme compliquée des indices spectraux (cf. fig. I.25) associée à la primauté du codage fréquentiel exercé par la cochlée, rend les études entre localisation et indices spectraux difficiles à mener et il n'existe pas encore de modèle de localisation basé sur les informations fréquentielles. De plus, les nombreuses tentatives de théorie unifiée sur le sujet se heurtent aux variabilités inter-individuelles des performances de localisation qui laissent supposer que différentes stratégies sont utilisées en fonction du sujet et de la tâche qui lui est demandée.

2.3.4 Indices monauraux et indices interauraux

Le paradigme de la localisation auditive monaurale est que le système auditif peut localiser des sons avec seulement les informations provenant d'une oreille. De nombreuses expériences ont ainsi montré qu'au moins les performances de localisation sur la dimension haut-bas sont conservées même avec une seule oreille, et que des sujets sourds d'une oreille arrivaient à localiser des sources en azimut. Cependant, Wightman et Kistler [Wightman and Kistler (1997)] ont examiné les différents problèmes qui peuvent perturber les résultats de telles expériences et ont montré, grâce à la synthèse binaurale, qu'en réalité, le paradigme de localisation monaurale n'est pas adapté à l'étude des indices spectraux monauraux pour les mécanismes de la localisation sonore. En effet, d'une part l'écoute avec une seule oreille est très difficile à réaliser et d'autre part, elle correspond à une écoute artificielle manquant de cohérence. Il est alors difficile de conclure sur la pertinence d'indices monauraux. De plus, une étude sur l'importance relative des indices interauraux et monauraux [Jin et al. (2004)] a montré que les réponses des sujets à un test de localisation sont liées aux variations des indices interauraux et non aux variations des indices monauraux. Dans la suite du document, il ne sera considéré que l'importance perceptive des indices interauraux.

2.3.5 Les HRTF

L'ensemble des modifications spectrales que subit une source sonore pour arriver à nos oreilles constitue les filtres liés à la tête, ou **HRTF pour Head Related Transfert Function**. Les HRTF traduisent les phénomènes physiques qui résultent du parcours de l'onde sur notre corps, principalement le torse, la tête et le pavillon. De ce fait, les HRTF sont reliées à la morphologie d'un auditeur et sont donc individuelles, spécifiques à chaque oreille et dépendent également de la position de la source sonore. Toutes les informations de spatialisation sont contenues dans les HRTF, et les indices perceptifs, ITD, ILD et indices spectraux peuvent en être extraits. L'ILD est calculée par le rapport de l'énergie entre le signal présenté à l'oreille droite et le signal à l'oreille gauche. L'estimation de l'ITD est plus complexe et reste une question ouverte. Une partie des travaux de thèse est consacrée à l'estimation de l'ITD (cf. chapitre II). Dans la suite du document il est considéré que le module des HRTF contient l'ILD.

3 LES SYSTÈMES DE SPATIALISATION SONORE

3.1 Stéréophonie, quadriphonie et systèmes n.m

Utilisant les indices psychoacoustiques de localisation que sont la différence de temps et la différence de niveau entre les deux oreilles d'un auditeur, la *stéréophonie* fut la première technique restituant des effets de spatialisation. Le concept de source virtuelle est né avec la stéréophonie. Les effets rendus sont principalement la latéralisation de la source virtuelle (d'autre effets peuvent être obtenus par l'ajout d'un effet de salle). Le contrôle du retard et du gain entre les deux hauts-parleurs permet de déplacer la source virtuelle entre les deux hauts-parleurs (cf. fig. I.15). Ce contrôle se réalise simplement sur les consoles de mixage par l'intermédiaire du *panoramique d'intensité*. L'effet droite-gauche est rendu de manière optimale si les positions des deux hauts-parleurs et de l'auditeur forment un triangle équilatéral. Les informations de retard et de gain peuvent

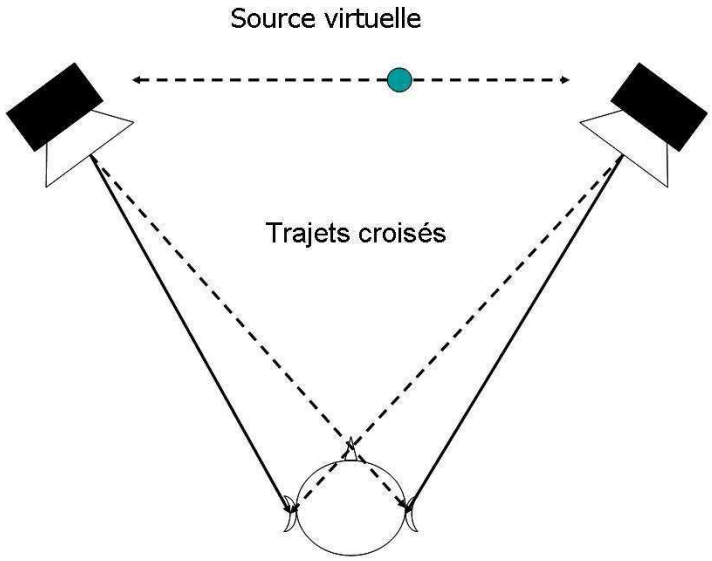


FIG. I.15 – Principe de la source virtuelle.

être captées directement par un couple de microphones (cf. [Mercier (1993)] pour les techniques de prise de son.) :

- stéréophonie de temps : couple de microphones omnidirectionnels écartés de quelque centimètres,
- stéréophonie d'intensité : couple XY, stéréosonic, MS (moyennant un décodage),
- stéréophonie mixte : couple AB.

Dans les années 1970, la quadriphonie fit brièvement une apparition. Basée sur le même principe que la stéréophonie, la quadriphonie offre une spatialisation sur le plan horizontal : la spatialisation sonore passe du 1D au 2D. Cette technique ne s'est pas développée principalement pour des problèmes de format propriétaire : aucun n'a fait l'unanimité. La quadriphonie est restée au stade d'expérience d'écoute particulière. A noter toutefois que le groupe de rock Pink Floyd fut parmi les premiers à proposer des

albums en quadriphonie (*Atom Heart Mother*, *The Dark Side of the Moon*, et *Wish you Were Here*), ces albums se sont vendus essentiellement en stéréophonie. Cette technique fut aussi utilisée lors de concerts (M. Jonasz, Pink Floyd) et certains artistes sont encore intéressés par ce système de reproduction (le groupe M).

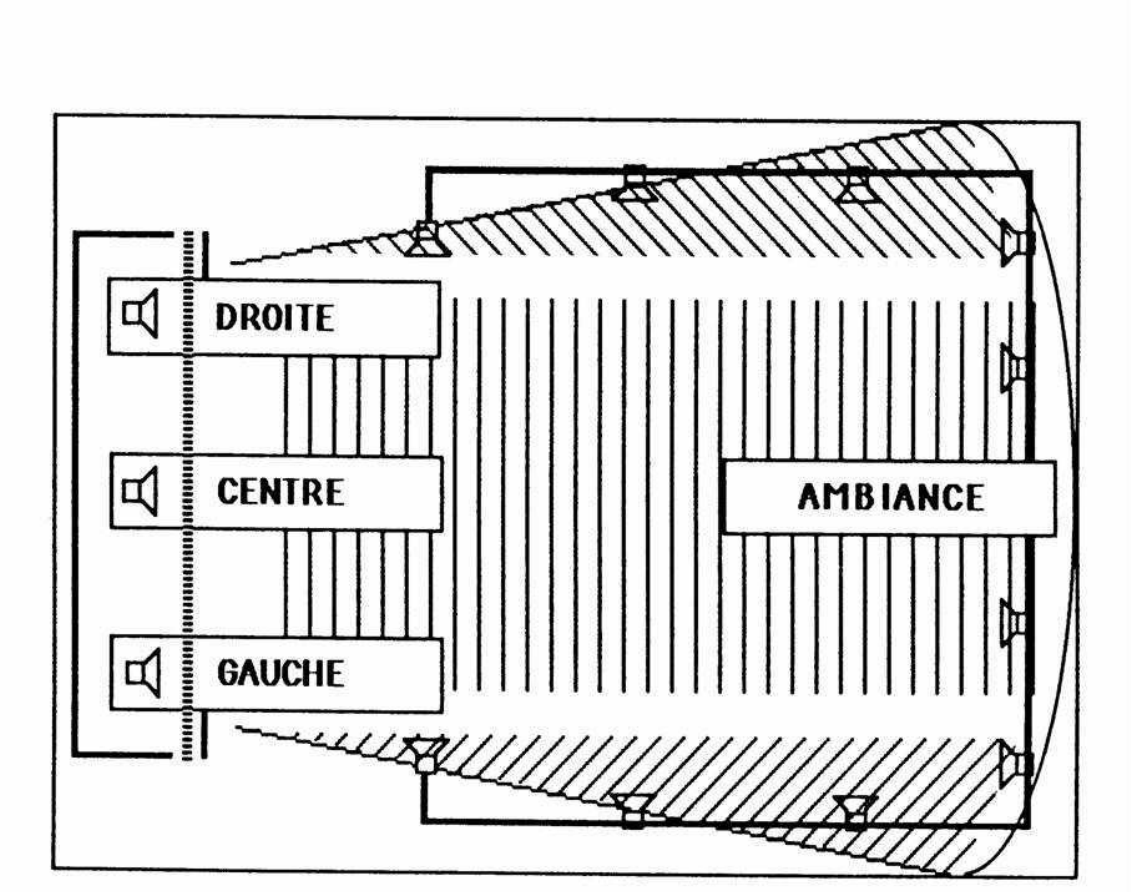


FIG. I.16 – Restitution en *Dolby stéréo* dans une salle de cinéma.

Si la quadriphonie n'a pas su trouver un standard pour se développer, le système *Dolby stéréo* s'est imposé à l'inverse dans les salles de cinéma comme standard de diffusion 2D. En 1965, l'ingénieur R. Dolby a proposé sur l'ajout deux voies audio supplémentaires sur les pellicules 35 mm (les deux canaux stéréophoniques se trouvent sur la bande noire de la pellicule) grâce à un encodeur matriciel basé sur la phase des signaux. Quatre voies peuvent alors être diffusées : une à droite, une au centre, une à gauche et une centrale à l'arrière qui restitue une ambiance sonore (cf. fig.I.16) . Le *Dolby stéréo* eut un immense succès car les salles ont pu s'équiper à moindre coût, le système étant adaptable à un rendu monophonique, stéréophonique sur deux canaux et stéréophonique sur quatre canaux. Depuis les années 1980, la plus grande partie des salles de cinéma sont équipées de système *Dolby stéréo*.

Au début des années 1990, les formats de diffusion sonore sont devenus numériques. Ainsi en 1992, le *Dolby stéréo* devient le *Dolby SRD* (Spectral Recording Digital), ou format *AC3*. Ce format numérique propose, outre un algorithme de réduction de bruit de fond, deux autres pistes audio (qui sont rajoutées entre les perforations d'entraîne-

ment des pellicules cinématographiques). Six canaux discrets, sans matriçage mais avec compression des données, peuvent alors être diffusés : c'est le système 5.1 conçu avec 3 canaux en façade (droite-centre-gauche) dans la bande [3 Hz - 20 kHz], deux canaux arrières (droite-gauche) pour l'ambiance et un canal (généralement central) pour les basses fréquences dans la bande [3 Hz - 120 Hz] (cf. fig. I.17). Ce système de diffusion connaît aujourd'hui un développement considérable et des installations domestiques à moindre coût sont disponibles (< 100 €). Par la suite, d'autres systèmes basés sur des canaux discréts ont vu leur apparition : 6.1, 7.1, 7.2, 9.2, 2.2.2 (2 canaux surélévés.) Le principal

FIG. I.17 – Différents systèmes de restitution basés sur la technique *Dolby Digital*.

avantage de ces méthodes est d'offrir au grand public des systèmes à moindre coût et facile d'installation. De plus, un nombre important de contenus est disponible (DVD et DVD musicaux, jeux).

L'utilisation de systèmes multi-haut-parleur discrèts rencontrent deux inconvénients majeurs. Premièrement, la zone de diffusion ou les effets de spatialisation sont les plus intenses (ou *sweet spot*), est très réduite. Par exemple, pour un dispositif domestique, le sweet spot ne peut contenir qu'un seule personne. De plus les effets de spatialisation sont principalement obtenus par mixage des différentes voix : l'emploi d'un ingénieur du son est alors nécessaire et le prix de production augmente. Un *bon* enregistrement 5.1 relève plus de l'art que de la science. Deuxièmement, ces systèmes sont pour l'instant cantonnés à une restitution en deux dimensions.

3.2 Ambisonic et holophonie

3.2.1 Ambisonic

Apparue dans les années 1970 grâce aux travaux de Gerzon, la technique *Ambisonic* repose sur une décomposition / recomposition du champ acoustique en un point correspondant au centre de la tête de l'auditeur. La décomposition s'effectue sur la base des harmoniques sphériques[4] (cf. équation I.2) tronquée à un certain ordre et la recomposition

[4]Les harmoniques cylindriques ont aussi été utilisées.

dépend du nombre et du positionnement des hauts-parleurs autour de l'auditeur.

$$p(\vec{r}) = \sum_{m=0}^{+\infty}(2m+1)j^m \sum_{0\leq n\leq m,\sigma=\pm 1} B^{\sigma}_{mn} Y^{\sigma}_{mn}(\theta,\phi) j_m(kr)$$

$$avec$$

$$Y^{\sigma}_{mn}(\theta,\phi) = P_{mn}(\sin\theta) \times \begin{cases} \cos(n\phi) & \text{si } \sigma = 1 \\ \sin(n\phi) & \text{si } \sigma = -1 \end{cases} \tag{I.2}$$

où B^{σ}_{mn} sont les coefficients ambisoniques, $Y^{\sigma}_{mn}(\theta,\phi)$ les harmoniques sphériques d'ordre m et j_m les fonctions de Bessel sphériques d'ordre m. La relation entre la décomposition et la reproduction est donnée par la matrice de décodage M_d :

$$\mathbf{g} = M_d \cdot B^{\sigma}_{mn} \tag{I.3}$$

où le vecteur **g** représente l'ensemble des gains à appliquer aux hauts-parleurs. Les coefficients de la matrice de décodage sont trouvés par des méthodes d'identification dans le cas d'un espacement régulier des hauts-parleurs et dans les autres configurations par des méthodes d'optimisation privilégiant soit l'énergie, soit la phase du vecteur résultant de la reproduction. Un système ambisonique est indépendant du système de diffusion car la matrice de décodage s'y adapte. La contrainte sur le nombre N de hauts-parleurs pour une restitution sur le plan horizontal est $N \geq 2m+1$, avec m ordre de décomposition sur les harmoniques sphériques. L'utilisation de *shell-filter* permet d'appliquer un décodage différent selon la fréquence (souvent un décodage basse fréquence et un décodage haute fréquence).

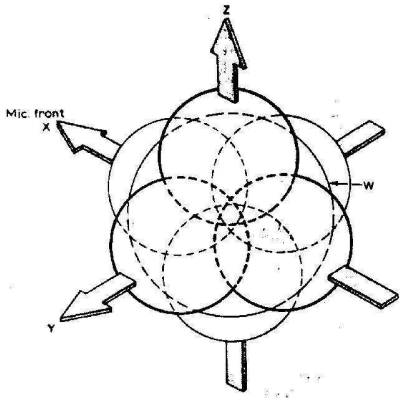
Cette technique a d'abord été proposée à l'ordre 1 de décomposition puis étendue aux ordres supérieurs par Bamford puis par Daniel [Daniel (2000)]. La prise de son à l'ordre 1 consiste en un système microphonique contenant quatre capsules cardioïdes (cf. fig. I.18.A). Il est démontré que les signaux ambisoniques peuvent être obtenus simplement par une combinaison linéaire des signaux des quatre capsules.

L'avantage de cette technique est la possibilité d'une reconstruction exacte du champ de pression autour d'un auditeur. Elle offre de plus une indépendance entre encodage et décodage ce qui la rend bien appropriée pour la diffusion sur un réseau de télécommunication. Le principal défaut est la taille du *sweet spot* qui est fortement réduit pour les hautes fréquences. Les recherches actuelles sont axées sur une prise de son aux ordres supérieurs qui devrait assurer une meilleur reconstruction des hautes fréquences (cf. fig. I.21).

3.2.2 Holophonie

Comme l'holographie optique qui permet la visualisation d'une image tri-dimensionnelle, l'holophonie reproduit un champ de pression tri-dimensionnel. S'inspirant du principe de Huygens, l'holophonie consiste à reproduire un champ acoustique en trois dimensions à partir d'un enregistrement sur une surface (fermée ou plane) et se formule comme un problème aux limites. Le champ sonore à l'intérieur d'un volume s'exprime sous la forme de l'intégrale de Kirchoff-Helmholtz :

A) Microphone *Soundfield*.



B) Décomposition d'une onde plane à l'ordre 1 sur les harmoniques sphériques.

FIG. I.18 – Réalisation pratique d'une prise de son ambisonique à l'ordre un. Figure du haut : microphone *Soundfield*, figure du bas : décomposition d'une onde à l'ordre un sur les harmoniques sphériques.

FIG. I.19 – Prototype de microphone ambisonique pour les ordres de décompositions supérieurs.

$$p(\vec{r},\omega) = \iint_{\partial\Omega}\left[\left(\vec{\nabla}p_0\cdot\vec{n}\right)\frac{e^{-jkR}}{4\pi R} - p_0\cos\alpha(1+jkR)\frac{e^{-jkR}}{4\pi R^2}\right]\mathrm{d}S$$

avec

$$\begin{cases} \partial\Omega & \text{la surface entourant le volume considéré} \\ \left(\vec{\nabla}p_0\cdot\vec{n}\right) & \text{dérivée normale à } \partial\Omega \text{ de la pression surfacique} \\ \frac{e^{-jkR}}{4\pi R} & \text{monopôle acoustique} \\ cos\alpha(1+jkR)\frac{e^{-jkR}}{4\pi R} & \text{dipôle acoustique} \\ p_0 & \text{pression surfacique} \end{cases} \tag{I.4}$$

Enoncée par Jessel en 1973, l'holophonie fut appliquée à la visioconférence par Nicol en 1999 à partir des travaux menés à l'Université de Delft sur le système de Wave Field Synthesis (WFS). Une des réalisations pratiques de cette technique est présentée sur la figure I.20 et I.21.

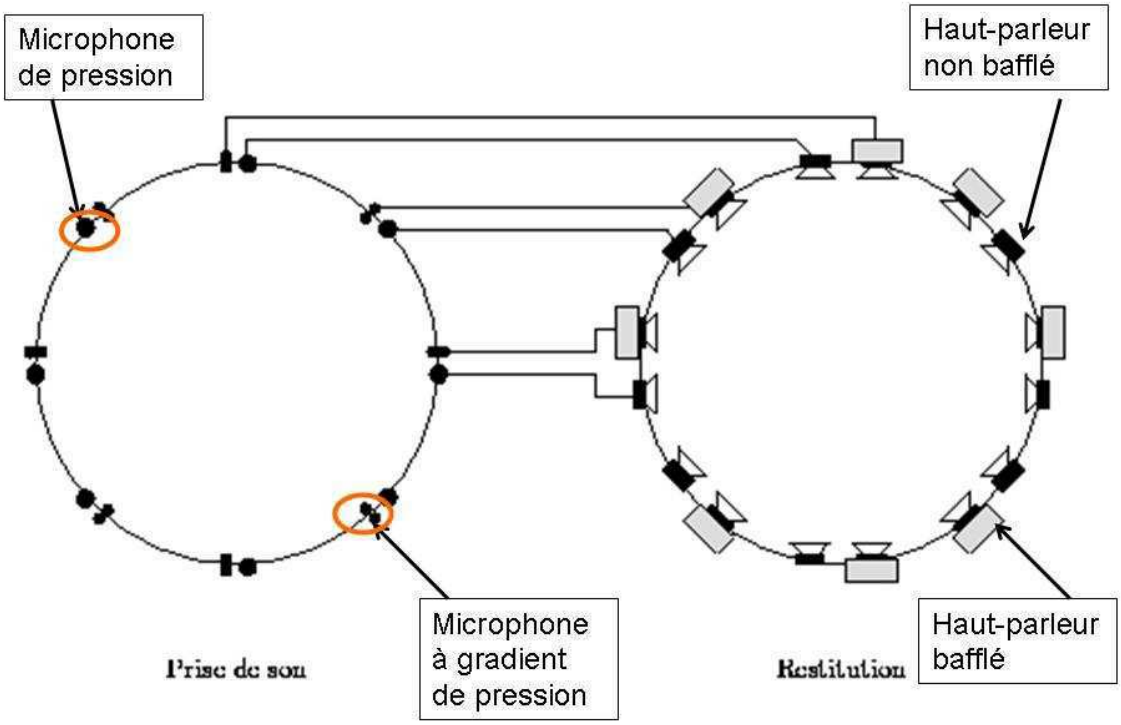


FIG. I.20 – Réseaux de captation et de restitution pour l'holophonie.

L'holophonie permet une restitution à l'identique du champ sonore et contrairement à la méthode ambisonique, le champ est reconstruit sur une vaste zone englobant plusieurs

FIG. I.21 – Réseau de 48 hauts-parleurs pour une restitution holophonique.

auditeurs qui peuvent se déplacer dans le champ sonore. L'inconvénient principal de cette technique repose dans le nombre important de microphones et de hauts-parleurs ce qui la rend très coûteuse et difficile à installer et à contrôler. Elle pose également des problèmes en termes de d'échantillonnage et de troncature spatiale.

3.3 La technique binaurale

La technique binaurale repose sur une reproduction au niveau du conduit auditif de l'auditeur, des informations acoustiques nécessaires à la construction par le système auditif d'une image sonore spatialisée. Enregistrements binauraux et synthèse binaurale sont deux techniques distinctes. L'enregistrement binaural consiste à placer des microphones dans les oreilles d'une tête, humaine ou artificielle, et à enregistrer une scène sonore. La diffusion de l'enregistrement se fait au casque sans aucun matriçage des signaux (cf. fig. I.22). Le principe des enregistrements binauraux a été développé dans les années 1960 et a surtout contribué à la définition de têtes artificielles (cf. fig I.23).

La synthèse binaurale consiste à simuler un enregistrement binaural à l'aide de filtres : les HRTF. Une fois les HRTF connues, il suffit d'effectuer la convolution entre un son monophonique et les HRTF associées à l'auditeur, à la position simulée et de diffuser le résultat de la convolution à l'aide d'un casque d'écoute. Le principe de la synthèse binaurale est illustré en figure I.24. N'importe quel contenu peut être ainsi spatialisé et notamment l'effet de salle[5]. Les premières études sur la synthèse binaurale ont montré l'équivalence perceptive entre une écoute champ libre et une écoute binaurale [Wightman

[5]La simulation binaurale d'un effet de salle fait notamment appel à la notion de HRTF diffuse pour le champs réverbéré dont la mesure reste complexe

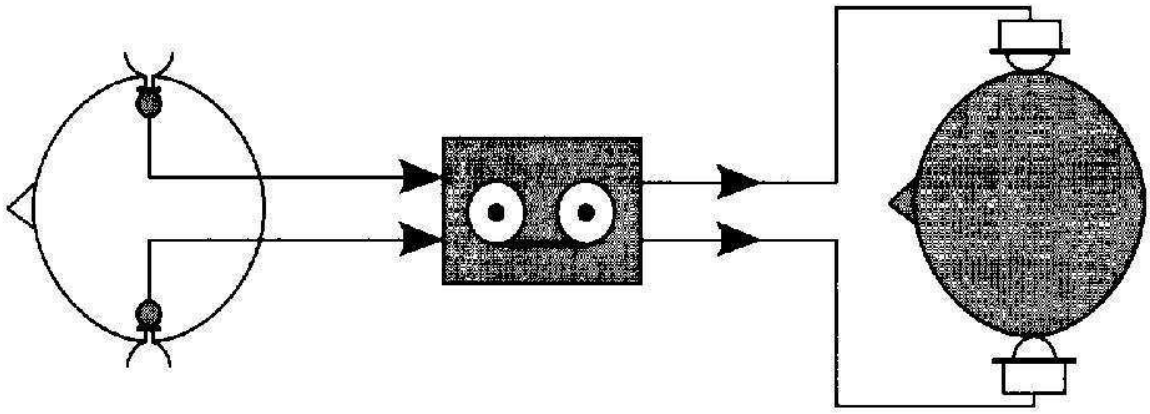


FIG. I.22 – Principe d'un enregistrement binaural.

and Kistler (1989a,b)]. Soit $x(t)$ un signal monophonique et $[h_r(t)\,;h_l(t)]$ un couple de HRIR correspondant à une position de l'espace. L'illusion d'une écoute tri-dimensionnelle sera alors donnée par la présentation des signaux $L(t)$ et $R(t)$ aux oreilles respectivement gauche et droite d'un auditeur :

$$L(t) = \int_0^t h_l(\tau)x(t-\tau)\mathrm{d}\tau \tag{I.5}$$

$$R(t) = \int_0^t h_r(\tau)x(t-\tau)\mathrm{d}\tau \tag{I.6}$$

Le principal avantage de la technique binaurale et de proposer une reconstitution 3D parfaite du champ sonore avec seulement deux signaux. Cette reconstitution ne nécessite aucun traitement supplémentaire sur les deux signaux alimentant le casque d'écoute. De plus, l'extraction des indices de localisation à partir des HRTF autorise un contrôle aussi précis que la localisation en champ libre. Les techniques binaurales ont ainsi permis de nombreuses études psychophysiques sur les mécanismes perceptifs et cérébraux. En pratique, les limitations se situent dans les mesures des HRTF. Les HRTF sont par définition individuelles et l'utilisation d'une synthèse binaurale non individualisée peut engendrer des défauts plus ou moins critiques comme les confusions avant / arrière et une perception intra-crânienne (comme une écoute stéréophonique au casque). Or, la

FIG. I.23 – Têtes artificielles pour la technique binaurale.

mesure de HRTF est très délicate, coûteuse et longue (le sujet doit rester plusieurs heures immobile sur une chaise dans une chambre sourde) ce qui pose des problèmes pour appliquer la synthèse binaurale au grand public. De plus, dans le cas idéal, deux filtres doivent être associés pour chaque source sonore et pour chaque position. Le coût d'implémentation devient rapidement prohibitif quand plusieurs sources sont simulées simultanément.

La synthèse binaurale peut aussi être reproduite sur deux haut-parleurs. Il s'agit alors de la technique *transaurale*. Afin de reproduire un rendu binaural sur haut-parleur il faut prendre soin d'annuler les trajets croisés (cf. fig. I.15). Cette opération est assurée par des filtres dont l'implémentation reste délicate.

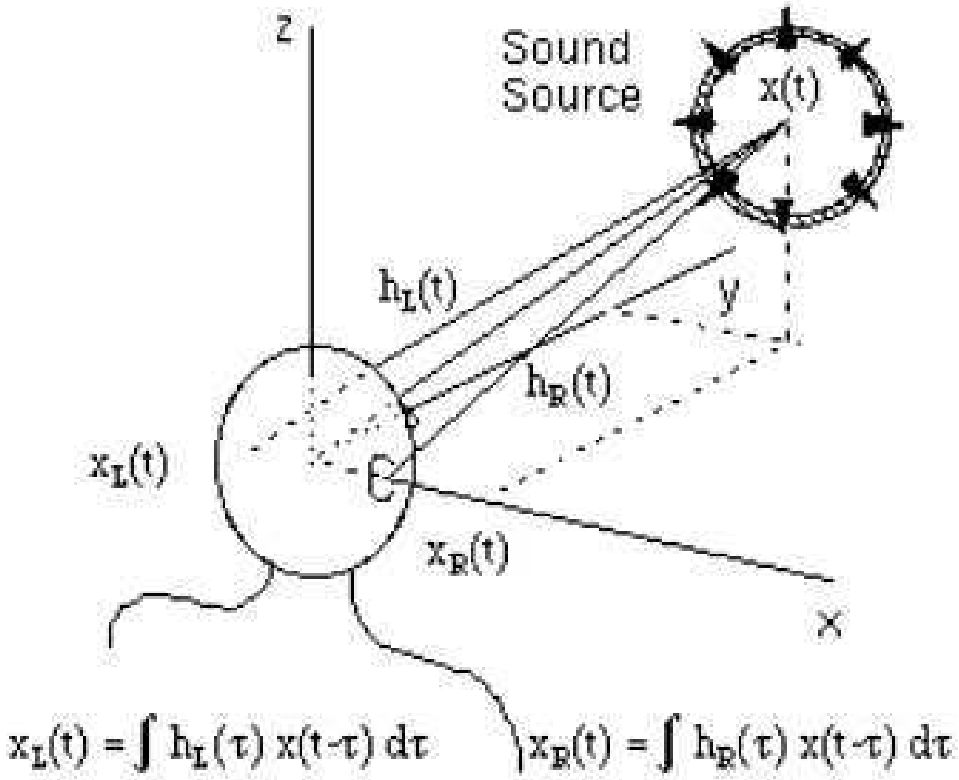


FIG. I.24 – Principe de la synthèse binaurale.

4 INTÊRET ET SPÉCIFICITÉ DE LA SYNTHÈSE BINAURALE

4.1 Décomposition des HRTF

Les HRTF peuvent être représentées soit par des fonctions complexes soit par une approche *traitement du signal* qui les représentent comme des filtres ou des réponses impulsionnelles. Les HRTF sont alors considérées comme des systèmes linéaires et invariants dans le temps. De plus, provenant de mesures physiques, elles sont représentées par des filtres RIF (Réponse Impulsionnelle Finie), ou FIR en anglais (Finie Impulse Response), causables et stables. Ces filtres sont appelés HRTF à phase mixte ($HRTF_{mixte}$). Il est alors possible de les décomposer en une composante à phase minimale $HRTF_{min}$ et une composante à excès de phase $HRTF_{excess}$. La composante à excès de phase est un filtre passe-tout qui contient les principales informations de phase. La composante à phase minimale a un module égal à celui de la composante à phase mixte. La phase de la composante à phase minimale φ_{min} se déduit du module de la $HRTF_{mixte}$ par une transformée de Hilbert (cf. équation I.9 [Oppenheim and Schafer (1989)]). Pour la plupart des positions mesurées, la phase de la composante à excès de phase φ_{excess} peut être considérée comme linéaire et peut être remplacée par un retard pur sans artefact audible [Kistler and Wightman (1992)]. L'implémentation la plus commune des HRTF est alors composée d'un retard pur et d'une composante à phase minimale. La différence entre les retards purs droit et gauche donne accès à l'ITD. Cette implémentation qui résulte de

plusieurs hypothèses est désignée par $\{ITD \bigoplus HRTF_{min}\}$ dans la suite du document. Les différentes décompositions des HRTF sont récapitulées ici et sont représentées sur la figure I.25 :

$$HRTF_{mixte} = H \cdot e^{\varphi} \tag{I.7}$$

$$|HRTF_{min}| = H \tag{I.8}$$

$$\varphi_{min} = Im\big[\text{Hilbert}(-\log(|H|))\big] \tag{I.9}$$

$$|HRTF_{excess}| = 1 \tag{I.10}$$

$$\varphi_{excess} = \varphi - \varphi_{min} \tag{I.11}$$

$$\widehat{\varphi}_{excess} \approx 2\pi f\tau \tag{I.12}$$

$$\varphi_{res} = \varphi_{excess} - 2\pi f\tau \tag{I.13}$$

$$ITD = \tau_{droit} - \tau_{gauche} \tag{I.14}$$

La phase résiduelle φ_{res} résultant de l'approximation linéaire de l'excès de phase est le plus souvent négligée car elle n'est pas audible. Cependant, il existe des HRTF dont les positions se situent principalement près de l'axe interaural et pour lesquelles cette information doit être reproduite. La figure I.26 (en bas) représente la phase résiduelle pour les positions ($\theta = 90°$, $\phi = 0°$) et ($\theta = 0°$,$\phi = 0°$). Dans [Plogsties et al. (2000)] il est montré que la phase résiduelle peut être premplacée par un retard pur additionnel sans l'introduction d'artefact audible. La figure I.26 fait apparaître une composante linéaire qui crée une avance de phase pour certaines fréquences. Ceci illustre les difficultés d'estimations de la composante linéaire de l'excès de phase. La composante linéaire de la phase résiduelle devrait être prise en compte par la méthode d'estimation. Cette observation pose aussi la question de la validité du modèle $\{ITD \bigoplus HRTF_{min}\}$. Cette question sera abordée au chapitre consacré à l'ITD.

4.2 Egalisation des HRTF

Quand les HRTF sont issues de mesures réalisées sur des sujets humains ou des têtes artificielles, il convient d'éliminer les contributions de la chaîne de mesure ainsi que toutes les contributions non spatiales. Pour enlever les contributions de la chaîne de mesure, une mesure de référence est effectuée au centre du système. Toutes les HRTF sont ensuite déconvoluées par cette mesure de référence. La suppression des autres contributions non spatiales est réalisée grâce à une égalisation par rapport à un champ sonore de référence. *L'égalisation champ libre* consiste à déconvoluer les HRTF par la HRTF à une position donnée, principalement la position ($\theta = 0°$,$\phi = 0°$). Cette égalisation optimise la reproduction des sources frontales. *L'égalisation champ diffus* (cf. fig. I.27) ne privilégie aucune direction et permet une optimisation de l'effet de salle. Les HRTF sont alors déconvoluées par une HRTF *moyennée* sur toutes les directions. Différentes méthodes d'estimations d'un champ diffus sont disponibles [Larcher (2001)]. Une HRTF *moyenne* comporte moins de fluctuations qu'une HRTF mesurée. L'égalisation champ diffus réduit davantage les colorations que l'égalisation champ libre.

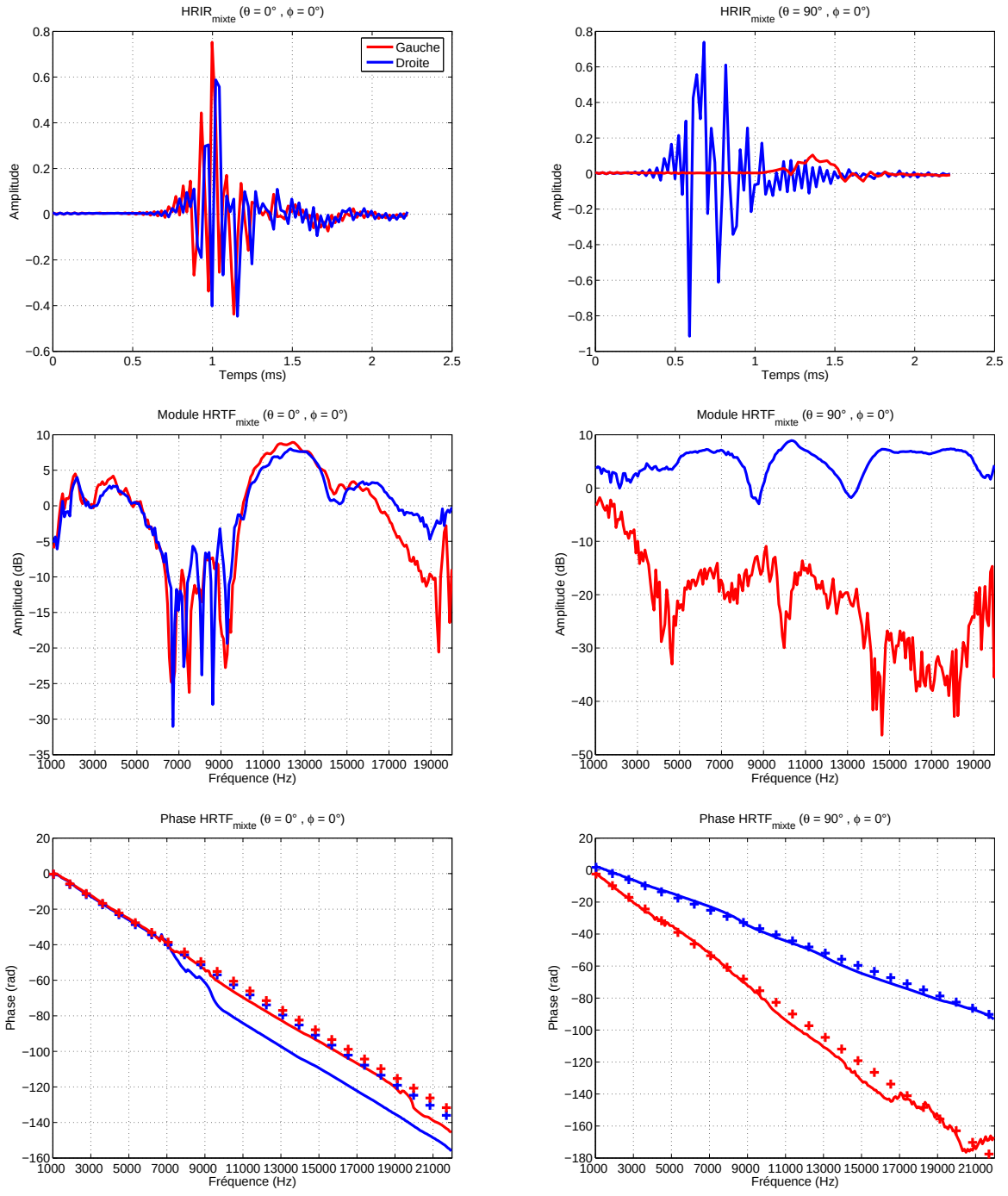


FIG. I.25 – $HRTF_{mixte}$ et $HRIR_{mixte}$. En rouge : oreille droite, en bleu : oreille gauche. Colonne de gauche : position $\theta = 0°$ $\phi = 0°$, colonne de droite position $\theta = 90°$ $\phi = 0°$. Figures en haut : HRIR. Figures au milieu : module de la $HRTF_{mixte}$. Figures en bas : phase de la HRTF. Les courbes en pointillé représentent l'estimation linéaire sur φ_{exces}.

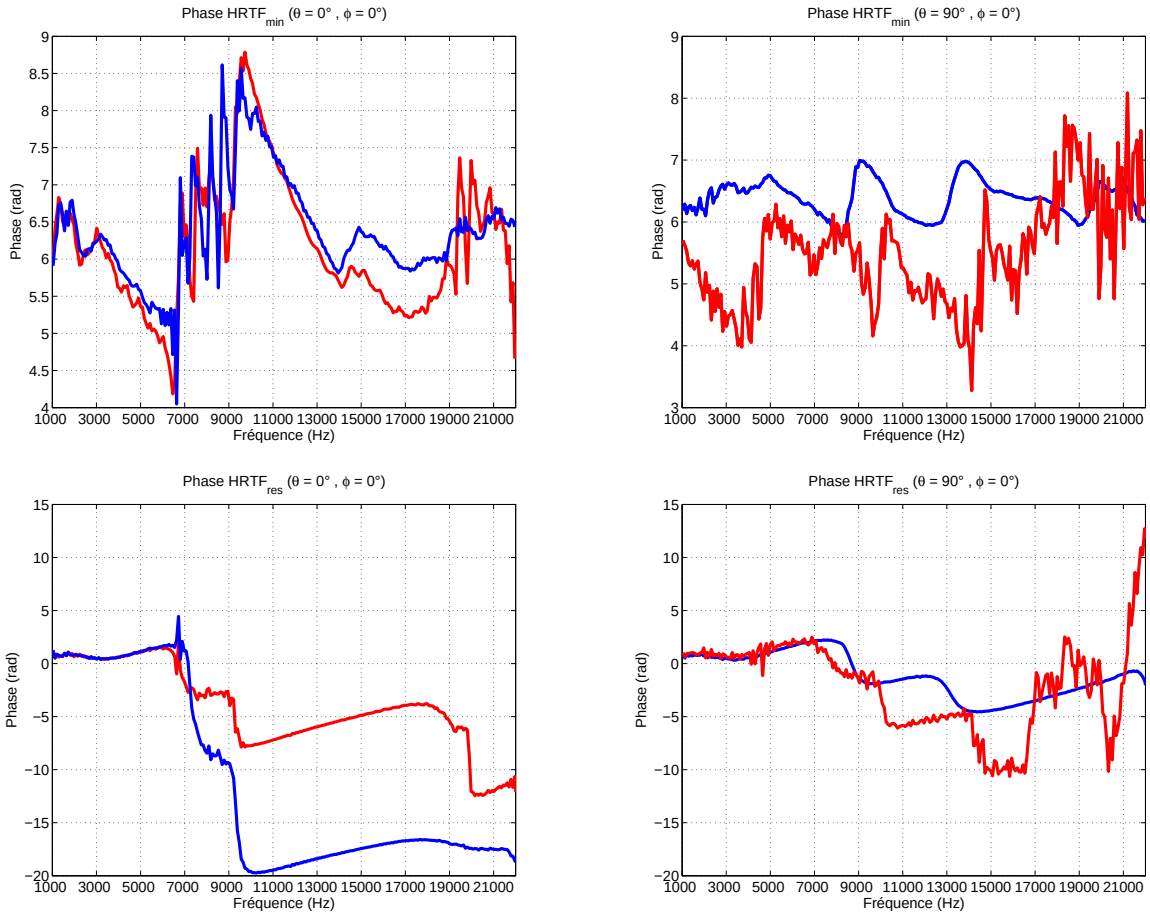


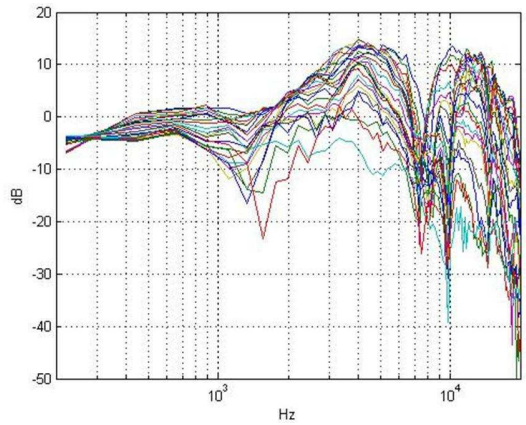
FIG. I.26 – HRTF et HRIR. En rouge : oreille droite, en bleu : oreille gauche. Colonne de gauche : position $\theta = 0° \ \phi = 0°$, colonne de droite position $\theta = 90° \ \phi = 0°$. Figures en haut : $phase_{min}$. Figures en bas : $phase_{res}$.

4.3 Lissage fréquentiel des HRTF

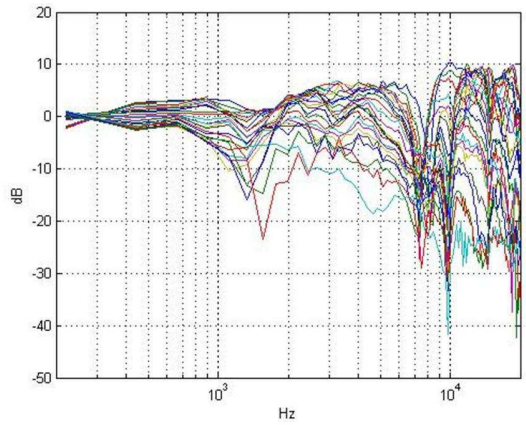
L'opération de lissage permet de réduire les variations des HRTF (cf. fig. I.27) pour prendre en compte la résolution fréquentielle du système auditif. Cette opération est réalisé par des moyennes glissantes du module des HRTF sur des tailles de fenêtres fréquentielles variables (cf. § 3.3.2). La taille des fenêtres fréquentielles est définie selon des échelles liées à la résolution fréquentielle du système auditif. Les HRTF *lissées* présentent des variations plus douces de leur module ce qui est bénéfique pour les opérations de modélisation des HRTF.

4.4 Réduction du coût d'implémentation

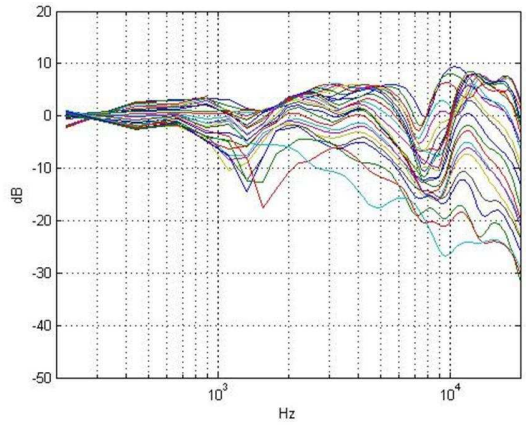
Les HRTF mesurées sont stockées sous la forme de filtres numériques RIF (Réponse Impulsionnelle Finie) généralement d'ordre 256, 512, 1024 ou 2048. Ces informations doivent être réduites pour de nombreuses applications (temps réel, multi-sources). Les HRIR (cf. fig I.25 du haut) sont majoritairement composées de zéros et l'information est concentrée dans les premiers échantillons. Sandvad et Hammershøï montrent que l'ordre



A) HRTF non égalisées



B) HRTF égalisées champ diffus



C) HRTF égalisées champ diffus et lissées

Fig. I.27 – HRTF d'un même individu pour un angle d'azimut variant entre -80° et 80°. Figures du haut : HRTF non égalisées. Figures du milieu : HRTF égalisées champ diffus. Figures du bas : HRTF égalisées champ diffus et lissées.

du filtre RIF peut être réduit à 72 [Sandvad and Hammershøi (1994)].

La spatialisation sonore à l'aide d'indices spectraux est gouvernée par les motifs, principalement les creux et les bosses, présents dans les HRTF. Asano et al. utilisent une modélisation *pole-zero*, ou modélisation RII (Réponse Impulsionnelle Infinie) pour rendre compte des motifs macroscopiques [Asano et al. (1990)]. Les résultats psychoacoustiques de cette étude indiquent qu'un modèle pôle-zero d'ordre 40, c'est-à-dire 81 coefficients, est perceptivement équivalent à un modèle tout-zero d'ordre 512. Sandvad et Hammershøï montrent qu'un modèle pôle-zero d'ordre 48 est nécessaire pour atteindre l'équivalence perceptive. Blommer et Wakefield indiquent qu'un modèle pôle-zéro d'ordre 60 est comparable à un filtre RIF d'ordre 2048 [Blommer and Wakefield (1997)]. Bien que la modélisation pôle-zéro puisse entraîner des artefacts de type confusion avant / arrière [Asano et al. (1990)], les algorithmes de modélisation des HRTF par des filtres RII sont toujours d'actualité [Hasegawa et al. (2000); Kulkarni and Colburn (2004)].

Le coût d'implémentation de la synthèse binaurale peut aussi être réduit grâce à l'implémentation *multi-canal*. Cette implémentation permet la reproduction des HRTF par un jeu de filtres de reconstruction et de gains spatiaux. Les filtres de reconstruction peuvent être obtenus par une analyse en composantes principales [Kistler and Wightman (1992)] ou par des techniques de projection des HRTF dans une base où leurs expressions peuvent être simplifiées [Larcher (2001)] (cf. chapitre 1 § 1.2). Cette implémentation permet la réduction du coût de calcul surtout pour des synthèses qui impliquent plusieurs sources (car le nombre de filtres est constant quelque soit le nombre de sources), mais introduit des artefacts audibles. Les techniques de décomposition doivent alors être adaptées pour réduire l'erreur perceptive [Rio and Warusfel (2002)].

4.5 Perception et technique binaurale

La technique binaurale permet en théorie de reproduire la perception naturelle. Le coût d'implémentation et les mesures des HRTF sont les principaux inconvénients de la méthode. De plus, l'utilisation de HRTF non-individuelles engendrent de nombreux artefacts audibles dont les confusions avant/arrière. L'utilisation d'un système de suivi de mouvement de la tête, ou *Head-Tracker* en anglais, allié à une implémentaion temps réel de la synthèse binaurale permet de réduire le taux de confusion avant/arrière et améliore la perception des sons en élévation [Bronkhorst (1995); Wightman and Kistler (1999)]. Ce système de reproduction qui permet la prise en compte des mouvements de la tête de l'auditeur par rapport à une scène sonore est appelé *synthèse binaurale dynamique* par opposition à la *synthèse binaurale statique* qui offre une scène sonore fixe. L'apport de la synthèse binaurale dynamique serait tel qu'il permettrait la réduction des artefacts audibles liées à des HRTF non-individuelles [Mackensen (2004)]. Cependant l'écoute statique est indispensable pour le contrôle précis des stimuli lors d'expériences psychoacoustiques : elle sera utilisée pour les tests d'écoute réalisés dans le présent travail de thèse.

Une scène sonore est généralement composée de nombreuses sources sonores et l'effet de salle peut être considéré comme une présentation simultanée des multiples sources-images liées aux réflexions sur les parois autour de l'auditeur. Les tests d'écoute décrits dans ce document se sont tous déroulés avec une seule source à la fois et sans effet de salle. Ceci permet l'étude systématique du paramètre testé.

4.6 Un exemple de moteur de spatialisation : le *Spat˜*

Le ***Spat˜*** (cf. fig. 4.6[6]), issu d'une collaboration entre l'IRCAM et France Telecom R&D, a pour objet la conception de modèles et de programmes de traitement du signal dédiés à la spatialisation sonore. Son domaine d'application couvre la création musicale, les productions audio-visuelles, la réalité virtuelle et les télécommunications. Il se compose d'un ensemble de modules logiciels de traitement du signal en temps réel. Il intègre, dans un même environnement, la synthèse de la localisation des sources sonores et celle de l'effet de salle (réverbération artificielle). L'architecture modulaire du Spatialisateur permet de s'adapter à la puissance de calcul disponible sur l'ordinateur hôte et de couvrir les différents formats de restitution classiques ou récents (stéréo, panoramiques d'intensité 2D ou 3D, binaural, transaural, ambisonic, wave field synthesis). Cette librairie existe, d'une part, sous la forme d'objets compatibles avec les environnements temps réel Max et jMax, et d'autre part, sous la forme d'une librairie de fonctions écrites en C/C++. La technologie *Spat˜* a fait l'objet de deux brevets et est utilisée dans de nombreuses productions musicales en concert ou dans la post-production discographique.

Les $HRTF_{min}$ utilisées pour le format binaural sont implémentées sous la forme de filtre RII d'ordre 12 (25 coefficients) pour des raisons de compacité. Le format original des $HRTF_{mixte}$ est sous la forme de filtres RIF de 512 coefficients. Un test a été réalisé pour estimer les éventuelles dégradations apportées lors des étapes transformant les filtres RIF en filtres RII (les détails du test sont donnés en annexe A).

Etant donné le caractère informel de ce test, les résultats sont présentés sous la forme d'une synthèse des propos recueillis. Les différences perçues portent sur le timbre, la localisation, la largeur de la tache sonore et l'externalisation de la source. Dans la première phase, les sujets connaissant les types de traitement du son écouté, arrivent à décrire des différences par rapport à la synthèse directe à quasiment toutes les étapes. Lors de l'écoute en aveugle certaines différences n'apparaissent plus. L'importance des dégradations perçues semble dépendre de la position synthétisée. Certaines étapes introduisent très peu d'effets perceptibles. C'est le cas du lissage de l'amplitude des HRTF, du warping fréquentiel et de la symétrisation. La modélisation RII des HRTF entraîne parfois un rapprochement de la source et procure une tache sonore moins précise. De plus une légère différence de timbre entre les méthodes RII apparaît. Toutefois, lors du test en aveugle, quasiment aucun sujet ne reporte des différences entre les méthodes RII.

Les dégradations les plus importantes apparaissent avec la modélisation de l'ITD. Ces dégradations sont aussi celles qui persistent avec le test en aveugle, avec un effet moindre. Ainsi la modélisation de la phase par l'ITD entraîne les défauts suivants :

- modification du timbre avec un contenu plus basses fréquences
- décalage en localisation (sur ou sous-estimation de l'ITD), avec une légère élévation de la source
- tache sonore plus diffuse, plus large
- léger rapprochement de la source (pour les positions frontales).

Les modifications perçues peuvent venir de l'implémentation de l'ITD au moyen d'un retard fractionnaire. En effet la réponse en fréquence de ce retard montre qu'il effectue un filtrage passe-bas sur le signal avec une fréquence de coupure à -3 dB vers 13 kHz. Cette modification du contenu fréquentiel peut aussi être à l'origine d'une sensation

[6]Informations disponibles à l'adresse suivante : http://recherche.ircam.fr/equipes/salles/

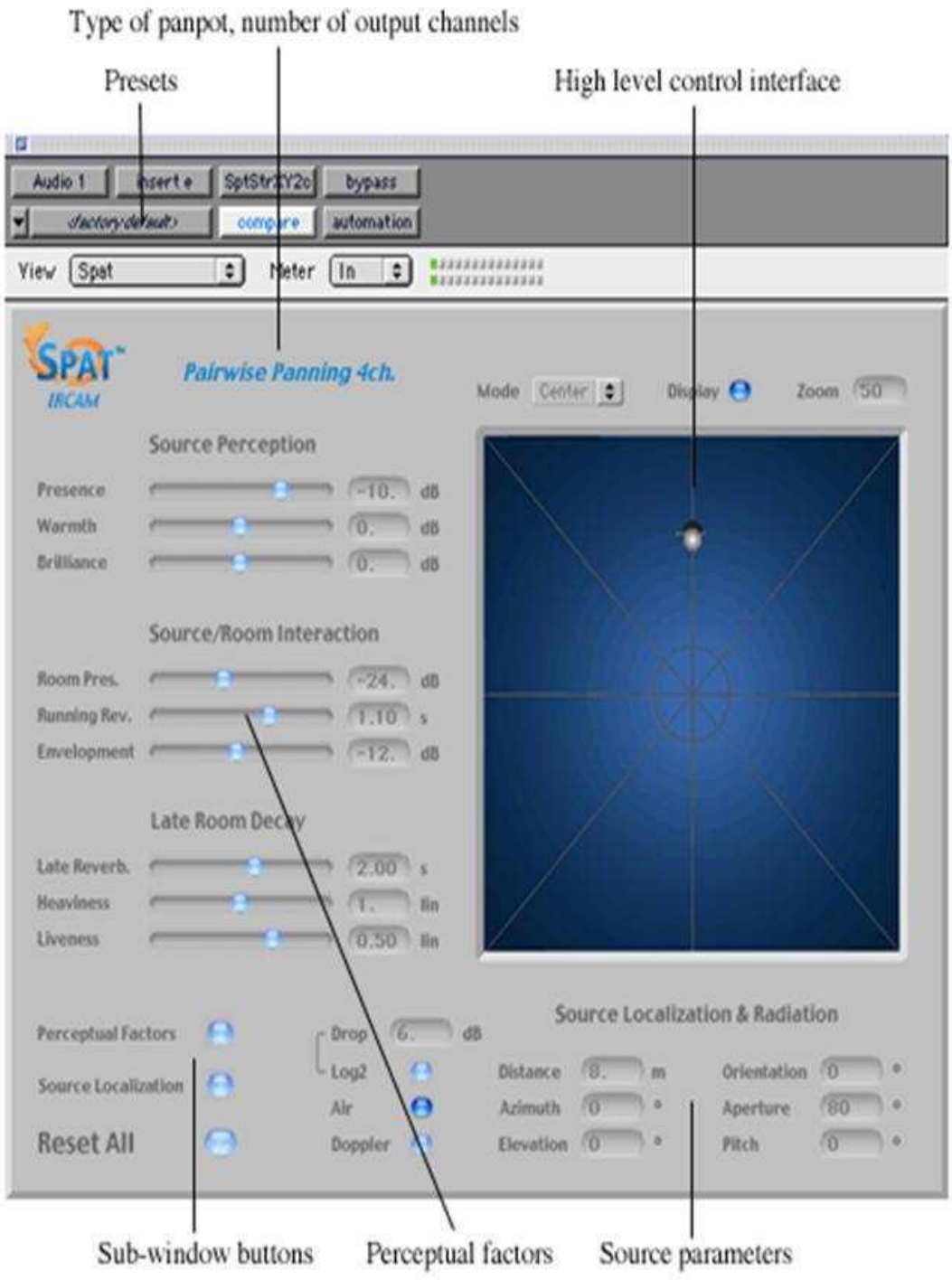


FIG. I.28 – Interface principale du *Spat˜*.

d'élévation de la source (cf. § 2.3.3) et d'un élargissement de la tache sonore.

Même si l'implémentation *Spat˜* engendre des modifications perceptibles par rapport à la convolution avec les HRIR mesurées et compensées champ diffus, ces modifications restent faibles et dépendent de la position synthétisée. Cependant, ce test n'a pas permis de savoir si ces transformations sont perçues comme des dégradations. Ce test permet toutefois d'illustrer l'importance d'une modélisation correcte de l'ITD et justifie l'orientation de la thèse dont une grande partie est consacrée à l'implémentation de l'ITD.

5 LE BINAUMATON

Pour s'approcher le plus de la perception naturelle, la synthèse binaurale doit utiliser des HRTF individuelles. Devant la difficulté de la mesure indivuelle de HRTF, la problématique globale de la synthèse binaurale peut être énoncée de la façon suivante :

Comment obtenir un rendu binaural de qualité sans mesurer les HRTF d'un auditeur ?

Plusieurs techniques sont proposées dans la littérature :

Jeu de HRTF "proches" : les HRTF sont principalement déterminées par la morphologie de l'auditeur. Les caractéristiques morphologiques de l'auditeur sont mesurées et sont comparées à celles associées aux HRTF stockées dans une base de donnée. Les HRTF les plus proches, au sens de la morphologie, sont alors sélectionnées pour cet auditeur.

Adaptation de HRTF : les distances inter-individuelles peuvent être réduites par une dilatation fréquentielle des HRTF [Middlebrooks (1999); Busson et al. (2004)]. Une correspondance entre facteur de dilatation optimal et paramètres morphologiques peut être trouvée afin d'adapter les HRTF d'un mannequin artificiel, par exemple, à n'importe quel auditeur.

Choix de HRTF : Plusieurs jeux de HRTF sont proposés et l'auditeur choisit celles qui lui conviennent le mieux. Cette procédure peut aussi être appliquée pour le choix du coefficient de dilatation fréquentielle.

Construction de HRTF : les techniques numériques de modélisation physique permettent une estimation des HRTF par le biais d'une représentation physique de l'auditeur.

D'autres approches hybrides peuvent être réalisées. L'idée du *binaumaton* s'inspire du *photomaton* pour offrir à n'importe quel auditeur ses HRTF individualisées à partir de sa morphologie identifiée d'après des prises de vues . Cette idée a fait l'objet d'un dépôt de brevet (FR 03 02467).

6 AXE DE RECHERCHE DE LA THÈSE

L'ensemble des travaux réalisés pour la thèse est dédié au problème de l'individualisation de la synthèse binaurale. Son implémentation $\{ITD \bigoplus HRTF_{min}\}$ est la plus commune : elle est utilisée comme base de travail. Les efforts d'individualisation se sont alors portés sur l'ITD, d'une part, et sur le module des $HRTF$[7], d'autre part.

[7]Comme la phase des $HRTF_{min}$ est entièrement déterminée par son module, travailler sur le module des HRTF revient à travailler sur les $HRTF_{min}$.

6.1 Individualisation de l'ITD

L'ITD a été très étudiée dans le passé comme indice de localisation isolé et peu d'études l'ont abordée dans le cadre de l'implémentation $\{ITD \bigoplus HRTF_{min}\}$. Afin de déterminer la meilleure technique d'estimation au sens de la perception, il convient d'établir les valeurs de l'ITD psychoacoustique et sa résolution en fonction de la position. Le premier axe d'étude (cf. chapitre II) est consacré à l'établissement des ces données. L'individualisation de l'ITD est alors abordée à la lumière de la sensibilité et de la variation inter-individuelle de l'ITD. Deux études expérimentales sont menées pour connaître d'une part l'ITD psychoacoustique dans le plan horizontal et d'autre part la sensibilité perceptive à l'ITD sur les cônes de confusion. La première expérience permet en outre la validation perceptive de l'implémentation $\{ITD \bigoplus HRTF_{min}\}$ même pour les positions où elle peut être remise en cause. Une nouvelle formulation de l'ITD du modèle de tête sphérique est aussi proposée. Cette formulation prend en compte les variations de l'ITD avec l'élévation et permet une estimation robuste et fidèle de l'ITD pour un grand nombre de sujets.

6.2 Acquisition de HRTF

L'évaluation d'un moteur de synthèse binaurale est souvent réalisée par comparaison avec l'utilisation de HRTF individuelles. L'acquisition des HRTF est une tâche difficile. L'axe de recherche concerne les techniques d'acquisition des HRTF par la mesure et par le calcul par éléments de frontière. Les différentes bases de données de HRTF utilisées pour ce travail de thèse sont décrites. Une méthode d'estimation perceptive de l'erreur introduite n'existe pas et le problème de la mesure de l'erreur est abordé. Grâce aux calculs par éléments de frontière, l'apport de modélisations géométriques simples, basées sur la morphologie de l'auditeur est évalué. La méthode de calcul par éléments de frontière donne accès aux informations basses fréquences contenues dans le module des HRTF.

6.3 Apprentissage de HRTF par la technique neuronale

Le troisième axe de recherche est l'étude d'une solution alternative à la mesure et aux calculs par éléments de frontière (cf. chapitre 1). Une étude de faisabilité est menée pour évaluer l'intérêt des réseaux de neurones pour réduire le nombre de mesures. L'idée est de mesurer un nombre restreint de positions et de prédire les autres positions. Le réseau de neurones *apprend* les relations de haut niveau entre les HRTF mesurées et les HRTF à prédire. La question de la position et du nombre de HRTF à mesurer est abordée. Les écoutes informelles réalisées avec des HRTF prédites par le réseau de neurones sont très encourageantes et le réseau est capable de reproduire des détails spectraux fins qui sont responsables du caractère idiosyncratique des HRTF.

BIBLIOGRAPHIE

Asano, F., Suzuki, Y., and Sone, T. (1990). Role of spectral cues in median plane localization. *J. Acoust. Soc. of Am.*, 88 :159–168.

Baumgarte, F. and Faller, C. (2003). Binaural cue coding-part i : psychoacoustic fundamentals and design principles. *IEEE Transactions on Speech and Audio Processing*, 11(6) :509–519.

Blauert, J. (1983). *Spatial Hearing*. The mit press edition.

Blommer, M. A. and Wakefield, G. H. (1997). Pole-zero approximations for head-related transfer functions using a logarithmic error criterion. *IEEE Transactions on speech and audio processing*, 5(3) :278–287.

Bronkhorst, A. W. (1995). Localization of real and virtual sound sources. *J. Acoust. Soc. of Am.*, 98(5) :2542–2553.

Busson, S., Nicol, R., and Warusfel, O. (2004). Influence of the ears canal location on spherical head model for the individualised interaural time difference. *Proceedings of CFA / DAGA Joint Meeting, Strasbourg, France*.

Chateau, N. (1996). *Localisation de sources sonores multiples dans l'hémisphère supérieur*. PhD thesis, Université de la méditerranée Aix Marseille II, laboratoire de mécanique et d'acoustique.

Cheng, C. I. and Wakefield, G. H. (2001). Introduction to head-related transfer functions (hrtfs) : representations of hrtfs in time, frequency and space. *J. Audio Engin. Soc.*, 49(4) :231–249.

Daniel, J. (2000). *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.

Faller, C. and Baumgarte, F. (2002). Binaural cue coding : a novel and efficient representation of spatial audio. volume 2, pages 1841–1844.

Faller, C. and Baumgarte, F. (2003). Binaural cue coding-part ii : Schemes and applications. *IEEE Transactions on Speech and Audio Processing*, 11(6) :520–531.

Grantham, D. W., Hornsby, B. W. Y., and Erpenbeck, E. A. (2003). Auditory spatial resolution in horizontal, vertical, and diagonal planes. *J. Acoust. Soc. of Am.*, 114(2) :3030–3038.

Han, H. L. (1994). Measuring a dummy head in search of pinna cues. *J. Acoust. Soc. of Am.*, 42 :15–37.

Hasegawa, H., Kasuga, M., Matsumoto, S., and Koike, A. (2000). Simple relaization of sound localization using hrtf approximated by iir filter. *IEICE Transactions fundamentals*, E83-A(6) :973–978.

Jin, G., Corderoy, A., Carlile, S., and van Shaik, A. (2004). Contrasting monaural and interaural spectral cues for human sound localization. *J. Acoust. Soc. of Am.*, 115(6) :3124–3141.

Kistler, D. J. and Wightman, F. L. (1992). A model of head-related transfert functions based on principal components analysis and minimum-phase reconstruction. *J. Acoust. Soc. of Am.*, 91(3) :1637–1647.

Kulkarni, A. and Colburn, H. S. (2004). Infinite-impulse-response models of the head-related transfer function. *J. Acoust. Soc. of Am.*, 115(4) :1714–1728.

Langendijk, E. H. A. and Bronkhorst, A. W. (2002). Contribution of spectral cues to human sound localization. *J. Acoust. Soc. of Am.*, 112(4) :1583.

Larcher, V. (2001). *Techniques de spatialisation des sons pour la réalité virtuelle.* PhD thesis, Université Paris VI.

Lord Rayleigh (1876). Our perception of the direction of sound. *Nature*, XIV :32–33.

Lord Rayleigh (1907). On our perception of sound direction. *Phylosophy Magazine*, 13 :214–232.

Mackensen, P. (2004). *Auditive localization. Head movements, an additional cue in localization.* PhD thesis, Technischen Universität Berlin.

Macpherson, E. A. and Middlebrooks, J. C. (2002). Listener weighting of cues for lateral angle : The duplex theory of sound localization revisited. *J. Acoust. Soc. of Am.*, 111(5) :2219–2236.

Mercier, D. (1993). *Le livre de techniques du son. Tome III.* Fréquences, paris edition. 458 p.

Middlebrooks, J. (1999). Individual differences in external-ear transfer function reduced by scaling in frequency. *J. Acoust. Soc. of Am.*, 106(3) :1480–1492.

Middlebrooks, J., Makous, J. C., and Green, D. M. (1989). Directional sensitivity of sound-pressure levels in the human ear canal. *J. Acoust. Soc. of Am.*, 86 :89–107.

Middlebrooks, J. C. (1992). Narrow-band sound localization related external ears acoustics. *J. Acoust. Soc. of Am.*, 92 :2607–2624.

Oldfield, S. R. and Parker, S. P. A. (1984). Acuity of sound localisation : a topography of auditory space. i. normal hearing. *Perception*, 13 :581–600.

Oppenheim, A. V. and Schafer, R. W. (1989). *Discrete-Time Signal Processing.* Prentice Hall, Englewood Cliffs, New Jersey.

Plogsties, J., Minnaar, P., Olesen, S. K., Christensen, F., and Moller, H. (2000). Audibility of all-pass components in head-related transfer function. Paris. AES 108$^{\text{th}}$ Convention.

Rio, E. and Warusfel, O. (2002). Optimizations of multi-channel binaural formats based on statistical analysis. Seville. FORUM ACUSTICUM SEVILLA.

Sandvad, J. and Hammershøi, D. (1994). Comparison of fir and iir filter representation of hrirs. pages 1–16. preprint 3862.

Shaw, E. A. G. (1982). *Localization of sound : theory and applications.* Amphora Press. chapitre External ear response and sound localization.

Shin-Cunningham, B. G., Santelli, S., and Kopco, N. (2000). Tori of confusion : binaural localization cues for sources within reach of a listener. *J. Acoust. Soc. of Am.*, 107(3).

Wightman, F. L. and Kistler, D. J. (1989a). Headphone simulation of free-field listening. i : Stimulus synthesis. *J. Acoust. Soc. of Am.*, 85(2) :858–867.

Wightman, F. L. and Kistler, D. J. (1989b). Headphone simulation of free-field listening. ii : Psychophysical validation. *J. Acoust. Soc. of Am.*, 85(2) :868–878.

Wightman, F. L. and Kistler, D. J. (1997). Monaural sound localization revisited. *J. Acoust. Soc. of Am.*, 101(2) :1050–1063.

Wightman, F. L. and Kistler, D. L. (1999). Resolution of front–back ambiguity in spatial hearing by listener and source movement. *J. Acoust. Soc. of Am.*, 105(5) :2841–2853.

Woodworth, R. S. and Schloesberg, G. (1962). *Experimental psychology.* New-York. pp 349-361.

II

Quelle différence interaurale de temps pour la synthèse binaurale ?

INTRODUCTION

La plupart des moteurs de rendu binaural effectuent une séparation entre ITD et $HRTF_{min}$ ce qui permet une réduction du coût d'implémentation de la synthèse binaurale. Les informations de phase droite et gauche sont regroupées dans une seule valeur l'ITD (cf. § I.4.1). Le présent chapitre s'intéresse au problème de l'individualisation de l'ITD, les aspects relatifs aux informations spectrales contenues dans le filtre à phase minimale seront abordés aux chapitres III et 1.

L'implémentation de l'ITD comme un retard pur impose une discrétisation en nombre entier d'échantillons[1]. Cette représentation est liée au contexte d'implémentation et peut sembler éloignée de la définition psychoacoustique de l'ITD comme premier indice perceptif de spatialisation ainsi que de la définition physique décrivant l'ITD comme la différence de temps d'arrivée entre l'oreille droite et l'oreille gauche d'une onde de pression acoustique. La problématique du travail présenté dans ce chapitre pose la question de la valeur d'ITD à appliquer à un filtre à phase minimale pour que le rendu de l'implémentation $\{ITD \bigoplus HRTF_{min}\}$ soit le plus proche, au sens de la perception, de la

[1] Les méthodes de délai fractionnaire peuvent introduire des artefacts audibles [Laakso et al. (1996)]

$HRTF_{mixte}$. Cette problématique a été abordée selon deux approches complémentaires :

- **Estimation la valeur de l'ITD** : Plusieurs méthodes d'estimation de l'ITD existent et font appel à des principes physiques, psychoacoustiques ou expérimentaux. Quelle méthode est la plus proche de la perception ?
- **Sensibilité perceptive à l'ITD** : L'estimation de l'ITD est soumise à de nombreuses variations et il convient d'estimer dans quelle mesure il faut les reproduire d'un point de vue perceptif. Quelle est notre sensibilité à l'ITD ?

La valeur du retard pur à associer aux filtres binauraux doit prendre en compte ces deux aspects pour déterminer l'effort d'individualisation à apporter à l'ITD au regard de l'application désirée (temps réel ou pré-calculé, public occasionnel ou sujet entraîné).

Pour apporter des éléments de réponse à cette question, la première partie de ce chapitre est consacrée à l'exposé et à la confrontation de l'état des connaissances sur l'ITD. Premièrement, les aspects liés à la perception seront présentés. Ensuite, les modèles de prédiction de l'ITD issus d'une modélisation physique de l'auditeur seront exposés et l'avantage du modèle de tête sphérique sera dégagé. Puis les trois familles de méthodes d'estimation de l'ITD à partir des mesures seront décrites. Enfin, les différentes dépendances de l'ITD seront comparées et une tentative de hiérarchisation de leur importance sera proposée. La première partie dégagera les lacunes des connaissances actuelles liées à la problématique générale du chapitre. Les autres parties du chapitre exposent le travail réalisé pour tenter de répondre aux questions ouvertes que sont l'estimation de l'ITD et de sa sensibilité.

La deuxième partie présente le travail théorique réalisé pour la définition d'une nouvelle formule d'estimation de l' ITD. Cette formule est basée sur le modèle de tête sphérique et offre une modélisation plus réaliste de l'ITD. Elle fournit, entre autre, la possibilité d'une individualisation plus pointue de l'ITD.

La troisième partie décrit une étude expérimentale permettant d'une part de valider le modèle d'implémentation $\{ITD \bigoplus HRTF_{min}\}$ et de déterminer d'autre part les méthodes de calcul les plus à même de reproduire les variation de l'ITD sur le plan horizontal. La sensibilité à l'ITD n'étant que très peu étudiée dans le cadre de l'implémentation $\{ITD \bigoplus HRTF_{min}\}$ de la synthèse binaurale, la cinquième partie présentera une expérience psychoacoustique qui a été menée pour mesurer les écarts perceptibles d'ITD. En fonction de cette sensibilité, il faudra adapter les modèles et les méthodes d'estimations. Enfin, une sixième et dernière partie regroupe les résultats obtenus au cours des deux expériences réalisées et propose des suites à donner à ces travaux.

1 ETAT DE L'ART DE LA CONNAISSANCE SUR L'ITD

Dans l'exposé de la théorie duplex de la localisation auditive, Rayleigh [Lord Rayleigh (1907)] indique que la perception de la direction des sources sonores est gouvernée par deux indices : l'ITD et l'ILD. Sans remettre en question l'importance perceptive de l'ILD, cette dichotomie n'est pas réalisée dans les travaux présentés ici et une des hypothèses de travail est de considérer que l'ILD est compris dans les différences spectrales interaurales résultant d'une présentation dichotique des HRTF.

1.1 ITD et localisation auditive

La différence interaurale de temps est l'indice psychoacoustique le plus important dans la localisation perceptive des sources sonores [Blauert (1983)]. Bien que l'ITD varie également avec l'élévation, la contribution principale de l'ITD à la localisation sonore est la perception des sources en azimut. Suivant cette définition, le principe de la mesure de l'ITD consiste alors à présenter des stimuli dont l'ITD est contrôlé et de demander au sujet d'indiquer la direction perçue. Ce principe se heurte à deux difficultés majeures : le contrôle de l'ITD et le report de la direction perçue. Le problème du report des réponses dans les tests d'écoute en synthèse binaurale reste une question ouverte. Cette question aborde des thématiques très diverses comme l'interaction multi-modale ou la kinesthésie. Le problème du report des réponses n'a pas été abordé dans ces travaux de thèse et le lecteur pourra se repporter aux travaux exposés dans [Pernaux et al. (2003)]. Ensuite le contrôle de l'ITD pose directement la question du type d'écoute réalisé. Le contrôle de l'ITD en continu ne peut se réaliser que dans le cas d'une écoute au casque. En effet, lors d'une écoute en champ libre, l'auditeur perçoit les sons à travers les HRTF et donc le contrôle de l'ITD en champ libre revient à extraire l'ITD de la même manière qu'il est extrait à partir des mesures. L'utilisation de la technique transaurale pose le même problème. Par contre, une écoute au casque permet de contrôler l'ITD avec une ligne à retard [Klump and Eady (1956); Domnitz (1973); Domnitz and Colburn (1977)], mais cela revient à effectuer des écoutes sans indices spectraux car il faudrait contrôler les HRTF de manière continue comme l'ITD. Ce type d'écoute réduit la spatialisation à une perception intra-crânienne : l'ITD n'est plus alors un indice de localisation mais un indice de latéralisation.

Un autre principe de mesure de l'ITD, qui rejoint la définition de l'ITD physique, serait de placer une source mobile émettant une impulsion de Dirac sur un arc avec en son centre deux microphones espacés d'une distance égale à la distance interaurale. Cette mesure est à la fois difficile à réaliser et presque hors de propos. En effet, d'une part, une impulsion de Dirac n'est pas réalisable avec un système haut-parleur du fait de l'énergie nécessaire et de l'inévitable dispersion introduite par les éléments mécaniques, et d'autre part, réduire l'oreille externe à deux capsules microphoniques espacées, c'est omettre d'importantes contributions de la morphologie de l'auditeur en terme de trajets acoustiques de la source aux tympans. La mesure de l'ITD nécessite alors d'utiliser d'autres signaux sources et une modélisation fidèle d'un auditeur, comme une tête artificielle, ce qui revient à extraire l'ITD de HRTF mesurées.

Si l'ITD au sens de la psychoacoustique semble difficilement mesurable, la connaissance de l'ITD relatif, ou plus précisément l'écart d'ITD non perçu, ou encore la Just Noticeable Difference (JND) en anglais, permettrait la définition d'une marge d'erreur acceptable sur l'ITD. La mesure de la JND de l'ITD semble plus simple que la mesure de l'ITD proprement dite et fait appel à des méthodes de mesure de seuil de la psychoacoustique. Comme décrit auparavant, cette mesure est dépendante du type d'écoute réalisé.

Etant donné la problématique centrale de ce chapitre, des tests d'écoute doivent être conduits pour répondre à cette question. Par contre, la modélisation physique de l'auditeur permet d'obtenir des valeurs déterministes de l'ITD, évitant ainsi de lourdes procédures psychoacoustiques.

1.2 Prédiction de l'ITD à partir d'une modélisation physique de l'auditeur

La modélisation physique de l'auditeur fut une des premières méthodes pour étudier les dépendances spatiales de l'ITD. Cette méthode considère la morphologie de l'auditeur comme étant composée de formes géométriques simples comme des sphères ou des ellipsoïdes. L'utilisation de formes simples permet la résolution des équations de propagation des ondes acoustiques. Des formules analytiques de l'ITD en fonction des coordonnées spatiales sont alors extraites. L'utilisation de formes simples permet aussi la formulation de relations analytiques entre les paramètres du modèle et l'ITD par le biais de considérations de tracé de rayon acoustique. La formulation la plus simple donnant l'ITD sur le plan horizontal, et aussi la plus utilisée, est ainsi obtenue par Woodworth [Woodworth and Schloesberg (1962)] qui considère la tête de l'auditeur comme une sphère rigide de rayon a et dont les oreilles sont diamétralement opposées (cf. fig.II.1) :

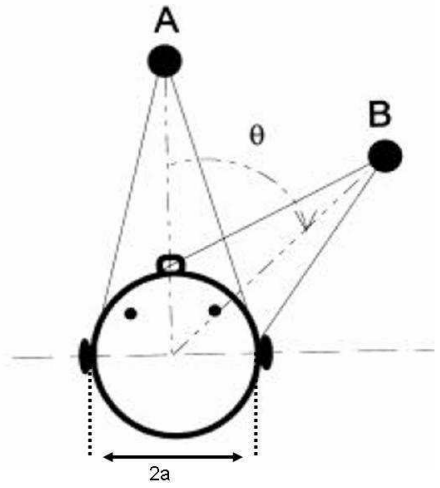


FIG. II.1 – Modèle de tête sphérique de rayon a.

$$ITD_{sphere}(\theta) = \frac{a}{c}(\sin(\theta) + \theta) \tag{II.1}$$

où c représente la célérité des ondes acoustiques et θ l'azimut de la source. Cette formulation a été largement étudiée et notamment certains travaux reportent sa bonne adéquation avec des mesures effectuées sur une tête artificielle [Kuhn (1977)] et avec une moyenne des ITD individuelles de 70 personnes [Plogsties et al. (2000)].

Cette formule a d'abord été proposée avec un rayon correspondant à une moyenne anthropométrique de 8.75 cm. Dans [Algazi et al. (2001b)] une étude empirique est menée pour individualiser le rayon du modèle sphérique. Un rayon optimal est obtenu au sens de la minimisation de l'erreur quadratique entre les valeurs de l'ITD données par la formule de Woodworth (cf. équation II.1) et l'ITD estimé à partir des mesures des 25 individus utilisés dans l'étude. La méthode d'estimation de l'ITD est une méthode de type *seuil* (cf. § 1.3.1) et est appliquée à des HRIR filtrées passe-haut à $f_c = 1.5$ kHz. Le seuil est pris à 10% de la valeur maximale de la HRIR (qui est suréchantillonnée d'un facteur 8). L'ITD obtenu est alors lissé par projection sur les harmoniques sphériques et troncature des ordres supérieurs de la projection. La procédure de minimisation vise une réduction de l'erreur quadratique par rapport à l'erreur introduite en utilisant le rayon anthropométrique moyen. Afin d'obtenir ce rayon, les auteurs réalisent une régression multi-linéaire entre différents paramètres anthropométriques et le rayon optimal. Le rayon optimal est alors calculé à partir de la demi-largeur, la demi-profondeur et la demi-hauteur de la tête respectivement X_1, X_2 et X_3 dans l'équation II.2 :

$$A_{opt} = 0,51X_1 + 0,019X_2 + 0,18X_3 + 3,2 \tag{II.2}$$

où A_{opt} désigne le rayon optimal en cm.

Plusieurs études ont permis d'ajouter la dépendance de l'ITD avec l'élévation. L'élévation est ici définie dans le système de coordonnées polaires verticales avec des mesures réalisées par plan d'azimut constant. Deux formulations co-existent sur la base de leur adéquation avec la base de données dont elles ont été extraites. [Larcher and Jot (1999)] proposent la formulation suivante :

$$ITD_{Larcher}(\theta) = \frac{a}{c}(\arcsin(\sin(\theta)\cos(\phi)) + \sin(\theta)\cos(\phi)) \tag{II.3}$$

tandis que [Saviojaa et al. (1999)] utilisent celle-ci :

$$ITD_{Savioja}(\theta) = \frac{a}{c}(\sin(\theta) + \theta).\cos(\phi) \tag{II.4}$$

Les erreurs dues à la modélisation de la tête de l'auditeur par une sphère avec oreilles centrées sont principalement une sous-estimation de l'ITD pour les positions proches de l'axe interaural et une ITD constant avec l'élévation. Les sources identifiées de ces erreurs sont la non sphéricité de la tête et la position des oreilles par rapport au centre de la tête [Algazi et al. (2001b); Duda et al. (1999)]. Les variations non reproduites par le modèle de tête sphérique sont situées sur les cônes de confusion (cf. § 2.3.2 et fig. II.9), où les variations de l'ITD en élévation, dans un système de coordonnées polaires-interaurales, peuvent atteindre 18 % de la valeur maximale de l'ITD sur un plan d'azimut constant.

Le modèle sphérique permet un prédiction robuste et assez fidèle de l'ITD mais qui ne varie pas avec l'élévation. Une autre modélisation physique de l'ITD a été alors créée en considérant la tête de l'auditeur comme une ellipsoïde rigide [Duda et al. (1999)]. L'ITD est obtenue de la même manière que pour la formule de Woodworth par la définition de trajet acoustique à la surface de l'ellipsoïde. Le modèle ellipsoïdal permet une meilleure prédiction en élévation que le modèle sphérique. Cependant le calcul est réalisé au moyen d'un algorithme prenant en compte la position relative du placement des oreilles ce qui se révèle beaucoup moins robuste et surtout moins pratique qu'une formule analytique.

La modélisation de l'ITD par la méthode de tracé de rayons est délicate pour des formes moins triviales. Pour étudier des modélisations physiques plus complexes, le calcul de HRTF par éléments de frontière peut être utilisé. L'ITD est alors estimée par les techniques d'extraction à partir des HRTF. Dans [Pernaux (2003)] une étude est menée sur l'ITD extraite à partir de HRTF calculée par la technique BEM. Plusieurs modélisations sont réalisées : modèle sphérique avec et sans modélisation du pavillon et modèle ellipsoïdal avec et sans décalage des oreilles. L'ITD est calculée par la méthode MaxIacc (cf. § 1.3.2) sur la composante à excès de phase des HRIR. Ces travaux montrent que ni une modélisation du pavillon ni le modèle de tête ellipsoïdale n'apporte des variations de l'ITD en élévation, ce qui est contradictoire avec [Duda et al. (1999)]. Par contre, les travaux montrent que le décalage des oreilles permet d'apporter des variations de l'ITD en élévation proches de celles observées sur les mesures.

Les modélisations physiques permettent une prédiction robuste et simple d'emploi de l'ITD. Cependant, la simplicité des modèles engendre des erreurs parfois importantes. Les méthodes d'estimation à partir des mesures offrent l'avantage de ne pas faire d'approximation de la morphologie de l'auditeur en travaillant directement avec les HRTF contenant tous les indices de localisation.

1.3 Estimation de l'ITD à partir des HRTF/HRIR

Les techniques d'estimation de l'ITD à partir des mesures sont nombreuses et les principes sur lesquels elles sont fondées sont hétéroclites. Comme toute méthode d'estimation, elles sont soumises à des problèmes de biais, de robustesse et de confiance. La présentation des estimateurs de l'ITD, qui ne saurait être exhaustive vu le nombre de techniques disponibles, est ici divisée en trois groupes :

- Les méthodes d'estimation de seuil
- Les méthodes de corrélation entre les HRTF, ou les HRIR, droits et gauches
- Les méthodes utilisant la phase des HRTF.

Au sein de ces trois groupes, de nombreuses variantes existent selon les auteurs et selon les propos.

1.3.1 Estimation de seuil

Le principe des méthodes d'estimation de l'ITD de type *seuil* est de déterminer le *temps d'arrivée*, ou encore le *retard initial* de l'onde sur l'oreille droite τ_d et sur l'oreille gauche τ_g. L'ITD est alors donnée par (cf. fig. II.2) :

$$ITD_{seuil} = \tau_d - \tau_g \tag{II.5}$$

La méthode la plus courante estime le *temps d'arrivée* comme l'instant où la HRIR dépasse un seuil donné. Par exemple, le temps d'arrivée peut correspondre au temps pour lequel la HRIR atteint 10 % de son maximum. Cette méthode considère donc que les HRIR sont constituées d'un pic principal situé au temps initial, ce qui revient à dire qu'une fois le retard initial retiré, les HRIR sont des réponses impulsionnelles de filtres à phase minimale. Bien que cette méthode semble être très proche à la fois de la définition de l'ITD et de la décomposition la plus commune des HRTF, elle souffre

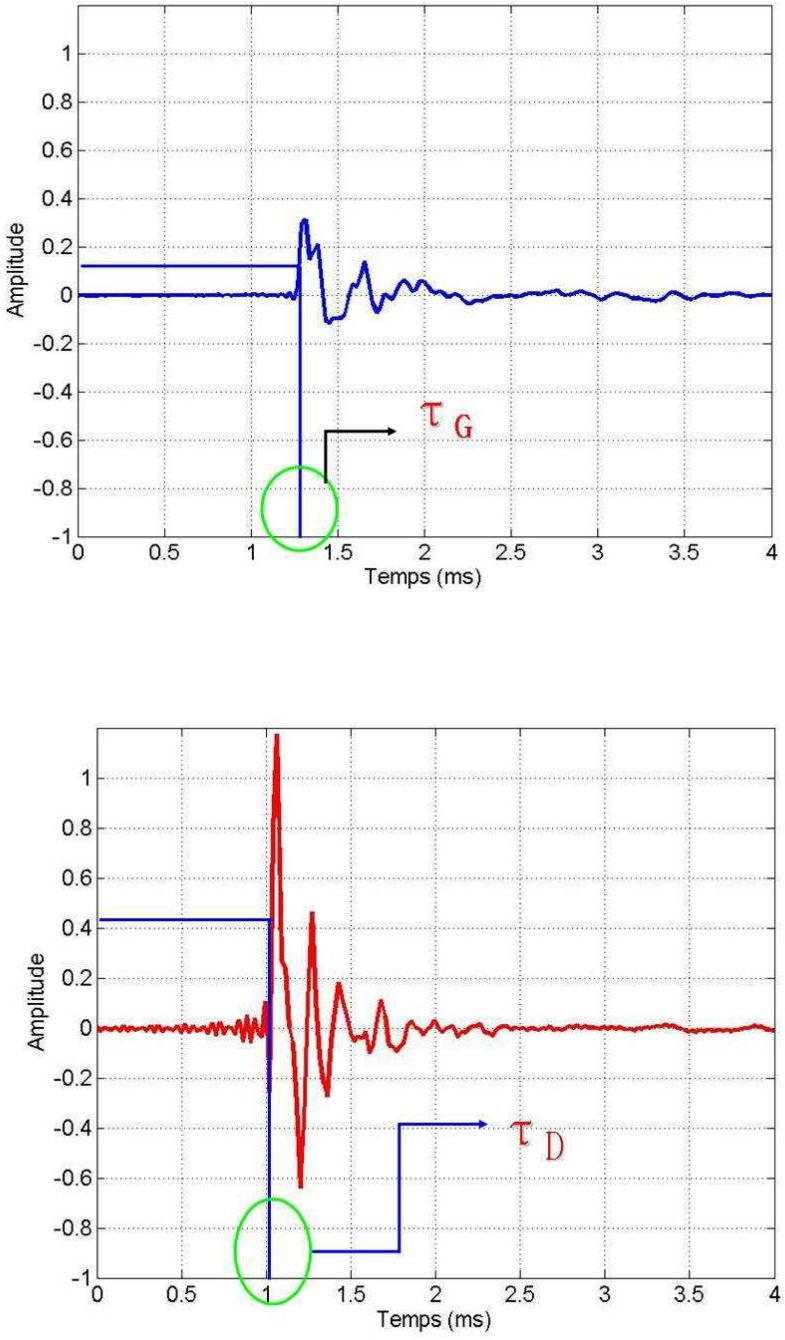


FIG. II.2 – Principe de l'estimation de l'ITD par le méthode de seuil : tracé des HRIR et du niveau de seuil. Figure du haut : estimation de τ_g, figure du bas : estimation de τ_d. L'ITD est alors la différence $\tau_d - \tau_g$.

premièrement de sa forte dépendance à la valeur du seuil choisie (cf. fig. II.3). Comme le rapport signal à bruit diminue pour les positions contralatérales, une valeur faible du seuil peut correspondre à une estimation au niveau du bruit (cf. fig. II.3 pour le seuil à 10 %). De plus, et toujours pour les positions contralatérales, les HRTF peuvent ne plus être bien décrites par un filtre à phase minimale. Ceci se traduit par l'apparition d'un pic secondaire de même niveau que le pic principal (cf. fig. II.3 pour le seuil à 90 % qui montre une sur-estimation de l'ITD aux positions proches de l'axe interaural). Enfin, certains auteurs utilisent non plus des HRIR mais l'énergie des HRIR ou encore des portions d'énergies des HRIR calculées sur certaines bandes de fréquences.

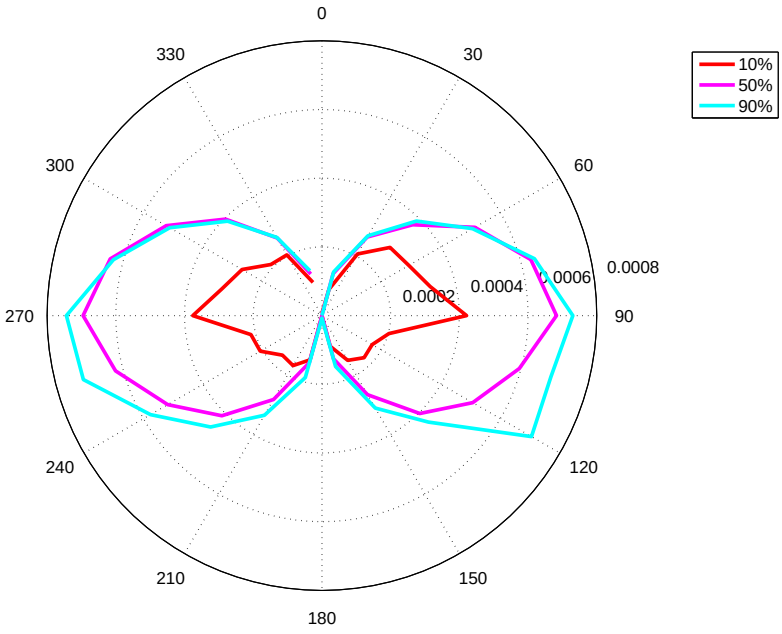


FIG. II.3 – Variation de la méthode d'estimation de ITD_{seuil} en fonction de la valeur du seuil sur le plan horizontal. Les ITD représentées sont moyennées sur toute la base LISTEN (cf. § III.1.3.1).

1.3.2 Corrélation droite-gauche

Initialement proposée par [Kistler and Wightman (1992)] la méthode MaxIACC (Maximum of the Interaural Cross Corrélation function) effectue le calcul de la fonction de corrélation interaurale, c'est-à-dire la corrélation entre le signal de l'oreille droite et le signal de l'oreille gauche. Cette méthode est basée sur l'hypothèse que le système auditif effectue la corrélation entre les signaux droit et gauche pour réaliser des tâches d'identification et de localisation de sources sonores. L'ITD est alors le temps pour lequel la fonction d'inter-corrélation atteint son maximum (cf. fig. II.2) :

$$ITD_{MaxIACC} = \max_{\tau} \frac{\int_{t_1}^{t_2} R(t).L(t-\tau)\mathrm{d}t}{\sqrt{\int_{t_1}^{t_2} R(t)^2\mathrm{d}t \int_{t_1}^{t_2} L(t)^2\mathrm{d}t}} \tag{II.6}$$

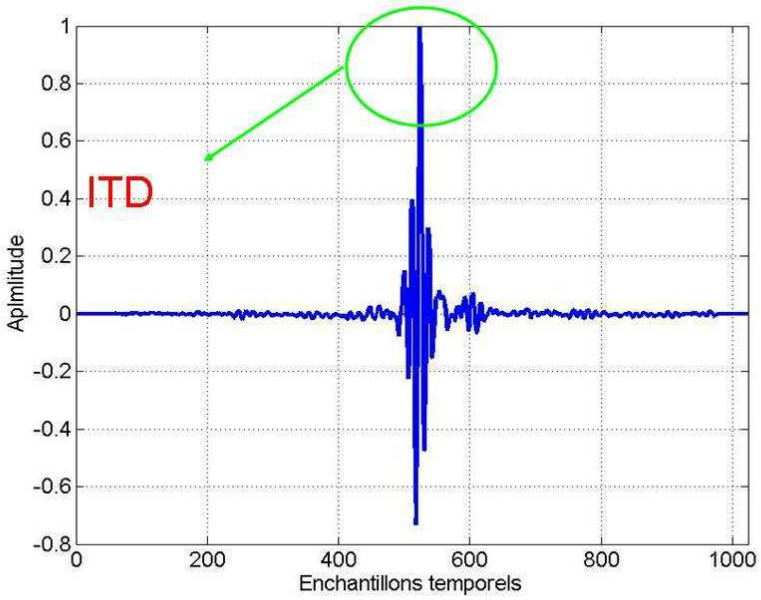


FIG. II.4 – Principe de l'estimation de l'ITD par la méthode de MaxIACC. Evolution de la fonction d'intercorrélation entre une HRIR droite et une HRIR gauche en fonction du retard. L'ITD est le valeur du retard où la fonction d'intercorrélation atteint son maximum.

avec $t_2 - t_1$ la fenêtre temporelle d'intégration. Afin d'améliorer la prédiction, les signaux R et L qui à l'origine désignent les HRIR droite et gauche, peuvent désigner l'enveloppe[2] des HRIR ou encore l'enveloppe de l'énergie des HRIR (cf. fig. II.5). Cette méthode est la plus robuste des méthodes d'estimation. Des modifications de cette technique ont considéré ce même calcul en différentes bandes fréquentielles. Les limites de l'estimation MaxIACC se situent au niveau des positions proches de l'axe interaural. Comme indiqué au paragraphe précédent, les HRIR contralatérales possèdent parfois deux pics principaux, alors que les HRIR ipsilatérales sont constitués d'un seul pic : le maximum de la fonction d'inter-corrélation peut alors être situé sur le deuxième pic de la HRIR contralatérale.

1.3.3 Hypothèse de linéarité de la phase

La troisième catégorie d'estimateurs de l'ITD utilise les informations de phase contenues dans la composante à excès de phase des HRTF. La phase de l'excès de phase étant quasi-linéaire : un retard pur associé à la pente de la phase peut être extrait. Une régression linéaire sur la phase de $HRTF_{excess}$ est effectuée sur un intervalle fréquentiel donné et le retard pur est la pente de la droite de régression. L'ITD est obtenue par la différence des pentes (cf. fig. II.6) :

[2]L'enveloppe d'un signal temporel $x(t)$ est donnée par le module de sa transformée de Hilbert : $Env_{x(t)}(t) = |\text{Hilbert}(x(t))|$

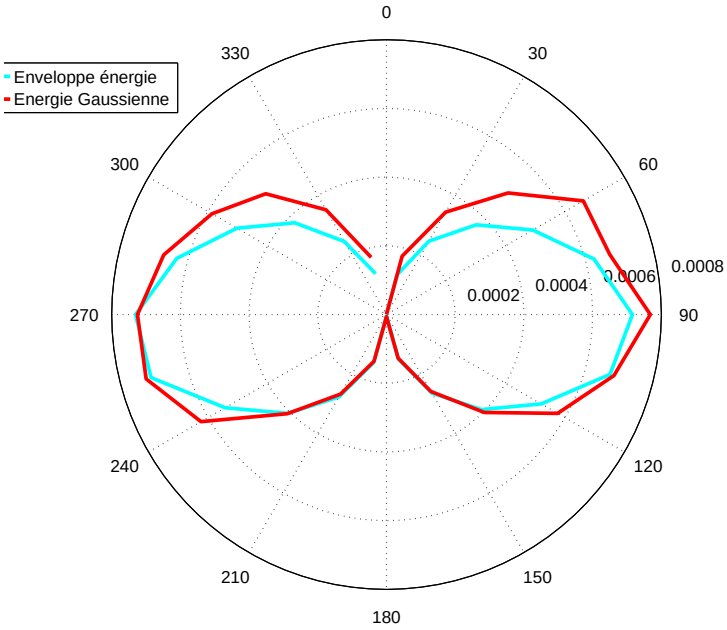


FIG. II.5 – Variation de la méthode du maximum de la fonction d'intercorrélation sur le plan horizontal pour un calcul sur l'enveloppe de l'énergie des HRIR (turquoise) et pour une modélisation gausienne de l'énergie (rouge). Les ITD représentées sont moyennées sur toute la base LISTEN.

$$ITD_{phase} = pente_{droite} - pente_{gauche} \tag{II.7}$$

Cette méthode d'estimation est à l'origine de la décomposition la plus commune des HRTF qui motive cette étude et qui fait l'hypothèse que les $HRTF_{excess}$ sont des filtres à phase linéaire. Cependant, si le caractère linéaire de la phase peut être considéré comme étant respecté à toutes les positions, c'est alors la composante à phase minimale, sur laquelle est reportée la phase résiduelle qui n'est plus négligeable pour les positions contralétarales, qui perd ses propriétés de filtre à phase minimale. Toutefois, il a été reporté que, malgré le caractère non-linéaire de la phase résiduelle, ce surplus de phase, qui est modélisé par un filtre passe-tout, peut être remplacé par un retard pur additionnel sans introduire d'artefacts audibles [Minaar et al. (2000)]. Enfin, l'ITD_{phase} est légèrement dépendante de la bande fréquentielle dans laquelle est effectuée la régression linéaire (cf. fig. II.7).

Une autre méthode d'estimation de l'ITD à partir des informations de phase a été proposée dans [Plogsties et al. (2000)]. L'ITD correspond au retard de groupe[3] interaural extrapolé à 0 HZ (cette technique est appelée IGD_0). Le retard de groupe de l'excès de phase est modélisé par des cellules d'ordre un et deux mises en cascades. Ces cellules ont toutes un comportement asympotique plat en basses fréquences, ce qui correspond à un retard pur. La valeur estimée du retard de groupe est alors le résultat de la sommation de tous ces comportements asympotiques des cellules, dont la valeur est lue à la fréquence 0 Hz. Dans [Minaar et al. (2000)], plusieurs méthodes sont décrites pour réaliser cette

[3]Le retard de groupe est la valeur opposée de la dérivée de la phase en fonction de la fréquence.

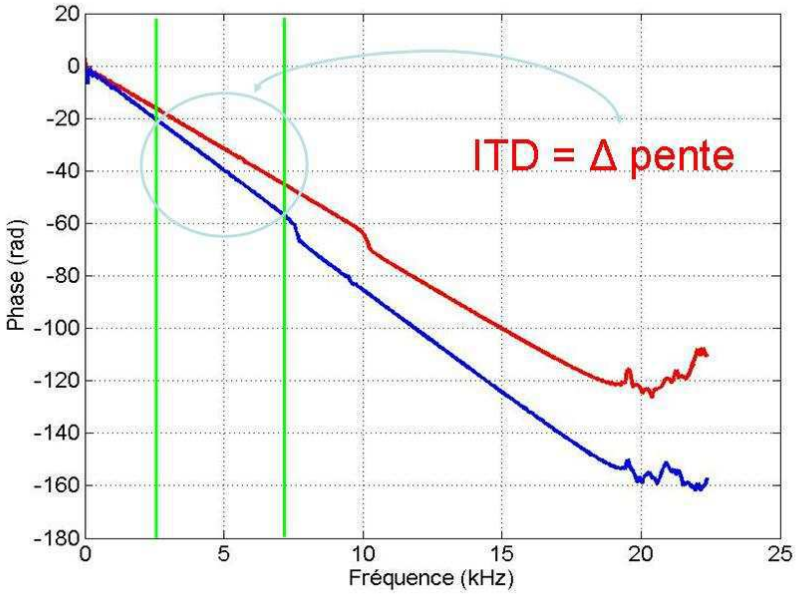


FIG. II.6 – Principe de l'estimation de l'ITD par la méthode de phase linéaire. Les deux barres verticales représentent la largeur de bande pour le calul de régression linaire. L'ITD est obtenue par la différence des coefficients directeurs droit et gauche.

estimation, faisant ainsi le lien avec toutes les autres méthodes existantes. Pour que cette méthode soit efficace, il faut forcer le module des HRTF à être égal à 1 à 0 Hz, ce qui peut poser des problèmes d'interpolation vu le peu de confiance qu'il faut accorder aux mesures des HRTF en basses fréquences. Cette méthode n'a pas été implémentée dans le reste de l'étude.

1.3.4 Conclusion sur les estimateurs de l'ITD

La description des différents types d'estimateurs illustre bien les difficultés rencontrées pour le choix d'une méthode pour le modèle $\{ITD \bigoplus HRTF_{min}\}$, tant les méthodes proposées diffèrent sur leur implémentation et sur leur fondement. Etant données les JND rapportées en écoute stéréophonique, de l'ordre de 10 μs pour des sources frontales, ainsi que la précision de localisation en azimut, de l'ordre du degré [Blauert (1983)], une erreur d'un seul échantillon temporel, qui à 44100 Hz vaut 22,8 μs, devrait conduire à des artefacts audibles. Bien plus qu'une simple erreur de localisation, c'est surtout le manque de cohérence entre ITD et composante à phase minimale qui risque de générer de plus grands défauts tels que la perception intra-crânienne. Une comparaison objective et subjective des estimateurs apparaît nécessaire. Avant de comparer les estimateurs, le travail présenté dans le paragraphe suivant permet de mieux appréhender les différentes dépendances de l'ITD.

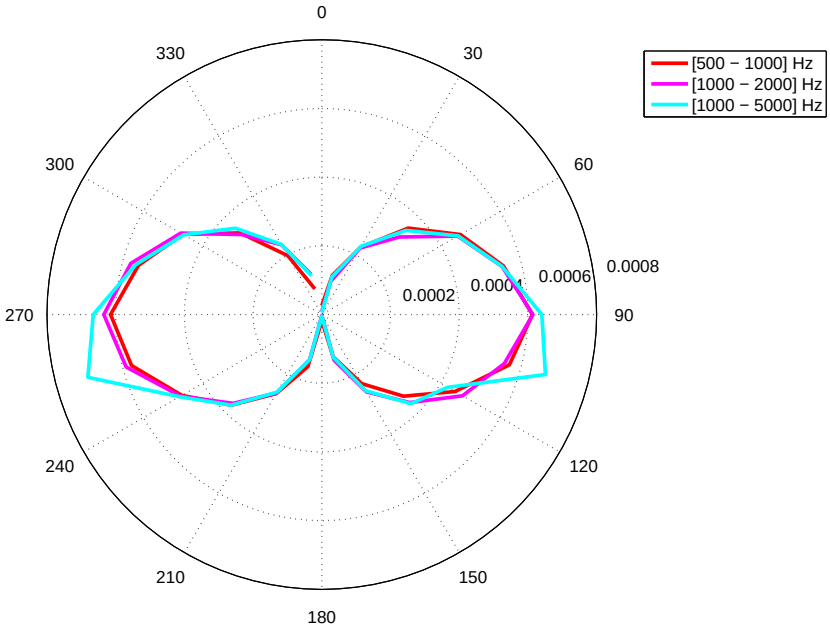


FIG. II.7 – Variation de la méthode de linéarité de la phase sur le plan horizontal en fonction de la bande de fréquence de régression. Les ITD représentées sont moyennées sur toute la base LISTEN.

1.4 Dépendances spatiales, fréquentielles et individuelles de l'ITD

L'ITD, premier indice de localisation, varie principalement avec l'azimut de la source sonore. L'ITD est contenue dans les HRTF à phase mixte. De ce fait elle varie aussi en élévation, d'une personne à une autre, avec la fréquence, et comme indiqué dans le paragraphe 1.3 avec la méthode d'estimation. Ce paragraphe montre ces variations en tentant d'en établir une hiérarchie.

1.4.1 L'ITD varie en azimut

La figure II.8 montre les variations d'une ITD moyenne en microseconde avec l'angle d'azimut en degré. L'ITD est calculée avec la méthode MaxIACC (cf. § 1.3.2) pour des positions décrivant le plan horizontal (plan d'élévation où les variations en azimut sont maximales) et moyennée sur tous les sujets de la base de HRTF LISTEN (cf. § 1.3.1). Avec ce calcul, l'ITD sur le plan horizontal est en moyenne de 6.1 μs. La valeur théorique est de 0 μs, ce qui correspond la valeur implémentée car 6.1 μs est inférieure au pas d'échantillonnage temporel (22,8 μs). L'ITD est globalement symétrique par rapport au plan médian et atteint ses extrema en $\theta = 90°$ et $\theta = 270°$. La figure II.8 fait apparaître une prédominance des valeurs l'ITD pour l'hémisphère arrière. Globalement, cette observation n'est pas vérifiée par un test de comparaison de moyenne[4]. Par contre, les positions

[4]Pour la comparaison des moyennes pour chaque hémisphère, un test de Kruskal-wallis est réalisé car la distribution n'est pas gaussienne. Pour la comparaison des positions symétriques, une ANOVA réalisée sur les positions. L'ANOVA indique un effet significatif de l'azimut.

105° et 255° ont des ITD supérieurs (en valeur absolue), aux positions symétriques de l'hémisphère avant, c'est-à-dire $\theta = 75°$ et $\theta = 285°$. Cette observation est confirmée par un test HSD (Honestly Significant Difference) de Tuckey qui montre que les différences observées sont significatives. Ces observations peuvent se justifier par la position relative des oreilles par rapport au plan frontal, qui se trouvent légèrement dans l'hémisphère arrière [Algazi et al. (2001a)]. Ceci montre la présence de différents morphotypes dans la base LISTEN. Si les oreilles sont diamétralement opposées, comme pour le modèle de tête sphérique de [Woodworth and Schloesberg (1962)], l'axe interaural est confondu avec un diamètre et les maxima de l'ITD sont atteints en ± 90°. Si les oreilles sont positionnées dans l'hémisphère arrière, les maxima sont situés, par exemple, en $\theta = 105°$ et $\theta = 255°$ et l'axe interaural est une corde, pour le modèle sphérique, et est déplacé par rapport au centre de la tête.

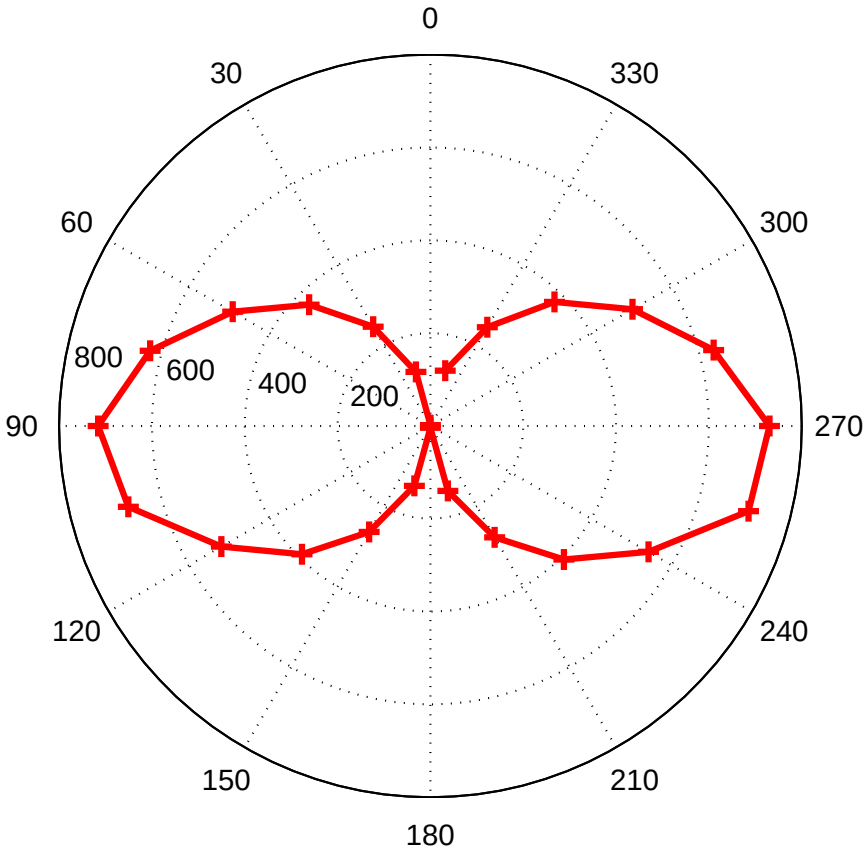


FIG. II.8 – $ITD_{MaxIACC}$ moyenne en valeur absolue pour la base LISTEN sur le plan horizontal. L'ITD est affichée en microseconde.

1.4.2 L'ITD varie en élévation

Comme le modèle de tête sphérique a longtemps été utilisé pour décrire la localisation auditive, les variations de l'ITD avec l'élévation, au sens des coordonnées polaires-interaurales, n'étaient pas prises en compte. En effet, le modèle sphérique, ainsi que tout

autre modélisation avec les oreilles diamétralement opposées, a ses cônes de confusions inclus dans des plans d'azimut constant (plans sagittaux). Pourtant le simple report de l'ITD sur des plans sagittaux fait apparaître des variations. Ces variations, dues à la non sphéricité de la tête et au décalage des oreilles par rapport en centre de la tête, peuvent atteindre 18 % de la valeur maximale de l'ITD, ce qui correspond à un décalage de la source de 15° en azimut [Duda et al. (1999)]. La figure II.9 représente l'ITD moyennée (en microseconde) en fonction de l'angle d'élévation en degré pour tous les sujets de la base CIPIC (cf. § III.1.3.2) pour trois plans sagittaux (25°, 45° et 65° d'azimut). L'ITD augmente avec l'élévation jusqu'à $\phi = 120°$ et diminue ensuite, ce qui peut être expliqué par la position des oreilles légèrement situées dans l'hémisphère inférieur [Duda et al. (1999); Algazi et al. (2001a)]. L'amplitude de variation dépend de l'azimut du cône et est maximale pour le cône à 45°.

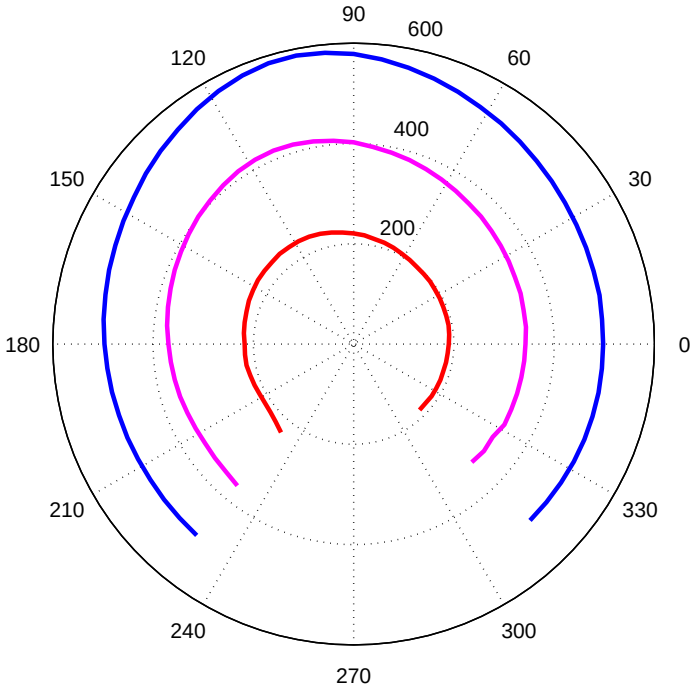


FIG. II.9 – ITD_{seuil} moyenne en valeur absolue de la base CIPIC sur des plans verticaux. Cone 65° en bleu, Cone 45° en rose et Cone 25° en rouge. L'ITD est affichée en microseconde.

1.4.3 L'ITD varie individuellement

Les variations en azimut et en élévation dépendent du sujet. Les variations individuelles sont directement liées à la morphologie du sujet et elles doivent être reproduites pour l'individualisation de la synthèse binaurale. Par exemple, les valeurs extrêmes augmentent avec le périmètre de la boîte crânienne ou la largeur de la tête [Middlebrooks

(1999)]. La position du canal auditif par rapport au centre de la tête joue aussi un rôle dans les variations de l'ITD [Duda et al. (1999); Algazi et al. (2001b)]. La figure II.10 décrit l'évolution de l'$ITD_{MaxIACC}$ en microseconde des 51 sujets de la base LISTEN en fonction de l'angle d'azimut pour le plan horizontal. La valeur maximale varie entre 650 μs et 900 μs et trois types de sujets se détachent :

- les sujets dont les extremum se situent en θ =90° et θ =270° (55 % des sujets),
- les sujets dont les extremum se situent dans l'hémisphère arrière en θ =105° et θ =255° (41 % des sujets).
- les sujets dont le maximum se situe en θ =285° (4 % des sujets).

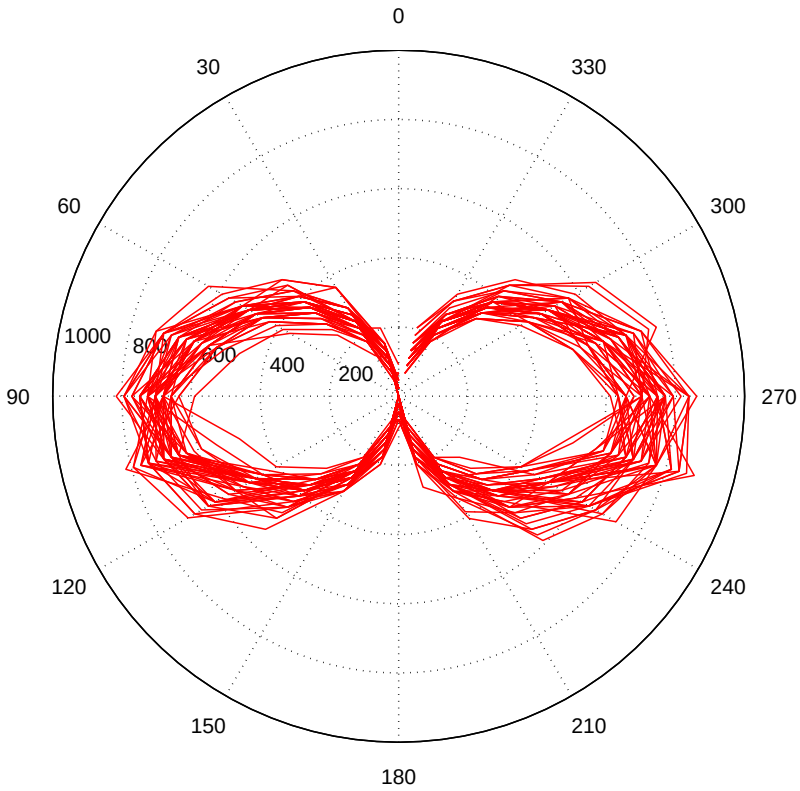


FIG. II.10 – Diagramme polaire de l'$ITD_{MaxIACC}$ en valeur absolue pour chaque sujet de la base LISTEN sur le plan horizontal. L'ITD est affichée en microseconde.

La figure II.11 décrit l'évolution de l'$ITD_{MaxIACC}$ en microseconde des 45 sujets de la base LISTEN en fonction de l'angle d'élévation en degré pour le plan vertical d'azimut 65°. Pour ce plan vertical, l'ITD varie entre 350 μs et 650 μs. L'ITD varie en valeur moyenne et la position du maximum se déplace d'une tête à l'autre.

1.4.4 L'ITD varie avec l'estimation

L'analyse du paragraphe 1.3 ayant montré les différences au sein même d'une famille d'estimateurs, il est intéressant de comparer globalement ces familles. Pour comparer

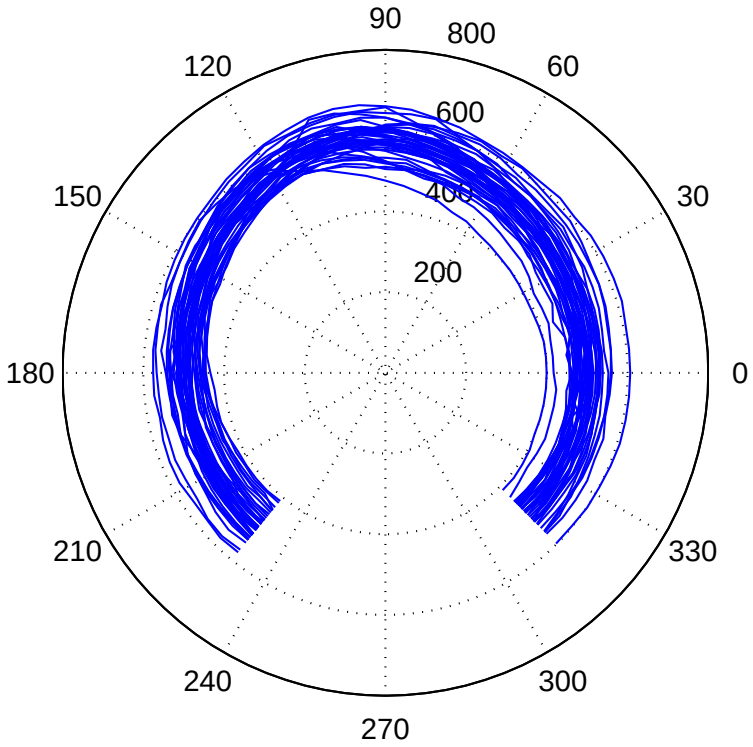


FIG. II.11 – Diagramme polaire de l'ITD_{seuil} en valeur absolue pour tous les sujets de la base CIPIC sur le plan vertical d'azimut 65°. L'ITD est affichée en microseconde.

ces méthodes sans trop introduire les variations individuelles, une ITD normalisée est calculée sur le plan horizontal de la manière suivante :

$$ITD_{norm}(\theta) = \frac{1}{Nb_{sujet}} \sum_i \frac{ITD_i(\theta)}{\max_\theta\{ITD_i(\theta)\}} \tag{II.8}$$

avec i indice du sujet. Cette ITD est calculée à partir de quatres méthodes différentes :

- ITD_{phase} = régression linéaire de la phase de l'excès de phase pour l'intervalle [1000-5000] Hz (vert),
- $ITD_{MaxIACC}$ = maximum de la fonction d'intercorrélation sur l'enveloppe des HRIR (rouge),
- ITD_{seuil} = seuil à 50 % du maximum de la HRIR (bleu),
- ITD_{sphere} = ITD du modèle de tête sphérique avec oreilles diamétralement opposées et $a = 87.5$ mm (noir),

La figure II.12 illustre les 4 ITD ainsi décrites. Les différences entre les méthodes sont bien visibles et apparaissent principalement pour les positions latérales. L'ITD_{phase} fait apparaître une plus grande variabilité inter-individuelle : le maximum éloigné de la valeur 1 signifie que le maximum se déplace d'un sujet à l'autre. De plus les autres

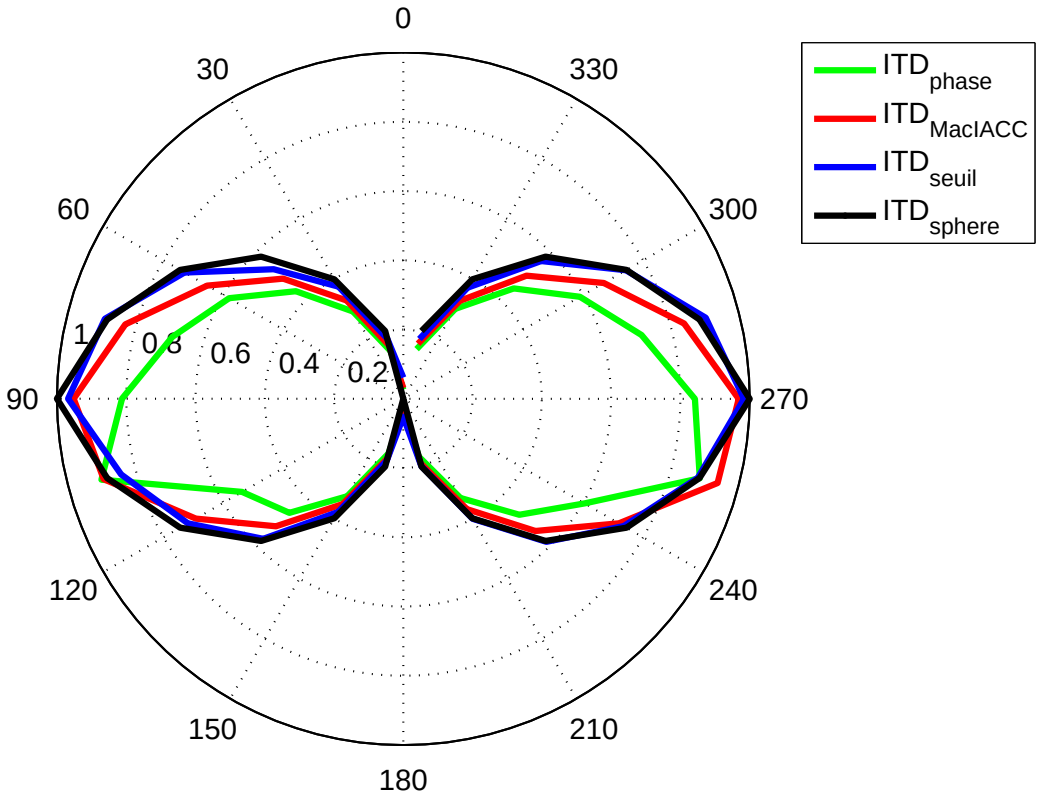


FIG. II.12 – ITD_{norm} moyenne en valeur absolue de la base LISTEN sur le plan horizontal pour trois estimations. Courbe bleue : ITD_{seuil}, Courbe rouge : $ITD_{MaxIACC}$, Courbe verte : ITD_{phase} et Courbe en pointillés rose ITD_{sphere}.

positions font apparaître de faibles valeurs de ITD_{norm} ce qui montre que le maximum de l'ITD est *sur-dimensionné* par rapport aux autres valeurs. Les autres méthodes ne font pas apparaître de tels comportements. L'ITD_{seuil} semble proche de ITD_{sphere} qui est parfaitement symétrique. La différence entre ITD_{seuil} et $ITD_{MaxIACC}$ reste assez faible.

1.4.5 L'ITD varie avec la fréquence

La définition mathématique de l'ITD en fonction de la fréquence est la différence du retard de groupe entre les HRIR droite et gauche [Algazi and Duda (2002)]. Dans [Kuhn (1977)], une étude expérimentale est menée sur des mesures effectuées sur un mannequin KEMAR et l'ITD est donnée par la différence de phase entre les deux microphones. Kuhn détermine l'existence de deux ITD asymptotiques : une ITD basse fréquence et une ITD haute fréquence, la transition s'effectuant entre 500 Hz et 3000 Hz. Un rapport de $\frac{1}{2}$ est alors observé entre l'ITD HF et l'ITD BF. L'ITD HF est bien reproduite par la formule de Woodworth (cf. équation II.1). Seulement cette observation est contradictoire avec l'hypothèse de linéarité de l'excès de phase, qui suppose un retard pour toutes les fréquences, et la zone de raccordement fréquentiel des deux ITD n'est pas bien définie.

Perceptivement, il semble difficile de considérer qu'une source large bande soit entendue à deux endroits différents. McFadden [McFadden (1981)] explique que le système auditif aurait besoin de léger retards entre ses fibres nerveuses, accordées sur différentes bandes fréquentielles, pour construire le même azimut entre les différentes bandes fréquentielles. Ainsi l'ITD objective varie avec la fréquence, et l'apprentissage fait que le traitement binaural de l'information regroupe les différentes ITD pour construire une seule position subjective. De plus, comme tous les travaux présentés ici sont réalisées dans le cadre de l'implémentation $\{ITD \bigoplus HRTF_{min}\}$ de la synthèse binaurale, la variation de l'ITD avec la fréquence n'est pas abordée et il est recherché une ITD adéquate pour toute la bande audible.

1.4.6 L'ITD varie avec la distance

Dans une étude comparative de différentes méthodes de calcul de l'ITD, Miller [Miller (2001)] montre que l'ITD varie très peu avec la distance de la source. Cette observation est cohérente avec [Brown and Duda (1998); Duda et al. (1999)] qui indiquent que la variation de l'ITD est plus importante avec la fréquence qu'avec la distance. Dans [Miller (2001)], l'auteur établit une formulation de l'ITD du modèle sphérique sur des considérations géométriques et introduit ainsi la dépendance avec la distance. Cette dépendance est importante surtout pour le champ proche mais semble artificielle comparée à des mesures effectuées à différentes distances sur un mannequin KEMAR. De plus, les mesures de HRTF sur lesquelles sont réalisés les travaux présentés ici, ont toutes été réalisées à distance fixe. La variation de l'ITD avec la distance n'est donc pas abordée ici.

1.4.7 Comparaison des dépendances

Le tableau II.1 présente l'amplitude de variation des différentes dépendances de l'ITD. Pour les variations en azimut et en élévation, une ITD moyennée sur les individus est considérée : pour l'azimut il s'agit de l'$ITD_{MaxIACC}$ sur le plan horizontal pour les sujets de la base LISTEN et pour l'élévation il s'agit de l'ITD donnée par la base CIPIC pour le plan d'azimut θ = -45°. Pour les variations selon la méthode, l'amplitude de variation correspond à l'ITD maximale par les trois méthodes de calculs (seuil, phase et MaxIACC) pour tous les sujets de la base LISTEN pour le plan horizontal. Pour la comparaison des variations selon les individus, l'intervalle reporté correspond à la variation entre les maxima d'ITD pour la méthode MaxIACC et pour la base LISTEN.

Tab. II.1 – Amplitude de variation des dépendances de l'ITD.

Dépendance	[Min - Max] (μs)
Azimut	[-716 ; 730]
Elevation	[333 ; 417]
Méthode	[748 ; 884]
Individu	[635 ; 884]

Une comparaison des dépendances de l'ITD ne saurait être exhaustive et demanderait de nombreux calculs et notamment le calcul d'ITD normalisée dont la nature du

terme normalisateur reste à définir. Le tableau II.1 permet tout de même une première évaluation de l'importance des variations de l'ITD. Ainsi, la variation de l'ITD est plus importante en azimut qu'en élévation. Selon les critères employés, la variation inter-individuelle est plus importante que les variations en élévation. De plus, les variations entre méthodes semblent plus faibles que les variations entre individus.

Le choix d'un estimateur de l'ITD tient en compte le fait qu'il reproduise correctement les différents variations spatiales et individuelles de l'ITD. Ce choix peut aussi être réalisé en considérant que l'erreur introduite par l'estimateur reste inférieure à l'acuité auditive de l'ITD. Cette acuité est caractérisée par la mesure du seuil d'audibilité d'un écart entre deux ITD (ITD calculée et ITD perçue). Il s'agit de la plus petite différence perçue entre deux ITD ou JND (Just Noticeable Difference) de l'ITD. Le paragraphe suivant présente les principales méthodes pour la mesure d'une JND et leur application à l'ITD.

1.5 Acuité auditive de l'ITD

L'ITD varie sur les plans sagittaux et les données présentes dans la littérature ne permettent pas de répondre, du moins dans l'état des connaissances rassemblées pour cette étude, à la question de la perception de ces variations. Le pouvoir de séparation angulaire de sources sonores a été établit, mais rien n'est connu sur l'acuité auditive de l'ITD sur les plans sagittaux. Des données psychoacoustiques sont manquantes : la connaissance de la JND de l'ITD pour des plans sagittaux permettrait de répondre. La démarche suivie (cf équation II.9) consiste à considérer qu'un écart est perçu si il est supérieur à la JND même si la définition d'une JND est liée à un pourcentage de détection qui peut varier entre 50% et 100 % selon les méthodes de mesure de la JND.

ITD_{calcul} est perceptivement correct si et seulement si

$$|ITD_{calul} - ITD_{ref}| < JND \quad \text{(II.9)}$$

où ITD_{ref} est une ITD de référence, par exemple l'ITD à $\phi = 0°$ par rapport aux autres élévations pour savoir si les variations de l'ITD en élévation sont perçues.

De nombreuses méthodes existent pour l'estimation d'un seuil de perception et elles ont été appliquées pour déterminer les JND de l'ITD mais, comme le montre la partie consacrée à l'état de l'art dans ce domaine, le cas de l'implémentation $\{ITD \bigoplus HRTF_{min}\}$ de la synthèse binaurale n'a pas été étudié. Le principe de base des méthodes d'estimation de seuil est de présenter des couples de stimuli, par exemple un stimulus avec une ITD de référence, ITD_{ref}, et un stimulus avec une ITD variable égale à $ITD_{var} = ITD_{ref} + \Delta ITD$. Le paramètre de l'expérience est alors ΔITD.

Cette section est divisée en deux parties. La première partie décrit l'état de l'art sur l'estimation de la JND de l'ITD. La deuxième partie expose les travaux expérimentaux qui conduisent au choix de deux méthodes. Ces deux protocoles sont ensuite utilisés dans la section consacrée à la mesure de la JND de l'ITD sur les cônes de confusion (cf. § 4).

1.5.1 Etat de l'art sur l'estimation de la JND de l'ITD

La JND de l'ITD ainsi que la résolution spatiale du système auditif humain, qui permet une interprétation en terme d'ITD, a été largement étudiée. Devant cette littérature abondante, il peut être utile de distinguer les études en **écoute champ libre** [Mills (1958);

Blauert (1983); Oldfield and Parker (1984); Perrot (1984); Perrot and Saberi (1990)], des **écoutes avec un casque stéréophonique** [Klump and Eady (1956); von Békésy (1960); Hershkowitz and Durlach (1969); Domnitz (1973); Hafter and Maio (1975); McFadden and Pasanen (1976); Domnitz and Colburn (1977); Zurek (1985); Tollin and B. (1998); Litovsky et al. (2000)] et des **écoutes en synthèse binaurale** où les stimuli sont filtrés par des HRTF [Ericson and McKinley (1989); Divenyi and Oliver (1989); Grantham et al. (2005); Best et al. (2004)]. Globalement, les expériences menées en champ libre et en écoute stéréophonique rapportent une JND similaire. La JND augmente avec l'angle d'azimut et atteint son minimum pour le plan médian. Certaines études ont même reporté l'impossibilité d'obtenir la JND pour les positions proches de l'axe interaural [Mills (1958)] ou quand l'écart angulaire entre deux stimuli n'est pas assez important [Divenyi and Oliver (1989)].

Klump et Eady [Klump and Eady (1956)] ont évalué la JND de l'ITD pour différents types de stimuli et pour différentes ITD de base, c'est-à-dire pour différentes positions dans le plan horizontal, en écoute stéréphonique. Les sujets écoutent deux sons consécutifs et il leur est demandé d'identifier lequel est perçu le plus à gauche. La JND est estimée à 75 % de la courbe psychométrique. Cette étude donne les variations de JND pour des bruits de différentes bandes de fréquences, pour des clicks uniques ou en série, pour des sons purs de fréquences comprises entre 90 Hz et 3200 Hz et ce pour différents ITD_{ref}. La JND la plus faible est obtenue pour $ITD_{ref} = 0$ μs (position frontale) avec un bruit dans une bande [100 ; 1700] Hz et vaut 9 μs. Dans une étude sur la discrimination angulaire (MAA en anglais pour Minimum Audible Angle) en champ libre[5], [Mills (1958)] demande au sujet de juger la position relative de deux stimuli consécutifs. La JND est estimée à 50 % de la courbe psychométrique par une interpolation linéaire entre les points représentant 25 % et 75 % . Bien que l'écoute champ libre contienne une ILD qui donne au sujet des indices de localisation supplémentaires, un seuil de 10 μs est trouvé pour un son pur à 750 Hz. Dans [Domnitz and Colburn (1977)], l'étude est consacrée à l'effet conjoint de l'ITD et l'ILD sur la JND de l'ITD. Le protocole utilisé est une méthode adaptative à choix forcé avec retour visuel sur la validité de la réponse. Le stimuli est un son pur à 500 Hz d'une durée de 300 ms et la JND obtenue la plus faible est de 10 μs. Cette JND est confirmée par Zurek [Zurek (1985)] pour des stimuli large bande et par Henning [Henning (1974)] pour un son harmonique modulé en amplitude. Cette valeur de 10 μs est aussi en accord avec Hafter et Maio [Hafter and Maio (1975)] qui utilisent des clicks de 20 μs de fréquence centrale variant entre 100 Hz et 2000 Hz et avec une procédure adaptative à choix forcé avec retour visuel des réponses. Même avec une synthèse binaurale non-individuelle, une JND de 16 μs est estimé [Grantham et al. (2003)]. La seule étude connue au moment de la rédaction se rapportant à la résolution de l'écoute en synthèse binaurale individuelle est [Best et al. (2004)]. Best *et al.* donne une JND de 50 μs pour des stimuli simultanés. Cette valeur est nettement supérieurs aux autres données des études précédentes. Une cause possible de cette différence est expliquée par le fait qu'il est plus difficile d'extraire de l'information de provenant de sons simultanés que d'une présentation consécutive. Les principales données de la littérature sont reportées dans le tableau II.2.

L'étude présentée ici a pour but d'estimer la JND de l'ITD associé à une implémentation $\{ITD \bigoplus HRTF_{min}\}$ de la synthèse binaurale, pour des positions variant en azimut et en élévation. La recherche bibliographique présentée dans le paragraphe précédent ne

[5]La relation entre MAA et JND est basée sur des mesures de phase réalisées sur une tête artificielle

TAB. II.2 – JND des études antérieurs pour $ITD_{ref} = 0\mu$s.

Référence	Condition d'écoute	JND (μs)
[Klump and Eady (1956)]	stéréphonique	9
[Mills (1958)]	champ libre	10
[Domnitz and Colburn (1977)]	stéréphonique	10
[Grantham et al. (2003)]	binaurale	16
[Best et al. (2004)]	binaurale individualisée	50

permet pas de dégager le protocole le mieux adapté à cette estimation. Le paragraphe suivant expose des tests d'écoute permettant de mieux appréhender les différences entre les protocoles.

1.5.2 Comparaison de différents protocoles : présentation

De nombreux protocoles expérimentaux existent pour la mesure de seuils de perception. En plus du protocole, plusieurs types de présentations des stimuli et plusieurs types de tâches existent et influent sur les valeurs des seuils, ou sur les moyens de les obtenir. Un seuil de perception est défini par rapport à la courbe psychométrique du paramètre testé et correspond à un niveau du paramètre qui donne X % de *bonnes réponses* (par exemple seuil à 75%). Une bonne réponse dépend du type de présentation des stimuli associé à la question posée au sujet. Il peut être distingué principalement deux sortes de protocoles pour la construction de la courbe psychométrique ou pour l'estimation directe d'une valeur de seuil.

Méthodes des constantes Dans ce genre d'expérience, l'expérimentateur doit créer à l'avance les groupes de stimuli correspondant à différents niveaux du paramètre testé. Il doit donc s'assurer que les valeurs testées couvrent une plage de paramètre assez large, par exemple pour couvrir la courbe psychométrique de 10 % à 90 % de bonnes réponses. Les différents groupes de stimuli sont présentés dans un ordre aléatoire. L'estimation du seuil se fait sur le tracé de la courbe psychométrique ou sur une fonction représentative des résultats [Levitt (1970)]. L'avantage de cette méthode est qu'elle couvre un large panel de valeurs du paramètre et que le nombre de présentations de chaque valeur de paramètre est le même. L'inconvénient majeur est qu'il faut connaître à l'avance un ordre de grandeur du seuil à estimer, sinon un nombre important de stimuli doit être présentés. De plus, si un seul seuil est désiré, le temps de test est beaucoup plus long que pour les méthodes adaptatives.

Méthodes adaptatives Ce sont des méthodes qui déterminent le niveau du paramètre en fonction du niveau précédent et de la réponse du sujet. Plusieurs protocoles adoptent ce principe de fonctionnement et de nombreuses familles de méthode adaptatives existent. La méthode Von Békésy, pour la détermination rapide des seuils d'audition, ou courbes isosoniques, est une méthode adaptative. La méthode adaptative consiste en un *parcours* de réponses. Un parcours type est représenté en figure II.13. Une série de réponses dans

un sens est appelé un *run*. Un changement de direction est nommé un *retournement*. Si le sujet donne un bonne réponse le niveau du paramètre baisse. En revanche, il augmente si la réponse est mauvaise. A chaque retournement, le pas est réduit et la convergence est assurée par un nombre de retournements fixé.

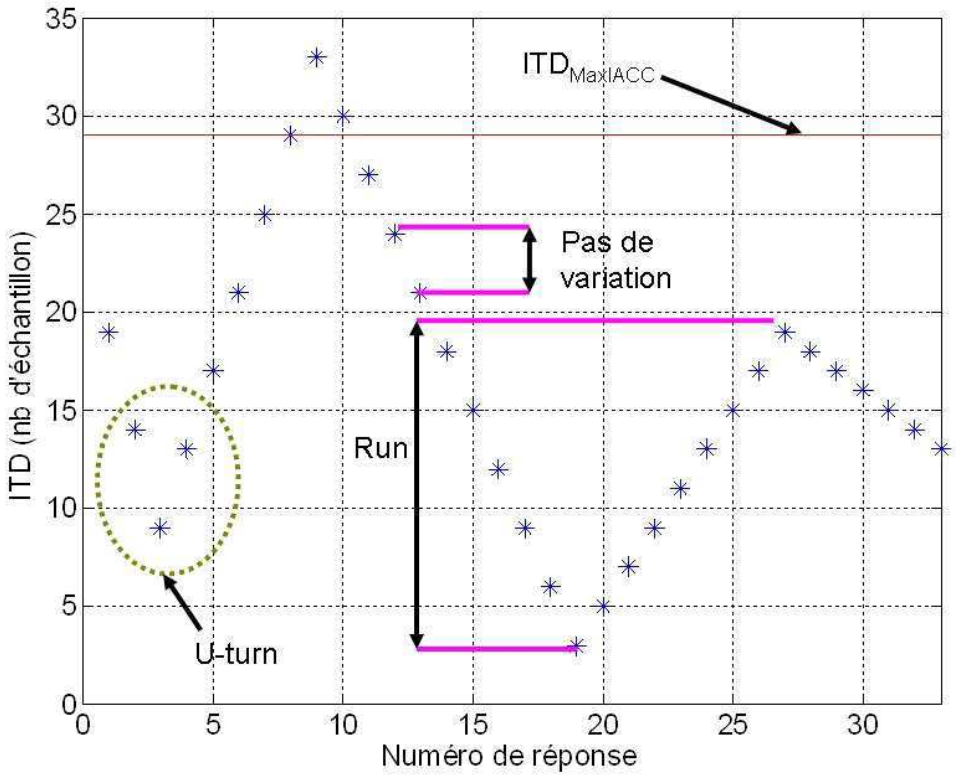


Fig. II.13 – Représentation des ITD en nombre d'échantillons au cours d'un essai. Les valeurs d'ITD modifiées par le sujet sont représentées à l'aide d'étoiles bleues. Le pas de variation est réduit après chaque retournement.

Ces méthodes ont l'avantage particulier de s'adapter aux réponses du sujet et donc peu de choses sont à connaître a priori. De plus, une estimation rapide du seuil est possible, sans avoir à tracer la courbe psychométrique. La méthode retenue pour l'estimation rapide du seuil, consiste a calculer la moyenne du deuxième *run*, ou *mid-run estimate*. Cette estimation du seuil a empiriquement donné de meilleur résultats que des méthodes complexes d'analyse de données. Une vérification de l'influence de la méthode d'estimation réalisée sur les données expérimentales exposées plus loin dans le document a montré la faible variation inter-méthodes (cf. fig. II.14). Par contre une attention particulière doit être portée sur la valeur des pas de variations : si ils sont trop grands l'estimation du le seuil risque d'être peu fiable, si il est trop petit de nombreuses observations risquent d'être inutiles. Dans [Levitt (1970)], des procédures adaptatives *modifiées* sont proposées. Ces méthodes proposent différentes stratégies de convergence. Par exemple, la méthode *2down-1up* fait augmenter le seuil, réponse *haute*, pour une mauvaise réponse et le fait

baisser, réponse *basse*, pour une série de deux bonnes réponses consécutives. Les méthodes adaptatives modifiées permettent ainsi d'estimer un seuil au niveau duquel la probabilité de réponse *haute* est égale à la probabilité de réponses *basses*. Par exemple, dans le cas d'une méthode *1down-1up*, le seuil est estimé à 50 % de la courbe psychométrique et dans le cas d'une méthode *2down-1up*, le seuil est estimé à 70 % de la courbe psychométrique.

Méthodes de présentation des stimuli Le type de présentation des stimuli peut avoir une influence sur la valeur des seuils de perception et une combinaison adéquate de procédure psychophysique et de présentation des stimuli doit être trouvée [Kollmeier et al. (1988)]. Les tâches de discrimination de sources en écoute binaurale sont particulièrement difficiles et il faut s'assurer que la tâche demandée au sujet maximise ses chances de réussites. Quatre procédures différentes, trois constantes et une adaptative, sont testées pour la détermination de la JND de l'ITD :

AB : Le sujet entend deux stimuli qui diffèrent seulement par leur ITD et il lui est demandé d'indiquer celui qui lui paraît le plus à gauche. La procédure employée est la méthode des constantes. Ce type de procédure est aussi nommée *2-intervalles 2-ACF*.

ABX : Le sujet entend trois stimuli et le dernier est une répétition du premier ou du deuxième. Il est demandé d'indiquer si X=A ou X=B. La procédure employée est la méthode des constantes. Ce type de procédure est aussi nommée *3-intervalles 2-ACF*.

ABC1 : Le sujet entend trois stimuli dont deux sont identiques. Il est demandé d'indiquer celui qui est différent des deux autres. La procédure employée est la méthodes des constantes. Ce type de procédure est aussi nommée *3-intervalles 3-ACF*.

ABC2 : Même procédure qu'avec ABC1 mais en employant une méthode adaptative modifiée de type *2down-1up*.

Ces méthodes sont réalisées avec retour visuel des réponses. Les combinaisons des stimuli sont présentées de manière aléatoire, sauf pour ABC2. Les stimuli sont des bruits blancs gaussiens de durée égale à 400 ms avec une attaque et une fin en $\cos^2$ de 10 ms chacune. La durée inter-stimuli est de 500 ms. Les stimuli sont échantilonnés à 96 kHz, ce qui donne un ΔITD_{min} de 10.4 μs. Les stimuli sont justes retardés par un ΔITD et aucun filtrage par des HRTF n'est réalisé : la perception est intra-crânienne. Les résultats attendus doivent être alors proches des valeurs issues de la littérature, c'est-à-dire 10 μs pour $ITD_{ref} = 0$ μs.

1.5.3 Comparaison de différents protocoles : étude expérimentale

1.5.4 Test ABC2

Procédure Un protocole adaptatif modifié est réalisé. Les paramètres du protocole sont les suivants :

- règles de parcours : 2down-1up
- convergence : 6 retournements ou 2 bonnes réponses pour $\Delta ITD = 1$ échantillon
- ΔITD de départ : choix aléatoire dans l'intervalle [1 - 5] échantillons
- variation du pas : [4 3 2 1 1 1] échantillon

28 sujets naïfs ont participé à cette expérience (11 femmes et 17 hommes). Les seuils sont mesurés autour de 4 ITD de référence : $ITD_{ref1} = 0$ μs (position frontale, $\theta = 0°$ et $\phi = 0°$), $ITD_{ref2} = 250$ μs ($\theta = 30°$ et $\phi = 0°$), $ITD_{ref3} = 500$ μs ($\theta = 60°$ et $\phi = 0°$) et $ITD_{ref4} = 750$ μs (oreille droite, $\theta = 90°$ et $\phi = 0°$)[6]. Chaque ITD_{ref} est répétée 5 fois. Un apprentissage de 10 min était donné aux sujets avant de passer le test.

Résultats Les résultats présentés ne tiennent pas compte des réponses de 5 sujets présentant des seuils pour ITD_{ref1} supérieurs à 120 μs et 3 sujets dont les réponses s'écartent de deux fois l'écart-type par rapport à la moyenne des réponses. Plusieurs méthodes d'estimation du seuil existent et évitent de devoir balayer une large plage de valeurs de paramètres pour la construction de la courbe psychométrique. Quatres méthodes d'estimations sont comparées :

- **mid-run** estime le seuil en calculant la moyenne du run compris entre le 1er et le 2^{ieme} retournement,
- **quatre dernières** calcul une moyenne sur les 4 derniers run,
- **meilleure dernière** donne la valeur la plus faible de la fin du parcours,
- **moyenne globale** évalue une moyenne sur toutes les valeurs du parcours.

Les valeurs des seuils pour différentes méthodes d'estimation de seuil sont reportées en figure II.14. Comme attendu [Klump and Eady (1956)], le seuil augmente avec ITD_{ref} et cette tendance est suivie par toutes les méthodes d'estimation. Les écarts entre les méthodes testent faibles, sauf pour ITD_{ref4}. Comme *mid-run* présente les valeurs les plus faibles, c'est cette technique qui est retenue dans la suite.

La figure II.15 représente la dispersion de la JND en fonction de ITD_{ref} pour la méthode d'estimation *mid-run*. Elle fait apparaître une augmentation du seuil moyen. La dispersion des réponses est quasiment constante sauf pour ITD_{ref4} qui présente une dispersion largement supérieure. Cette dispersion maximum pour ITD_{ref4} est cohérente avec une augmentation de la difficulté de la tâche pour des positions perçues proches d'une oreille [Mills (1958)]. Ces observations sont confirmées par les valeurs inscrites dans le tableau 1.5.4.

Tab. II.3 – Moyennes et écart-types des seuils mesurés pour les différents ITD_{ref}. Valeurs calculées sur la base de cent estimations du seuil.

ITD_{ref} (μs)	0	250	500	750
Moyenne (μs)	69.9	86.4	100.7	151.1
Ecart-type (μs)	23.6	26.0	30.3	68.4

Les valeurs moyennes du tableau 1.5.4 donnent des seuils largement supérieurs à ceux des études précédentes. Plusieurs points doivent être pris en considération pour comparer ces résultats à ceux de la littérature. Premièrement, les sujets n'ont reçu qu'un entraînement rapide d'une dizaine de minutes. Les études sur les JND de l'ITD font souvent mention de plusieurs heures d'entraînement, voir une centaine. Deuxièmement, la méthode utilisée est à choix forcé de 1 parmi 3 stimuli. La plupart des études précédentes

[6]Les positions d'azimut indiquées sont données à titre indicatif sur la base de la formule de Woodworth avec un rayon de tête égale à 875 mm et dépendent fortement de l'individu.

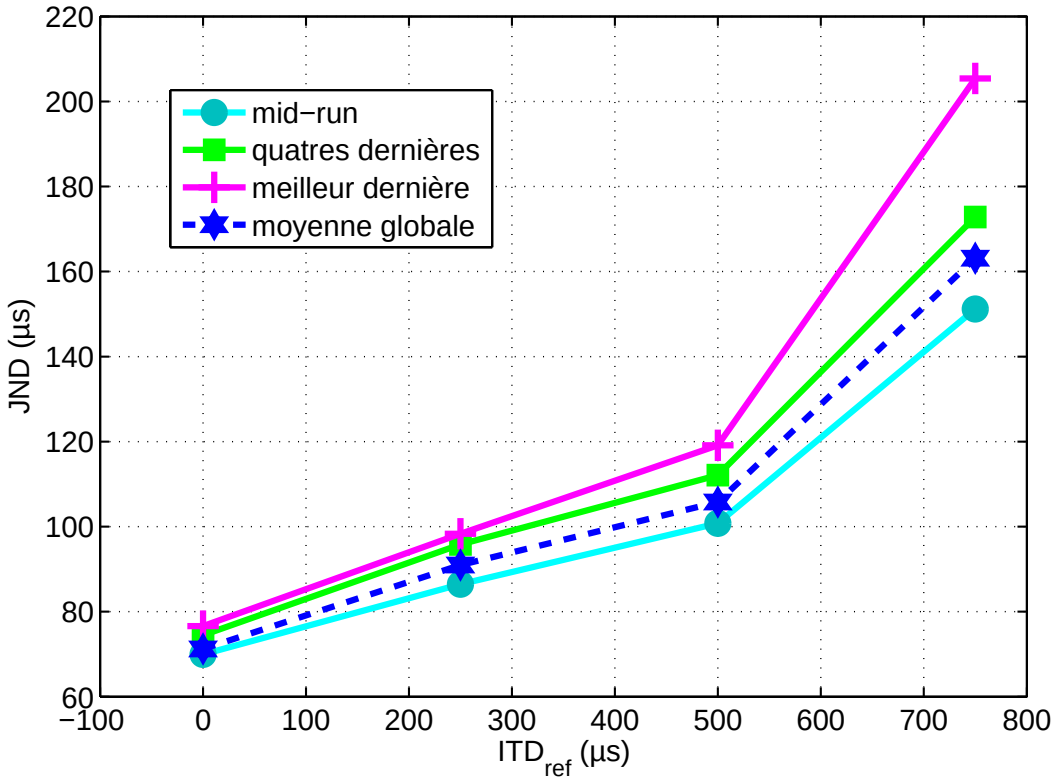


FIG. II.14 – JND en en fonction de l'ITD de base et pour plusieurs estimateurs du seuil. *mid-run* estime le seuil en calculant la moyenne du run compris entre le 1er et le 2^{ieme} retournement, *quatres dernières* calcul une moyenne sur les 4 derniers run, *meilleure dernière* donne la valeur la plus faible de la fin du parcours et *moyenne globale* évalue une moyenne sur toutes les valeurs du parcours.

ont eu recours à des méthodes à choix forcé de 1 parmi 2 stimuli. Par exemple, dans [Klump and Eady (1956)] les sujets doivent juger lequel des deux sons présentés est situé le plus à gauche. Ce type de jugement peut paraître plus facile à réaliser, vu qu'il ne fait appel ni aux performances de localisation, ni à la mémoire, le sujet jugeant d'un mouvement de la source. Les avantages de ces méthodes sont décrits dans [Yost et al. (1974)].

Les méthodes adaptatives sont très utilisées dans les expériences d'écoute et notamment en écoute binaurale car elles présentent de nombreux avantages par rapport aux autres techniques tels qu'une meilleure efficacité, une plus grande flexibilité et une moins grande dépendance aux hypothèses restrictives. Ainsi Perrot et Saberi [Perrot and Saberi (1990)] utilisent une méthode adaptative *1 intervalle 2 ACF* 3down-1up et un retour visuel des réponses (le sujet indique s'il entend le deuxième son à gauche ou à droite du premier son) pour mesurer des MAA en écoute en champ libre pour des positions qui varient en élévation et en azimut. Tolling *et al.* [Tollin and B. (1998)] se servent d'une

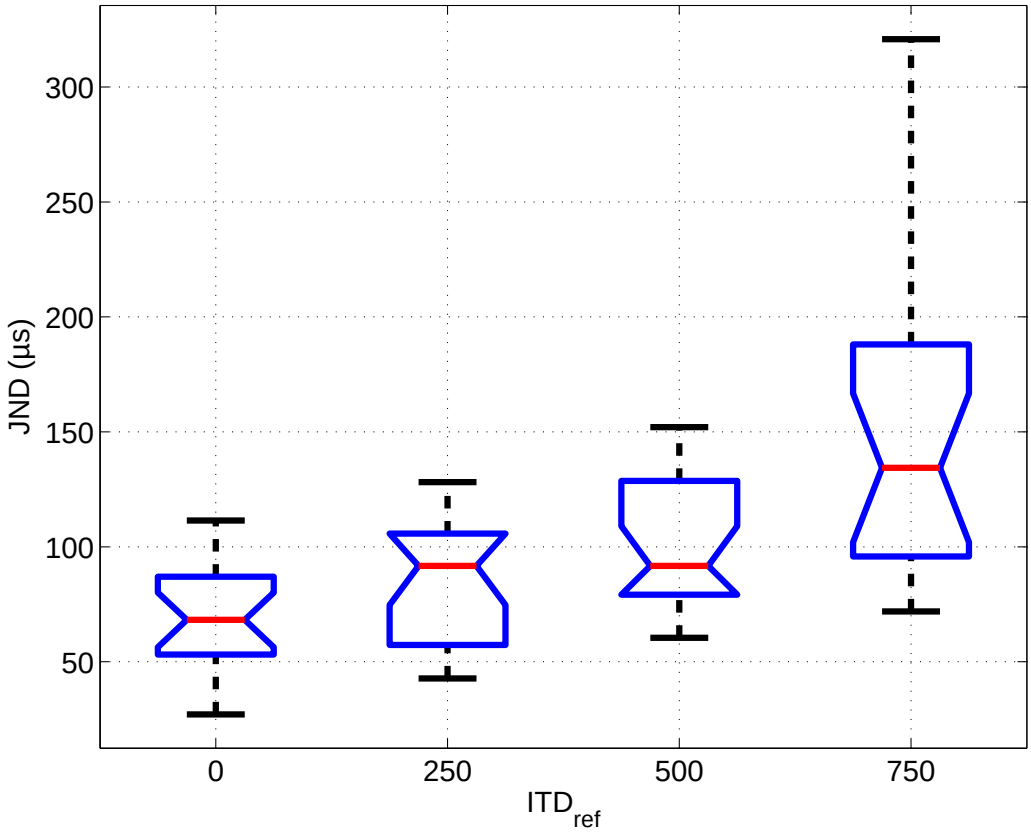


FIG. II.15 – JND en fonction de ITD_{ref}. Les parties inférieures et supérieures des boîtes bleues sont les 1^{er} et 3^{ime} quartiles. La ligne rouge au milieu de la boîte représente la valeur médiane. Les valeurs indiquées par les barres horizontales de part et d'autre de la boîte représentent une mesure de dispersion donnée par 150 % de la distance inter-quartiles.

méthode adaptive *2 intervalles 2 ACF* avec retour visuel des réponses, 3 down-1up pour étudier l'influence de la durée de l'intervalle entre deux clicks sur l'ITD. Trahiotis *et al.* [Trahiotis et al. (1990)] montrent même que ces méthodes permettent d'atteindre avec des sujets naïfs et dès la première session (pas d'effet d'apprentissage observé) des niveaux de seuils établis par des experts avec des méthodes à paramètres fixés. Cependant pour notre expérience, les résultats de ce premier test font apparaître des seuils de perceptions largement supérieurs à ceux rencontrés dans la littérature (10 μs contre 70 μs).

1.5.5 Tests AB, ABX et ABC1

Protocole Trois autres tests ont été réalisés pour évaluer les différences avec trois autres méthodes constantes. Quatre sujets ont participé à ce test comparatif (3 hommes et une femme) dont deux naïfs et deux experts. Le paramètre était l'ITD variant entre 0 μs et 52 μs par pas de 10.4 μs (valeur d'un échantillon à Fe = 96000 Hz). Chaque valeur d'ITD était répétée 50 fois. Le niveau $ITD_{ref} = 0$ μs était présenté à chaque essai. Aucun

entraînement spécifique n'a été délivré aux sujets avant d'aborder ce test.

Résultats Un écart ΔITD est désigné comme *perçu* si le pourcentage de réponses correctes de la condition $ITD_{ref} + \Delta ITD$ est significativement différent du pourcentage de réponses correctes de la condition ITD_{ref} (hypothèse nulle[7] rejetée à 95 %). La figures II.16, II.17 et II.18 présentent les pourcentages de réponses correctes en fonction de l'ITD présentée.

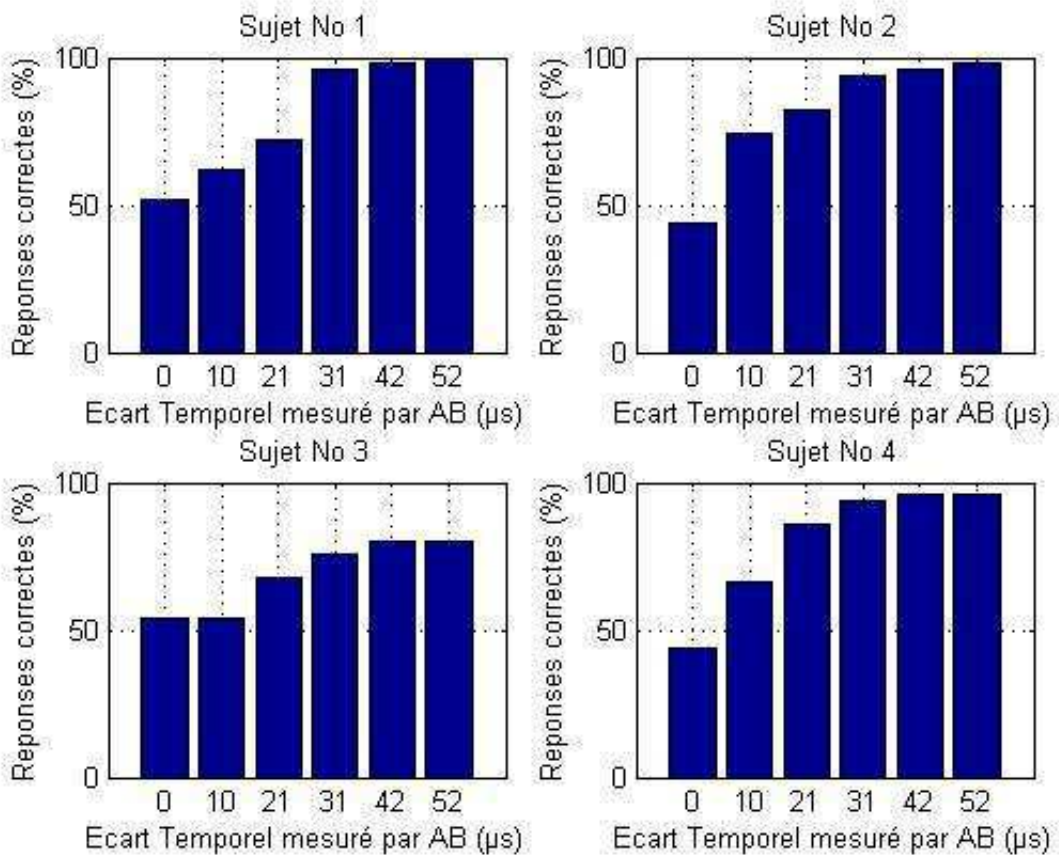


FIG. II.16 – Pourcentage de réponses correctes en fonction de ΔITD pour la procédure AB.

Les figures II.16, II.17 et II.18 montrent clairement des différences entre les seuils mesurés. Les seuils d'ITD perçus les plus bas sont obtenus par la méthode AB et les seuils les plus élevés par la méthode ABX. Une augmentation du pourcentage de réponses correctes avec l'ITD est observée globalement pour toutes les méthodes. La méthode ABX semble être la moins bien réalisée par les sujets : il n'y a pas d'évolution franche du pourcentage de réponses correctes et aucun ΔITD n'a pu être considéré comme perçu pour le sujet 3 avec cette méthode. D'après des entrevues informelles à la suite du test, il est apparu nettement que la tache ABX a été la plus dure à réaliser. Le tableau II.4

[7]Hypothèse nulle : la différence observée entre les pourcentages est due au hasard de l'échantillonnage.

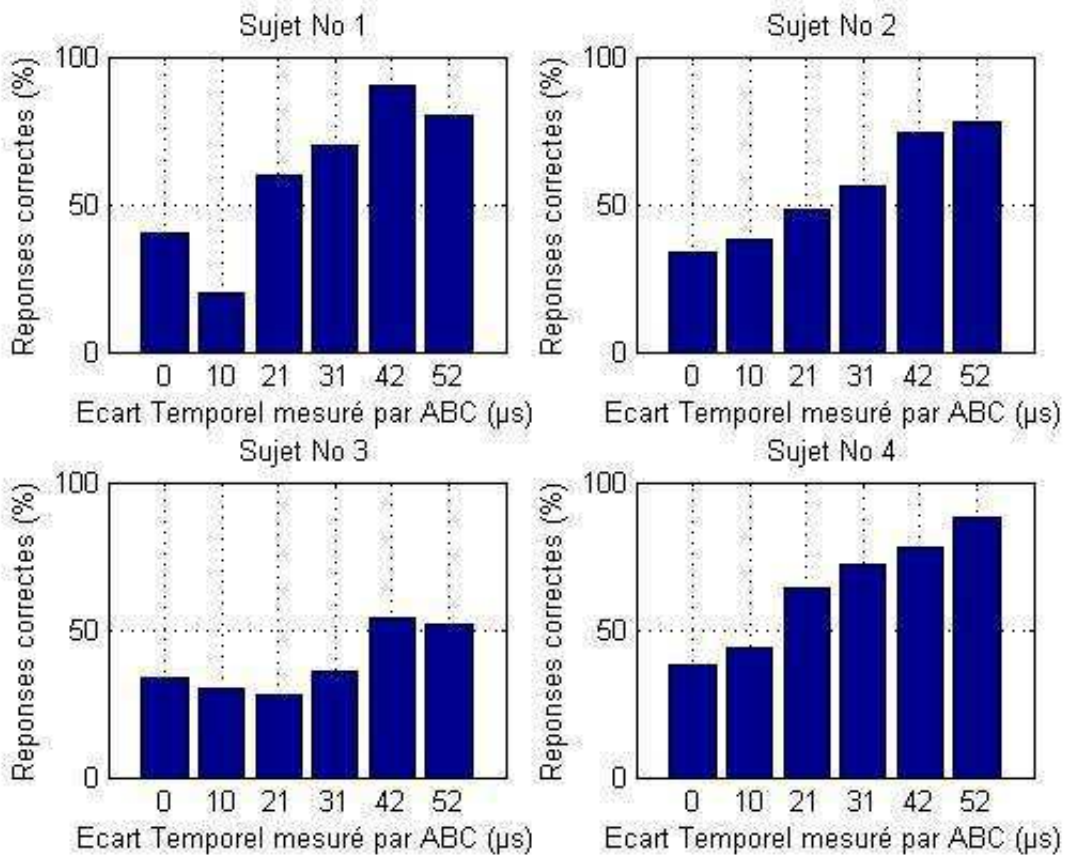


FIG. II.17 – Pourcentage de réponses correctes en fonction de ΔITD pour la procédure ABC.

donne les JND en fonction de la méthode utilisée. La dernière colonne reporte le seuil obtenu avec le test ABC2 pour la condition $ITD_{ref} = 0$ μs.

TAB. II.4 – Comparaison des JND en fonction des protocoles utilisés. Une croix indique qu'aucun ΔITD testé n'a pu être considéré comme perçu. Les JND sont données en μs.

Méthode / Sujet	ABX	AB	ABC1	ABC2
Sujet 1	31	21	21	44
Sujet 2	21	10	31	71
Sujet 3	X	31	42	117
Sujet 4	31	10	21	54

Le tableau II.4 fait clairement apparaître des différences sur les seuils de perception en fonction de la méthode utilisée. Seule la méthode AB procure des JND similaires à ceux de la littérature. La méthode ABC2 pour la condition $ITD_{ref} = 0$ μs donne des seuils plus de deux fois supérieurs à ceux obtenus par la méthode ABC1. Il semble donc que le

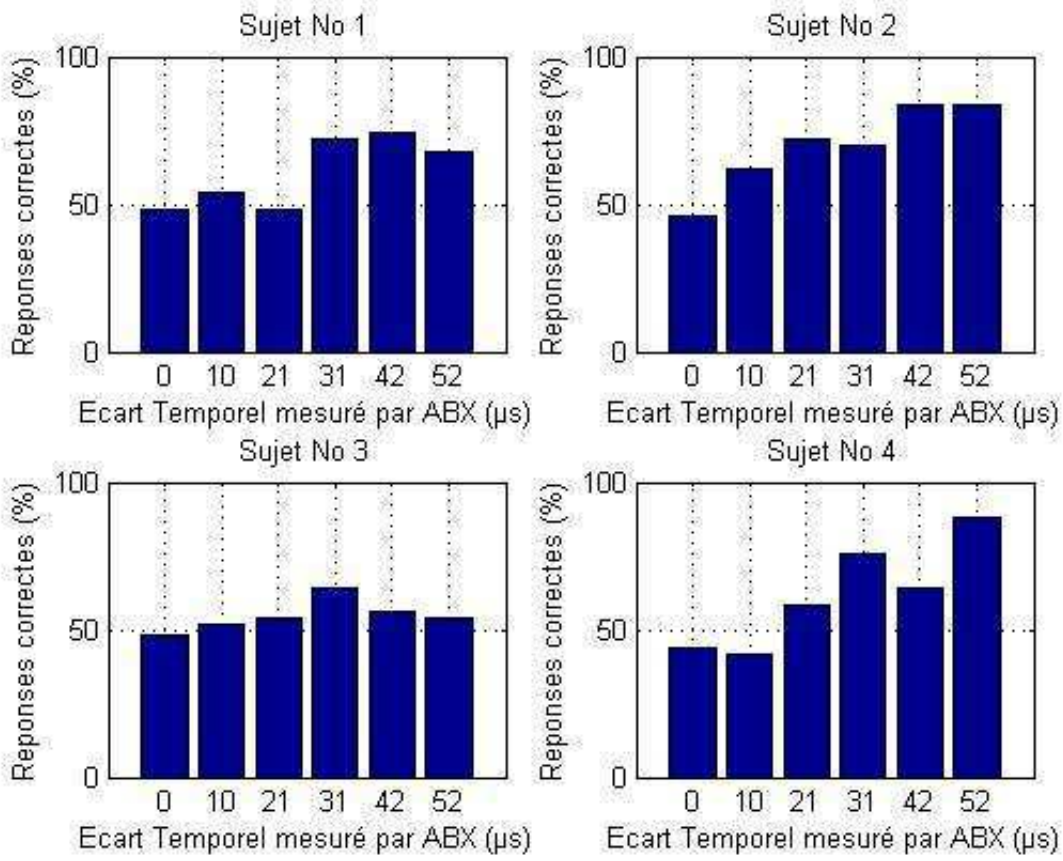


FIG. II.18 – Pourcentage de réponses correctes en fonction de ΔITD pour la procédure ABX.

procédé d'adaptation soit responsable de l'augmentation du seuil. Ceci peut notamment s'expliquer par le fait qu'une erreur d'attention est moins pénalisante pour les méthodes non adaptatives. En effet, une erreur est ensuite moyennée sur un grand nombre de présentations (50 ici), tandis que pour une méthode adaptative, une erreur fait augmenter le seuil et le sujet peut avoir beaucoup de difficulté à revenir au niveau du seuil avant la faute d'attention car cela lui demande une grande série de bonnes réponses. De plus, comme mentionné dans [Trahiotis et al. (1990)], une procédure adaptative permet un entraînement moins lourd, voir aucun, par rapport aux méthodes constantes. En contre partie, le sujet ne dispose pas assez d'écoutes pour les niveaux faibles du paramètre testés alors que les méthodes constantes offrent un nombre égal d'écoutes pour tous les niveaux du paramètre.

1.6 Conclusion sur l'estimation de la JND de l'ITD

Différentes méthodes et procédures détaillées et largement utilisées dans la littérature sont testées. La comparaison des méthodes c'est faite sur l'estimation de la JND de l'ITD seule, c'est-à-dire sans filtrage par les HRTF, qui est une valeur connue pour la position frontale ($ITD_{ref} = 0$ μs). Ce test met en évidence les différences dans les JND

obtenues. Dans le choix d'un protocole expérimental, bien d'autres aspects sont à prendre en compte, notamment la fatigue du sujet, sa disponibilité et son entraînement. Comme la mesure de la JND de l'ITD dans l'implémentation $\{ITD \bigoplus HRTF_{min}\}$ de la synthèse binaurale n'est pas reportée dans les études antérieurs, la JND sera alors mesurée avec deux méthodes différentes : la méthode ABC2 (la plus rapide mais donnant le seuil le plus haut) et la méthode AB (la plus longue mais donnant le seuil le plus bas).

1.7 Bilan des connaissances sur l'ITD et présentation des axes de recherches sur l'ITD

Devant la multitude des méthodes de calcul de l'ITD, leur complexité de mise en oeuvre, et surtout devant le fait qu'elles sont des méthodes d'estimations, dans le sens où les mesures sont nécessaires, une formulation simple comme celle du modèle de tête sphérique devient intéressante du point de vue pratique. Le modèle sphérique est individualisable et l'approche menée par [Algazi et al. (2001b)] semble une piste à suivre. Cependant, l'individualisation du modèle sphérique est limitée et ce modèle ne rend pas compte des variations de l'ITD sur des cônes de confusion. Dans [Duda et al. (1999)] l'hypothèse que le décalage des oreilles modifie l'ITD est faite. Cette hypothèse est vérifié dans [Pernaux (2003)] grâce à la formulation analytique des HRTF du modèle de tête sphérique. Le paragraphe 2.3 présente l'étude théorique réalisée pour les travaux de thèse décrivant une formulation analytique de l'ITD du modèle sphérique prenant en compte le décalage des oreilles sur la sphère. Cette étude évalue l'importance relative des paramètres de cette formule qui reproduit les variations observées de l'ITD sur des cônes de confusions. Outre une modélisation plus fine de l'ITD, un gain d'individualisation est aussi escompté. Ces travaux ont donné lieu à une présentation et un article pour le congrès CFA/DAGA 2004 [Busson et al. (2004)].

Les méthodes d'estimation et de prédiction sont très variées et leur comportement global et individuel est différent d'une méthode à l'autre. Cependant l'hypothèse d'une équivalence perceptive entre les HRTF à phase mixte et le modèle $\{ITD \bigoplus HRTF_{min}\}$ est validée pour toutes les positions [Minaar et al. (2000)]. Un doute persiste pour les positions contralatérales, positions ou le signal est très atténué et résulte des phénomènes complexes de diffraction autour de la tête [Avendano et al. (1999)]. Pour ces positions, la phase n'est plus linéaire [Møller et al. (1995)] et le modèle peut être remis en cause. Dans une série d'articles consacrés à la synthèse binaurale [Minnaar et al. (1999); Plogsties et al. (2000); Minaar et al. (2000)], l'équipe du laboratoire d'acoustique de l'Université d'Aalborg montre que le modèle $\{ITD \bigoplus HRTF_{min}\}$ reste valide même pour les positions latérales et que la phase de la composante à excès de phase des HRTF peut toujours être remplacée par un retard pur, même si les HRTF ne sont plus modélisables par des filtres à phase minimale pour ces positions. Ce retard pur est le retard de groupe de la composante à excès de phase cumulé à 0 Hz. Or, comme indiqué dans [Algazi et al. (2001b)], les HRTF sont peu valides en basses fréquences et les informations doivent y être extrapolées. Il apparaît alors nécessaire de mener des tests psychoacoustiques pour connaître d'une part la validité d'autres méthodes que celle du retard de groupe à 0 Hz et d'autre part vérifier l'adéquation de ces autres méthodes avec le modèle $\{ITD \bigoplus HRTF_{min}\}$. L'étude expérimentale décrite au paragraphe 3 a été menée dans ce sens et a donné lieu à une présentation et un article à la 118$^{\text{th}}$ convention AES [Busson et al.

(2005a)].

Dans la connaissance de l'importance d'un paramètre, il faut établir deux types de valeurs : la valeur absolue et la sensibilité à cette valeur. Les valeurs absolues subjectives et objectives de l'ITD évoluent principalement avec l'azimut de la source et l'individu et ces aspects ont été largement traités dans la littérature. Par contre, à l'inverse de la sensibilité de l'ITD étudiée comme un indice isolé, très peux d'études ont porté sur la sensibilité de l'ITD dans le modèle $\{ITD \bigoplus HRTF_{min}\}$. Devant le nombre de méthodes de calcul, la JND de l'ITD revêt un intérêt particulier pour la comparaison subjective de méthodes objectives et détermine le degré de précision *nécessaire* pour reproduire les variations de l'ITD. Le paragraphe 4 décrit l'étude expérimentale qui a permis la mesure de la JND de l'ITD sur des plans sagittaux dans le cadre de l'implémentation $\{ITD \bigoplus HRTF_{min}\}$ de la synthèse binaurale. Cette étude est aussi décrite dans [Busson et al. (2005b)].

2 INDIVIDUALISATION DU MODÈLE DE TÊTE SPHÉRIQUE POUR LA REPRODUCTION DES VARIATIONS DE L'ITD SUR DES CÔNES DE CONFUSION

Le modèle de tête sphérique avec oreilles centrées est la modélisation physique de l'auditeur la plus utilisée et la plus étudiée. Elle permet l'obtention à moindre coût d'implémentation d'indices monauraux et binauraux apportant des informations de spatialisation convenant globalement. Les indices spectraux sont même très bien reproduits jusqu'à 2 kHz [Katz (1998)]. L'ITD donnée par la formule de Woodworth (cf. équation II.1) convient en moyenne et offre des performances individuelles proches d'autres méthodes de calcul beaucoup plus coûteuse en terme d'implémentation (cf. tableau II.7). L'ITD du modèle sphérique peut être individualisée en calculant un rayon optimal minimisant l'erreur entre l'ITD issue des mesures et l'ITD du modèle. Cette démarche est suivie par [Algazi et al. (2001b)] qui obtiennent ainsi une formule empirique entre trois paramètres morphologiques et le rayon optimal. L'ITD_{sphere} calculée de cette manière offre une erreur faible (32 μs en moyenne) sur la prédiction de l'ITD hautes fréquences (f $>$ 1.5 kHz). Les erreurs rémanentes sont principalement localisées sur les positions latérales et sont dues au caractère non-sphérique de la tête de l'auditeur et au décalage des oreilles par rapport au centre de la tête [Algazi et al. (2001b); Duda et al. (1999)]. En effet, le tracé de l'ITD sur un plan sagittal fait apparaître des variations que les modèles de têtes avec oreilles centrées, c'est-à-dire diamétralement opposées, ne peuvent reproduire. La figure II.19 illustre ce propos : la ligne horizontale représente l'ITD du modèle sphérique avec oreilles centrées et la courbe rouge la moyenne des ITD des sujets de la base CIPIC pour le cône de confusion correspondant à $\theta = 45°$.

Afin d'améliorer la prédiction de l'ITD par le modèle de tête sphérique, une formule analytique ,nommée FDO pour Formule de Décalage de Oreilles, qui prend en compte un décalage en élévation et en azimut de la position des oreilles par rapport à une position diamétralement opposée a été mise au point. Cette formule est validée dans le cas où le décalage est nul, c'est-à-dire par rapport à la formule de Woodworth et la formule de Larcher et est ensuite utilisée pour décrire l'influence du décalage des oreilles sur une sphère pour le calcul de l'ITD. Suivant la procédure d'optimisation décrite dans [Algazi et al. (2001b)], le décalage des oreilles ainsi que les rayon optimaux pour les sujets de la

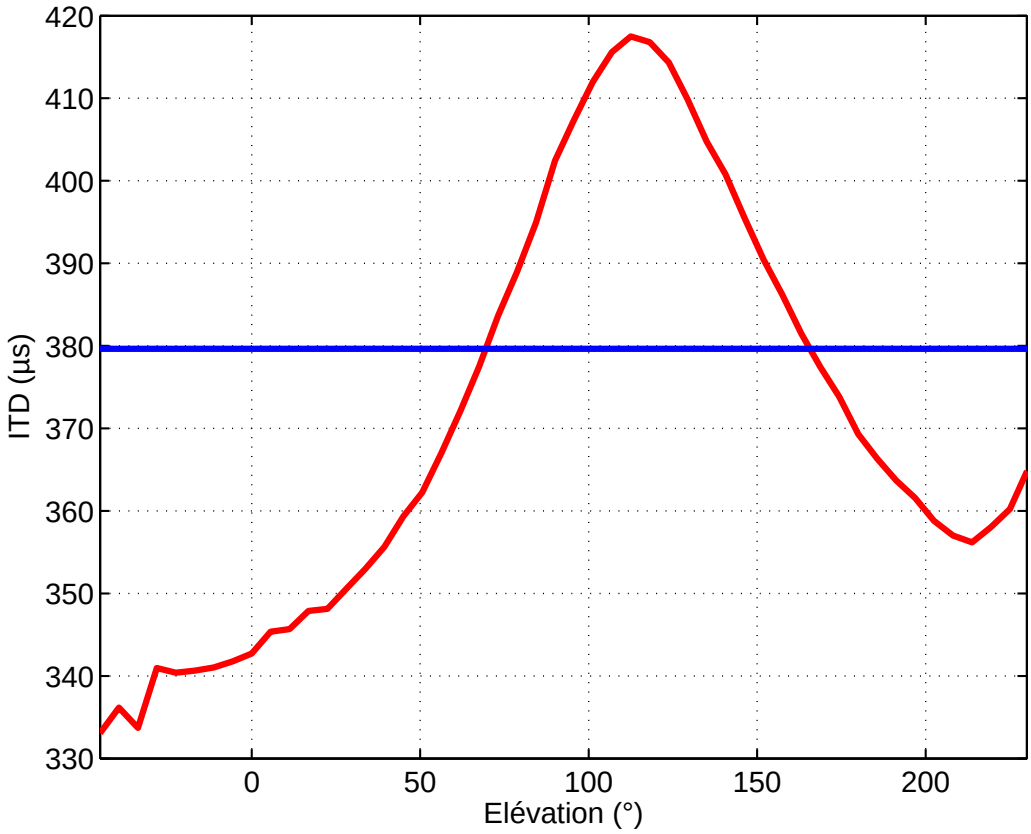


FIG. II.19 – ITD du plan sagittal $\theta = 45°$ en fonction de l'élévation pour la moyenne des ITD de la base CIPIC (courbe rouge) et pour la formule de Woodworth (ligne bleue).

base de données FTR&D ont été calculés. Enfin les résultats de la procédure d'optimisation sont analysés et permettent de dégager l'importance relative des paramètres de la formule FDO.

2.1 Formule de Décalage des Oreilles

Une formule analytique donnant l'ITD pour un modèle de tête sphérique avec prise en compte d'un décalage des oreilles par rapport au centre de la tête n'apparaît pas dans la littérature. La formule FDO considère une sphère rigide de rayon R, une vitesse du son c et, comme la formule de Woodworth, le masquage de la tête est exprimé en fonction de la distance parcourue sur la sphère. L'onde incidente et la position des oreilles sont repérées par l'intermédiaire de vecteurs unitaires décrivant la direction et le sens de l'onde incidente et l'orientation des oreilles par rapport au centre de la sphère. Le vecteur unitaire de l'onde, $\overrightarrow{U}_{inc}$, est orienté vers le centre de la sphère et les vecteurs unitaires des oreilles, respectivement $\overrightarrow{U}_d$ pour l'oreille droite et $\overrightarrow{U}_g$ pour l'oreille gauche, sont orientés vers l'extérieur de la sphère. Les distances d'arc sont calculées avec des produits scalaires entre les vecteurs d'ondes et les vecteurs d'oreille. Quatres cas sont distingués selon que

l'onde incidente éclaire ou pas les oreilles (cf. fig.II.20) :

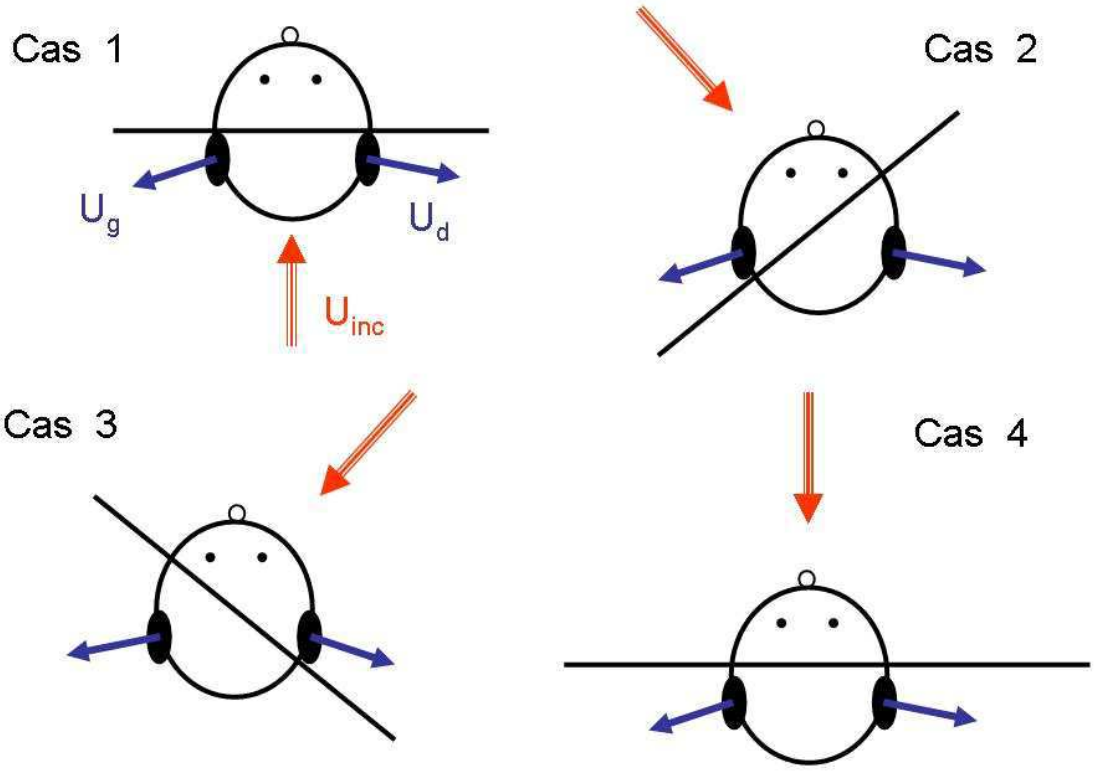


FIG. II.20 – Différents cas de figure pour le calcul de l'ITD du modèle sphérique avec oreilles décalées. La ligne noire indique la séparation entre la zone de masquage et la zone éclairée.

– **Cas 1** : L'onde incidente éclaire les deux oreilles,

$$ITD = -\frac{R}{c} * (\overrightarrow{U}_{inc} \cdot \overrightarrow{U}_g + \overrightarrow{U}_{inc} \cdot \overrightarrow{U}_r) \tag{II.10}$$

– **Cas 2** : L'onde incidente éclaire l'oreille gauche

$$ITD = -\frac{R}{C} * (\frac{\pi}{2} - \arccos(\overrightarrow{U}_{inc} \cdot \overrightarrow{U}_g) - \overrightarrow{U}_{inc} \cdot \overrightarrow{U}_d) \tag{II.11}$$

– **Cas 3** : L'onde incidente éclaire l'oreille droite,

$$ITD = -\frac{R}{C} * (\frac{\pi}{2} - \arccos(\overrightarrow{U}_{inc} \cdot \overrightarrow{U}_d) - \overrightarrow{U}_{inc} \cdot \overrightarrow{U}_g) \tag{II.12}$$

– **Cas 4** : L'onde incidente est masquée par la tête,

$$ITD = \frac{R}{C} * (\arccos(\overrightarrow{U}_{inc} \cdot \overrightarrow{U}_g) - \arccos(\overrightarrow{U}_{inc} \cdot \overrightarrow{U}_d)) \tag{II.13}$$

Cette formulation prends en compte l'azimut et l'élévation de l'onde incidente et reproduit ainsi les variations de l'ITD avec l'élévation dans un repère polaire-vertical. La validation de cette formule est effectuée en considérant le cas où les oreilles sont diamétralement opposées. La validation dans le plan horizontal est réalisée par comparaison avec la formule de Woodworth et par la formule de Larcher et Jot pour les autres plans d'élévation. Pour les deux validations, le rayon de la sphère est pris égal au rayon moyen anthropométrique ($R = 87{,}5$ mm).

Les figure II.21.A et II.21.B montrent une parfaite correspondance entre les formules de Woodworth et de Larcher et Jot. Il est démontré en annexe B que les formules de Woodworth et FDO sont équivalentes sur le plan horizontal. La formule FDO est donc validée dans le cas où il n'y a pas de décalage des oreilles. La validation pour les cas avec décalages ne peut se faire que par l'intermédiaire de méthodes de calcul de l'ITD à partir de HRTF ou HRIR. Ceci entraîne des problèmes liés à l'estimation et donc il ne pourrait y avoir une correspondance parfaite comme avec les formules analytique de Woodworth et de Larcher et Jot. Le travail rapporté au paragraphe 2.3 montre la capacité d'adaptation de la formule FDO à des ITD issues de mesures.

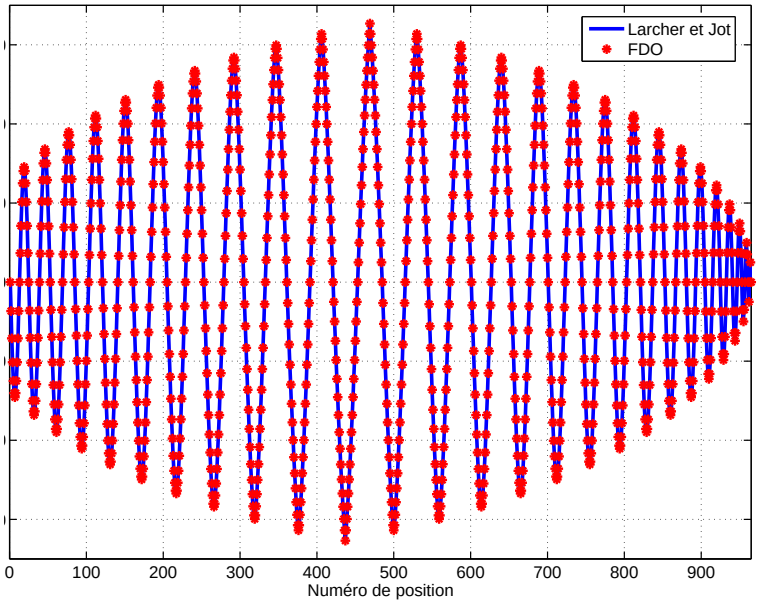
2.2 Influence du décalage des oreilles sur le modèle de tête sphérique pour le calcul de l'ITD

L'influence des paramètres de la formule FDO sur l'ITD est évalué sur le plan sagittal défini par $\theta = 60°$ en coordonnées polaires-interaurales. Une ITD normalisée est utilisée : les valeurs données par FDO sont normalisées par l'ITD_{sphere} pour $\theta = 60°$. La figure II.23.b représente l'influence d'un décalage des oreilles en azimut et la figure II.23.a l'influence d'un décalage des oreilles en élévation. Le décalage en azimut est compté positivement lorsqu'il correspond à un déplacement vers l'hémisphère avant, et compté négativement vers l'hémisphère arrière. Le décalage en élévation est compté positivement lorsqu'il correspond à un déplacement vers l'hémisphère supérieur, et compté négativement vers l'hémisphère inférieur (cf. fig. II.22).

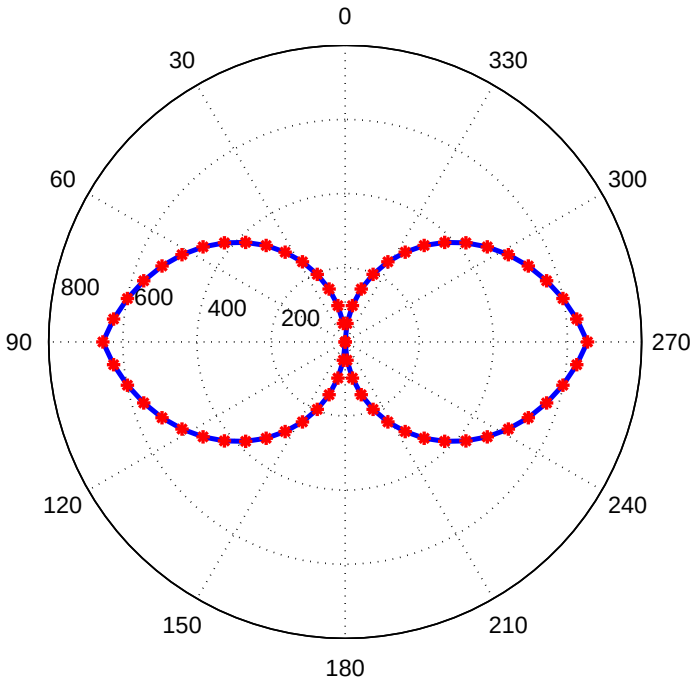
Les deux effets majeurs du décalage des oreilles sont d'une part l'apparition d'une bosse qui se transforme en pic quand le décalage augmente et d'autre part une baisse de la valeur moyenne de l'ITD. La position de la bosse ne change pas avec le décalage, c'est juste son amplitude qui évolue. Décalages en azimut et décalages en élévation ont des effets similaires mais dans des directions différentes. Un décalage positif en azimut entraîne l'amplification de la bosse pour l'élévation basses -50° et un décalage négatif entraîne l'amplification de la bosse pour l'élévation haute autour de 90°. En ce qui concerne le décalage des oreilles en élévation, un décalage positif entraîne l'augmentation de la valeur maximale d'une bosse autour de 20°. Un décalage négatif a un effet similaire mais pour une bosse autour de 165°. Les variations engendrées sont symétriques par rapport aux valeurs d'élévation indiquées (symétrie due à la sphère). Ces effets sont d'autant plus importants que l'azimut du plan sagittal considéré est proche de l'axe interaural.

2.3 Optimisation des paramètres de la formule FDO

La démarche décrite dans [Algazi et al. (2001b)] est appliquée aux huit sujets de la base de HRTF FTR&D. Une ITD haute fréquence est calculée à partir de HRIR et l'erreur



A) Comparaison avec formule de Larcher et Jot pour toutes les positions de la base FTR&D.



B)Comparaison avec la formule de Woodworth sur le plan horizontal.

FIG. II.21 – Validation de la formule FDO. A) Comparaison avec formule de Larcher et Jot pour toutes les positions de la base FTR&D. B) Comparaison avec la formule de Woodworth sur le plan horizontal.

rms entre cette ITD et l'ITD donnée par différents modèles est estimée. Le calcul se fait

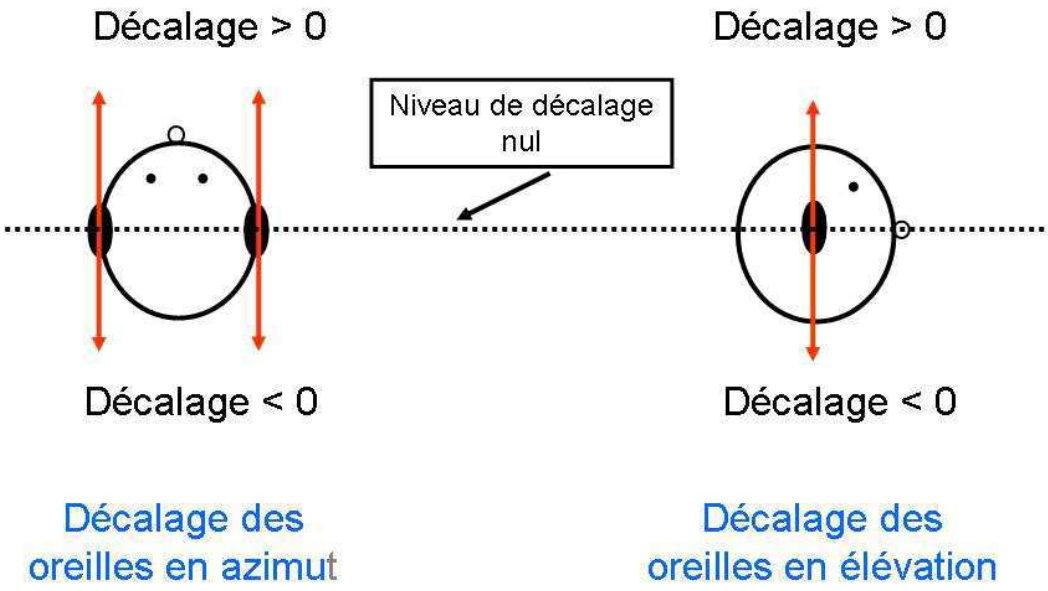
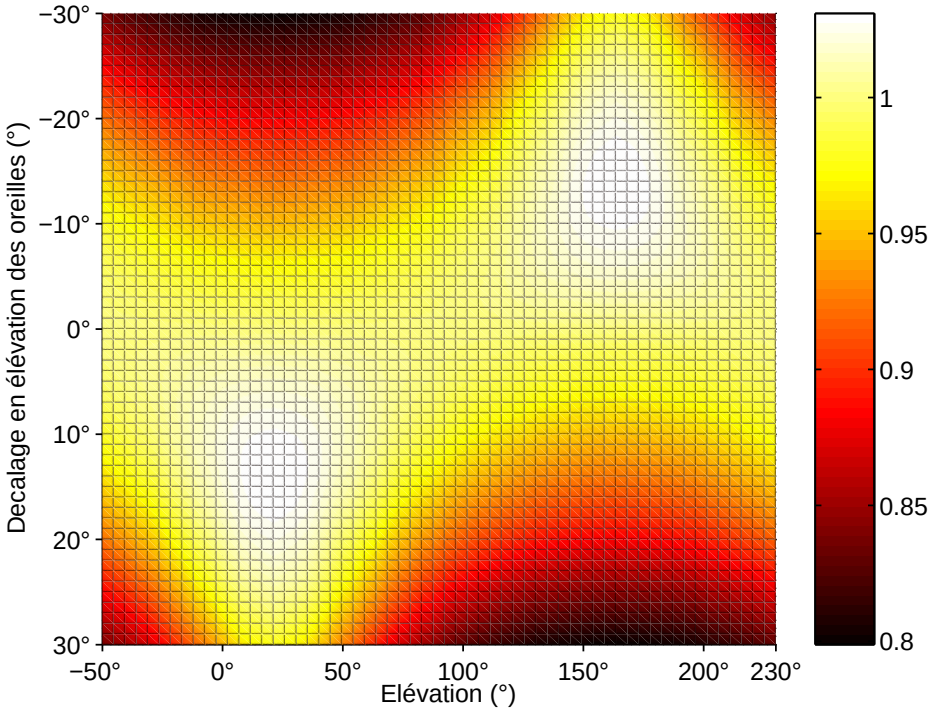


FIG. II.22 – Convention pour le décalage des oreilles.

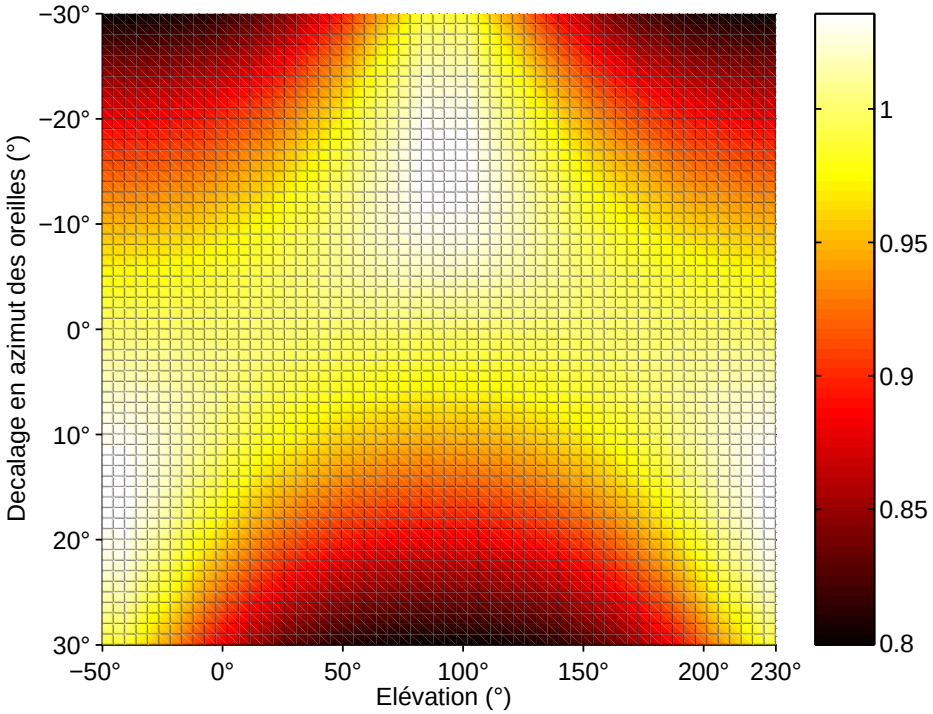
sur les 965 positions de la base FTR&D. Ensuite une procédure est menée de la même façon pour optimiser le rayon de la formule de Woodworth et aussi le décalage optimal des oreilles grâce à la formule FDO. Cette procédure consiste à calculer l'erreur pour toute une plage de variation des paramètres et récupérer les paramètres qui procurent une erreur minimale. Les valeurs des différentes erreurs rms sont reportées dans le tableau II.5.

Le tableau II.5 montre que globalement la formule d'Algazi reliant trois paramètres morphologiques et un rayon optimal permet une réduction de l'erreur de 11 % par rapport à la formule de Woodworth. Cependant, ce rayon n'est pas optimal pour la base de données FTR&D. En effet, la procédure de minimisation sur le rayon donne un autre rayon optimal. Ce nouveau rayon réduit l'erreur par rapport à la formule de Woodworth de 15.3 % en moyenne. Enfin, il apparaît que le décalage des oreilles permet de réduire l'erreur de prédiction de 27.1 % en moyenne, ce qui représente une amélioration de 14.3 % par rapport à la formule avec le nouveau rayon. Les valeurs des nouveaux rayons et des décalages des oreilles optimaux sont reportés dans le tableau II.6. Les décalages en azimut sont tous négatifs et tous positifs en élévation. Les valeurs restent dans des ordres de grandeur réalistes.

Une régression multi-linéaire est réalisée entre trois paramètres morphologiques, respectivement largeur X_1, hauteur X_2 et profondeur de la tête X_3, et le nouveau rayon.



a) Influence du décalage en élévation des oreilles.



b) Influence du décalage en azimut des oreilles.

FIG. II.23 – Influence du décalage des oreilles sur l'ITD normalisée.

Tab. II.5 – Erreurs rms entre les méthodes de calcul de l'ITD et l'ITD haute fréquence estimée à partir de HRIR. Les valeurs des erreurs sont données en μs. Les pourcentages sont calculés par rapport à l'erreur rms de la formule Woodworth.

Sujets FTR&D	Woodworth	Rayon Algazi	Nouveau rayon	FDO
Ma	58	43	42	37
Jm	54	44	34	32
Va	42	46	42	37
Je	45	41	41	35
Ro	41	45	39	30
Mo	34	36	36	28
Pa	54	49	38	34
No	40	38	37	33
Moyenne	**46**	**43**	**38**	**33**
	Amélioration par rapport à Woodworth	**11 %**	**15.3 %**	**27.1 %**

Tab. II.6 – Nouveaux rayons et décalages des oreilles pour les sujets FTR&D.

Sujets FTR&D	Nouveau rayon (mm)	décalage en azimut (°)	décalage en élévation (°)
Ma	99	-11.5	2
Jm	99	-13	2
Va	92	-11	5
Je	95	-11	5
Ro	95	-13.5	5
Mo	90	-13	5
Pa	99	-10.5	5
No	94	-9	5

Les poids suivants sont obtenus pour un rayon donné en millimètres (965 positions, 8 sujets) :

$$R_N = -0.0094 * X_1 + 0.5064 * X_2 + 0.2045 * X_3 + 17.5 \quad \text{(II.14)}$$

Algazi *et al.* trouvent les poids suivants (1250 positions, 25 sujets) :

$$R_a = 0.51 * X_1 + 0.019 * X_2 + 0.19 * X_3 + 32 \quad \text{(II.15)}$$

tandis que Larcher [Larcher (2001)] donne (825 positions de l'hémisphère supérieur, 17 sujets) :

$$R_l = 0.66 * X_1 - 0.04 * X_2 + 0.11 * X_3 + 33.3 \quad \text{(II.16)}$$

Si les poids changent peu entre la formule donnée par Algazi *et al.* et celle donnée par Larcher, la différence est nette avec la formule tirées des données FTR&D. L'importance des paramètres est redistribuée : ainsi la largeur de la tête influence beaucoup R_N. La

valeur de la constante et presque divisée par deux. Ceci est plutôt encourageant pour la formule du nouveau rayon car cela signifie que les paramètres morphologiques ont globalement plus d'influence sur le rayon optimal que dans les autres formules. Cette comparaison des formules montrent qu'une base de données se doit d'être *universelle* pour que des formules de ce types soient optimales pour tous. De plus, il a été montré que le rayon optimal dépend des positions sur lequel il est optimisé (r_{2d} et r_{3d} dans [Larcher (2001)]). C'est pourquoi, l'utilisation de la formule d'Algazi *et al.* pour la prédiction de l'ITD des sujets de la base FTR&D n'est pas satisfaisante vu que la formule de Woodworth donne de meilleurs résultats (cf. tableau II.7).

La formule FDO optimisée permet une amélioration de 27 % par rapport à la formule de Woodworth, ce qui réduit l'erreur moyenne de 13 μs. Cette réduction peut sembler faible, mais le décalage des oreilles sur une sphère est surtout intéressant pour reproduire les variations de l'ITD individuelle des cônes de confusion. Ainsi, avec un décalage approprié, il est possible d'adapter l'ITD d'une sphère de rayon équivalent afin que celui reproduise les variations sur les cônes de confusions. Seulement, l'ITD individuel, issu des mesures, fait apparaître des variations qui sont peu symétriques et surtout qui dépendent du cône de confusion. Ainsi, si on veut adapter une sphère équivalente de telle sorte qu'elle reproduise les variations de l'ITD, il faut choisir un cône en particulier. En effet, comme décrit auparavant, l'effet du décalage des oreilles est à la fois l'apparition d'une bosse dans le profil de l'ITD avec l'élévation et une réduction de la valeur moyenne. Cette réduction de la valeur moyenne entraîne une augmentation du rayon équivalent et donc un rayon optimal est spécifique à chaque cône de confusion.

Une fonction implémentée avec Matlab est écrite pour trouver les décalages des oreilles qui minimisent l'erreur de reconstruction pour un cône en particulier. Etant donné que les variations de l'ITD sont plus importantes au fur et à mesure que l'on s'éloigne du centre de la tête, il est préférable de se focaliser sur un cône dont l'amplitude de variation de l'ITD est la plus forte. La figure II.24 représente l'ITD moyenne de la base CIPIC pour le cône à 65° d'azimut et l'ITD donnée par FDO optimisée pour cette ITD. FDO s'adapte remarquablement bien aux données qui lui sont présentées. Le rayon optimal et alors de 91 mm et les décalages des oreilles en azimut et en élévation sont respectivement de -3° et de 19°.

2.4 Conclusion

La formule FDO du modèle de tête sphérique avec décalage des oreilles est validée. Cette formule prend en compte les variations de l'ITD avec l'élévation et permet de réduire l'erreur avec les mesures par un positionnement approprié du placement des oreilles. L'hypothèse émise par Duda [Duda et al. (1999)], c'est-à-dire que le déplacement des oreilles permet de faire varier l'ITD sur un plan sagittal, est vérifiée. Le décalage des oreilles permet ainsi à l'ITD du modèle sphérique de s'adapter à l'ITD d'un cône de confusion. La question de savoir si ces variations doivent être reproduites est abordée au paragraphe 4. Une étude expérimentale y est menée pour connaître la sensibilité perceptive à l'ITD sur les cônes de confusion.

L'étude sur la formule FDO a montré que le modèle sphérique pouvait être amélioré pour la prédiction de l'ITD. La formule FDO permet une estimation de l'ITD rapide et sans mesure de HRTF. Le paragraphe qui suit présente une étude expérimentale pour la

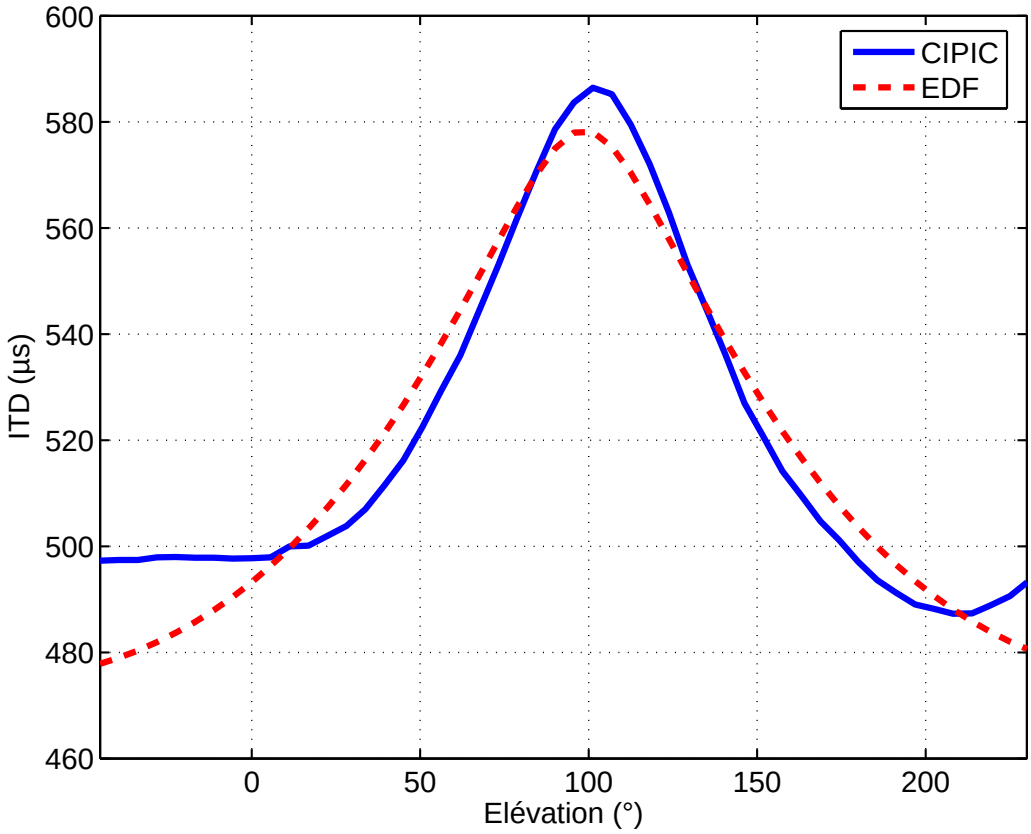


FIG. II.24 – ITD moyen du cône à 65° de la base CIPIC et FDO optimisée en fonction de l'élévation.

mesure de l'ITD psychoacoustique. Cette ITD permet la comparaison des performances des méthodes de calcul de l'ITD.

3 ESTIMATION SUBJECTIVE DE L'ITD SUR LE PLAN HORIZONTAL

3.1 But du test

Les amplitudes de variation de l'ITD sont maximales sur le plan horizontal et les méthodes de calcul diffèrent le plus pour les positions latérales (cf. § 1.4). Le but du test est de déterminer les valeurs psychoacoustiques de l'ITD sur le plan horizontal et ainsi d'estimer l'adéquation des méthodes de calcul avec les données psychoacoustiques. Par *valeurs psychoacoustiques* il faut comprendre ici **les valeurs de l'ITD qui associées avec un filtre à phase minimale procurent le même rendu perceptif que le filtre à phase mixte**. Le rendu n'est toutefois évalué que par l'intermédiaire de la précision de localisation et **il est demandé au sujet d'ajuster l'ITD d'un son test pour que sa position perçue corresponde à la position perçue d'un son cible**. Le sujet manipule presque directement le son test : il s'agit alors d'un *pointeur acoustique*. Une telle procédure évite de nombreux artefacts. Les procédures utilisant un système de report non-acoustique in-

troduisent des biais psychomoteurs et kinesthésiques en plus de l'apprentissage spécifique dans la manipulation du système.

3.2 Choix d'un protocole expérimental

Dans certaines études sur la JND de l'ITD [Domnitz and Colburn (1977); Klump and Eady (1956); Moushegian and Jeffress (1959)], les sujets ajustent l'ITD par l'intermédiaire d'un bouton rotatif qui est relié à une ligne à retard *introduisant* l'ITD entre les signaux envoyés aux deux oreilles. Des écoutes préliminaires ont mis en évidence les difficultés pour faire spatialement[8] correspondre un son convolué par une HRIR à phase mixte avec un son convolué avec une HRIR à phase minimale pour les positions latérales. Si par exemple, le son cible correspond à la position $\theta = 90°$ et que le sujet essaie d'ajuster l'ITD d'un son test de manière continue, comme avec un bouton, il se peut que le sujet n'ait jamais satisfaction. Cet effet peut s'expliquer par la perte d'information liée à l'élimination de la composante passe-tout des HRTF mais est aussi lié à la notion de trajectoire et à des phénomènes complexes de présentation des stimuli. Le sujet augmentera l'ITD avec l'impression que la position perçue ne change plus jusqu'à ce que l'ITD atteigne des valeurs ne permettant plus la fusion des informations droites et gauches (ITD proche de 1 ms). Pour éviter ce problème, une méthode qui assure la convergence de la valeur d'ITD est utilisée et la plage de variation de l'ITD est restreinte. Cette procédure est issue des méthodes adaptatives décrites dans [Levitt (1970)].

Comme il est proposé une adaptation d'un protocole existant, il convient de s'assurer que le nouveau protocole n'introduit que des artefacts audibles acceptables. C'est pourquoi une condition expérimentale de contrôle est créée. Dans cette condition, le sujet effectue la même tâche, mais cette fois-ci avec des stimuli de même nature, c'est-à-dire un son convolué par une HRIR à phase minimale. L'ITD du son test doit être alors très proche de l'ITD utilisé pour le son cible.

3.3 Description du protocole expérimental

Stimuli Le stimulus de base est un bruit blanc gaussien d'une durée de 400 ms filtré pass-bas à $f_c = 3$ kHz. Cette bande de fréquence est choisie pour favoriser la localisation à partir de l'ITD et amoindrir l'effet d'une ILD perturbateur constant par rapport à l'ITD qui varie. Une pente de croissance et de décroissance en cosinus d'une durée de 5 ms est appliquée au stimulus de base pour obtenir une attaque douce du son reproduit sur casque. Ce stimulus est ensuite soit convolué avec une paire de HRIR à phase mixte, pour le son cible de l'expérience principale, soit un son convolué avec une HRIR à phase minimale, pour tous les autres stimuli. Les filtres sont issus de bases de données et sont implémentés sour la forme de filtre RIF. Les HRIR sont propres à chaque sujet et sont égalisées en champ diffus.

Positions testées Les positions testées sont restreintes au plan horizontal et sont indiquées sur le schéma de la figure II.25. Elles correspondent aux 12 angles d'azimut suivants : 0°, 45°, 75°, 90°, 105°, 135°, 180°, 235°, 255°, 270°, 285°, 315°. Chaque posi-

[8] Au sens de la perception

tion est testée cinq fois (cinq essais) et fait l'objet d'une présentation aléatoire, pour un total de 60 essais.

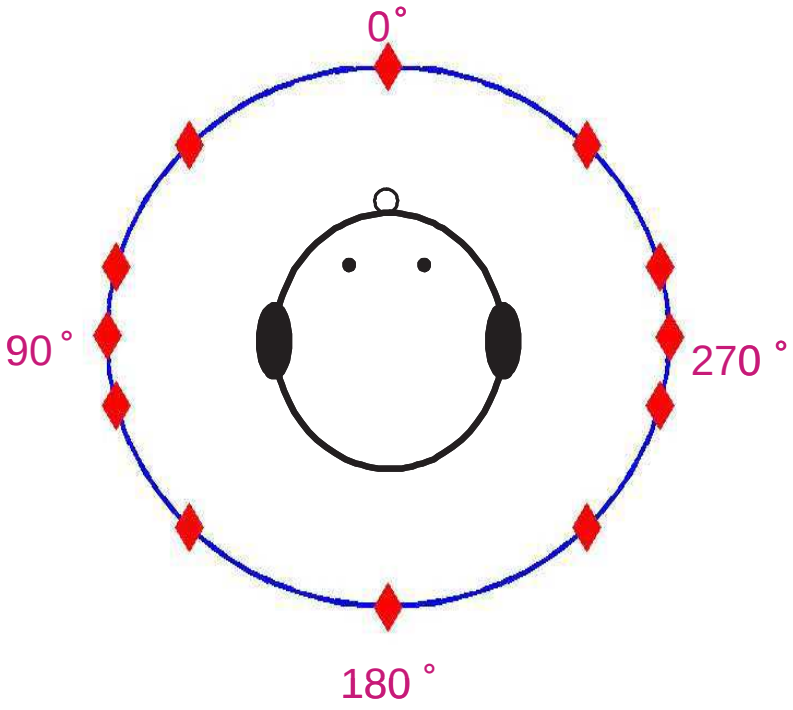


FIG. II.25 – Positions tests pour l'estimation subjective de l'ITD sur le plan horizontal. Les positions sont indiquées au moyen de losanges rouges. Vue de dessus.

Protocole expérimental Le protocole est inspiré d'une procédure adaptative à choix forcé [Levitt (1970)] (cf. § 1.5.2). Le sujet entend une séquence jouée deux fois et composée de deux stimuli : le son cible puis, 500 ms après, le son test. A la fin de chaque séquence, il est demandé au sujet de répondre à la question suivante :

*Utilisez les touches **u** et **o** pour rapprocher le plus possible le 2ème son du 1er. Pour écouter une nouvelle fois, tapez **i**. Si les positions vous semblent identiques, validez leur position en ré-écoutant trois fois.*

Il est indiqué au préalable la direction la plus *fréquemment* provoquée par une pression sur les touches **u** ou **o**, c'est-à-dire droite ou gauche, du clavier situé devant le sujet. Les flèches directionnelles n'ont pas été utilisées car des confusions avant / arrière peuvent se produire et perturber le sujet quant au déplacement perçu du son test. Le sujet a la possibilité de répéter deux fois la séquence et la troisième fois sert de validation de la position si une correspondance exacte entre le son cible et le son test est perçue.

Au début de l'essai, le son cible est *positionné* aléatoirement. Le positionnement du son test est réalisé avec une $ITD_{MaxIACC}$ auquel est rajouté une ITD supplémentaire variant aléatoirement entre un et quatre échantillons temporels (un échantillon temporel correspond à $22.8\mu s$). Après l'écoute de la séquence, le sujet presse une touche du clavier

situé devant lui et indique *la direction que doit prendre le son test pour se rapprocher du son cible*. Cette direction est prise en compte et l'ITD du son test est modifié dans ce sens. La convergence de l'essai est assurée par la prise en compte des changements de direction, limité à six, et par la réduction du pas de variation de l'ITD du son test à chaque changement de direction [Levitt (1970)]. Pour le premier run, le pas de variation est fixé à cinq échantillons et pour le dernier run un échantillon. La valeur d'ITD perçue est prise comme la valeur de la moyenne du deuxième run [Levitt (1970)] (cf. § 1.5.2).

Comme indiqué précédemment, des difficultés peuvent apparaître pour les positions latérales où une ITD trop importante entraîne un manque de cohérence avec les indices spectraux. Pour éviter cette situation, un interval de variation fixé à $[1.5*ITD_{MaxIACCmin} - 1.5*ITD_{MaxIACCmax}]$ où les extremum sont calculés pour chaque sujet, est calculé. Si le sujet arrive à une ITD en dehors de cet intervalle, un message apparaît à l'écran lui indiquant qu'il a atteint les bornes de variation.

Avant le début du test, le sujet est familiarisé à la tâche avec quatre essais d'entraînement. Le sujet a la possibilité de faire autant de pauses qu'il (elle) le désire et un message s'affiche tous les 20 essais pour lui rappeler qu'il (elle) doit faire une pause. Le test dure en moyenne une heure.

Sujets 21 sujets (6 femmes et 15 hommes) participent à ces deux tests : 10 pour la condition de contrôle et 11 pour la condition expérimentale. Les HRTF individuelles des sujets appartiennent à deux bases de HRTF différentes : la base LISTEN (15 sujets) et la base FTR&D (6 sujets). Une ANOVA réalisée sur toutes les données n'a pas mis en évidence un effet de la base de données.

Dispositif expérimental Le sujet est assis dans une cabine insonorisée offrant un niveau de bruit de fond de 20 dBA. L'interface graphique, l'algorithme de test et la création des stimuli sont réalisés avec Matlab®installé sur une station de travail UNIX (700 MHz). Le signal est traité par une carte son RME ADI-8 Pro®reliée à un amplificateur Yamaha P2075®. Le sujet écoute les stimuli à un niveau sonore de 78 dBA sur un casque d'écoute AKG K240®.

3.4 Résultats de la condition de contrôle

La condition de contrôle donne la possibilité au sujet d'entendre exactement les deux mêmes stimuli, c'est-à-dire des stimuli convolués par un modèle $\{ITD \bigoplus HRIR_{min}\}$ des HRTF. Il est donc attendu que les valeurs perçues de l'ITD soient égales à l'$ITD_{MaxIACC}$ utilisé pour *positionner* le son cible. Les écarts observés sont des indications sur la difficulté globale de la tâche. Les contributions des différentes erreurs (localisation, manque de cohérence du modèle, algorithme de convergence) ne peuvent cependant pas être dissociées. Pour les positions latérales, si les valeurs obtenues restent cohérentes cela signifie que d'une part la tâche est réalisable même pour ces positions particulières et que d'autre part cela apporte des premières informations sur la validité du modèle $\{ITD \bigoplus HRIR_{min}\}$. En effet pour ces positions, la littérature rapporte que les HRTF sont mal modélisables par un filtre à phase minimale [Avendano et al. (1999); Kulkarni et al. (1999)]. De plus, le positionnement relatif de sources sonores sur les positions latérales

est une tâche difficile même en écoute champ libre [Mills (1958); Braasch and Hartung (2002)].

La figure II.26 représente les valeurs d'ITD perçues en fonction des valeurs d'ITD cible. La droite rouge indique le cas idéal, c'est-à-dire une droite d'un coefficient directeur égal à 1. Globalement les réponses sont très proches des valeurs présentées et peu de points semblent aberrants. La pente de la droite de régression passant par tous les points est égale à 0.954, ce qui confirme l'observation globale. Ces observations permettent de conclure que globalement la tâche est réalisée et que le modèle est cohérent pour toutes les positions.

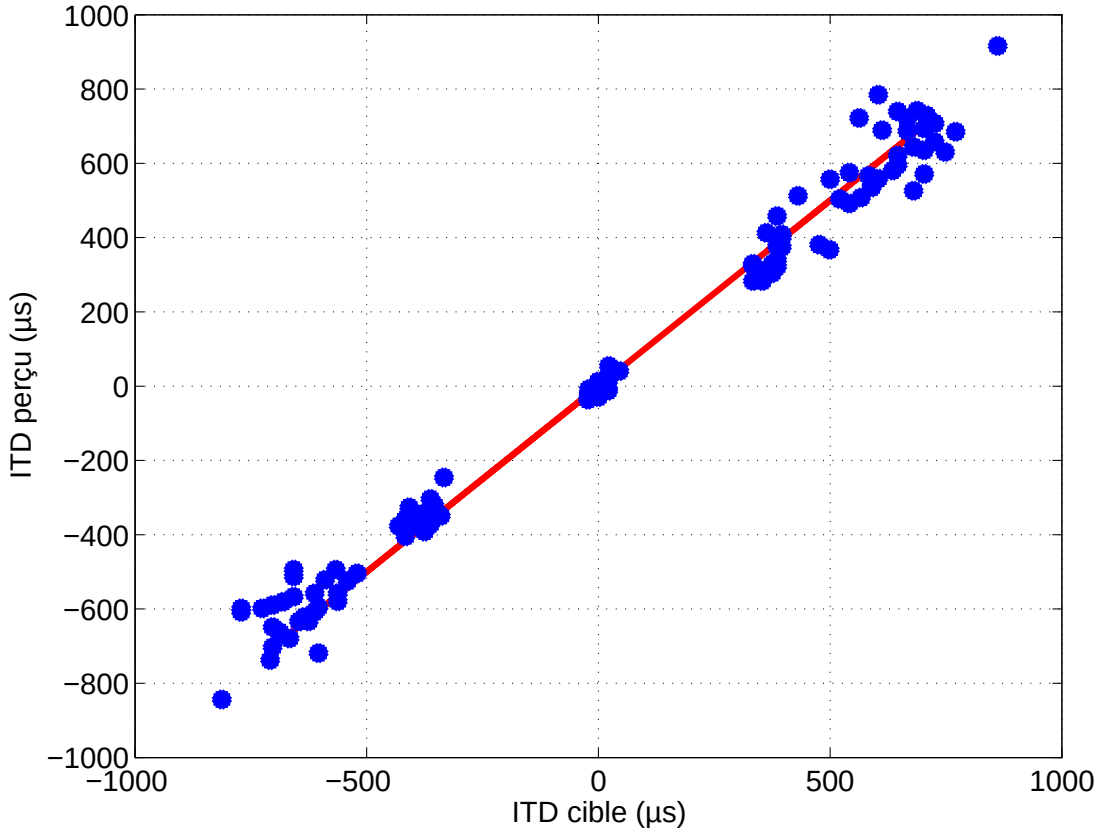


Fig. II.26 – ITD perçue en fonction de l'ITD cible pour les réponses de la condition contrôle. La droite rouge représente le cas idéal.

Pour évaluer globalement la faisabilité de la condition de contrôle, une ITD normalisée est calculée avec l'équation II.8 utilisant la moyenne des cinq réponses par position. La figure II.27 représente les variations de l'ITD moyen normalisé pour chaque sujet ainsi que l'intervalle correspondant à deux fois l'écart-type. L'écart-type ainsi calculé est alors indépendant des différences inter-individuelles qui modifient considérablement la valeur absolue de l'ITD (cf. § 1.4.3). L'écart-type dépend de l'azimut. Il est plus important pour les positions latérales et atteint son minimum pour les positions dans le plan médian ($\theta = 0°$ et $\theta = 180°$). L'écart-type peut être relié, entre autres, à la difficulté de la tâche. Celle-ci

est donc plus difficile à réaliser pour les positions latérales que pour les positions frontales. Cette observation est cohérente avec l'évolution des performances de discrimination de sources sonores en azimut [Blauert (1983)].

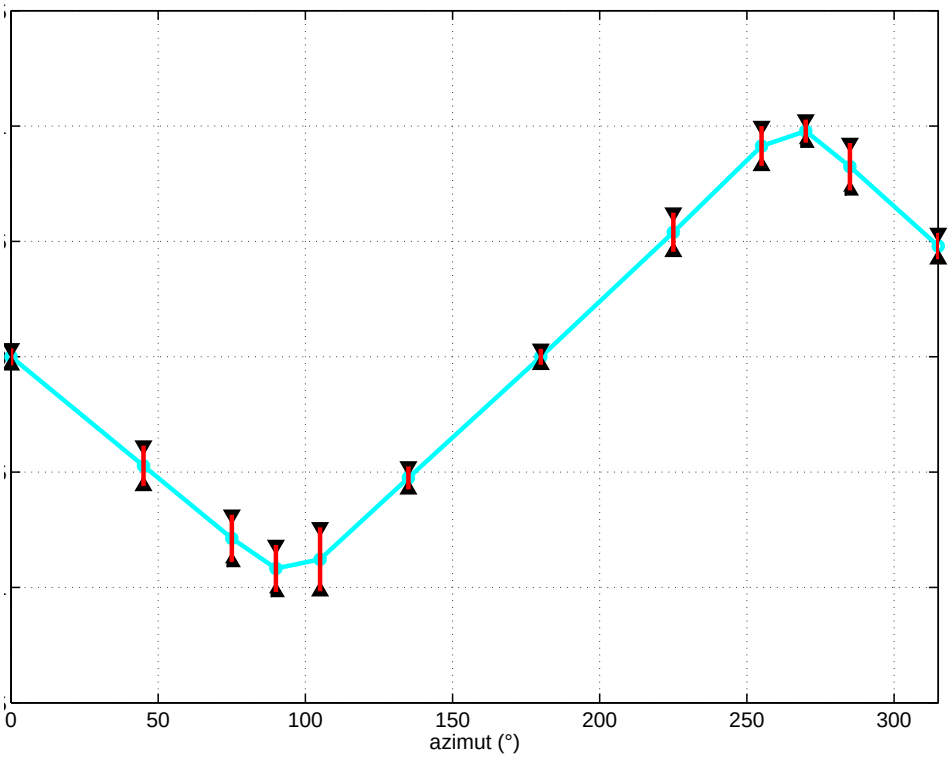


FIG. II.27 – ITD moyenne normalisée en fonction de l'azimut de la position cible. Les barres rouges verticales bornées par des triangles noires représentent deux fois l'écart-type.

Une réponse erronée est comptée comme telle quand l'ITD à la fin de l'essai est au-delà de l'intervalle $[1.5 * ITD_{MaxIACCmin} - 1.5 * ITD_{MaxIACCmax}]$. La figure II.28 représente le pourcentage de réponses erronées en fonction de l'azimut du son cible. Ces réponses erronées sont localisées sur les positions latérales et ne dépassent jamais 6 % du nombre de réponse total par position de son cible. Sur l'ensemble des réponses, ce taux n'excède pas 1 %.

La faisabilité de la tâche est maintenant évaluée grâce à l'erreur absolue moyenne entre l'ITD cible et l'ITD perçue. L'erreur est calculée de la manière suivante :

$$E(\theta) = \frac{1}{N} \sum_{i=1}^{N} |ITD(\theta, i) - \widehat{ITD}(\theta, i)| \qquad \text{(II.17)}$$

où la valeur d'ITD, moyennée sur les cinq essais, reportée par le i^{me} sujet est $\widehat{ITD}(\theta, i)$ et celle utilisée pour le son cible est $ITD(\theta, i)$, N est le nombre de sujet (égal à 10). La figure II.29 décrit les variations de cette erreur en fonction de l'azimut de la position cible.

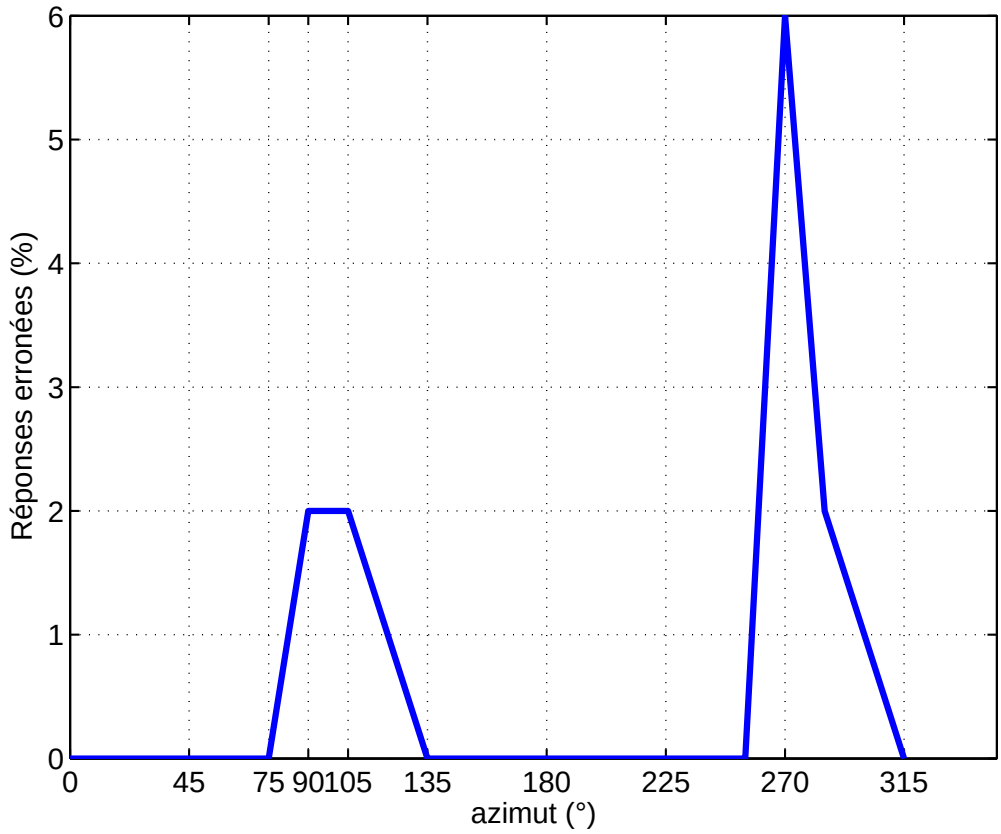


FIG. II.28 – Pourcentage de réponses erronées en fonction de l'azimut de la position cible.

Cette erreur dépend de l'azimut et est plus faible pour les positions frontales que pour les positions latérales, ce qui est en accord avec la perception. L'erreur vaut 14 μs pour la position $\theta = 0°$ et 12 μs pour $\theta = 180°$ ce qui, comparé au pas d'échantillonnage est très faible. De plus, comme l'indique la figure II.30 qui représnte la fonction de répartition empirique[9] de l'erreur absolue, 68 % des erreurs sont inférieures ou égales à 22.3 μs. Dans l'expérience décrite dans [Mills (1958)], une JND de 10 μs[10] est obtenu pour la position $\theta = 0°$.

Ces observations permettent de conclure sur la validité globale du protocole expérimental. Les sujets réalisent correctement la tâche. Les écarts d'estimation de l'ITD restent faibles, et pour les positions du plan médian ils sont inférieurs au pas d'échantillonnage. Les erreurs sont proches à la fois des seuils de perception reportés dans la littérature et de la précision temporelle des ITD présentés. Les résultats de la condition contrôle assurent la validité des résultats de la condition de test.

[9]La fonction de répartition empirique d'une loi de probabilité associe à un réel x la probabilité cumulée F(x) des valeurs inférieures ou égales à x.

[10]JND estimée à 50 % de la courbe psychométrique

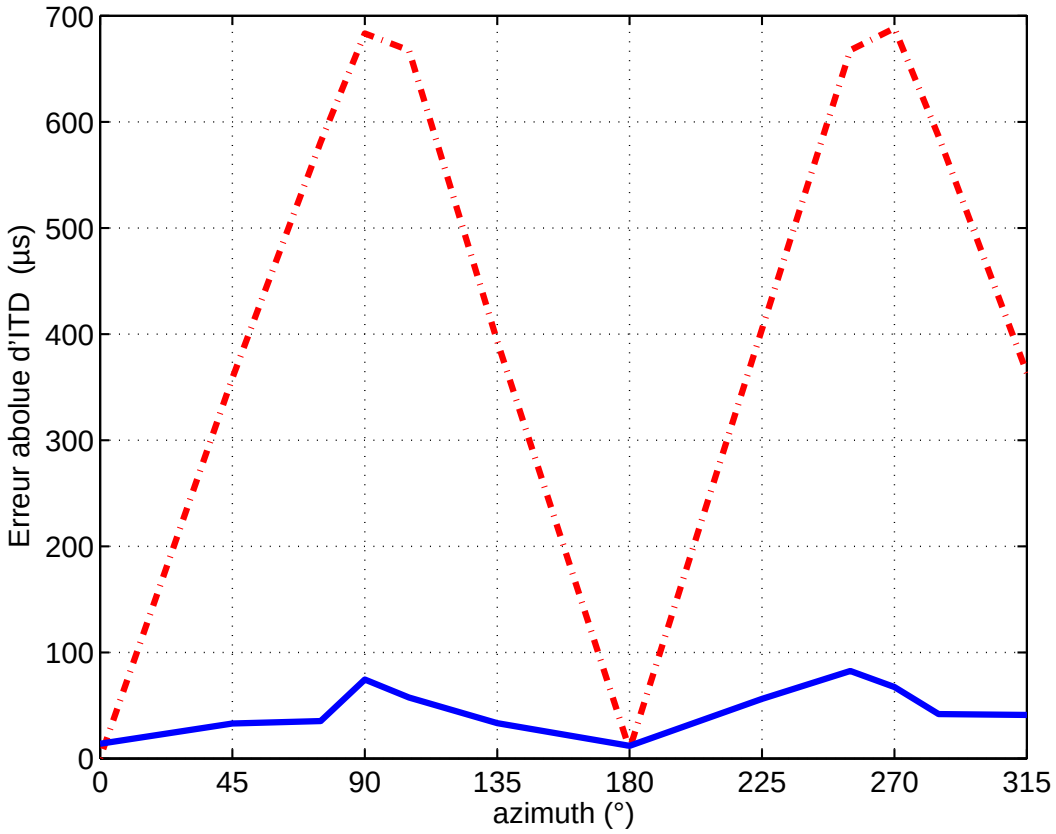


FIG. II.29 – Erreur absolue moyenne d'ITD en fonction de l'azimut de la position cible. La courbe en pointillés rouges indique la valeur moyenne de l'ITD cible pour une comparaison visuelle.

3.5 Résultats de la condition de test

Avant d'analyser les résultats, les réponses des sujets sont examinées pour éviter la prise en compte de points aberrants. Si la dernière valeur d'un essai est en dehors de l'intervalle $[1.5ITD_{MaxIACCmin} - 1.5ITD_{MaxIACCmax}]$ la valeur de l'essai est remplacée par la moyenne des autres réponses pour la même position. Seulement 17 (2.6%) réponses sur un total de 660 sont ainsi corrigées.

Une ANOVA réalisée sur la totalité des réponses des deux conditions expérimentales, c'est-à-dire contrôle et test, n'indique pas d'effet de la condition expérimentale ($F(1, 176) = 0.3, p = 0.605$). Qui plus est un *t-test* effectué sur les deux conditions montre que la différence des résultats entre les deux tests n'est pas significative. Comme les résultats de la condition de contrôle ont montré que la tâche est réalisée avec succès ceci indique que la tâche de la condition test est aussi réalisée correctement. Cette analyse confirme une fois de plus la bonne correspondance perceptive entre des HRIR à phase mixte et le modèle $\{ITD \bigoplus HRIR_{min}\}$.

Les réponses des sujets en fonction de l'azimut du son cible sont représentées sur la figure II.31. Les réponses sont données sous la forme de boîtes à moustaches dont

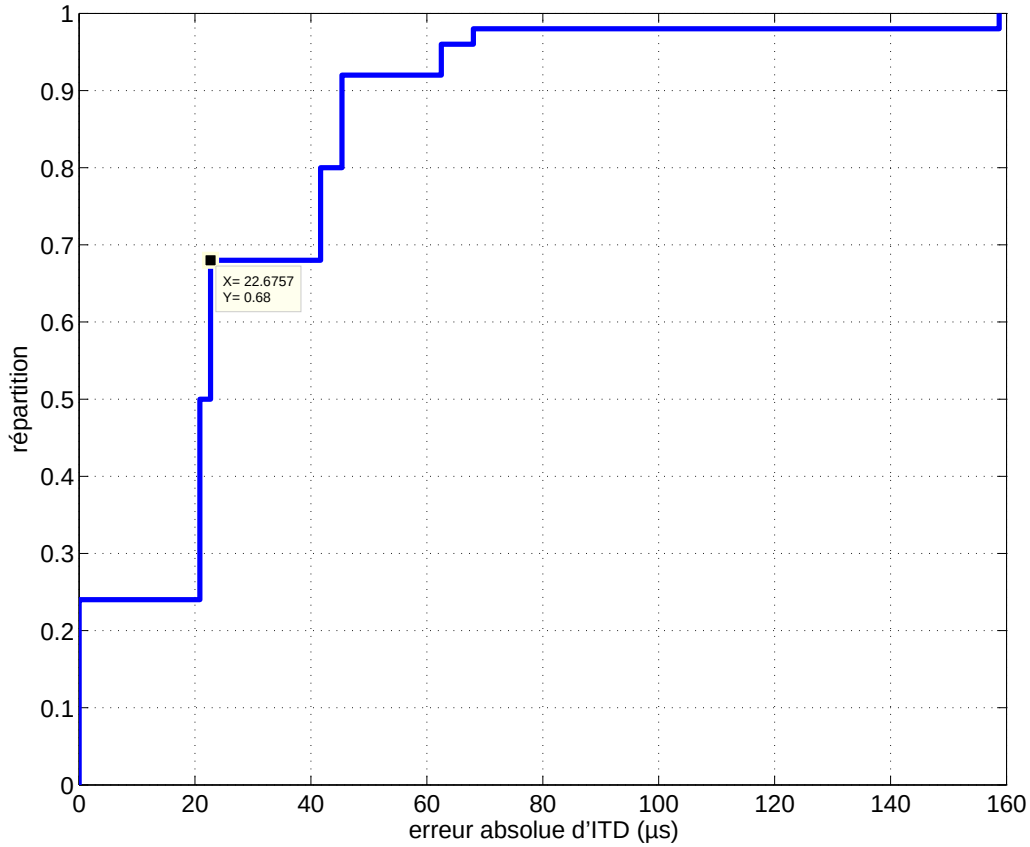


FIG. II.30 – Fonction discrète de répartition empirique de l'erreur absolue d'ITD pour la position $\theta = 0°$.

le descriptif est donné dans la légende de la figure II.31. Comme attendu, l'ITD perçu varie avec l'azimut et comme pour les valeurs données par les différents calculs, l'ITD est plus importante sur les côtés que sur le plan médian. Cette observation est confirmée par une analyse de variance ANOVA qui montre comme seul effet, l'effet de l'azimut ($F(11, 120) = 511.6, p < 0.001$). Les variations de la dispersion des réponses sont cohérentes avec, à la fois l'évolution de la précision de localisation en champ libre [Blauert (1983); Mills (1958)] et les variations de la JND de l'ITD avec l'azimut [Klump and Eady (1956); Hershkowitz and Durlach (1969)]. Pour chaque position, l'ITD est donnée avec une dispersion comparable aux différences inter-individuelles. La figure II.31 ne fait apparaître que trois points aberrants. Ces observations confirment une fois de plus que le modèle $\{ITD \bigoplus HRIR_{min}\}$ est valide et que la précision de localisation associé à ce modèle est comparable à la précision en écoute champ libre.

Les variations perçues de l'ITD avec l'azimut comportent plusieurs asymétries (par rapport à un modèle de tête sphérique avec oreilles centrées). Premièrement, une dissymétrie gauche-droite est observée. Par exemple, la réponse moyenne à 90° et de -657μs alors qu'elle est égale 683 μs pour $\theta = 270°$. Cette observation n'est toutefois pas signi-

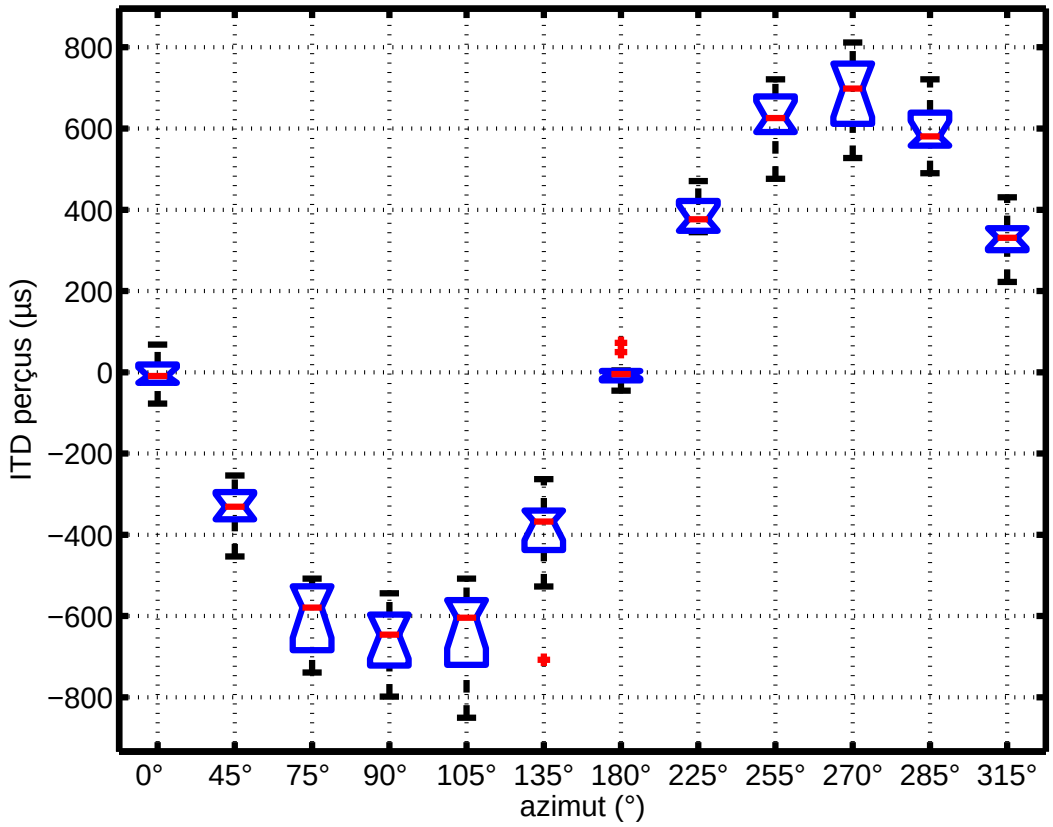


FIG. II.31 – ITD perçue en μs en fonction de l'azimut du son cible. Les parties inférieures et supérieures des boîtes bleues sont les 1^{er} et 3^{ime} quartiles. La ligne rouge au milieu de la boîte représente la valeur médiane. Les valeurs indiquées par les barres horizontales de part et d'autre de la boîte représentent une mesure de dispersion donnée par 150 % de la distance inter-quartiles. Les points rouges montrent la présence de données isolées.

ficative (au sens d'un test de Tuckey). Cependant, ce décalage est aussi observé sur les diagrammes polaires de l'ITD normalisée pour la comparaison des méthodes d'estimation (cf. fig. II.12) ce qui laisse penser à un décalage des mesures vers la droite et donc cette dissymétrie serait indépendante de la perception. Par contre, la dissymétrie par rapport à l'axe interaural semble être un élément à reproduire. La valeur moyenne à 75° est -608 μs et celle à 105°, la position symétrique de 75° par rapport à l'axe interaural, est -647 μs. Cette dissymétrie est reproduite pour 255° et 285°. Un test de Tuckey montre que pour les résultats du test, ces dissymétries ne sont pas significatives. Mais l'étude menée au paragraphe 1.4.1 a montré que ces dissymétries existent et sont significatives pour l'ITD issue de mesures. La figure II.12 fait aussi apparaître de tels comportement spatiaux pour certains estimateurs. La méthode ITD_{phase} indique même des extremum atteints en 105° et 255°. Par contre l'ITD_{seuil}, comme l'ITD du modèle sphérique, ne présente pas de dissymétrie. La présence de ce type de variations spatiales peut ainsi devenir un critère de comparaison entre les différentes méthodes de calculs de l'ITD pour autant

que l'asymétrie soit vérifiée.

3.6 Comparaison des estimateurs sur le plan horizontal

Les valeurs moyennes des réponses des sujets ainsi que les moyennes des différents types de méthodes de calcul décrites auparavant sont reportées sur les figures II.32 A), B), C), et D). Les méthodes semblent toutes reproduire correctement les valeurs des réponses des sujets. Seul le modèle ITD_{sphre} avec rayon individualisé sous-estime toutes les valeurs sauf pour les positions $\theta = 90°$ et $\theta = 270°$, au contraire de l'ITD_{sphere} avec le rayon moyen qui sous-estime l'ITD perçue pour les valeurs latérales. Cette observation est cohérente avec [Miller (2001); Minaar et al. (2000)] qui montrent aussi une sous-estimation de l'ITD aux positions latérales. Il convient alors de rappeler que la formule reliant le rayon optimal et trois paramètres morphologiques est obtenue pour la base de données CIPIC. L'étude menée au paragraphe 2.3 a montré que cette formule dépend de la base de HRTF utilisée. Il se peut alors que l'erreur observée soit due à une mauvaise correspondance entre le rayon optimal obtenu pour les sujets de la base CIPIC et le rayon optimal pour les sujets qui ont passé le test. De plus ce rayon est optimal pour toutes les positions et pas seulement pour le plan horizontal.

Les autres méthodes montrent des résultats très proches des réponses des sujets et seul leur comportement pour les positions latérales semblent pouvoir les départager. Dans l'analyse des résultats (cf. §3.5), il est mentionné une dissymétrie par rapport à l'axe interaural. Cette dissymétrie est aussi présente pour les valeurs données par $ITD_{MaxIACC}$ et ITD_{phase}, c'est-à-dire que les valeurs à 75°, respectivement 285°, sont inférieures aux valeurs à 105°, respectivement 255°. ITD_{phase} exhibe même des extrema délocalisés (105° et 255° au lieu de 90° et 270°). Cependant, si ce comportement est proche des réponses des sujets l'erreur introduite semble trop importante pour être perceptivement correcte.

Le calcul de l'erreur absolue permet de mieux appréhender la teneur des performances des estimateurs. L'erreur est calculée selon la formule II.17, mais cette fois-ci $ITD(\theta, i)$ désigne l'ITD calculée par la méthode en question, et N est le nombre de sujets de la condition test (11). La figure II.33 représente $EC(\theta)$ pour les quatre types de méthodes de calcul en fonction de l'azimut. Cette figure montre clairement que ITD_{phase} et ITD_{sphere} avec rayon individualisé introduisent des erreurs importantes par rapport aux autres méthodes : ITD_{sphere} donne des erreurs réparties sur toutes les positions tandis que ITD_{phase} donne des erreurs faibles sauf pour les positions latérales qui peuvent atteindre 200 μs d'erreur absolue. Ce critère permet de distinguer les autres méthodes comme proches de la perception.

La comparaison des performances des méthodes de calcul est maintenant analysée avec trois autres critères. Le premier, appelé $Ec1$ est simplement la moyenne de $EC(\theta)$ sur toutes les positions :

$$EC1 = \frac{1}{N}\frac{1}{M}\sum_{\theta_j=1}^{\theta_M}\sum_{i=1}^{N}|ITD(\theta_j, i) - \widehat{ITD}(\theta_j, i)| \tag{II.18}$$

avec M le nombre de positions (12). Le second critère permet la comparaison de la répartition empirique des erreurs. Pour chaque $EC(\theta)$, la fonction de répartition empirique est déterminée et le seuil correspondant à 75 % de la répartition empirique des erreurs

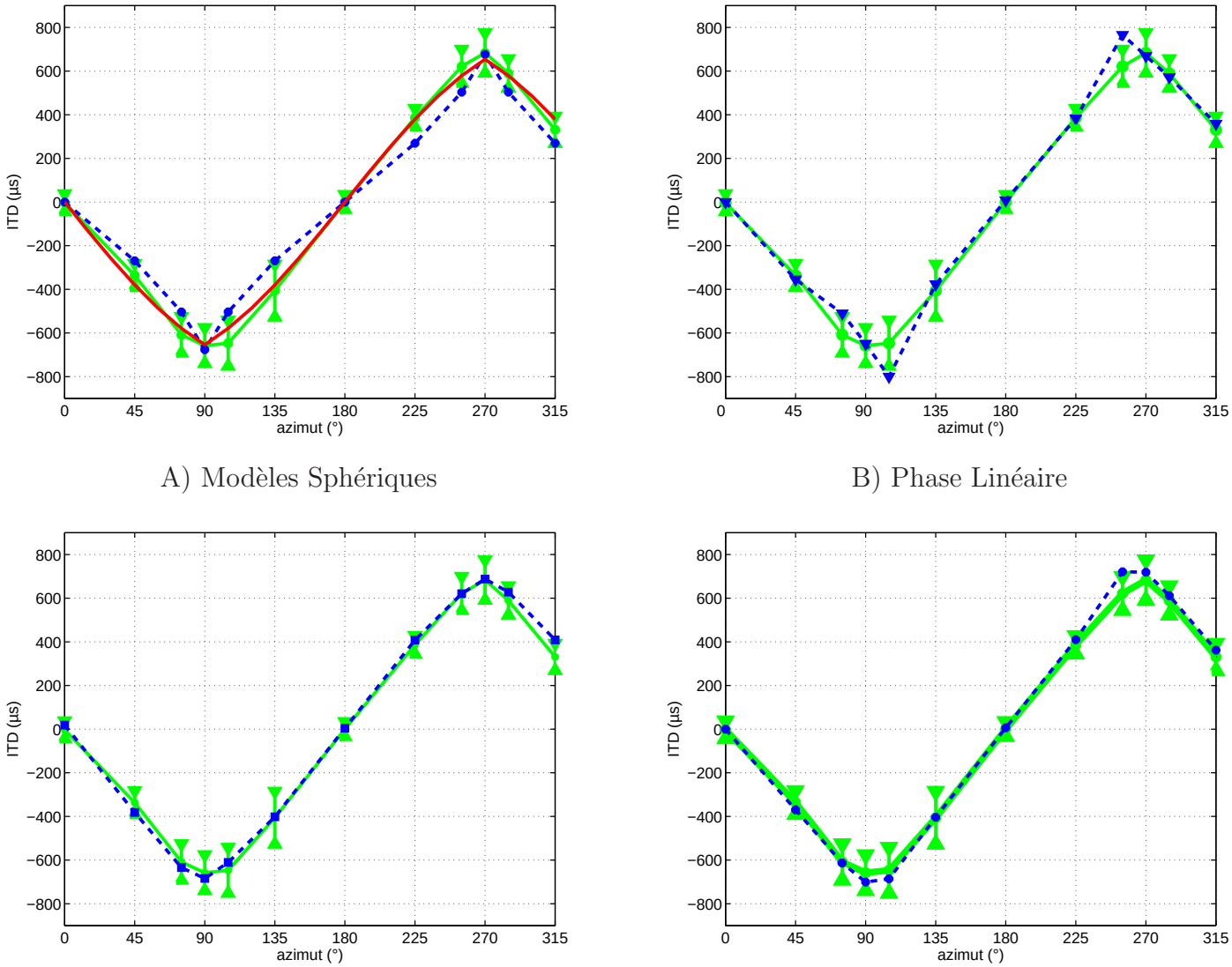


FIG. II.32 – Comparaison entre ITD perçue (vert) et méthodes de calcul (bleu et rouge) en fonction de l'azimut. Les barres verticales indiquent l'écart-type des réponses. A) Modèles Sphériques ; en rouge modèle de Woodworth avec a = 875 mm et en bleu moyenne des ITD individualisés selon la formule d'Algazi (cf. § 1.2), B) Régression linéaire de la phase de l'excès de phase sur [1000 - 5000] Hz, C) Estimation du Seuil des HRIR à 50% de leur maximum et D) Maximum de la fonction de corrélation entre les enveloppes des HRIR droites et gauches.

est retenu. Ce critère est nommée $EC2$. Le troisième et dernier critère, le plus restrictif, fait le comptage des valeurs calculées qui sortent de l'intervalle de dispersion définit par $[ITD(\theta) - S, ITD(\theta) + S]$, où $ITD(\theta)$ représente les réponses moyennes individuelles en fonction de l'azimut et où S représente l'écart-type associé à $ITD(\theta)$. Cet intervalle

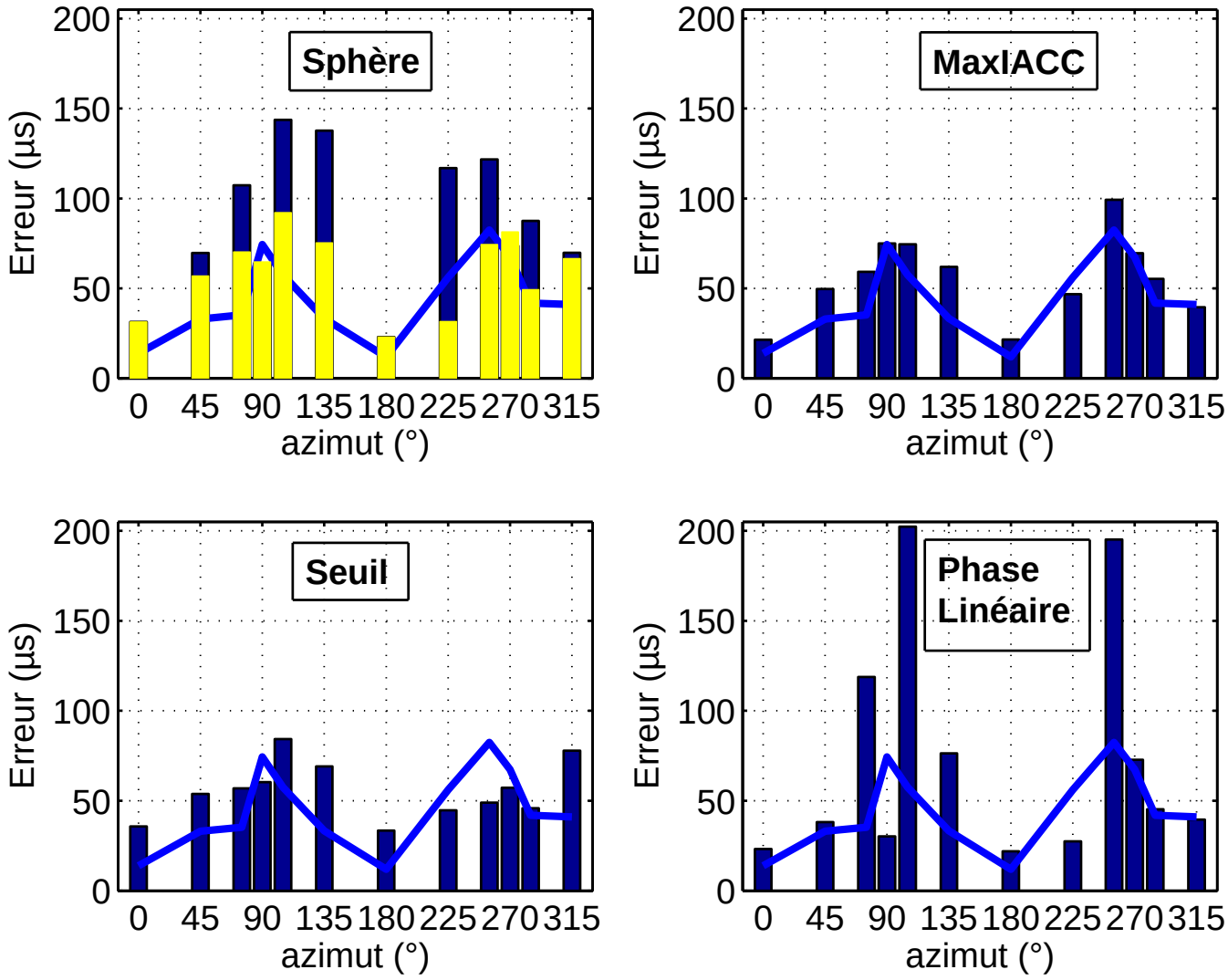


FIG. II.33 – Erreur absolue moyenne entre ITD perçue et ITD calculée en fonction de l'azimut. En haut à gauche : modèles sphériques, en jaune modèle de Woodworth avec a = 87,5 mm et en bleu ITD individualisée selon la formule d'Algazi (cf. § 1.2) ; en haut à droite : maximum de la fonction de corrélation entre enveloppe des HRIR droites et gauches ; en bas à gauche : estimation du seuil des HRIR à 50% de leur maximum ; en bas à droite : régression linéaire de la phase de l'excès de phase sur [1000 - 5000] Hz.

englobe alors 68 % des réponses[11]. Ce dernier critère est appelé $EC3$ et correspond au pourcentage par rapport au nombre de réponses moyennées, c'est-à-dire 252 (11 sujets* 12 positions).

Tab. II.7 – Critères d'erreurs et méthodes de calcul de l'ITD.

	Sphère	Sphère individualisée	Phase linéaire	Seuil	MaxIACC
EC1 (μs)	60	87	75	55	56
EC2 (μs)	89	118	91	77	79
EC3 (%)	19	28,2	16,7	19,4	17,5

Le tableau II.7 fait ressortir ITD_{seuil} et $ITD_{MacIACC}$ comme étant les deux méthodes produisant les erreurs moyennes les plus faibles avec 55 μs et 56 μs respectivement. Ces méthodes présentent aussi des valeurs faibles pour $EC2$. Par contre, du point de vue individuel, ITD_{phase} semble la méthode de calcul la plus pertinente. Seulement, $EC1$ et $EC2$ sont assez importants pour cette méthode ce qui indique que la méthode introduit moins d'erreurs individuelles mais que ces erreurs sont plus importantes par rapport aux autres méthodes. Le tableau II.7 montre aussi que ITD_{sphere} avec un rayon moyen offre de bonnes performances, surtout en considérant qu'il s'agit d'une formule analytique. Par contre, ITD_{sphere} avec rayon individualisé selon la formule II.2 montre que les poids calculés avec la base CIPIC ne réduisent pas l'erreur de l'ITD sphérique sur le plan horizontal.

3.7 Conclusion

Cette étude a permis de valider une nouvelle fois le modèle $\{ITD \bigoplus HRTF_{min}\}$ pour l'implémentation de la synthèse binaurale. Malgré les différences analytiques et perceptives qui peuvent apparaître pour les positions latérales, les sujets ont globalement réussi à faire correspondre un stimulus avec la modélisation et un stimuli avec des HRTF à phase mixte. L'analyse des résultats montre que la difficulté de la tâche, associée à la dispersion des réponses, est plus importante pour les positions proches de l'axe interaural ce qui est cohérent avec les résultats en écoute champ libre [Blauert (1983); Mills (1958)] et en écoute au casque stéréophonique [Klump and Eady (1956); Domnitz and Colburn (1977)]. En effet, ces études montrent que la résolution angulaire ou les JND de l'ITD augmentent avec l'azimut. La comparaison des méthodes de calcul de l'ITD a permis de mettre en avant les méthodes ITD_{seuil} et $ITD_{MacIACC}$ comme étant les plus proches de la perception. L'ITD de la formule de Woodworth (cf. équation II.1) offre cependant un compromis intéressant entre facilité d'usage et performance. Enfin cette étude permet de valider un protocole expérimental pour la détermination de l'ITD. Des améliorations peuvent être apportées à ce protocole, notamment en modifiant les paramètres de la procédure adaptative utilisée (variation du pas, nombre de retournement pour la convergence).

[11] Le pourcentage donné correspond à une distribution normale des réponses, qui est vérifié mais non représenté ici.

4 ESTIMATION DE LA JND DE L'ITD SUR DES CÔNES CONFUSION

L'ITD varie principalement avec l'angle d'azimut de la source sonore et, dans une moindre mesure, avec l'angle d'élévation (cf. tableau II.1). L'étude présentée dans la partie précédente du chapitre a montré que le modèle de tête sphérique offre des performances presque similaires, dans le plan horizontal, comparées aux performances des techniques d'estimation de l'ITD nécessitant des mesures de HRTF/HRIR. L'utilisation de la formule FDO permet de plus, et contrairement au modèle classique de tête sphérique, d'obtenir des variations de l'ITD avec l'élévation proche de celles observées à partir des mesures. Seulement, la question de l'audibilité des variations de l'ITD en élévation est une question ouverte. De plus, ces variations semblent individuelles (cf. fig. II.9) et la capacité de la formule FDO à reproduire ces variations est alors aussi posée.

L'étude expérimentale présentée dans ce paragraphe a pour but de déterminer si les variations de l'ITD le long d'un cône de confusion sont audibles. Les résultats permettront de dégager des tendances pour l'optimisation de la précision nécessaire des systèmes de spatialisation sonore au regard de la résolution auditive. L'objectif est l'évaluation des JND de l'ITD pour les cônes de confusion. Cette évaluation se fera dans le cadre d'une écoute séquentielle des stimuli ce qui est pertinent pour la simulation de trajectoire sonore.

Les sujets doivent détecter des variations d'ITD pour des positions de sources simulées par la synthèse binaurale décrivant des cônes de confusion. Pour chaque position, l'ITD variable (ITD_{var}) est comparée par rapport à une ITD de référence (ITD_{ref}) qui est définie comme l'ITD calculée à partir des HRTF mesurées sur les sujets selon la méthode décrite dans [Algazi et al. (2001b)]. La JND de l'ITD est estimée pour différents cônes de confusion définis par leur angle d'azimut à $\phi = 0°$.

L'analyse des résultats de cette expérience donnera des éléments de réponse aux questions suivantes :

Premièrement : Les variations de l'ITD pour des sources situées le long de cônes de confusion sont-elles audibles ?

Deuxièmement : Si ces variations sont audibles, c'est-à-dire si elles sont supérieures aux JND, est-ce que le caractère individuel de ces variations doit-être reproduit ? En d'autres termes, est-ce qu'une ITD qui présente des variations moyennes est acceptable du point de vue de la JND ?

Troisièmement : Est-ce que la formule de Woodworth individualisée ou la formule FDO permet une reproduction satisfaisante des variations de l'ITD et éventuellement des variations individuelles ?

4.1 Protocoles expérimentaux

Les études préliminaires (cf. §1.5.2) ont montré que le choix d'une procédure psychophysique est un paramètre expérimental. Ces études sont axées sur l'estimation de la JND de l'ITD en l'absence d'indices de localisation supplémentaires comme l'ILD. Les deux procédures testées ici sont respectivement celles qui donnent les résultats les plus faibles (méthode des constantes, 2-intervalles 2 ACF), protocole 2, et celle qui donne les seuils les plus importants (méthode adaptative modifiée 2down-1up, 3-intervalles 3 ACF), protocole 1 (cf. § 1.5).

4.1.1 Stimuli

Chaque stimulus est créé en effectuant la convolution d'un bruit blanc gaussien d'une durée de 400 ms avec une attaque et une fin en cos^2, c'est-à-dire le même bruit que celui utilisé dans les études préliminaires (cf. § 1.5.2), avec la $HRIR_{min}$ retardée correspondant à la position testée. A chaque présentation, les deux stimuli différents sont le stimulus retardé par ITD_{ref} et le stimulus retardé par $ITD_{ref} + \Delta ITD$. Le pas de variation du paramètre ΔITD est obtenu en suréchantillonnant les HRIR à f_e =96000 Hz, ce qui donne un pas de 10,41 μs, c'est-à-dire un pas correspondant au JND de la littérature. Le suréchantillonnage est préféré aux techniques de délais fractionnaires qui introduisent des artefacts audibles (filtrage passe-bas)[Laakso et al. (1996)].

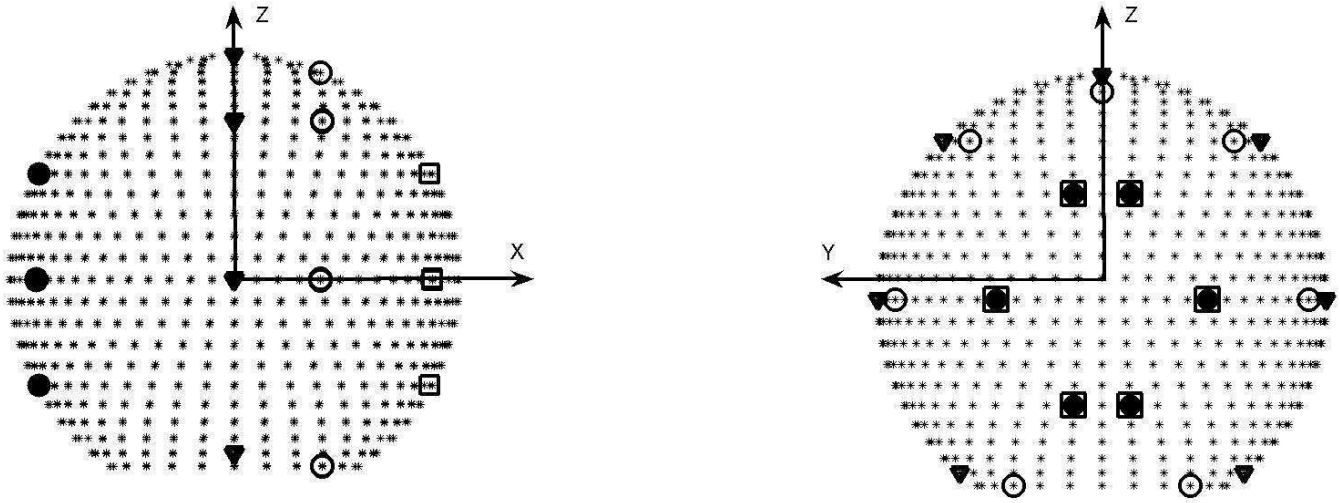
4.1.2 Positions simulées

Le JND de l'ITD est évalué en fonction de l'angle d'azimut et d'élévation de la position simulée. La répartition empirique spatiale en azimut, dans un système de coordonnées polaire-vertical, des points de mesures de la base de HRTF FTR&D est constante pour une élévation donnée mais varie avec l'élévation (cf. § 1.3.3). Cet échantillonnage spatial n'est pas adapté pour la définition de cônes de confusion qui sont des plans d'azimut constant dans un système de coordonnées polaires-interaurales. C'est pourquoi, les positions sélectionnées pour cette expérience ne décrivent pas parfaitement des cônes de confusion, mais des trajectoires qui peuvent s'en écarter de quelques degrés. La création de trajectoires qui approchent des cônes de confusion est préférée à l'utilisation de technique d'interpolation susceptible d'introduire des artefacts audibles. L'erreur introduite est de l'ordre de grandeur des erreurs causées par un mouvement de la tête du sujet pendant une session de mesures (cf. § 1.2). De plus, comme la tâche du sujet est d'effectuer une comparaison de stimuli ne différant que par leur ITD, ces erreurs ne sont pas pénalisantes.

Les points de mesure de la base de HRTF FTR&D ainsi que les points sélectionnés pour l'expérience sont représentés en figure II.34. Pour des raisons pratiques, le système de coordonnées polaires-interaurales est utilisé dans la suite de l'étude et les positions testées sont référencées par rapport à l'azimut du cône de confusion auquel elles appartiennent. Quatre cones de confusion qui correspondent aux angles d'azimut 0°, 22°, 61° et -61° sont testés. Les quatre cônes sont nommés respectivement cône 0°, cône 22°, cône 61° et cône -61°. La comparaison des résultats entre les cônes 61° et -61° donnera des éléments de réponse quant à la symétrie de la JND par rapport au plan médian. A cause de la répartition empirique spatiale des points de mesure, les cônes n'ont pas tous le même nombre de positions. Les 26 positions testées sont répertoriées dans le tableau II.8.

4.1.3 Procédures

Protocole 1 Il s'agit d'une méthode adaptative à choix forcé parmi trois propositions (3AFC) avec une règle d'évolution du paramètre de type 2down-1up et un retour visuel sur les réponses. Chaque séquence sonore est composée de trois stimuli : deux sont identiques, un est différent. La tâche du sujet est d'identifier le stimulus différent. L'essai est stoppé quand le parcours a effectué un sixième retournement, ou quand le sujet donne deux bonnes réponses consécutives pour la valeur minimale du seuil ΔITD_{min}, c'est-à-



Vue de face Vue de profil

FIG. II.34 – Positions des HRTF mesurées (*) et positions des points sélectionnés pour la définition de cônes de confusion. Vue de face à gauche et vue de profil à droite. Les positions testées pour le cône à 0° sont représentées par des triangles, celles pour le cône à 22° par des cercles, celles pour le cône à 61° par des carrés et celle pour le cône à -61° par des points.

TAB. II.8 – Angles d'azimut et d'élévation dans un système de coordonnées polaires-interaurales des positions testées.

Azimut du cône	en bas de-vant	niveau des yeux de-vant	en haut de-vant	zénith	en haut der-rière	niveau des yeux der-rière	en bas der-rière
0°	-50.5°	0°	45°	90°	135°	180°	230.5°
22°	-56°	0°	45°	67.5°	135°	180°	236°
61°	-28°	0°	28°		118°	180°	208°
-61°	-28°	0°	28°		118°	180°	208°

dire $10,41$ μs. Au début de chaque essai, ΔITD est égal à quatre ΔITD_{min} et décroît après chaque retournement par pas de ΔITD_{min}. L'estimation du seuil est réalisée avec la méthode *mid-run* et compte-tenu de la règle d'évolution choisie, le seuil correspond à 70,7 % de bonnes réponses [Levitt (1970)]. Le sujet reçoit un entraînement de quatre essais dont les seuils ne sont pas pris en compte par la suite. Le sujet peut effectuer des pauses quand il (elle) le désire. Chaque position est testée trois fois et est présentée dans

un ordre aléatoire. De plus, comme il a été reporté [Domnitz and Colburn (1977)] que la JND décroît lorsque l'ITD est en compétition avec l'ILD[12], deux sens de variation, correspondant à un $\Delta ITD > 0$ et un $\Delta ITD < 0$, sont présentés au sujet. Au total, un sujet réalise 156 essais (26 positions * 2 signes de ΔITD * 3 répétitions).

Protocole 2 C'est une méthode à paramètres constants à choix forcé parmi deux (2AFC) avec retour visuel sur les réponses. Chaque séquence sonore est composée de deux stimuli : un avec ITD_{ref} et l'autre avec ITD_{var}. Un ensemble de 10 valeurs de ΔITD est utilisé : de $\Delta ITD = 0$, pour une mesure du niveau de chance, à $\Delta ITD = 9\Delta ITD_{min}$. La tâche du sujet est d'indiquer lequel des deux stimuli est perçu le plus à gauche. Chaque paire de stimuli est présentée trente fois et l'ordre de présentation et de combinaison des couples ITD_{ref}/ITD_{var} est aléatoire. Le seuil est le plus petit ΔITD qui atteint 75% de bonnes réponses. Selon le test du χ^2, pour trente répétitions, le niveau de 75% est significativement différent du niveau de chance (50%). Pour ce protocole, seuls les cônes à 0°, 22° et 61° sont testés, l'hypothèse de symétrie des performances est testée avec le protocole 1. Chaque sujet réalise 6000 essais (20 positions * 10 ΔITD * 30 répétitions), répartis en deux sessions.

4.1.4 Sujets

Cinq sujets, deux hommes et trois femmes, ont participé à l'expérience gouvernée par le protocole 1. Leur courbe d'audition a été vérifiée avant l'expérience. Les sujets 1 à 4 écoutent leurs propres HRTF et le sujet 5 écoute avec les HRTF d'une autre personne. Seuls les sujets 4 et 5 ont participé à l'expérience utilisant le protocole 2.

4.2 Résultats

Le but premier de l'étude menée ici est d'estimer la JND de l'ITD pour des positions évoluant en azimut et en élévation dans le contexte d'une implémentation $\{ITD \bigoplus HRTF_{min}\}$ de la synthèse binaurale. Trois facteurs expérimentaux sont étudiés : l'azimut du cône de confusion, l'élévation le long d'un cône de confusion et le sens de variation de l'ITD. Les résultats du protocole 1 sont d'abords présentés et ensuite comparés aux résultats du protocole 2.

4.2.1 Protocole 1

Une analyse de variance (ANOVA) est réalisée sur les JND obtenues pour chaque sujet en considérant trois facteurs expérimentaux : l'azimut du cône, l'élévation le long d'un cône et le signe de ΔITD. Elle montre que le seul effet significatif sur la JND est l'azimut du cône ($F(3,192) = 4.44$, $p < 0.05$) et que ni l'élévation sur un cône de confusion ($F(5,192) = 0.51$, $p = 0.76$), ni le signe de ΔITD ($F(1,192) = 0.01$, $p = 0.93$) n'ont un effet significatif. L'indépendance de la JND avec l'élévation est cohérente avec les observations de [Oldfield and Parker (1984)] qui montrent, dans une expérience en champ libre, que la précision de localisation en azimut, qui peut être reliée avec la JND de l'ITD,

[12]Par exemple, pour une positions située à $\theta = 61°$, un ΔITD négatif déplace le son vers le centre ce qui est contradictoire avec l'ILD contenu dans les HRTF et qui indiquent une position située dans l'hémisphère droit, tandis qu'un ΔITD positif renforce la latéralisation de la position simulée.

ne dépend quasiment pas de l'élévation. Le fait que la JND ne dépende pas du signe de ΔITD semble contradictoire avec les observations issues de la littérature [Moushegian and Jeffress (1959); Domnitz and Colburn (1977); Blauert (1983)] qui indiquent que la JND de l'ITD augmente quand ITD et ILD sont en opposition, c'est-à-dire par exemple quand l'ITD indique une position située sur la droite et l'ILD une position située sur la gauche. Cette différence avec les travaux présentés ici et les études antérieures peut s'expliquer par le fait que dans la présente expérience ITD et ILD peuvent être en léger désaccord, de quelques degrés, mais qu'elles indiquent rarement des positions perçues dans deux régions très différentes comme c'est le cas dans les travaux antérieurs.

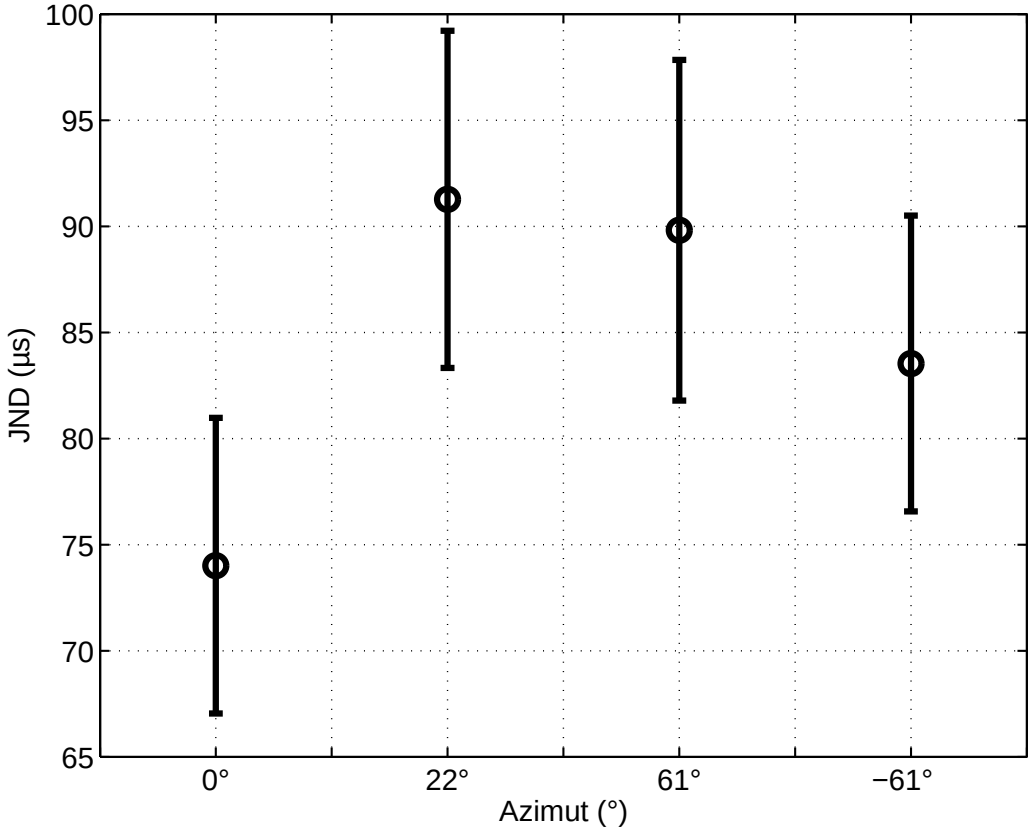


Fig. II.35 – JND moyenne issue du protocole 1 en fonction de l'azimut du cône de confusion. Les barres verticales indiquent l'intervalle de confiance à 95%.

Les variations de la JND avec le cône de confusion sont reportées en figure II.35. Les valeurs affichées correspondent aux moyennes des JND réalisées en regroupant les variations avec l'élévation et les variations avec le signe de ΔITD puisque l'ANOVA montre que ces variations n'ont pas d'effet significatif sur la JND. Les barres verticales qui traversent chaque point représentent l'intervalle de confiance à 95%. Cet intervalle est quasiment le même d'un cône à l'autre et varie entre $\pm 7\mu$s et $\pm 8\mu$s. Ces valeurs sont faibles comparées à la précision du test (10,41 μs) ce qui indique que la tâche est correctement réalisée par les sujets. Les valeurs les plus faibles sont issues du cône 0°, ce

qui est cohérent avec les études antérieures, avec une moyenne de 74 μs. Les JND des autres cônes de confusion ne sont pas significativement différentes les unes des autres, au sens d'un test HSD de Tuckey, ce qui indique notamment que la JND semble symétrique par rapport au plan médian, car les JND moyennes des cônes 61° et -61° ne sont pas significativement différents. Les valeurs des JND et leurs intervalles de confiance à 95% pour chaque sujet sont reportés dans le tableau II.9. Globalement, les JND obtenues avec le protocole 1 sont largement supérieurs à ceux des études antérieurs.

Tab. II.9 – JND1 moyenne et intervalle de confiance à 95% pour les 5 sujets et pour les 4 cônes de confusion.

Cône \ Sujet		1	2	3	4	5	Moyenne
0°	Moyenne	62	71	114	55	67	74
	Intervalle de confiance	± 8.6	± 17.4	± 22.1	± 8.1	± 10.1	± 7
22°	Moyenne	88	81	138	73	76	91
	Intervalle de confiance	± 10.1	± 22.5	± 22.2	± 9.8	± 11.3	± 7.9
61°	Moyenne	83	89	103	79	95	90
	Intervalle de confiance	± 15.9	± 18.7	± 19.9	± 13.4	± 20.7	± 8
-61°	Moyenne	76	83	118	63	77	84
	Intervalle de confiance	± 11.8	± 20.8	± 14.5	± 7.7	± 14.4	± 7

Les résultats entre les différents sujets sont affichés sur la figure II.36 qui représente les valeurs du tableau II.9. Des différences importantes existent entre les sujets, observation déjà reportée précédemment dans des expériences similaires [Domnitz (1973); Best et al. (2004)]. Ces différences sont principalement dues au sujet 3 qui contrairement aux autres sujets, qui obtiennent des JND comprises entre 60 μs et 100 μs, reporte des JND toujours supérieures à 100 μs. Cette différence peut s'expliquer par le niveau d'expertise des sujets. En effet les sujets 1, 2, 4 et 5 sont experts en tests d'écoute. De surcroît le sujets 1, 4 et 5 sont experts en son 3D. Une autre remarque concernant les sujets, est que le sujet 5 qui ne bénéficiait pas d'une synthèse binaurale individualisée, obtient des JND semblables aux sujets 1 et 4.

4.2.2 Protocole 2

Les résultats des deux protocoles sont présentés en figure II.37 en fonction de l'azimut du cône, c'est-à-dire 0°, 22° et 61°, et pour les deux sujets ayant participé aux deux protocoles (sujet 4 et 5 de l'expérience précédente). La figure II.38 représente quant à

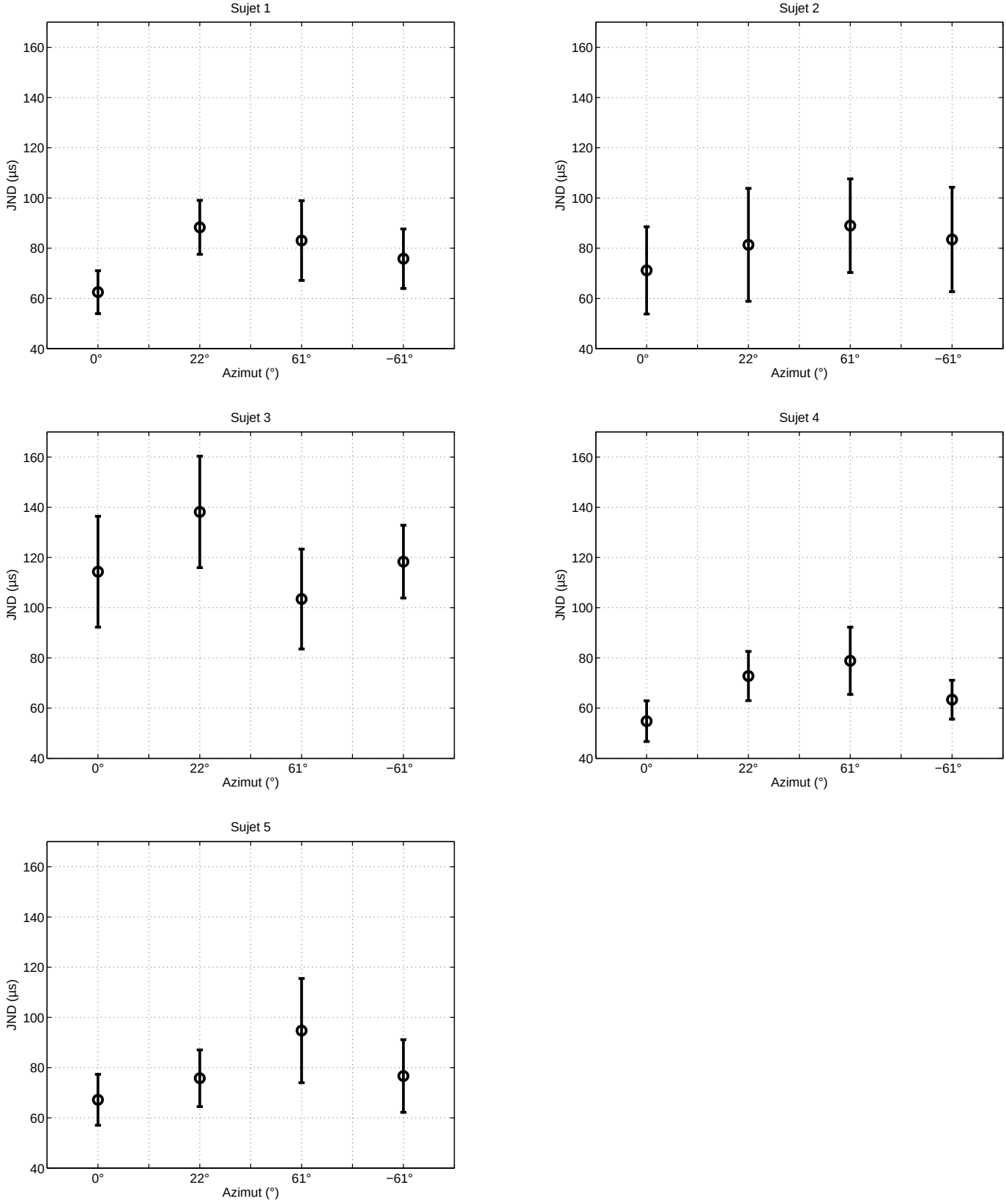


FIG. II.36 – JND moyenne issue du protocole 1 en fonction de l'azimut du cône de confusion pour les 5 sujets. Les barres verticales indiquent l'intervalle de confiance à 95%.

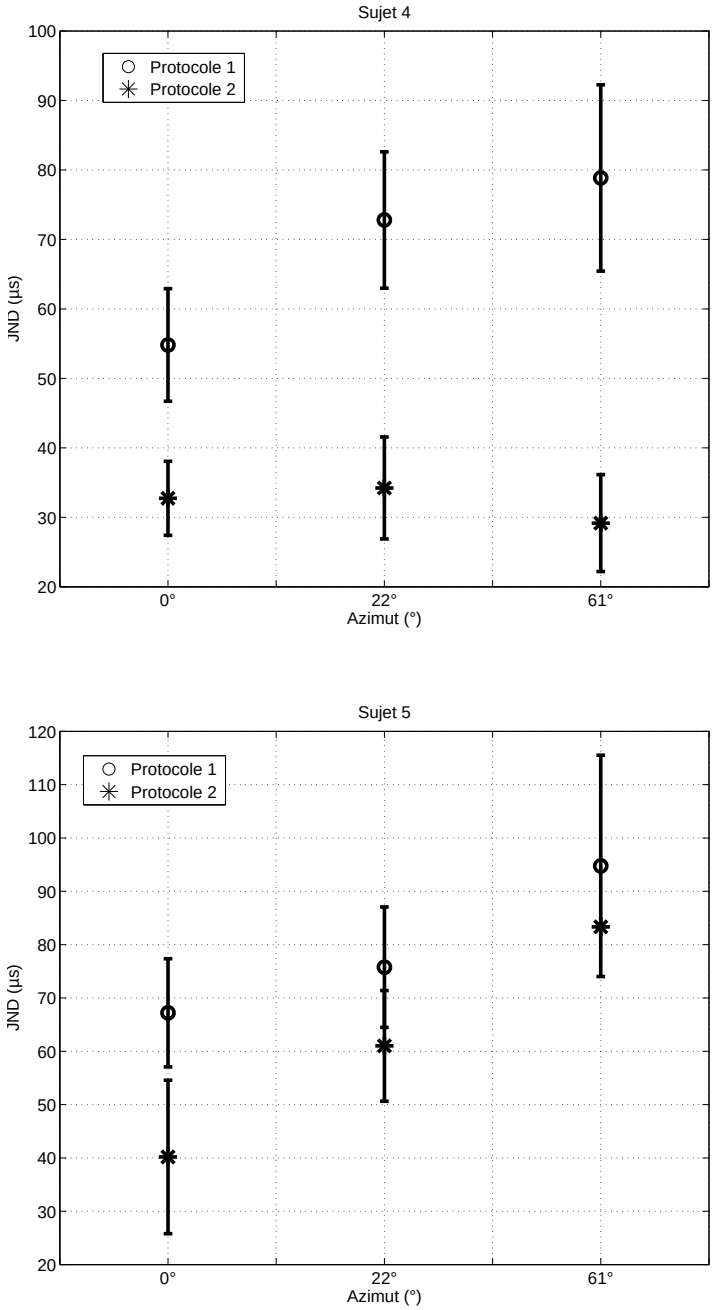


FIG. II.37 – Comparaison des JND moyennes issues des deux protocoles en fonction de l'azimut du cône de confusion pour les sujets 4 er 5. Les barres verticales indiquent l'intervalle de confiance à 95%. Les lignes reliées par des cercles représentent les valeurs du protocole 1 et celles reliées par des étoiles indiquent les valeurs du protocole 2.

elle les variations de la JND avec l'élévation pour les trois cônes de confusion et pour les deux protocoles. Comme indiqué par les expériences préliminaires (cf. § 1.5.2), les JND obtenues avec le protocole 2 sont globalement inférieures à ceux du protocole 1 et notamment, les JND du sujet 4 pour le protocole 2 sont deux fois plus bas. Pour le protocole 2, les JND des deux sujets montrent des tendances différentes. Les JND du sujet 5 suivent globalement la même variation avec l'azimut du cône, c'est-à-dire une augmentation de la JND avec l'azimut, tandis que les JND du sujet 4 restent constantes, au contraire de ses JND obtenues au protocole 1. Les JND du sujet 4 pour le protocole 2 varient entre 20 μs et 40 μs : ces valeurs sont cohérentes avec la littérature (cf. tableau II.2). Le fait que la JND ne varie pas avec l'azimut est cohérent avec les observations reportées dans [von Békésy (1960)], qui indiquent que la JND est constante pour un $ITD_{ref} < 600$ μs, mais contradictoire avec la plupart des autres études [Klump and Eady (1956); Mills (1958); Grantham et al. (2003)]. Les JND du sujet 5 augmentent fortement avec l'azimut, pour atteindre même les JND du protocole 1. De plus, pour certaines positions fortement latéralisées, les seuils obtenus ne sont pas significativement différents du niveau de chance. Pour le cône à 61°, le sujet 5 n'a obtenu qu'un seul seuil significatif sur les 6 élévations du cône. Le fait que le sujet 5 utilisait les HRTF d'un autre personne peut expliquer cette observation. L'utilisation de HRTF non-individuelles entraîne des artefacts audibles comme une perception intra-crânienne et une augmentation des confusions avant/arrière. Qui plus est, il est reporté dans des études précédentes des difficultés pour recueillir des JND ou des MAA pour les positions fortement latéralisées [Mills (1958); Grantham et al. (2003)]. Lors d'une perception intra-crânienne, ou lors d'une perte d'externalisation des sources, le sujet peut alors avoir de grande difficulté à percevoir les positions relatives des stimuli car la perception de leur espacement est fortement réduit. Cette difficulté n'a apparemment pas été relevée pour le protocole 1. Dans la partie consacrée à l'analyse des résultats, seules sont considérées les JND du sujet 4 car elles sont les plus restrictives pour la comparaison des modélisations de l'ITD.

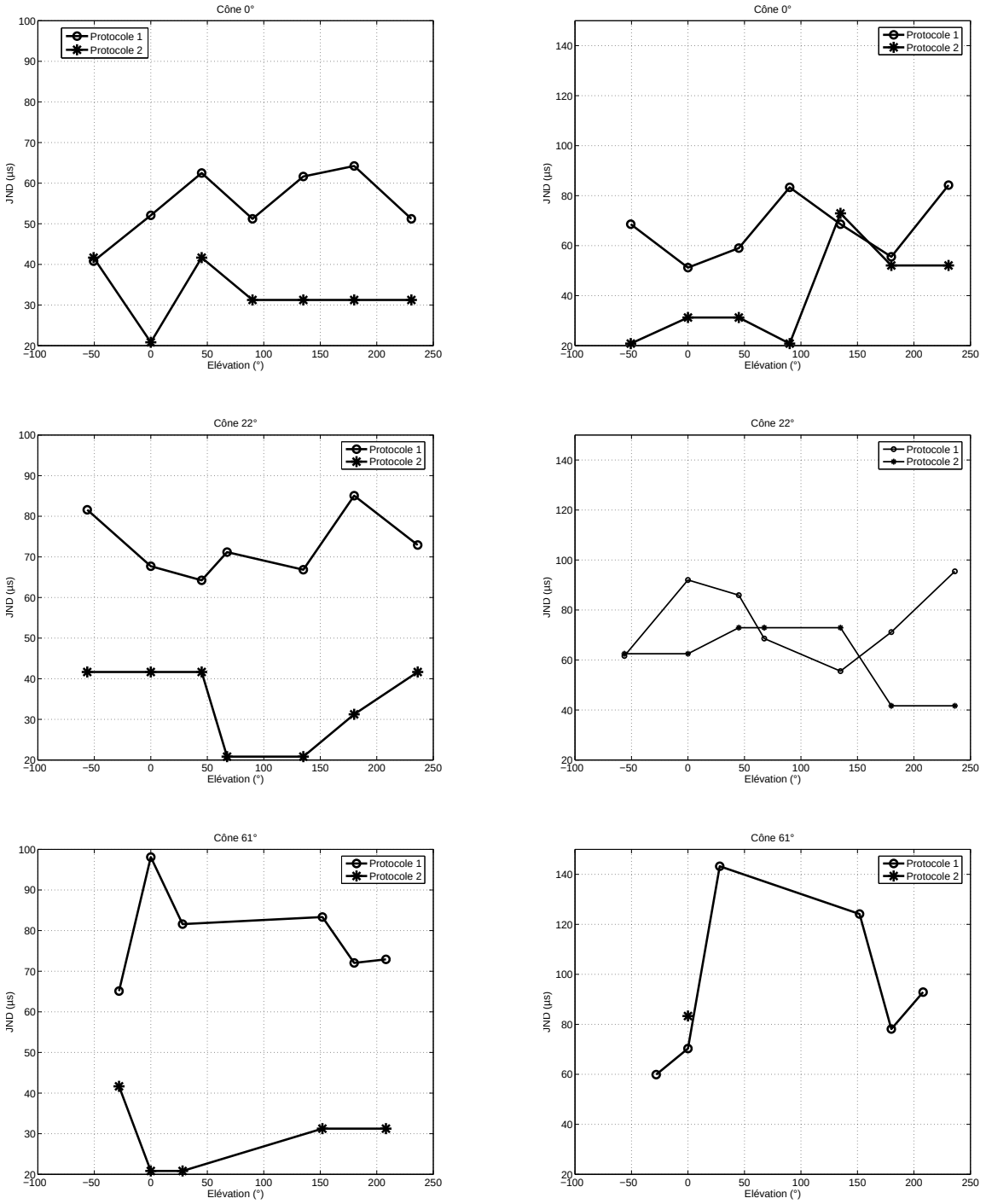
4.3 Analyse

La comparaison des protocoles utilisés conduit à la définition de deux JND. La première partie de l'analyse compare les deux protocoles et donne des éléments de réponse quant à leur signification et leur utilisation.

Les valeurs de la JND définissent une *zone de flou* de l'ITD, c'est-à-dire un ensemble de valeur d'ITD qui ne sont pas discriminées par un individu. La mesure de cette *zone de flou* permet de déterminer l'audibilité des variations de l'ITD sur les cônes de confusion dans le but de connaître le degré de précision nécessaire aux moteurs de rendu sonore spatialisé. La question de l'audibilité des variations de l'ITD avec l'élévation est abordée dans la deuxième partie de l'analyse.

Les variations de l'ITD avec l'élévation dépendent de l'individu. Grâce à la mesure de la *zone de flou*, il peut être déterminé si une modélisation de l'ITD avec des variations moyennes en élévation est satisfaisante. Dans la troisième partie de l'analyse, l'individualisation des variations de l'ITD sur les cônes de confusion est examinée.

Les variations fines de l'ITD qui ne sont pas perçues, au sens de la JND, sont perceptivement inutiles et peuvent être supprimées ce qui entraîne une réduction de la complexité et du coût d'implémentation de la synthèse binaurale. Deux méthodes de calcul de l'ITD



a) Sujet 4 b) Sujet 5

FIG. II.38 – JND moyennes issues des deux protocoles pour chaque cône de confusion pour le sujet 4, colonne de gauche, et pour le sujet 5, colonne de droite, en fonction de l'élévation. Les lignes reliées par des cercles représentent les valeurs du protocole 1 et celles reliées par des étoiles indiquent les valeurs du protocole 2.

sont examinées pour connaître leur capacité à reproduire les variations perçues :

- **le modèle de tête sphérique avec rayon individualisé** [Algazi et al. (2001b)]
- **le modèle FDO** [Busson et al. (2004)].

L'avantage principal de ces méthodes est qu'elles permettent une estimation de l'ITD d'un individu sans avoir à mesurer ses HRTF. La validité perceptive de ces deux méthodes est étudiée dans la quatrième partie de l'analyse.

4.3.1 Comparaison des deux protocoles

L'étude de comparaison des protocoles (cf. §1.5.2) met en évidence des différences importantes. Ce paragraphe reprend en partie l'analyse commencée et rajoute des éléments de discussion à la lumière des JND.

Les avantages des méthodes adaptatives par rapport aux méthodes des constantes sont nombreux [Trahiotis et al. (1990)]. Notamment, il n'est pas nécessaire de connaître à l'avance avec précision l'intervalle des valeurs du paramètre à balayer car le protocole s'adapte automatiquement à la perception du sujet. La méthode d'estimation du seuil avec la méthode des contantes est basée sur la construction de la courbe psychométrique du paramètre en question. La construction de la courbe psychométrique requiert un nombre important de réponses à collecter pour obtenir une bonne précision dans l'estimation du seuil, d'où un test long. La méthode d'estimation du seuil utilisée pour la procédure adaptative est au contraire à la fois rapide et fiable [Levitt (1970)]. Cette première analyse permet d'expliquer la différence de durée du test entre les deux protocoles : 2 heures pour la méthode adaptative et 7 heures pour la méthode des constantes.

Comme indiqué dans la comparaison des protocoles (cf. §1.5.2), les méthodes adaptatives ne nécessitent pas d'entraînement spécifique. Cependant, les études psychoacoustiques en synthèse binaurale sont souvent décrites avec un entraînement intensif des sujets qui écoutent plusieurs centaines, voir plusieurs milliers de stimuli avant de commencer l'expérience [Trahiotis et al. (1990)]. Dans la phase d'entraînement, le nombre de stimuli à faible niveau du paramètre doit être important pour que les sujets apprennent les indices nécessaires à la constitution d'une stratégie de réponse efficace. Dans les méthodes adaptatives, il se peut que le sujet n'entende pas assez de stimuli à faible niveau du paramètre pour établir sa stratégie de réponse, ce qui peut expliquer en partie la grande différence observée dans le mesure de la JND. De plus, une erreur d'inattention est plus pénalisante dans une procédure adaptative (cf. § 1.5.2).

La tâche du sujet est plus complexe lors d'une présentation des stimuli du type ABC (analyse des différences de 3 stimuli et identification, et donc mémoire et appariement, de deux stimuli complexes) que pour une présentation de type AB (détection d'un mouvement). De plus, en raison du nombre important de stimuli pour le protocole 2, un effet d'apprentissage est observé entre les deux sessions de 3000 stimuli chacune. Les valeurs de la JND pour la première session sont supérieures à ceux du deuxième bloc. Vu le nombre de stimuli, les valeurs de la JND obtenues avec le protocole 2 doivent conduire aux plus petites valeurs que l'on puisse mesurer. Le protocole 1 semble alors donner des valeurs proches d'une perception moyenne. En effet, les seuils issus du protocole 2 sont extraits à partir d'un parcours de réponses où le sujet entend différents niveaux du paramètre (cf. fig. II.13). Le protocole 1 semble alors suffisant pour des applications grand public ou des écoutes occasionnelles, telle une visite sonore de musée (avec spatialisation des commentaires en fonction de l'orientation relative du visiteur et des objets), alors que le

protocole 2 donnerait des valeurs intéressantes pour des données psychoacoustiques ou des applications spécifiques comme l'entraînement des pilotes d'avion de chasse ou le jeu.

4.3.2 Perception des variations de l'ITD

L'analyse de l'audibilité des variations de l'ITD avec l'élévation est effectuée sur les ITD de 43 sujets[13] de la base de données CIPIC (cf. chapitre III § 1.3.2). L'analyse est effectuée sur les cônes de confusion à 0°, 20° et 65°. Pour examiner l'audibilité des variations de l'ITD sur ces cônes de confusion, un paramètre adimensionnel est calculé. Ce paramètre permet la comparaison des variations de l'ITD des 43 sujets par l'utilisation d'un facteur de normalisation propre au sujet. Ce critère rend compte des variations individuelles de l'ITD en prenant l'ITD individuelle à l'élévation $\phi = 0°$ comme référence au sein d'un même cône de confusion. Il est défini par la formule suivante :

$$e_{var^i}(\theta) = \frac{ITD_i(\theta,\phi) - ITD_i(\theta,0)}{ITDmax_i} \tag{II.19}$$

où $ITD_i(\theta,\phi)$ représente l'ITD de l'individu i à l'azimut θ et à l'élévation ϕ, $ITDmax_i$ est l'ITD maximale de l'individu i. De la même manière, une valeur normalisée de la JND est calculée :

$$JND_{norm}(\theta) = \frac{JND(\theta)}{\frac{1}{M}\sum_{i=1}^{M} ITDmax_i} \tag{II.20}$$

où $JND(\theta)$ est la JND en fonction de l'azimut du cône et $ITDmax_i$ est l'ITD maximale pour le sujet i. Le critère de dispersion est calculé pour chaque individu et pour chaque élévation. 2150 valeurs du critère sont alors estimées pour chaque cône de confusion et les fonctions de répartitions empiriques des critères $e_{var^i}(\theta)$ sont évaluées pour les trois cônes. Les trois fonctions correspondant aux trois cônes de confusion sont représentées en figure II.39. Les valeurs adimensionnelles des deux JND y sont représentées. Ainsi, la probabilité d'audibilité, au sens d'une JND, est déterminée par $p_{JND_{norm}} = 1 - F(JND_{norm})$. En effet, $F(JND_{norm})$ donne la probabilité que $e_{var^i}(\theta)$ soient inférieurs ou égales à JND_{norm} ou en d'autre termes, que les variations individuelles soient non perçues au sens de l'équation II.9, et donc la probabilité que les variations individuelles soient perçues est donnée par $p(audible) = 1 - F(JND_{norm})$. La probabilité d'audibilité est reportée dans le tableau II.10 pour les trois cônes de confusion et pour les deux JND_{norm}. D'après la JND1, il est clair que les variations de l'ITD par rapport à $ITD_{\theta,0}$ ne sont pas audibles. Le pourcentage d'audibilité reste inférieur à 16 % quel que soit l'azimut du cône. Par contre, les pourcentages d'audibilité par rapport aux JND2 sont largement supérieurs et atteignent 73 % pour le cône à 65° ce qui suggère que les variations doivent être reproduites pour une synthèse binaurale fine.

Il semble que pour des applications grand public, les variations de l'ITD sur les cônes de confusions n'ont pas besoin d'être reproduites. Une ITD constante en élévation suffit, si la valeur qu'elle donne en azimut est correcte. Par contre, les variations de l'ITD en élévation sont perçues selon la JND2. Ces variations dépendent de l'individu. Au paragraphe suivant, les variations moyennes sont examinées pour savoir s'il faut qu'elles soient individualisées.

[13] La base de HRTF du CIPIC contient 45 sujets, mais deux sujets qui présentent des valeurs aberrantes d'ITD ont été supprimés de l'analyse

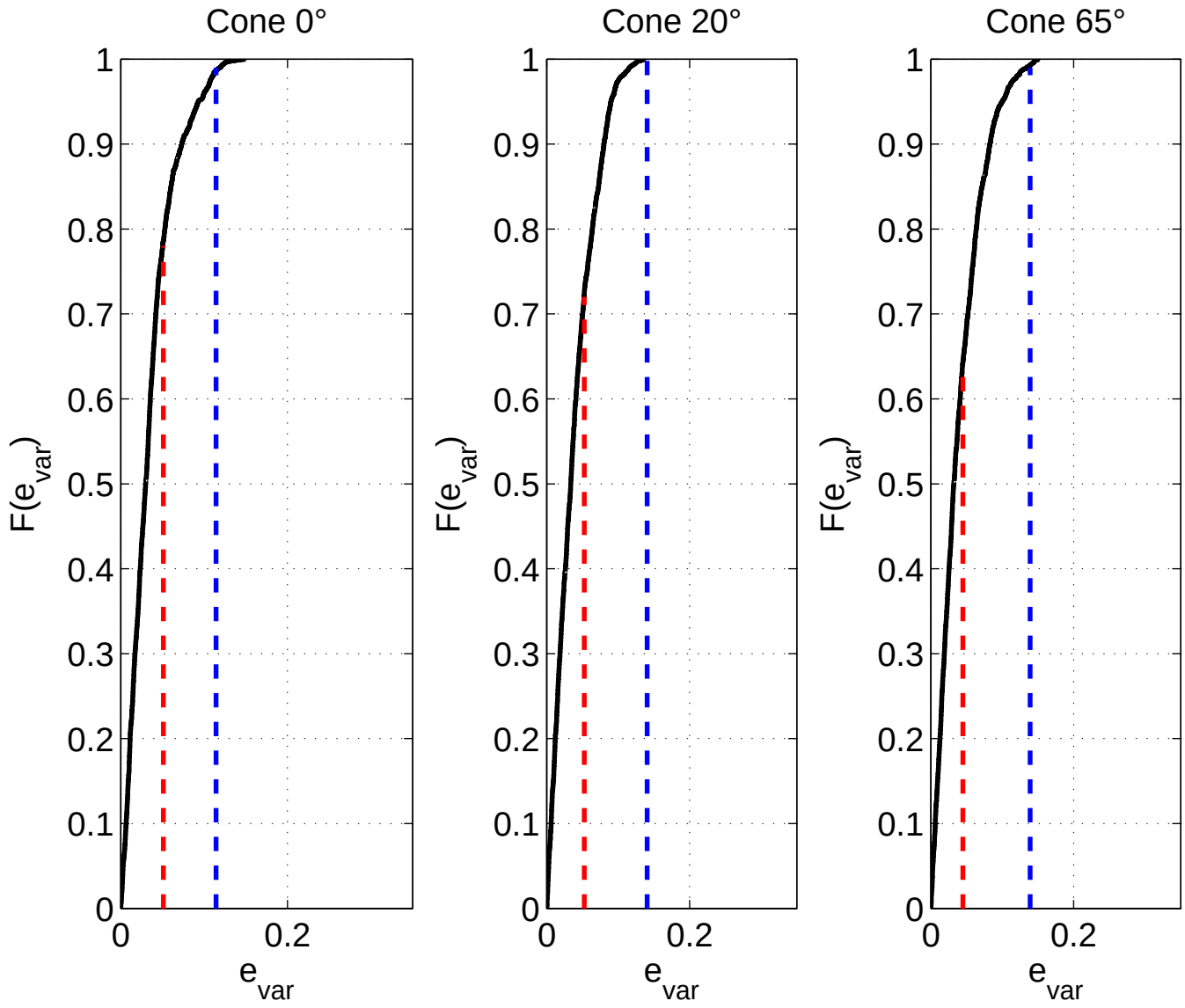


FIG. II.39 – Fonctions cumulatives de distribution du critère $e_{var^i}(\theta)$ pour les trois cônes de confusions et pour tous les sujets. Les lignes verticales rouges correspondent à $JND2_{add}(\theta)$ et les lignes verticales bleues correspondent à $JND1_{add}(\theta)$.

TAB. II.10 – Pourcentage d'audibilité des variations de l'ITD sur les cônes de confusion en fonction des deux protocoles.

Cône	$p_{JND1_{norm}}$	$p_{JND2_{norm}}$
0°	7	53
20°	1	33
65°	16	73

4.3.3 Individualisation de l'ITD sur les cônes de confusion

Afin d'examiner si les variations de l'ITD sur les cônes de confusion doivent être individualisées, un autre critère adimensionnel est calculé. Ce critère est évalué de la même manière que pour le paragraphe précédent mais en considérant cette fois-ci l'écart par rapport à une ITD moyenne. L'ITD moyenne correspond à la moyenne pour les 43 sujets de la base CIPIC de l'ITD en fonction de l'azimut et de l'élévation :

$$ITD_{moy}(\theta,\phi) = \frac{1}{M}\sum_{i=1}^{43} ITD_i(\theta,\phi) \tag{II.21}$$

Le critère adimensionnel est donné par :

$$e_{ind^i}(\theta,\phi) = \frac{ITD_i(\theta,\phi) - ITD_{moy}(\theta,\phi)}{ITDmax_i} \tag{II.22}$$

Les valeurs de $e_{ind^i}(\theta,\phi)$ sont calculées pour tous les individus et toutes les positions et ensuite regroupées par cônes de confusion. L'audibilité des variations individuelles est déterminée par le tracé de la fonction de répartition empirique du critère $e_{ind^i}(\theta,\phi)$ et par la valeur de $1 - F(JND_{norm})$. Les pourcentages d'audibilité des variations individuelles par rapport à une ITD moyenne sont reportés dans le tableau II.11. Selon JND1, les variations individuelles n'ont pas besoin d'être reproduites et la présentation d'une ITD moyenne est perceptivement satisfaisante. Même avec JND2, les pourcentages d'audibilité restent très faibles et ne dépassent pas 36 %.

Tab. II.11 – Pourcentage d'audibilité des variations individuelles de l'ITD sur les cônes de confusion en fonction des deux protocoles.

Cône	$p_{JND1_{norm}}$	$p_{JND2_{norm}}$
0°	1	22
20°	0	28
65°	1	36

Le tableau II.11 indique que les variations individuelles de l'ITD sur les cônes de confusion ne sont pas audibles et une ITD qui reproduit des variations moyennes suffit. Cependant, pour des systèmes de spatialisation sonore demandant une très grande précision, il peut être nécessaire de réduire les taux d'audibilité. Toujours dans une optique de réduction du coût d'implémentation de la synthèse binaurale, la précision apportée par des formules analytiques décrivant l'ITD et n'ayant pas besoin des mesures de HRTF est analysée dans le paragraphe suivant.

4.3.4 Validation perceptive des modélisations sphériques de l'ITD

Deux méthodes de calcul de l'ITD sont comparées. Ces deux méthodes correspondent à une modélisation de la tête de l'auditeur par une sphère et correspondent d'une part à la formule de Woodworth avec un rayon individualisé selon la méthode proposée dans [Algazi et al. (2001b)], nommée ici MSI (Model Sphérique Individualisé) et d'autre part

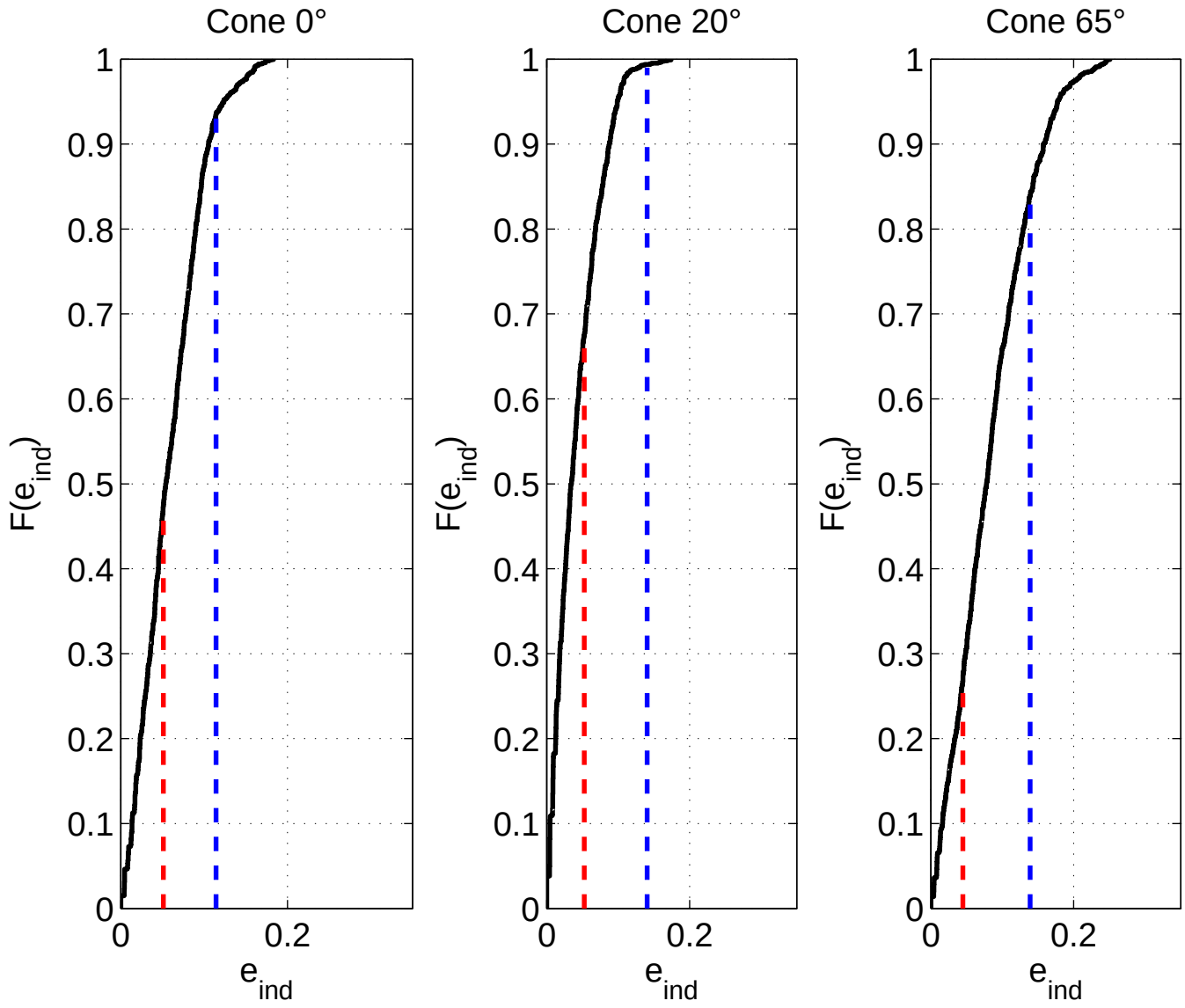


FIG. II.40 – Fonctions de répartition empirique du critère $e_{ind^i}(\theta)$ pour les trois cônes de confusions et pour tous les sujets. Les lignes verticales rouges correspondent à $JND2_{norm}(\theta)$ et les lignes verticales bleues correspondent à $JND1_{norm}(\theta)$.

à la formule FDO qui apporte des variations de l'ITD sur des cônes de confusion grâce au décalage des oreilles sur la sphère. Comme décrit au paragraphe 2.3, la formule FDO peut être optimisée par une procédure de minimisation de l'erreur quadratique pour reproduire certaines valeurs de l'ITD. Ainsi, la formule FDO est adaptée à l'ITD de chaque sujet et ce uniquement pour les positions décrivant le cône à 65°. Les variations du cône à 65° étant supérieures à celles des autres cônes il est attendu que si FDO reproduit correctement les variations de l'ITD pour le cône à 65°, elle reproduise aussi les variations pour les autres cônes. La méthode d'évaluation d'audibilité est réitérée ici avec le calcul d'un nouveau critère adimensionnel :

$$e_{mod^i}(\theta,\phi) = \frac{ITD_i(\theta,\phi) - ITD_{mod}(\theta,\phi)}{ITDmax_i} \quad \text{(II.23)}$$

où l'indice *mod* est relatif à MSI ou FDO. Les fonctions de répartitions empiriques des critères $e_{mod^i}(\theta,\phi)$ sont affichées en figure II.41. Le troisième modèle utilisé est une formule FDO qui a été optimisée sur l'ITD moyenne des 43 sujets pour le cône à 65° (cf. fig.II.24). Les paramètres de cette FDO moyenne sont un rayon de tête de 91 mm, un décalage des oreilles en azimut de -3° et en élévation de 19°. L'intérêt de cette FDO moyenne est de pouvoir donner une formule moyenne pour l'ITD qui apporte des variations de l'ITD sur les cônes de confusion contrairement aux autres formulations (par exemple la formule de Woodworth). Les pourcentages d'audibilité sont reportés dans le tableau II.12 pour les deux modélisations et la figure II.41 représente les fonctions de répartitions empiriques relatives à e_{MSI}, e_{FDO} et e_{FDOmoy} pour les trois cônes de confusion.

D'après les résultats du tableau II.12 les modélisations MSI et FDO sont perceptivement correctes au sens de la JND1. L'individualisation réalisée pour obtenir le modèle MSI est suffisante pour une précision moyenne de la synthèse binaurale. Les résultats relatifs à JND2 montrent la nécessité de l'individualisation du modèle FDO. L'utilisation d'une FDO moyenne permet de réduire l'audibilité pour le cône où la formule a été optimisée mais ne montre pas de meilleures performances que la formule FDO pour les autres cônes. La formule FDO moyenne montre des performances très encourageantes : même pour les cônes où elle n'a pas été optimisée, les taux d'audibilité restent faibles ($<$ 30%).

Tab. II.12 – Pourcentage d'audibilité des modélisations de l'ITD sur les cônes de confusions.

	MSI		FDO_{ind}		FDO_{moy}	
Cône	$p_{JND1_{norm}}$	$p_{JND2_{norm}}$	$p_{JND1_{norm}}$	$p_{JND2_{norm}}$	$p_{JND1_{norm}}$	$p_{JND2_{norm}}$
0°	2	22.5	2	23	2	23
20°	1	36	0	28	0	34
65°	2	51	0	4	0	36

4.4 Conclusion

L'étude préliminaire ayant montré les différences existantes entre différents protocoles expérimentaux, l'audibilité des variations de l'ITD sur les cônes de confusion est étudiée en fonction de deux procédures expérimentales. L'analyse des différences entre les deux protocoles choisis pour l'étude psychoacoustique incite à considérer les seuils obtenus par la procédure adaptative, c'est-à-dire JND1, comme des indicateurs d'une perception moyenne, qui sont mieux adaptés à un utilisateur occasionnel de la synthèse binaurale. Par contre, les seuils obtenus avec la méthode des constantes, c'est-à-dire JND2, semblent plus proches de niveaux de perception asymptotique. L'analyse menée sur les résultats des tests suit cette dichotomie.

Les performances moyennes correspondent à une JND entre 74 μs et 91 μs tandis que les performances asymptotiques se situent entre 20 μs et 40 μs ce qui est proche des

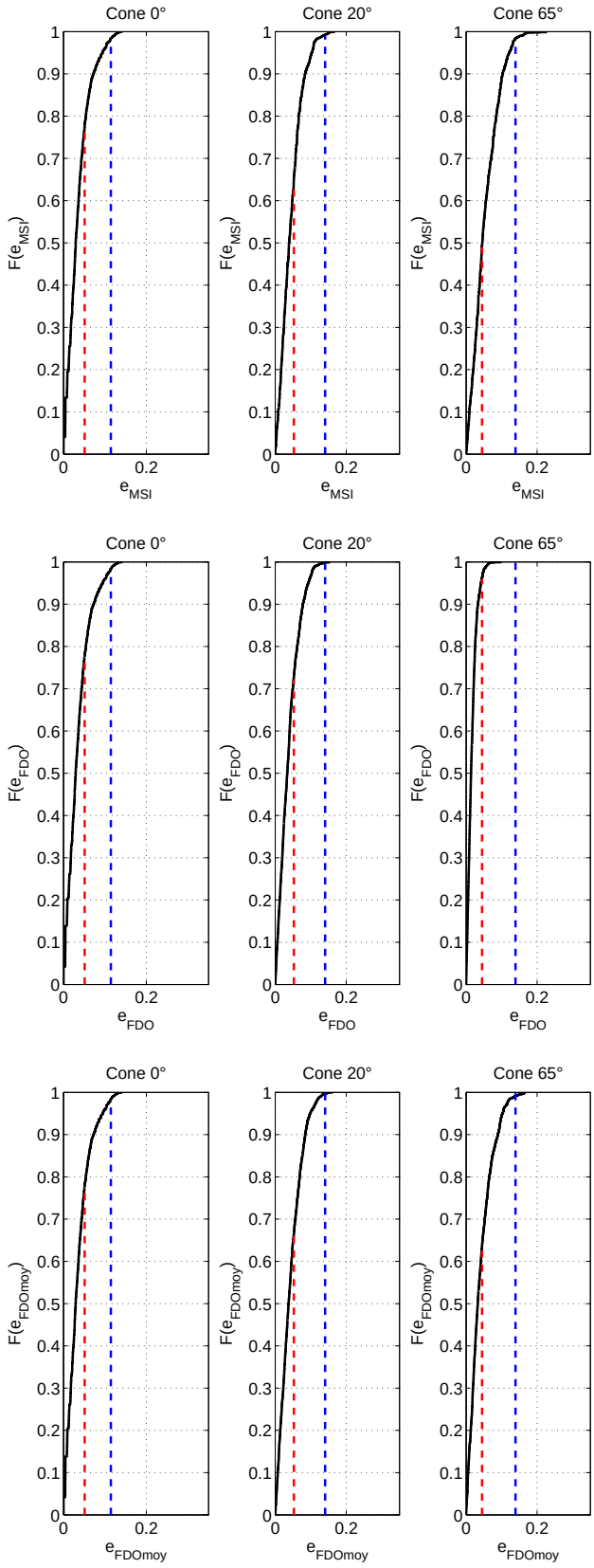


FIG. II.41 – Fonctions de répartition empirique du critère $e_{mod}(\theta)$ pour les trois cônes de confusion et pour les trois modélisations. Les lignes verticales rouges correspondent à $JND2_{add}(\theta)$ et les lignes verticales bleues correspondent à $JND1_{add}(\theta)$. Les figures du haut représentent $e_{MSI}(\theta)$, celles du mileu $e_{FDO}(\theta)$ et celles du bas $e_{EDFmoy}(\theta)$.

performances du système auditif en écoute champ libre et en écoute stéréophonique. Globalement, la JND de l'ITD ne varie pas avec l'élévation et dépend seulement de l'azimut du cône de confusion considéré. D'autres expériences sont nécessaires pour établir une cartographie plus complète de la JND de l'ITD. Cette première étude permet néanmoins de restreindre les champs d'investigation de la JND, ce qui est indispensable pour une étude expérimentale utilisant la procédure des constantes.

Au regard d'une perception moyenne, les variations sur les cônes de confusion ne sont pas perçues et n'ont donc pas besoin d'être reproduites. Une modélisation $\{ITD \bigoplus HRTF_{min}\}$ de la synthèse binaurale avec une ITD ne dépendant que de l'azimut de la source est alors suffisante. L'ITD variant avec l'azimut doit toutefois être fidèle à l'ITD du sujet sur le plan horizontal (cf. § 3). A ce titre, l'ITD de la formule de Woodworth convient. L'ITD MSI doit par contre être adaptée en fonction de la base de donnée (cf. § 2.3).

Au vu d'une synthèse binaurale recherchant un bon compromis entre précision et coût d'implémentation, les variations de l'ITD sur les cônes de confusion doivent être reproduites. Une ITD qui reproduit des variations moyennes permet de réduire l'erreur introduite et est perceptivement valide. La formule FDO individualisée et optimisée sur un cône de confusion ayant un maximum de variation d'ITD représente une modélisation très intéressante des variations individuelles de l'ITD. L'utilisation de la formule FDO est donc encourageante et une relation multi-linéaire entre décalage optimal des oreilles et paramètres morphologiques, de la même manière que la formule du rayon optimal dans [Algazi et al. (2001b)], permettrait une ITD individuelle reproduisant les variations de l'ITD sur les cônes de confusion de manière acceptable sans avoir à mesurer les HRTF d'un individu.

5 CONCLUSION

Les études expérimentales et théoriques présentées dans ce chapitre permettent de mieux appréhender le choix d'une méthode de calcul de l'ITD dans le cadre de l'implémentation $\{ITD \bigoplus HRTF_{min}\}$ de la synthèse binaurale.

La mesure de l'ITD psychoacoustique dans le plan horizontal, c'est-à-dire là où l'ITD varie le plus, a permis de valider l'implémentation la plus couramment utilisée de la synthèse binaurale et ce même pour les positions où les HRTF sont mal représentés par des filtres à phase minimale. Cette étude expérimentale a mis en évidence les performances globales et individuelles des méthodes d'estimation à partir des mesures que sont la méthode *seuil* à 50 % des HRIR et la méthode *MaxIACC* sur les enveloppes des HRIR.

Cependant, l'utilisation d'une formule analytique qui ne requiert pas de mesures a priori procure un intérêt en terme de coût d'implémentation. La formule FDO développée avec le travail théorique reporté dans ce chapitre offre une prédiction plus fine et une capacité d'individualisation supérieure aux formules classiques. L'adaptation des paramètres de la formule FDO permet une reproduction fidèle des variations de l'ITD et notamment les variations en élévation qui n'étaient pas reproduites avec les formules classiques.

La mesure du pouvoir de discrimination de l'ITD, ou JND de l'ITD, a montré que les variations de l'ITD sur des plans sagittaux doivent être reproduites. Les JND obtenues sont constantes avec l'angle d'élévation. Elles ne dépendent que de l'azimut du plan

sagittal et les valeurs les plus faibles sont proches des performances de localisation en écoute champ libre. L'optimisation de la formule FDO permet d'obtenir une ITD qui reproduit les variations individuelles de l'ITD sur des plans sagittaux.

L'application d'une démarche d'optimisation entre paramètres de la formule FDO et certaines données morphologiques propres à chaque sujet permettra l'obtention d'une relation d'individualisation des paramètres sans avoir recours aux mesures de HRTF. Cette relation devra être étendue à plusieurs bases de données pour obtenir une relation la plus universelle possible.

BIBLIOGRAPHIE

Algazi, V., Duda, R., and Thompson, D. (2001a). The cipic hrtf database.

Algazi, V. R., Avendano, C., and Duda, R. O. (2001b). Estimation of a spherical-head model from anthropometry. *J. Acoust. Soc. of Am.*, 49 :472–478.

Algazi, V. R. and Duda, R. O. (2002). Approximating the head-related transfer function using simple geometric models of the head and torso. *J. Acoust. Soc. of Am.*, 112(5) :2053–2064.

Avendano, C., Duda, R. O., and Algazi, R. (1999). Modelling the contralateral hrtf.

Best, V., van Schaik, A., and Carlile, S. (2004). Separation of concurrent broadband soun sources by human listeners. *J. Acoust. Soc. of Am.*, 115(1) :324–336.

Blauert, J. (1983). *Spatial Hearing*. The mit press edition.

Braasch, J. and Hartung, K. (2002). Localization in the presence of a distracter and reverberation in the frontal horizontal plane. i. psychoacoustical data. *Acta Acoustica united with Acoustica*, 88 :942–955.

Brown, C. P. and Duda, R. O. (1998). A structural model for binaural sound synthethis. volume 6. IEEE Trans. Speech and Audio Proc.

Busson, S., Nicol, R., and Katz, B. F. G. (2005a). Subjective investigations of the interaural time difference in the horizontal plane. *Presented at the* 118$^{\text{th}}$ *AES Convention, Barcelona, Spain*. Convention Paper 6324.

Busson, S., Nicol, R., and Warusfel, O. (2004). Influence of the ears canal location on spherical head model for the individualised interaural time difference. *Proceedings of CFA / DAGA Joint Meeting, Strasbourg, France*.

Busson, S., Nicol, R., Warusfel, O., and Gros, L. (2005b). Just noticeable difference of the interaural time difference on cones of confusion. to be summited to J. Acoust. Soc. of Am.

Divenyi, P. L. and Oliver, S. K. (1989). Resolution os steady-state sound in simulated auditory space. *J. Acoust. Soc. of Am.*, 85(5) :2042–2052.

Domnitz, R. H. (1973). The interaural time jnd as a simultaneous function of interaural time and interaural amplitude. *J. Acoust. Soc. of Am.*, 53(6) :1549–1552.

Domnitz, R. H. and Colburn, H. S. (1977). Lateral position and interaural discrimination. *J. Acoust. Soc. of Am.*, 61(5) :1586–1598.

Duda, R. O., Avendano, C., and Algazi, V. (1999). An adaptable ellipsoidal head model for the interaural time difference. In *Proc. IEEE International Conference on Acoustics Speech and Signal Processing*, volume II, pages 965–968. ICASSP'99.

Ericson, M. and McKinley, R. L. (1989). Auditory localization cue synthesis and human performance. pages 718–725, Dayton, OH. NAECON 89.

Grantham, D. W., Hornsby, B. W. Y., and Erpenbeck, E. A. (2003). Auditory spatial resolution in horizontal, vertical, and diagonal planes. *J. Acoust. Soc. of Am.*, 114(2) :3030–3038.

Grantham, D. W., Whillhite, J. A., Frampton, K. D., and Ashmead, D. H. (2005). Reduced order modeling of head related impulse responses for virtual acoustics displays. *J. Acoust. Soc. of Am.*, 117(5) :3116–3152.

Hafter, E. R. and Maio, J. D. (1975). Difference thresholds for interaural delay. *J. Acoust. Soc. of Am.*, 57(1) :181–187.

Henning, G. B. (1974). Detectability of interaural delay in high-frequency complex waveforms. *J. Acoust. Soc. of Am.*, 55(1) :84–90.

Hershkowitz, R. M. and Durlach, N. I. (1969). Interaural time and amplitude jnds for a 500-hz tone. *J. Acoust. Soc. of Am.*, 46(6, Part 2) :1464–1467.

Katz, B. F. G. (1998). *Measurement and calculation of individual head-related transfer function using a boundary element model incuding the measurement and effect of skin and hair impedance.* PhD thesis, Pennsylvania State University.

Kistler, D. J. and Wightman, F. L. (1992). A model of head-related transfert functions based on principal components analysis and minimum-phase reconstruction. *J. Acoust. Soc. of Am.*, 91(3) :1637–1647.

Klump, R. G. and Eady, H. R. (1956). Lateral position and interaural discrimination. *J. Acoust. Soc. of Am.*, 28(5).

Kollmeier, B., Gilkey, R. H., and Sieben, U. K. (1988). Adaptative staircase techniques in psychoacoustics : A comparison of human data and a mathematical model. *J. Acoust. Soc. of Am.*, 83(5) :1852–1862.

Kuhn, G. F. (1977). Model for the interaural time differences in the azimuthal plane. *J. Acoust. Soc. of Am.*, 62(1).

Kulkarni, A., Isabelle, S., and Colburn, H. (1999). Sensitivity of human subjects to head-related transfer function phase spectra. *J. Acoust. Soc. of Am.*, 105(5) :2821–2840.

Laakso, T. I., Välimäki, V., Karjailanen, M., and Laine, K. U. (1996). Splitting the unit delay : tools for fractional delay filter design. *IEEE Signal Processing Magazine*, 12(1) :30–60.

Larcher, V. (2001). *Techniques de spatialisation des sons pour la réalité virtuelle.* PhD thesis, Université Paris VI.

Larcher, V. and Jot, J. M. (1999). Techniques d'interpolations de filtres audionumériques : Application à la reproduction spatiale des sons sur écouteurs. Marseilles. Présenté au Congès Français d'Acoustique.

Levitt, H. (1970). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. of Am.*, 49 :467–477.

Litovsky, R. Y., Hawley, M. L., and Fligor, B. J. (2000). Failure to unlearn the precedence effect. *J. Acoust. Soc. of Am.*, 108(5) :2345–2352.

Lord Rayleigh (1907). On our perception of sound direction. *Phylosophy Magazine*, 13 :214–232.

Møller, H., rensen, M. F. S., i, D. H., and Jensen, C. B. (1995). Head related transfer functions of human subjects. *J. Audio Engin. Soc.*, 43(5) :300–321.

McFadden, D. (1981). The problem of different interaural time differences at different frequencies. *J. Acoust. Soc. of Am.*, 69(6) :1586–1598. Letters to the Editor.

McFadden, D. and Pasanen, E. G. (1976). Lateralization at high frequnecies based on interaural time differences. *J. Acoust. Soc. of Am.*, 59(3) :634–639.

Middlebrooks, J. (1999). Individual differences in external-ear transfer function reduced by scaling in frequency. *J. Acoust. Soc. of Am.*, 106(3) :1480–1492.

Miller, J. D. (2001). Modelling iteraural time difference assuming à spherical head.

Mills, A. W. (1958). On the minimum audible angle. *J. Acoust. Soc. of Am.*, 30 :237–248.

Minaar, P., Plosgsties, J., Olesen, S. K., Christensen, F., and Moller, H. (2000). The interaural time difference in binaural synthesis. *J. Acoust. Soc. of Am.*, 92 :207.

Minnaar, P., Christensen, F., Moller, H., Olesen, S. K., and Plogsties, J. (1999). Audibility of all-pass components in binaural synthesis. Munich. AES 106$^{\text{th}}$ Convention.

Moushegian, G. and Jeffress, L. A. (1959). Role of the interaural time and intensity differences in the lateralization of low-frequency tones. *J. Acoust. Soc. of Am.*, 31(11) :1441–1445.

Oldfield, S. R. and Parker, S. P. A. (1984). Acuity of sound localisation : a topography of auditory space. i. normal hearing. *Perception*, 13 :581–600.

Pernaux, J. (2003). *Spatialisation du son par les tehcniques binaurales : application aux services de télécommunication*. PhD thesis, Institut National de Polytechnique de Grenoble.

Pernaux, J.-M., Emerit, M., and Nicol, R. (2003). Perceptual evaluation of binaural sound synthesis : the problem of reporting localization judgments. *Presented at the* 114$^{\text{th}}$ *AES Convention, Amsterdam, The Netherlands*. Convention paper.

Perrot, D. R. (1984). Concurrent minimum audible angle : a re-examination of the concept of auditory spatial acuity. *J. Acoust. Soc. of Am.*, 75(4) :1201–1206.

Perrot, D. R. and Saberi, K. (1990). Minimum audible angle thresholds for sources varying in both elevation in azimuth. *J. Acoust. Soc. of Am.*, 87(4) :1728–1731.

Plogsties, J., Minnaar, P., Olesen, S. K., Christensen, F., and Moller, H. (2000). Audibility of all-pass components in head-related transfer function. Paris. AES 108$^{\text{th}}$ Convention.

Saviojaa, L., Huopaniemi, J., Lokki, T., and Vaanen, R. (1999). Creating interactive virtual acoustic environnements. *J. Acoust. Soc. of Am.*, 47(9) :675–705.

Tollin, D. J. and B., H. G. (1998). Some aspects of the lateralization of echoed sound in man. i. the classical interaural-delay based precedence effect. *J. Acoust. Soc. of Am.*, 104 :3030–3038.

Trahiotis, C., Bernstein, L. R., Buell, T. N., and Spektor, Z. (1990). On the use of adaptative procedures in binaural experiments. *J. Acoust. Soc. of Am.*, 87(3) :1359–1361.

von Békésy, G. (1960). *Experiments in hearing.* Acoustical Society of America, mcgraw-hill book company edition.

Woodworth, R. S. and Schloesberg, G. (1962). *Experimental psychology.* New-York. pp 349-361.

Yost, W. A., Turner, R., and Bergert, B. (1974). Comparison among four psychophysical procedures used in lateralization. *Perception & Psychophysics*, 15(3) :483–487.

Zurek, P. M. (1985). Spectral dominance in sensitivity to interaural delay for broadband stimuli. *Presented at the* 110^{th} *meeting of the Acoustical Society of America.*

III

Acquisition de HRTF

INTRODUCTION

Le but de ce chapitre est de présenter deux méthodes d'acquisition de HRTF : la mesure empirique et le calcul numérique par éléments de frontière. La mesure permet l'obtention de $HRTF_{mixte}$ tandis que le calcul n'a été utilisé ici que pour l'acquisition du module des $HRTF_{mixte}$, la modélisation de l'ITD est abordée au chapitre II. Cette étude part du principe qu'une reproduction fidèle de l'écoute naturelle grâce à la synthèse binaurale est obtenue avec les HRTF individuelles de chaque auditeur : les mesures sont alors le mètre étalon. Cependant, la mesure des HRTF est un problème majeur de la synthèse binaurale. C'est une tâche délicate soumise à de nombreuses sources d'erreurs. La mesure est longue et demande un matériel coûteux, notamment le système mécanique pour le positionnement relatif microphone/haut-parleur. La mesure nécessite aussi un environnement spécifique, comme l'utilisation d'une chambre anéchoïque. Le choix même des positions à mesurer est loin d'être évident. Le calcul de HRTF revêt ici toute son importance : pour la faisabilité de l'application grand public de la synthèse binaurale, il est impératif de se doter d'une procédure simple d'acquisition des HRTF.

La résolution numérique du problème acoustique correspondant à la mesure des HRTF permet de s'affranchir de la lourde tâche de la mesure. L'approche abordée dans les travaux réalisés est celle des éléments de frontière, ou *BEM* pour Boundary Element Method. Cette méthode permet l'obtention d'un champ de pression dans un volume par le calcul du champ de pression sur la surface entourant le volume. La surface considérée est, pour notre étude, une modélisation géométrique de la morphologie d'un auditeur.

L'avantage principal de la BEM par rapport aux mesures est une répétabilité assurée. De plus cette technique permet le placement des microphones et des sources avec une très grande précision, sans souci d'encombrement (microphone placé dans le conduit auditif). Enfin, il est possible de contrôler l'impédance des surfaces et les caractéristiques du fluide entourant l'objet d'analyse. La BEM permet ainsi d'effectuer un travail systématique sur l'influence des paramètres de modélisation de l'objet d'analyse (morphologie de l'auditeur). Par exemple, un travail qui examine l'influence sur les HRTF, de la position du canal auditif par rapport au centre d'une sphère représentant la tête d'un auditeur peut être mené. Cependant, la méthode BEM appliquée au problème particulier de l'acquisition de HRTF requiert des ressources informatiques considérables, mémoire vive, vitesse du processeur, pour réaliser un calcul dans des temps acceptables.

Ce chapitre s'articule autour de deux parties. La première partie est consacrée à la comparaison de différentes bases de HRTF mesurées. Les trois principales bases de données utilisées au cours des ces travaux sont décrites et comparées. Le principe général de la mesure est présenté et les erreurs introduites lors des sessions de mesures sont exposées.

La deuxième partie de ce chapitre présente le travail réalisé pour l'acquisition de HRTF en utilisant la méthode BEM. Ces travaux représentent une avancée dans la compréhension de l'influence de la morphologie de l'auditeur sur les HRTF. L'auditeur est ici modélisé par des formes géométriques simples. Le première section est consacrée à l'énoncé du problème et à la description du principe de la méthode BEM. Les aspects pratiques sont analysés et les limites techniques de la BEM sont mises en évidence. Les travaux antérieurs consacrés à l'application de méthodes numériques pour le calcul de HRTF sont ensuite rapportés. L'apport du travail réalisé sur l'application de la méthode BEM à des géométries simplifiées de la morphologie de l'auditeur pour le calcul de HRTF est exposé. Enfin, la dernière partie conclut sur l'utilisation de la BEM pour la prédiction de HRTF.

1 LA MESURE DE HRTF

1.1 Principe de la mesure de HRTF

Une HRTF est une fonction de transfert complexe entre la pression délivrée par une source sonore en champ libre P_1, c'est-à-dire la pression au lieu du centre de la tête de l'auditeur, et la pression résultante au niveau du tympan P_2 [Møller (1992); Møller et al. (1995); Hammershøi and Møller (2002)]. L'auditeur est vu comme un objet qui diffracte une onde en champ libre. Les HRTF sont les fonctions de diffraction, réflexion et résonance de cet objet. Cette fonction de transfert dépend de la position relative de la source sonore par rapport à l'auditeur, de la morphologie de l'auditeur et du contenu fréquentiel de la source sonore. La mesure de la position de la source sonore est effectuée en utilisant des systèmes de coordonnées, le plus souvent sphériques, inclus dans le *référentiel auditeur* (cf. fig.I.5). Le centre de la tête étant difficile à définir et à mesurer, il est alors utilisé des systèmes de coordonnées dont un des axes est confondu avec l'axe interaural (cf. fig.I.4) : le centre de la tête, et donc du référentiel sujet, est le milieu de l'axe interaural.
Le principe de la mesure est de placer la source sonore à une position de l'espace repérée

par ses coordonnées Θ, de mesurer P_2 grâce à un microphone situé au niveau des tympans, de mesurer P_1 au lieu du centre de la tête, auditeur absent, pour obtenir la HRTF aux coordonnées Θ :

$$HRTF(\Theta, f, \lambda) = \frac{P_2(\lambda)}{P_1(\Theta, f)} = \frac{\text{Pression au niveau du tympan}}{\text{Pression délivrée par une source sonore}} \quad \text{(III.1)}$$

L'espace auditif est souvent représenté à l'aide de coordonnées sphériques qui sont bien adaptées aux performances de localisation : la perception de la distance absolue, qui plus est en condition anéchoïque, est très peu développée chez l'homme. Par contre le système auditif discrimine très bien des sources en azimut. La variation des HRTF, et des indices monoraux et interauraux, avec la distance n'a pas été abordée dans les travaux présentés ici[1]. Les mesures de HRTF sont pour la plupart effectuées à rayon constant.

Pour des raisons physiologiques évidentes, il est impossible de mesurer un signal acoustique au niveau du tympan. La position du point de mesure dans le conduit auditif a longtemps été une question ouverte [Møller (1992); Hammershøï and Møller (2002); Algazi et al. (1999)]. Aujourd'hui, la méthode de mesure au niveau de l'entrée du conduit auditif avec le conduit bouché (cf. fig. III.2), *blocked ear canal* en anglais, est utilisée au lieu des méthodes en conduit ouvert (cf. fig. 1.1). En effet il a été reporté que les modifications introduites par le conduit auditif sont indépendantes de la position de la source sonore. Les variabilités inter-individuelles, autre source d'erreurs indépendantes de la position, sont atténuées en conduit bloqué [Møller (1992); Chateau (1996)].

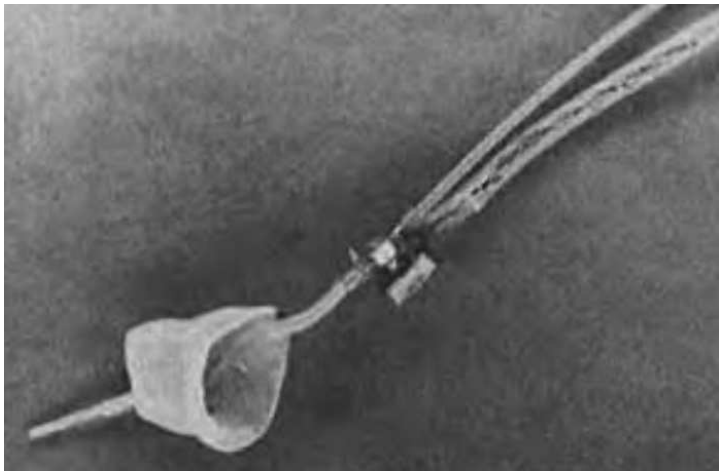
Les HRTF étant largement dépendantes de la fréquence, une attention particulière doit-être apportée à la nature du signal source. Différents types de signaux ont été testés et une revue des avantages et inconvénients de quatre méthodes (impulsion, bruit blanc gaussien, séquence de longueur maximale et codes de Golay) est présentée dans [Cheng and Wakefield (2001)]. L'article ne fait cependant pas mention d'un type de signal très avantageux : le *sinus glissant* (*chirp* ou encore *sweep* en anglais) (cf. figure III.3).

$$y(t) = sin(f(t).t) \quad \text{(III.2)}$$

Ce type de signal permet l'obtention de toute la bande audio en une seule mesure et comporte assez d'énergie pour une utilisation linéaire du haut-parleur. De plus ce signal étant déterministe, il est facilement reproductible et les traitements pour récupérer l'information sont simplifiés.

Les HRTF sont finalement obtenues par la division fréquentielle du signal reçu au niveau du tympan et du signal mesuré, avec la même configuration sujet absent, au lieu du centre de sa tête (souvent confondu avec le centre du dispositif de mesure.) Cette opération permet d'enlever la contribution de la chaîne de mesure (le choix de la position relative du microphone et du Haut-parleur est arbitraire). Enfin, une dernière étape permet d'enlever toutes contributions non spatiales : c'est l'égalisation par rapport à un champ de référence. Cet aspect est abordé au paragraphe 4.2 qui décrit les deux types d'égalisation les plus utilisés.

[1] Le lecteur pourra se reporter aux travaux exposés dans [Brungart and Rabonowitz (1998)] et [Duraiswami et al. (2004)] pour avoir plus d'informations à ce sujet



a) Moule pour microphone

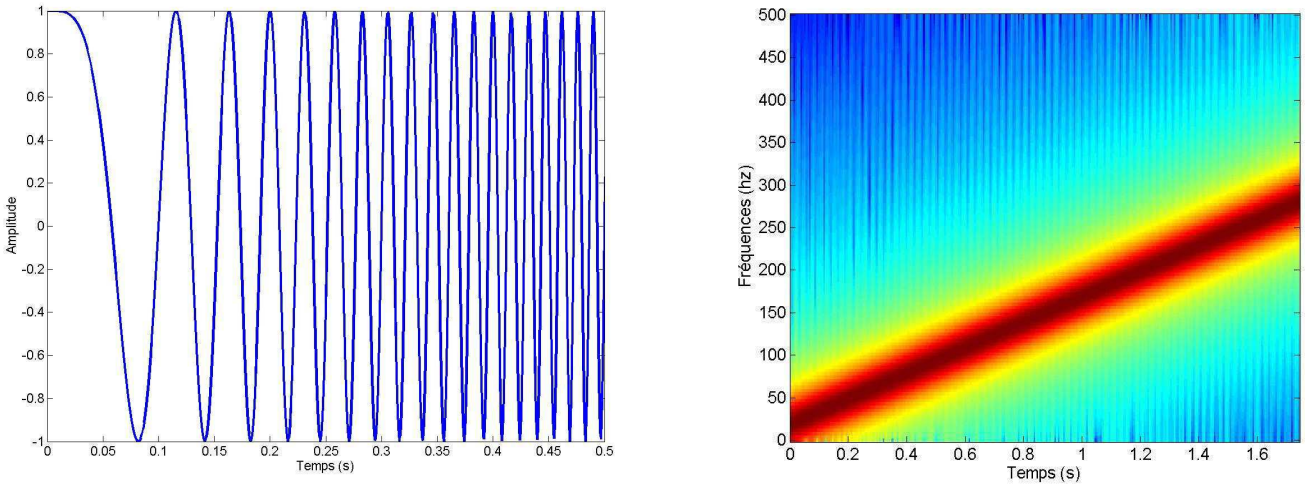
b) Placement du moule

Fig. III.1 – Dispositif expérimental de placement de microphone en conduit ouvert [Wightman and Kistler (1989)].

1.2 Sources d'erreur lors des campagnes de mesures

La mesure acoustique est globalement une chose délicate et qui doit être le fruit de nombreuses précautions. Par exemple la condition de mesures anéchoïques est difficile à réaliser à cause des multiples réflexions sur le système de mesure. Plutôt que de rechercher la mesure la plus *propre*, une estimation des erreurs introduites par le système de mesure est réalisée. Les erreurs sont ensuite *retirées* de la mesure. C'est le but de la mesure sujet absent que d'estimer les contributions de la chaîne de mesure. Cependant, cette opération nécessite que les erreurs restent limitées : c'est le problème de toute opération de déconvolution. Cette étape est encore plus sensible pour enlever la contribution du couplage pavillon-écouteur lors d'écoutes au casque. Un article sur les méthodes de déconvolution de réponse de casque conseille même de ne pas effectuer de telle opération sur des données aussi sensibles à la coloration spectrale que les HRTF [Kulkarni and

FIG. III.2 – Utilisation de moulage du conduit auditif pour le placement des microphones en conduit fermé [Vandernoot].



a) Représentation temporelle b) Spectrogramme

FIG. III.3 – Sinus glissant linéaire de fréquence d'échantillonnage 1 kHz.

Colburn (2000)]. Cette étape est considérée comme une opération de post-traitement et n'est pas détaillée ici.

Plusieurs types d'erreurs sont référencés dans la littérature. La liste ci-dessous décrit les principales sources d'erreur.

- **Réflexions sur le système de mesure** : le système de mesure de HRTF est souvent complexe et comporte de nombreux éléments mécaniques, notamment une arche devant supporter le HP et éventuellement le siège du sujet. Tous ces éléments doivent être recouverts d'un absorbant acoustique. La moindre réflexion peut contaminer les mesures. Le fenêtrage des mesures est aussi indispensable et toutes les précautions doivent y être apporté (utilisation de fenêtre appropriée au rapport signal à bruit).
- **Erreur de positionnement du microphone** : la méthode de mesure avec conduit auditif bloqué permet entre autre d'enlever la variance sur la position du microphone dans le canal auditif. Cependant, selon la taille de la capsule microphonique, une erreur de positionnement persiste à l'entrée au niveau du meatus (cf. fig. III.4).

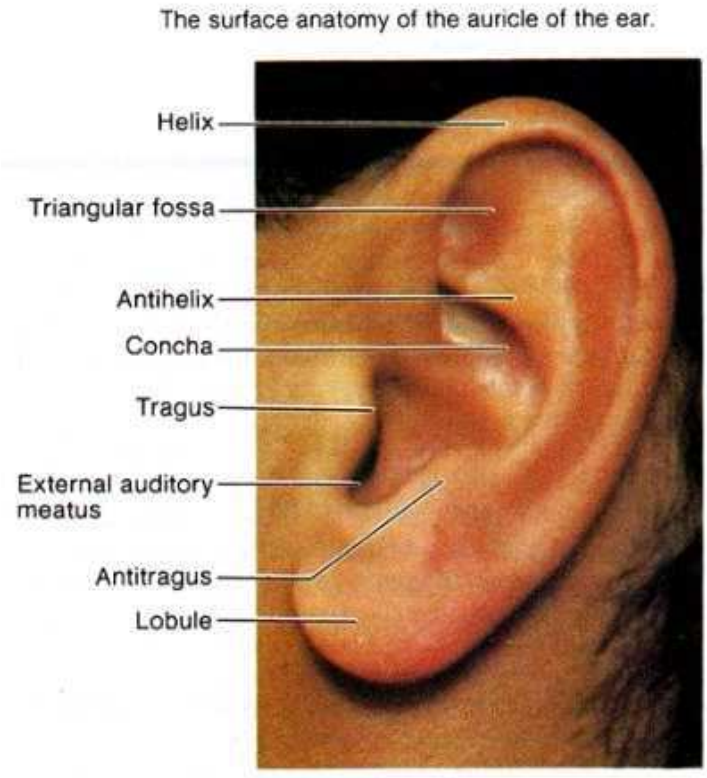


Fig. III.4 – Photographie d'un pavillon de l'oreille externe (cf. http://www.iha-online.co.uk/).

Une solution pratique pour résoudre ce problème est l'utilisation de moulage du conduit (cf. fig. III.2). Le moulage est percé pour y incorporer les capsules : le placement du microphone est alors moins variant. Cette technique a été utilisée pour la base de HRTF LISTEN [Vandernoot].

- **Erreur de positionnement du sujet** : dans certains dispositifs [Møller et al. (1995)], le sujet est debout lors de la mesure, ce qui correspond à une situation naturelle d'écoute. Même avec un dispositif aidant le sujet à se placer au centre du repère des erreurs sont introduites. Møller et al. estiment que pour leur système de mesure, une erreur de placement du sujet de 10 mm entraîne un décalage temporel de 30 μs sur les HRIR [Møller et al. (1995)]. Le placement du sujet sur un siège au centre du repère permet de réduire ces erreurs.
- **Mouvements de tête** : la position relative du haut-parleur et de la tête du sujet est cruciale. Pour les HRTF frontales, un problème de positionnement peut être à l'origine d'une mauvaise perception des sources devant ($\theta = 0°$ et $\phi = 0°$) : les sources sont alors perçues plus proches de la tête et en hauteur. Certains protocoles utilisent un support permettant de maintenir la tête des sujets tout au long de la séance de mesures. Ceci réduit les mouvements de la tête mais ne les élimine pas et le maintien prolongé de la tête rajoute de l'inconfort. Møller et al. analysent ce type d'erreur en terme d'ITD pour $\theta = 0°$ et $\phi = 0°$ et indiquent qu'une rotation de la tête de 5° entraîne une ITD de $80\mu s$ alors que la valeur 0 est attendue pour cette position [Møller et al. (1995)]. Seulement cette mesure peut être noyée dans les dissymétries morphologiques des sujets.
- **Basses fréquences** : pour assurer une condition de champ libre, les systèmes de mesure sont souvent placés dans des chambres anéchoïques. L'absorbtion des ondes y est garantie sur une large bande de fréquences. Seulement la condition anéchoïque est difficilement réalisable pour les basses fréquences. De plus, la taille des haut-parleurs se doit d'être réduite pour limiter l'encombrement et leur poids faible pour soulager le support et éviter une forte inertie lors du déplacement du haut-parleur. C'est pourquoi la partie basses fréquences des HRTF est souvent peu fiable et donc ignorée. Une solution à la mesure des basses fréquences, est l'utilisation de modèle numérique de HRTF. Ainsi la partie basses fréquences des HRTF de la base CIPIC est reconstruite à partir d'un filtre RII [Algazi et al. (2001a)] pour f $<$ 400 Hz.
- **Reproductibilité** : la contribution globale des erreurs précédentes est estimée par des mesures répétées dans les mêmes conditions à différents intervalles de temps. Møller et al. effectuent trois mesures dans ce sens et indiquent une très faible variation jusqu'à 20 kHz (de l'ordre de 1 db en basses fréquences) [Møller et al. (1995)]. Des variations plus importantes sont observées pour les positions contralatérales. Dans [Miller (2001b)] l'erreur quadratique moyenne E_{rms} entre les puissances des mesures sert de critère d'analyse pour la mesure de l'erreur de reproductibilité :

$$E_{rms} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} \left| \quad 20 * log_{10}(\frac{HRTF_1(i)}{HRTF_2(i)}) \quad \right|^2} \tag{III.3}$$

avec N nombre de fréquences des HRTF, $HRTF_1(i)$ représente une HRTF à une position donnée et $HRTF_2(i)$ la même HRTF mesurée soit un jour plus tard, soit une semaine plus tard. Trois mesures sont ainsi effectuées. Sur trois sujets différents, l'erreur de répétabilité est en moyenne sur toutes les positions (24 positions réparties sur le plan horizontal) de 2.63 dB pour le délai de 1 jour et de 4.43 dB pour le délai d'une semaine. Cette erreur est importante mais reste heureuse-

ment en-dessous des erreurs introduites par l'utilisation de HRTF non-individuelles (comprises entre 7 et 13 dB) [Miller (2001a)].

1.3 Les trois bases de données utilisées

De nombreuses bases de données de HRTF, publiques ou privées, existent. Ce paragraphe présente celles utilisées au cours des travaux de thèse que ce soit pour des tests perceptifs (cf. chapitre II) ou pour la modélisation (cf. chapitre 1). La description de ces trois bases met en évidence des différences entre les systèmes de mesure. Ces différences rendent difficile la réunion de bases de HRTF. Or, les modèles de prédiction de HRTF (cf. chapitre 1) doivent être construits sur des bases de données qui comportent le plus de mesures possible pour assurer l'universalité du modèle. Le tableau III.1 récapitule les principales caractéristiques de chaque base de données.

Tab. III.1 – Principales caractéristiques des trois bases de données de HRTF utilisées dans la thèse.

Base de HRTF	**Positions mesurées par oreille**	**Fréquence d'échantillonnage (kHz)**	**Signal**	**Avantages**	**Inconvénients**
Listen **187** positions **50** sujets	élévation constante $\Delta az_{min} = 15°$ $\Delta el_{min} = 15°$	44,1	sweep	sujets disponibles[2], rapidité	peu de points de mesures
FTR&D **965** positions **9** sujets	élévation constante $\Delta az_{min} = 5.625°$ $\Delta el_{min} = 5.625°$	48	bruit large bande	répétabilité, bon échantillonnage spatial, sujets disponibles	Séances de mesures longues, peu de sujets
CIPIC **1250** positions **50** sujets	azimut constant $\Delta az_{min} = 5°$ $\Delta el_{min} = 5.625°$	44.1	Code de Golay	échantillonage spatial décrivant des cônes de confusion	Basses fréquences modélisées, sujets non disponibles

1.3.1 La base LISTEN

La base de HRTF LISTEN correspond aux mesures effectuées à l'IRCAM[3] dans le cadre du projet européen LISTEN[4]. Un total de 187 positions ont été mesurées pour les

[3]Institut de Coordination Acoustique/Musique, 1 place Igor Stravinsky, 75004 Paris
[4]http://recherche.ircam.fr/equipes/salles/listen/context.html

51 sujets de la base. Le protocole est décrit en détail dans [Vandernoot]. Les principaux éléments sont rappelés ici. Le sujet est assis sur une chaise fixée à une table tournante : l'angle de la table tournante définit l'azimut de la mesure. La tête du sujet est attachée à un repose-tête. Le haut-parleur est fixé à une structure métallique dont l'axe de rotation est confondu avec l'axe interaural. La rotation du haut-parleur décrit alors un arc fixe définissant des angles d'élévation variant entre -40 et $+90$. Les mesures sont réalisées par plan d'élévation constante.

La totalité des mesures est réalisée en 1 heure. Le déclenchement des mesures n'est effectué que si la position de la tête, contrôlée par un head-tracker magnétique, est correcte.

FIG. III.5 – Système de mesure de HRTF pour la base de données LISTEN (cf. http://recherche.ircam.fr/equipes/salles/listen/system/protocol.html).

1.3.2 La base CIPIC

Les données de la base CIPIC [Algazi et al. (2001b)] comportent 1250 positions (50 élévations * 25 azimuts) mesurées sur 43 sujets humains et 2 têtes artificielles (KEMAR avec 2 moulages d'oreilles différents). Les mesures sont réalisées en conduit auditif fermé. Le sujet est assis au centre du repère. Un arc comportant 5 HP et dont l'axe de rotation est perpendiculaire à l'axe interaural du sujet,

FIG. III.6 – Système de mesures de HRTF pour la base de données CIPIC.

La partie Basses-Fréquences (jusqu'à 400 Hz) est modélisée par un filtre RII prenant en compte la diffraction de l'onde acoustique sur le torse.

1.3.3 La base FTR&D

La base de HRTF de FTR&D comporte 965 mesures réalisées sur 8 sujets humains et une tête artificielle (HMSIII, Head Acoustics). Le système de mesure utilisé pour constituer la base de HRTF de FTR&D [Pernaux (2003)] est celui décrit dans [Bronkhorst (1995)] et se situe au TNO Human Factors Research Institut à Soesterberg (Pays-Bas). Les mesures sont effectuées en conduit auditif fermé. Le sujet, dont la tête est maintenue, est assis sur une chaise pivotante et son axe interaural est confondu avec l'axe de rotation du rail supportant le haut-parleur (pour la position az = 0° et el = 0°). Le haut-parleur se déplace sur un rail en forme d'arc de rayon 1.40 m. Les positions relatives de la tête du sujet et du haut-parleur sont contrôlées par un head-tracker magnétique. Un message sonore automatique indique au sujet les corrections à apporter à sa posture et la position du HP est ajusté (le positionnement du haut-parleur par rapport au sujet est obtenu avec une précision de 1°). La séquence de mesure est gouvernée par un algorithme minimisant les trajets du haut-parleur.

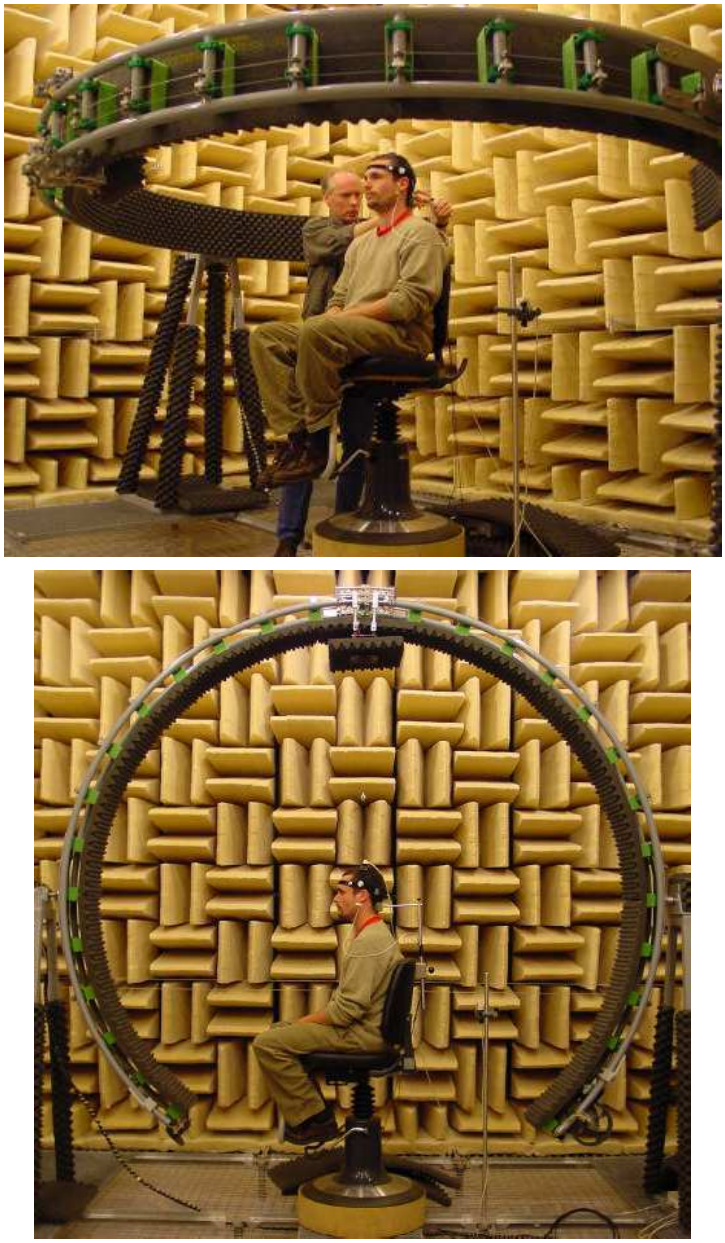


FIG. III.7 – Système de mesure de HRTF pour la base de données FTR&D.

1.3.4 conclusion

Le tableau III.1 récapitule les données exposées aux paragraphes précédents. La réunion de ces trois bases de données est problématique sous plusieurs aspects. Premièrement, le matériel, la salle et le processus sont spécifiques à chaque base de données. Deuxièmement, les points de mesure sont différents (cf. fig. III.8). Pour l'étude sur la modélisation, cela est rédhibitoire quant à l'utilisation de réunion de base car la création du modèle doit se faire avec des instances de vecteurs qui représentent les mêmes données. Pour obtenir les mêmes points de mesure, les points manquants peuvent être interpolés. Cependant cette option n'a pas été retenue car elle peut entraîner des modifi-

cations audibles. Il ne semble donc pas possible de créer un modèle à partir de différentes bases de données, c'est-à-dire un modèle *transverse* aux bases et ceci pose le problème de l'universalité des modèles de prédiction binauraux. De plus, vu les ressources nécessaires à la mesure de HRTF, la création d'une base comportant 1000 sujets (nombre d'instances parfois nécessaire à un système de prédiction) est difficilement envisageable.

Dans le processus de construction d'un modèle de prédiction il est alors intéressant d'estimer la capacité de généralisation. Par exemple, les données sont partagées en deux ensembles : un pour la création du modèle, l'autre pour sa validation. L'erreur introduite par le modèle est calculée sur le premier ensemble et le deuxième ensemble permet d'évaluer la capacité du modèle à s'adapter à des nouvelles instances qui n'ont pas servi à la construction du modèle. Cette approche est notamment utilisée dans la création d'un réseau de neurones (cf. chapitre 2).

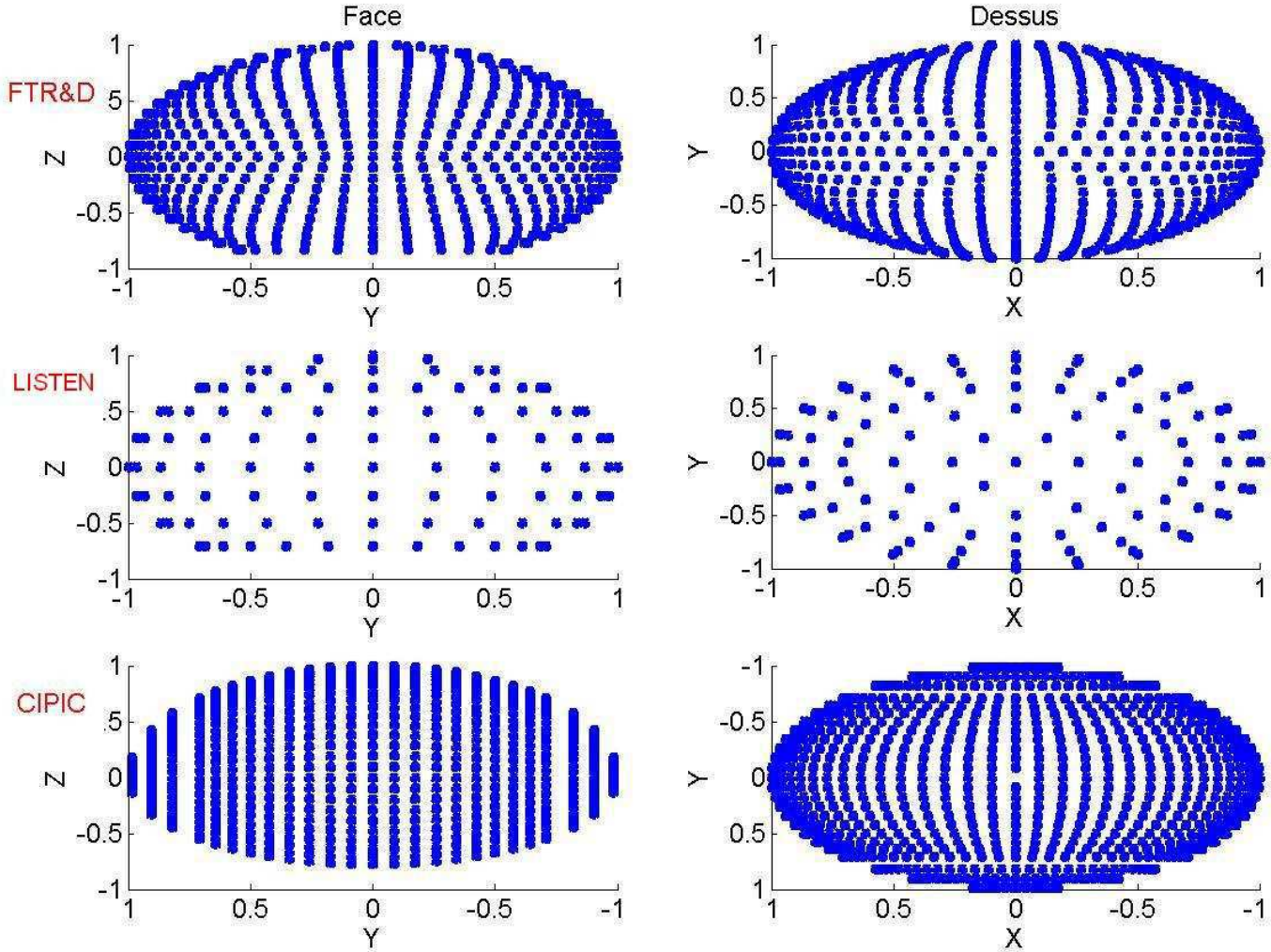


Fig. III.8 – Points de mesure de trois bases de données de HRTF. Figures de gauche : vue de face, Figures de droite : vue du dessus. Les points de mesures ont été ramenés à un rayon unitaire.

1.4 La mesure de HRTF est un problème qui reste d'actualité

Que ce soit pour la validation perceptive ou pour la modélisation de HRTF, et malgré les sources d'erreurs inhérentes, les mesures sont nécessaires car elles sont la seule signature des informations qui arrivent aux tympans. De plus, les tests perceptifs nécessitent la disponibilité d'un grand nombre de sujets. Une autre alternative est l'utilisation de têtes artificielles (cf. fig.I.I.23). Elles permettent une meilleur reproductibilité des mesures. Des séances de mesures plus longues sont réalisables et le nombre de points de mesures peut être augmenté. Cependant l'utilisation de HRTF non-individuelles entraîne des artefacts perceptifs.

Les bases de données de HRTF continuent de se constituer et les systèmes de mesure sont un sujet d'étude. Dans [Zotkin et al. (2004)] un système de mesure utilisant une formulation réciproque de la mesure de HRTF est décrit : un haut-parleur miniature est placé dans les oreilles du sujet et une matrice sphérique de microphones est disposée autour du sujet. Ce système permet de mesurer jusqu'à 64 HRTF simultanément (cf. fig. III.9). Les principales limitations de ce système sont liées à la taille des haut-parleurs. Les mesures ne sont valides qu'à partir de 2.5 kHz et le niveau de sortie faible conduit à un faible rapport signal à bruit.

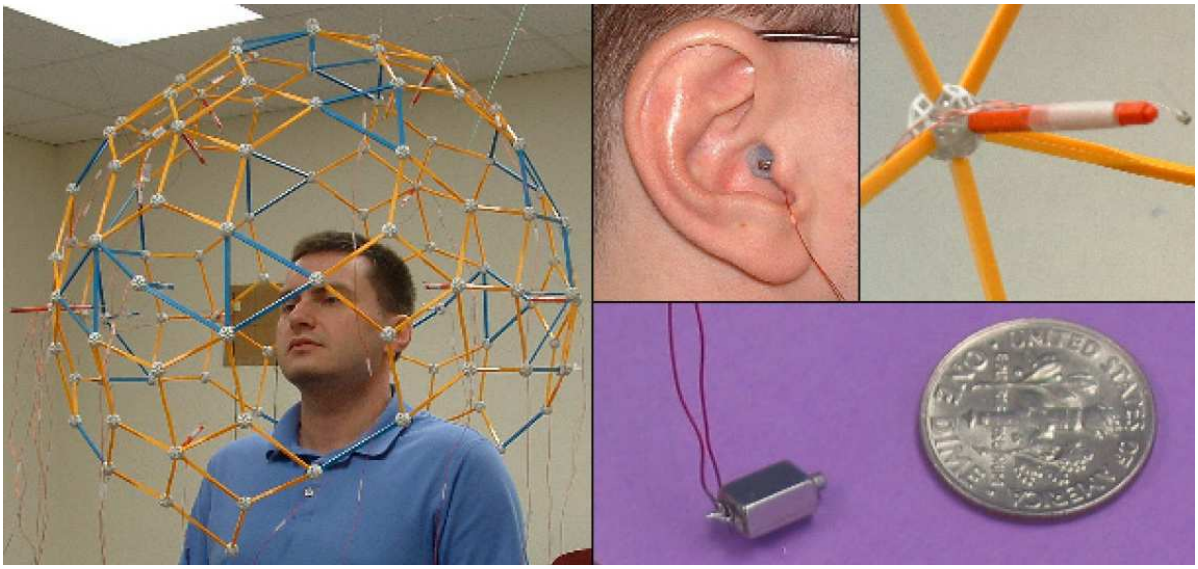


Fig. III.9 – Système de mesures réciproque de HRTF [Zotkin et al. (2004)].

Le calcul de HRTF par BEM représente une alternative à la mesure. L'acquisition de grandes bases de données est plus facilement réalisable et la répétabilité est assurée. Cependant, la BEM se base sur une modélisation physique de l'auditeur. La puissance de calcul requise pour obtenir les HRTF d'une représentation fine, telle que la forme compliquée du pavillon, sont considérables. Une modélisation qui utilise des formes géométriques simples, comme des sphères, demande moins de ressources informatiques. Il convient alors de tester la capacité de modélisation de formes simples. La suite du chapitre décrit le travail d'application de la BEM aux calculs de HRTF. Des formes géométriques simples sont utilisées pour rendre compte de la morphologie d'un auditeur. Un nouveau modèle est calculé et permet de mieux appréhender la plage de travail de la BEM pour le calcul HRTF.

2 ACQUISITION DE HRTF PAR MODÉLISATION NUMÉRIQUE

INTRODUCTION

Un rendu optimal de la synthèse binaurale est assuré en utilisant des HRTF individuelles. Un des moyens permettant de s'affranchir de la lourde tâche de mesure de HRTF, est la modélisation numérique du problème acoustique correspondant. Plusieurs méthodes numériques sont disponibles au sein desquelles plusieurs formulations existent. A la différence des mesures acoustiques, ces méthodes garantissent la reproductibilité des résultats. De plus ces techniques permettent le placement des microphones et des sources avec une très grande précision, sans souci d'encombrement (microphone placé dans le conduit auditif). Enfin, il est possible de contrôler l'impédance des surfaces, les caractéristiques du fluide entourant l'objet d'analyse. Ces conditions idéales permettent un travail systématique sur l'influence des paramètres de modélisation de l'objet d'analyse. Les modèles étudiés dans la présente étude sont composés de formes simples : sphère, cylindre, ellipsoïde. Les solutions analytiques existent pour ces géométries car les fonctions propres des systèmes de coordonnées sphériques, cylindriques et ellipsoïdales sont connues. Mais leur utilisation peut être complexe et les solutions analytiques pour des combinaisons de formes simples sont difficiles à obtenir.

L'objet des paragraphes suivants est de présenter le travail réalisé grâce à ces méthodes. L'étude ayant été faite grâce à la technique BEM, une plus grande partie lui sera consacrée. La première partie est consacrée à l'énoncé du problème et au principe de la BEM. Ensuite, les principes de trois autres méthodes couramment utilisées pour le calcul de HRTF sont expliqués. La quatrième partie de ce chapitre analyse les aspects pratiques et met en évidence les limites techniques de la BEM pour la prédiction de HRTF. La cinquième partie est consacrée à l'exposé des résultats des travaux antérieurs. Ensuite, la sixième partie présente les travaux réalisés au cours de cette étude. Enfin, la dernière partie conclura sur l'utilisation de la BEM pour la prédiction de HRTF.

2.1 Principe de la modélisation par éléments de frontières

La BEM (Boundary Element Method), ou méthode par élément de frontière, est une méthode de calcul numérique très employée pour la résolution de problèmes de diffraction. La méthode permet l'obtention du champ de pression dans un volume à partir du calcul de la pression à la surface entourant le volume. Les solutions sont obtenues par la résolution de la formulation intégrale de l'équation d'onde. Pour le calcul de HRTF, le problème physique équivalent est la diffraction d'une onde par un volume représentant la tête et éventuellement le cou, le torse et l'oreille externe.

2.2 Enoncé du problème

Le principe de la BEM est de remplacer un problème différentiel aux limites du domaine d'étude par un problème intégral sur les bords du domaine.

Soit un domaine D fermé et délimité par une surface S_D contenant un fluide homogène, la variation de pression $p(M,t)$ pour un point M appartenant à D de coordonnées

(x, y, z) au temps t est donnée par l'équation d'onde :

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \Delta p = f \tag{III.4}$$

avec c la vitesse du son dans le fluide, Δ l'opérateur de Laplace et f terme de source acoustique.

L'équation d'onde (III.4) décrit le comportement de l'onde acoustique à l'intérieur du domaine D. Par contre rien n'est connu du comportement de l'onde sur S_D, c'est-à-dire sur les parois entourant le fluide. Il est alors nécessaire de définir des conditions de surface qui décrivent le comportement de l'onde sur S_D. Dans le cas particulier du problème acoustique lié à la modélisation de HRTF, un volume V délimité par une surface S est introduit : ce volume représente la modélisation du corps de l'auditeur. Ainsi la surface délimitant le domaine D est la réunion de S et de S_U qui délimite le domaine d'étude (cf. fig.III.10). En pratique, S_U est rejetée à l'infini et les conditions de type Sommerfeld sont appliquées (pas de réflexion de l'onde sur S_U). Il est souvent plus

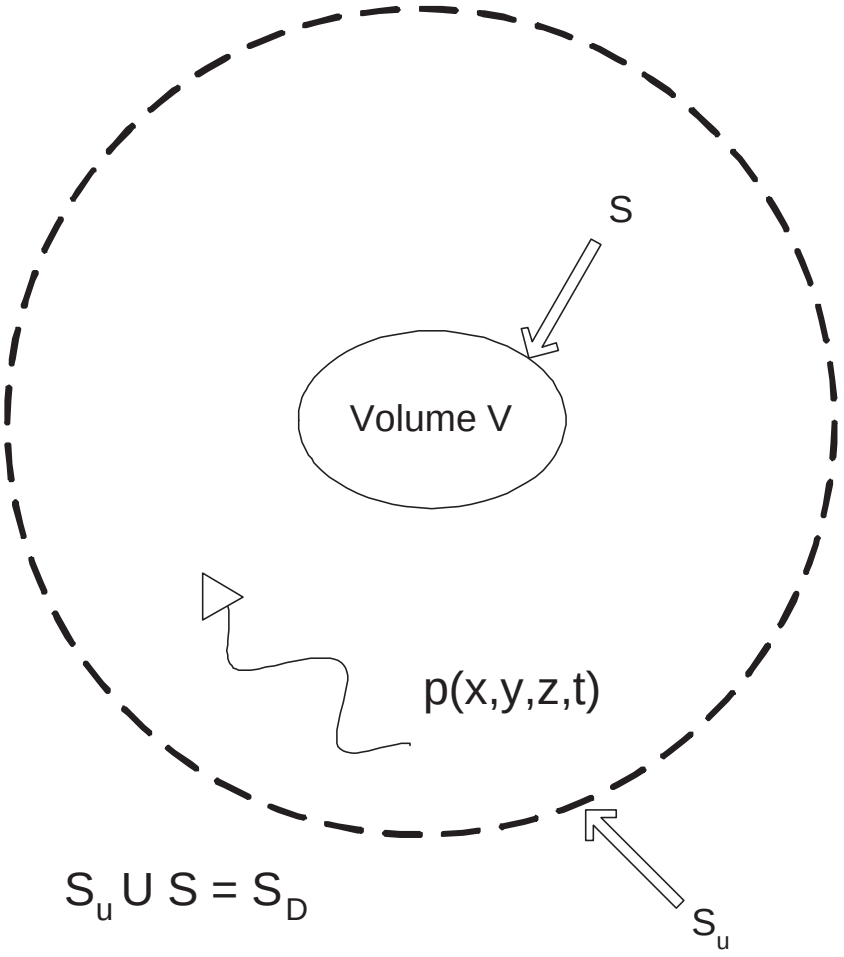


Fig. III.10 – Schéma du domaine D d'étude. Le volume D représente une modélisation de l'auditeur.

pratique d'exprimer le problème en régime harmonique. L'expression du problème avec

les conditions de surfaces devient alors :

$$\begin{cases} \Delta\mathbf{p} + k^2\mathbf{p} = -f & \forall\, M \in V \\ \frac{\partial \mathbf{p}}{\partial n_0} + jk_0\beta\mathbf{p} = U_0 & \forall\, M \in S \end{cases} \tag{III.5}$$

où $k = \frac{2\pi f}{c}$ est le nombre d'onde, β désigne l'admittance spécifique des parois et $\frac{U_0}{\mathbf{j}k_0\beta}$ la vitesse vibratoire imposée à la surface.

2.3 Résolution analytique

Comme le problème est linéaire (cf. équation III.5), la solution peut être recherchée comme la superposition de champs élémentaires créés par chaque élément de source f. Le problème élémentaire associé à un élément de source est le suivant :

$$\begin{cases} [\Delta + k^2]\chi(\vec{r}, \vec{r_0}; \omega) & = -\delta(\vec{r} - \vec{r_0}) & \forall\, M \in V \\ \frac{\partial \chi}{\partial n} + \mathbf{j}k_0\beta\chi & = 0 & \forall\, M \in S \end{cases} \tag{III.6}$$

avec $\delta(\vec{r} - \vec{r_0})$ source élémentaire située en $\vec{r} = \vec{r_0}$, avec $\vec{r}$ les coordonnées d'un point de D et $\vec{r_0}$ les coordonnées d'un point de S, dont la solution χ en espace infini est la fonction de Green :

$$G(\vec{r}, \vec{r_0}) = \frac{e^{\mathbf{j}kr}}{4\pi r} \tag{III.7}$$

avec $r = \mid \vec{r} - \vec{r_0} \mid$.

La formulation intégrale du problème III.5 est alors obtenue grâce à la fonction de Green [Bruneau (1998)] :

$$p(\vec{r}) = \iiint_D G(\vec{r}, \vec{r_0}) f(\vec{r_0})\, \mathrm{d}D_0 + \iint_S \left[\frac{\partial G(\vec{r}, \vec{r_0})}{\partial n_0} p(\vec{r_0}) - G(\vec{r}, \vec{r_0}) \frac{\partial p(\vec{r_0})}{\partial n_0} \right] \mathrm{d}S_0 \tag{III.8}$$

où le terme $\frac{\partial}{\partial n_0}$ désigne la dérivée par rapport à $\vec{r_0}$. La solution s'exprime alors comme la superposition de sources réparties dans D et de sources localisées sur les bords du domaine. Les *sources de frontières* se présentent sous la forme d'une combinaison linéaire de G (source monopolaire) et $\frac{\partial G}{\partial n_0}$ (source dipolaire) [Bruneau (1998)]. En considérant une source n'émettant plus, le champ de pression peut alors être écrit comme suit :

$$p(\vec{r}) = \iint_S \left[\frac{\partial G}{\partial n_0} p - G \frac{\partial p}{\partial n_0} \right] \mathrm{d}S_0 \tag{III.9}$$

Une attention particulière doit être portée à la direction de la normale à S qui définit l'intérieur du volume V. Ainsi pour les problèmes internes à V, la normale doit pointer vers l'intérieur du volume alors que pour les problèmes externes, modélisation de HRTF par exemple, elle doit pointer vers l'extérieur. En BEM, les normales aux surfaces doivent toutes avoir la même orientation. Pour le calcul de HRTF, les normales doivent indiquer l'extérieur de la surface : il s'agit donc d'un problème externe.

2.4 Résolution numérique

Pour résoudre numériquement l'équation III.9, les expression de p et $\frac{\partial p}{\partial n_0}$ sur S sont discrétisées :

$$\begin{cases} p = \sum_{i=1}^{n_p} p_i f_i \\ \frac{\partial p}{\partial n_0} = \sum_{i=1}^{n_p} \frac{\partial p}{\partial n_0 i} f_i \end{cases} \tag{III.10}$$

avec f_i fonction d'approximation du champ de pression, n points de discrétisation des inconnues, ou *éléments de frontière*. L'ensemble des points de dicrétisation de la surface forme un **maillage** surfacique du volume d'étude. Il faut alors évaluer pour chaque point de la discrétisation l'équation (III.9) :

$$p(\vec{r}) = \sum_{i=1}^{n} \int_S \left[\frac{\partial G}{\partial n_0} p_i - G \frac{\partial p}{\partial n_0 i} \right] \mathrm{d}S_0 \tag{III.11}$$

En introduisant des conditions aux frontières de type Neumann :

$$\frac{\partial p}{\partial n_i} = -j\rho\omega V_i \tag{III.12}$$

avec V_i la vitesse vibratoire de l'élément de surface S_i, il ne reste plus alors que les n $\mathbf{p_i}$ inconnues. Chaque p_i est ensuite exprimée dans un système de coordonnées locales et grâce à des fonctions d'interpolation du champ de pression. Le système d'équation donnant accès aux p_i peut être écrit de manière matricielle [Ciscowski and Brebbia (1991)] :

$$\{p\} = [K^1]\{p\} - [K^2]\{V\} \tag{III.13}$$

où $\{p\} = \{p_1, p_2, \ldots, p_n\}$, $\{V\} = \{V_1, V_2, \ldots, V_n\}$, $[K^1]$ et $[K^2]$ correspondent respectivement aux évaluations de $\int_S G$ et $\int_S \frac{\partial G}{\partial n_0}$ dans le système de coordonnées locales. Le travail principal des codes de calcul BEM est d'estimer et surtout d'inverser les matrices $[K^1]$ et $[K^2]$.

2.5 Choix des conditions aux limites

Pour résoudre un problème de propagation d'onde, les conditions aux limites sont indispensables. Elles traduisent les effets d'interaction de l'onde sonore avec la surface du modèle d'étude et apportent les équations nécessaires à la résolution du problème. Les modèles étudiés ici, décrivent la morphologie humaine de la tête et du haut du corps et sont utilisés pour rendre compte des effets de diffraction d'une onde sonore sur un auditeur. Dans ce problème, le volume discrétisé n'émet pas d'onde sonore, n'a pas de mouvement et ne résonne pas. Le choix d'une condition aux frontières se fera donc sur la base du choix d'une impédance de surface pour tenir compte, ou pas des effets de l'impédance de la peau et/ou des cheveux. De plus, le choix d'une impédance complexe pour la peau peut permettre d'atténuer les effets d'artefacts mathématiques liées à la méthode BEM. Ce paragraphe décrit deux sortes de conditions aux limites :

- Impédance infinie
- Impédance complexe

2.5.1 Condition de Neumann

Cette condition a été utilisée pour décrire la démarche de mise en forme matricielle de la technique BEM. Elle décrit des éléments de frontière soumis à une vitesse vibratoire :

$$\frac{\partial \mathbf{p}}{\partial n} = -j\rho\omega \mathbf{V_n} \tag{III.14}$$

Cette équation relie la dérivée normale à la surface de la pression acoustique à la vitesse vibratoire normale de la structure. Dans le cas ou la structure est fixée, la vitesse vibratoire est nulle et la relation III.14 devient :

$$\frac{\partial \mathbf{p}}{\partial n} = 0 \tag{III.15}$$

Cette relation sera utilisée en première hypothèse. Elle permet la simplification des relations matricielles. Elle correspond à considérer la partie du corps en question comme complètement rigide.

2.5.2 Condition de Robin

Dans une étude sur la modélisation de HRTF par BEM [Katz (2001b)], B. Katz a introduit l'impédance des cheveux dans un modèle BEM (l'impédance de la peau est considérée comme rigide). Le maillage a été obtenu par un scanner laser sur une personne et modifié pour les spécificités de la BEM. Pour prendre en compte l'effet des cheveux, il a utilisé un modèle simple : il définit une impédance pour une zone du maillage correspondant à la partie de la tête du sujet recouverte de cheveux. Bien que peu réel[5], ce modèle a montré que l'effet des cheveux sur le spectre des HRTF dépend de la fréquence et de la position. Son influence majeure se situe pour des HRTF contralatérales : par rapport à un modèle rigide, les cheveux atténuent le signal pour $f < 3.5kHz$ et l'amplifie pour $f > 3.5kHz$. Bien que l'influence de l'impédance des cheveux et de la peau semble peu modifier le spectre des HRTF, rajouter une condition d'impédance peut faciliter la convergence du calcul et atténuer l'effet des artefacts mathématiques.

2.6 Fréquences irrégulières en problème externe

Lorsque l'on travaille avec un problème externe, des artefacts mathématiques peuvent apparaître dans la résolution du problème. L'apparition des ces artefacts dépend du degré de symétrie du volume diffractant et de sa taille. Ces artefacts se traduisent par des valeurs de pression erronées, voire des impossibilités de recherche de solution pour certaines fréquences. Ces fréquences sont appelées *fréquences irrégulières* et correspondent aux fréquences des modes de résonance du volume diffractant en considérant le problème dual au problème externe. Par exemple, si l'on considère un problème externe avec des conditions de Neumann nulles aux frontières ($\frac{\partial p}{\partial n} = 0$), alors les fréquences irrégulières apparaîtront pour les résonances du problème interne en condition de Dirichlet $p = 0$. Ce problème est dû à la non unicité de la formulation intégrale en problème externe [Schenck (1968)].
Principalement deux méthodes ont été développées pour tenter de résoudre ce problème. L'approche de Burton et Miller [Burton (1973)] consiste à écrire une combinaison linéaire de deux équations pour chaque noeud : l'équation de Helmholtz (III.5) et la dérivée normale de l'équation de Helmholtz. Cette approche permet de travailler avec un système d'équation de Kramer et le problème peut être résolu avec des méthodes classiques. Cependant, l'intégrale de la dérivée normale peut être délicate à calculer et il n'y a pas

[5]Comme le préconise Katz, il aurait fallu définir un fluide avec l'impédance des cheveux entourant une partie de la tête et ce pour tout angel d'incidence, mais le logiciel utilisé pour l'étude ne permettait que la définition d'incidence non normale

de méthodes formelles pour trouver les coefficients de la combinaison linéaire des deux équations. L'autre approche est appelée méthode *CHIEF* pour Combined Helmholtz Integral Equation Formulation. Cette méthode consiste à placer des noeuds supplémentaires à l'intérieur du volume et à y adjoindre la condition $p = 0$. De cette manière, et si les points rajoutés ne sont pas placés sur un noeud de pression d'un des modes du volume, la nullité de la pression interne est assurée. Les avantages principaux de la méthode CHIEF sont sa rigueur théorique et sa simplicité d'emploi. Par contre, l'ajout de noeuds au problème rend le système sur-déterminé ce qui complique l'étape d'inversion des matrices.

2.7 Autres méthodes de modélisation numérique

L'objet de ce paragraphe est de présenter brièvement les autres méthodes de résolution numérique d'un problème de diffraction acoustique. L'objet de l'étude n'est ni de déterminer la faisabilité du calcul de HRTF avec ces méthodes, ce qui a été largement fait par Katz [Katz (1998)] et Kahana [Kahana (2000)], ni d'élire la meilleure technique en terme de rapidité, de précision ou de robustesse.

2.7.1 Méthode indirecte par éléments de frontière

La méthode décrite précédemment, précisément appelée Direct Collocationnal BEM, ne peut être utilisée que pour les surfaces closes : le problème est alors soit intérieur, soit extérieur au volume V délimité par S. La méthode IBEM, Indirect Boundary Element Method, pour permet de travailler avec des surfaces ouvertes ou fermées. Ceci peut s'avérer particulièrement avantageux pour le calcul de diffraction sur un volume à symétrie de révolution : une intégrale est calculée sur un arc et le résultat pour tout le volume est obtenue par rotation [Kahana (2000)]. La méthode indirecte travaille avec des potentiels de couche qui sont différents de part et d'autre de la surface. Les quantités situées sur l'extérieure de V sont notées p^+ et $\frac{\partial p^+(r_0)}{\partial n}$ et les quantités à l'intérieur de V sont notée p^- et $\frac{\partial p^+(r_0)}{\partial n}$. Les potentiels de couches sont définis comme suit :

$$\mu(r_0) = p^+(r_0) - p^-(r_0)$$

μ est appelé *saut de pression* ou *potentiel de double couche* et représente une distribution de sources dipolaires sur la surface S. Pour la dérivée normale de la pression, le potentiel de couche est :

$$\sigma(r_0) = \frac{\partial p^+(r_0)}{\partial n} - \frac{\partial p^-(r_0)}{\partial n}$$

σ est généralement appelé *saut de la dérivée normale de la pression* ou *potentiel de simple couche* et représente une distribution de sources monopolaires sur la surface S. La formulation intégrale (III.9) prend alors la forme :

$$p(\vec{r}, t) = \int_S \left[\frac{\partial G}{\partial n} \mu(r_0) - G\sigma(r_0) \right] \mathrm{d}S \tag{III.16}$$

Le problème des fréquences irrégulières est traité en ajoutant des éléments de surface avec des conditions de frontière absorbante.

2.7.2 Méthode par éléments finis

La méthode FEM requiert la division du volume d'étude V en un grand nombre de petits éléments de volume. Cette étape est appelée *discrétisation du domaine*. Le champ de pression est discrétisé sur tout le volume d'étude et non plus sur la surface. Les points de dicrétisation doivent être proches les uns des autres pour obtenir une description satisfaisante de la distribution de pression dans V. Une équation est obtenue pour chaque point de dicrétisation et le système d'équations ainsi créé peut-être résolu. Le résultat est alors la pression pour chaque point. Il existe un grand nombre de méthodes à éléments finis et toutes utilisent la discrétisation du volume d'étude. La technique FEM permet la définition de couplage entre le fluide entourant la surface d'étude et l'objet immergé.

Cette formulation pourrait être utilisée pour prendre en compte l'influence des cheveux sur les HRTF. Les cheveux pourraient être modélisés comme un fluide entourant la tête. Cependant, la définition d'un volume sphérique entourant la tête d'un rayon d'environ 1 m, correspondant aux points de mesures de HRTF, rend cette technique difficilement applicable pour le calcul de HRTF.

2.7.3 Méthode par éléments finis étendue à l'infini

La technique FEM est utilisée pour résoudre des problèmes intérieurs. Cependant il existe des formulations qui permettent d'étendre les solutions à un domaine extérieur infini, ce qui est requis pour la définition de conditions en champ libre. Le volume extérieur est alors divisé en deux. Le champ proche est modélisé par un maillage FEM comme décrit dans le paragraphe précédent, tandis que le champ lointain est modélisé avec une couche simple d'éléments spéciaux s'étirant à l'infini. Les fonctions d'interpolation pour ces éléments spéciaux doivent décroître à l'infini[6].

2.8 La modélisation par éléments finis en pratique

Comme il a été décrit dans les paragraphes précédents, de nombreuses solutions techniques sont disponibles pour la prédiction de HRTF. L'élaboration d'un code de calcul n'étant pas un axe de recherche retenu pour cette étude, la solution technique choisie est alors celle des logiciels de calcul à disposition. Au cours de ces travaux deux logiciels ont été retenus pour le calcul et un pour la création des maillages :

- un logiciel commercial : VNOISE2.0
- un code de recherche : Front3D
- un logiciel commercial pour la creation des maillages : 3DSMAX4.2

Cette étude a donc commencé par une validation des codes de calculs[7] grâce à des résultats de Kahana publiés sur internet[8] et à la formulation analytique de la diffraction d'une onde plane sur une sphère.

[6]De nombreuse autres formulations existent. Le lecteur pourra se reporter à *Acta Acustica United with Acustica, vol 89, pp 1-85, 2003*

[7]La validation du logiciel VNOISE a déjà été réalisée dans [Pernaux (2003)] mais pour des versions plus anciennes (1.0 et 1.1)

[8]http ://www.isvr.soton.ac.uk/FDAG/VAP/index.htm

2.8.1 Le mailleur

Un maillage est une surface discrétisée. Le but de l'étude étant de connaître les performances d'individualisation de modèles géométriques simples (pour une paramètrisation de la synthèse binaurale) l'utilisation de techniques de maillages, telles que le scanner laser, donnant accès à des détails très fins n'a pas été retenue. De plus, les maillages issue d'un scanner nécessitent des manipulations pour rendre le maillage viable en entrée d'un code de calcul BEM [Katz (1998); Kahana (2000)]. Les éléments du maillage doivent être des formes géométriques simples et connues (triangulaire de préférence) pour la définition des fonctions d'interpolation. Enfin, la résolution du maillage doit tenir compte de plusieurs contraintes reliées entre elles et qui définissent le cadre d'application de la technique BEM pour le calcul des HRTF. Trois éléments fondamentaux sont alors à prendre en considération :

- la limitation fréquentielle du calcul
- la taille du modèle
- les capacités de stockage mémoire disponible

Limitation fréquentielle du calcul La fréquence maximale est déterminée par la finesse du maillage. Pour reconstruire un signal issu de calculs par éléments de frontière, la pratique recommande une résolution de 4 à 10 noeuds par longueur d'onde [Katz (1998); Kahana (2000)]. Cette résolution dépend des fonctions d'approximation du champ de pression à la surface. Ainsi, si une résolution de 6 éléments par longueur d'onde, notée $\lambda/6$, est utilisée et que la taille maximale du maillage entre deux noeuds est de 14 mm, la fréquence maximale de calcul est :

$$f_{max} = \frac{340}{0.014 * 6} = 4046Hz \tag{III.17}$$

Katz indique que l'on peut travailler avec une condition de $\lambda/4$ pour des positions frontales, $\lambda/6$ pour des positions latérales et $\lambda/10$ pour des positions contralatérales où peu d'énergie est disponible. Cette condition imposée par le calcul en éléments de frontière est prépondérante dans la définition d'un maillage.

Le mailleur utilisé pour cette étude est 3DSMAX4.2. Ce logiciel est dédié à la construction d'environnement graphique en trois dimensions (jeu, clip, film). La définition de maillage y est très rapide. Des vérifications sont parfois nécessaires, comme l'absence d'éléments pathologiques (éléments non triangulaires, etc...). La distance inter-noeuds doit être mesurée car non paramètrable. L'exportation d'un maillage du mailleur vers le logiciel de calcul BEM se fait par l'intermédiaire de fichiers textes où sont listées les coordonnées des noeuds ainsi que le numéro des noeuds définissant chaque élément.

La taille du modèle De la taille du modèle dépend le nombre de noeuds. Plus les surfaces discrétisées sont grandes, plus le nombre de noeuds augmente et ce de manière non-linéaire. Par exemple, une sphère de rayon 16 cm avec une condition de $\lambda/6$ pour une fréquence maximale de 12000 Hz avec des éléments triangulaires fait 12252 noeuds tandis qu'une sphère de 8 cm avec les mêmes conditions fait 3242 noeuds. Ainsi, l'ajout d'un torse, qui est modélisé par une ellipsoïde, abaisse notablement la fréquence maximale de calcul.

La capacité de stockage mémoire Les calculs en éléments finis, une fois un maillage réalisé et les coefficients issus de l'intégrale de Green calculés, peuvent se résumer à un calcul d'inversion de matrice. Dans le cas d'une formulation directe de la BEM avec calcul de la pression à chaque noeud du maillage, n_n étant le nombre de noeuds, la matrice à inverser est de taille $n_n * n_n$. La précision nécessaire aux calculs requiert l'utilisation de coefficients de type *double*, ce qui correspond à 8 octets par élément de matrice. Les calculs s'effectuent de plus avec des coefficients complexes (les HRTF sont des fonctions complexes), donc les calculs demandent 16 octets par élément de matrice. La taille mémoire pour stocker une matrice de 10 000 * 10 000 noeuds est donc : $10000^2 * 16 = 1,6$ Go. Le système d'exploitation utilisé est Windows XP, que ce soit pour Front3D ou VNOISE2.0, ce qui autorise au maximum 2 Go dédiés pour une application. Les limites de mémoire vive, RAM (Read Only Memory), sont atteintes assez rapidement surtout pour des modèles disposant de grandes surfaces (torse).

2.9 Les codes de calculs

Chaque code utilise des formulations différentes que ce soit pour la mise en forme matricielle du problème acoustique à résoudre, pour la gestion des fréquences irrégulières ou pour l'inversion de la matrice. Il est donc intéressant de comparer les deux formulations.

VNOISE2.0 Logiciel développé par la société STS[9], VNOISE2.0 est un logiciel dédié aux calculs BEM. Seules les fonctionnalités de calcul, d'importation et d'exportation de données ont été utilisées. VNOISE2.0 utilise la formulation direct de la BEM. L'inversion de la matrice se fait de manière itérative en utilisant des algorithmes de calcul basés sur la méthode des sous-espaces de Krylov [Francescantonio (2003)]. Outre un gain éventuel de vitesse dans l'inversion de la matrice, VNOISE2.0, grâce aux solveurs itératifs, propose une parallèlisation du calcul : le calcul est réparti sur plusieurs machines reliées par un réseau (5 au maximum). Les ressources mémoires des machines sont alors additionnées, la rapidité du calcul est alors largement augmentée. VNOISE2.0 propose aussi d'utiliser non plus de la RAM pour stocker les coefficients de la matrice mais de spécifier un espace disque. Ceci présente l'avantage de pouvoir travailler avec un nombre de noeuds beaucoup plus important qu'avec de la RAM. Cependant, l'accès en écriture d'un disque dur est 1000 fois moins rapide que l'accès à la RAM : les temps de calcul en configuration *disk solver* sont prohibitifs (environ 3 semaines de calcul pour un maillage de 12000 noeuds). Un inconvénient à l'utilisation d'algorithmes itératifs est que le calcul doit être recommencé pour chaque fréquence et pour chaque position de mesure.

Les positions testées lors de cette étude étant les mêmes que celles des mesures de la base de HRTF TNO, 965 au total, et en considérant qu'un calcul d'un maillage de 2000 noeuds demande 5 heures, le temps de calcul pour toutes les positions atteint vite des proportions déraisonnables. Pour résoudre ce problème, le principe de réciprocité [10] applicable aux équations de l'acoustique a été utilisé pour réduire considérablement le temps de calcul. Ainsi, si le problème initial était la diffraction d'une onde plane sur le modèle pour des points sources situés sur une sphère de 1,40 m de rayon (comme pour

[9] www.sts-soft.com

[10] La pression produite en un point A par une source située en B est égale à la pression en B produite par la même source située en A

les mesures de HRTF) centrée sur le modèle et le point de mesure situé au niveau du canal auditif, le problème réciproque est la diffraction d'une source monopolaire située au niveau du canal auditif avec des points de mesure situés sur une sphère de 1,40 m de rayon [11]. A la distance de 1,40 m, l'onde sonore est considérée comme plane. Ainsi le calcul pour 965 points de mesure nécessite l'ajout de 965 noeuds supplémentaires au modèle. La formulation réciproque nécessite des modifications du maillage. Kahana a mis en évidence que la proximité de la source avec le maillage entraîne des erreurs de calcul. C'est pourquoi il faut raffiner le modèle à proximité de la source : la distance inter-noeuds doit être de l'ordre de grandeur de la distance source-surface.

Les résultats de la formulation réciproque doivent aussi être égalisés afin d'éliminer les contributions de la chaîne de mesure [Kahana (2000)]. Pour les mesures de HRTF, l'égalisation par une mesure en champ libre est parfois utilisée. Cette égalisation consiste à mesurer la réponse au centre de la tête en l'absence de l'auditeur pour une incidence donnée et de diviser toutes les mesures par cette mesure en champ libre. Pour la formulation réciproque, le champ libre consiste à placer une source monopolaire au centre du repère et de calculer la fonction de transfert pour les 965 points de mesure. Seulement, ne travaillant qu'avec des sources et des microphones idéaux, cette opération consiste à soustraire un gain constant correspondant à l'atténuation de l'onde et un retard pur pour le temps de trajet.

Front3D Issu d'une collaboration entre France Telecom R&D et le Laboratoire d'Acoustique et de Mécanique UPR7051 du CNRS, le logiciel Front3D programmé en Fortran90 réalise des calculs BEM utilisant la formulation directe. La principale différence entre VNOISE2.0 et Front3D est que ce dernier calcule la pression non plus pour chaque noeud mais pour chaque élément du modèle. De plus, Front3D inverse complètement la matrice de calcul et donc un calcul fournit la pression en différents points de mesure simultanément : la formulation réciproque n'a donc pas été implémentée. L'écriture en Fortran90 permet l'utilisation de librairies mathématiques optimisées et gratuites telles que Lapack. Le calcul sur les éléments du modèle et non sur les noeuds a l'inconvénient de multiplier les ressources mémoires nécessaires par 4. En effet, le nombre d'éléments d'un maillage est à peu près deux fois plus grand que son nombre de noeuds. Ainsi pour un calcul avec un maillage de 5 000 noeuds, la taille des matrices à inverser est d'environ 1.6 Go de RAM. Le problème des fréquences irrégulières n'est pas spécifiquement traité, mais les algorithmes utilisés semblent plus robustes que les solveurs itératifs. Une version de ce logiciel est aussi disponible avec une intéraction fluide-fluide, en considérant la tête comme de l'eau, mais cela nécessite le calcul de vitesse acoustique des éléments et donc requiert 4 fois plus de RAM.

2.10 Travaux antérieurs

La formulation de la méthode BEM est apparue au cours du 20^e siècle et a été d'abord utilisée pour résoudre des problèmes en électromagnétisme. Cette méthode a connu un essor considérable durant les années 80 et 90 grâce à l'avènement des micro-ordinateurs et s'est étendue à de nombreux domaines de la physique. L'évolution rapide de la puissance des ordinateurs ainsi que leurs capacités de stockage a permis l'étude

[11] Des bases de données de HRTF ont été mesurées de cette manière [Zotkin et al. (2004)]

de modèles physiques de plus en plus complexes. Pour un historique des applications de la BEM à l'acoustique, le lecteur pourra consulter l'ouvrage de Ciscowski et Brebbia [Ciscowski and Brebbia (1991)]. En ce qui concerne la modélisation de HRTF par BEM, Weinrich [Weinrich (1984)] fut le premier à montrer la validité de la méthode. Il a calculé la réponse en champ proche d'un modèle de tête de 212 éléments, sans pavillon qu'il a comparé avec des mesures sur une réplique du maillage. Etant donnée la résolution du maillage (5 cm entre chaque noeud) la fréquence maximale de calcul est de 1,7 kHz avec la condition de 4 noeuds par longueur d'onde. La comparaison entre mesure et simulation est satisfaisante. Malgré l'utilisation d'un modèle grossier sans pavillon, et donc une simulation de HRTF bien éloignée d'une HRTF de tête humaine, ce travail a montré la validité de la méthode.

Poursuivant la même approche, c'est-à-dire l'étude de l'influence des différentes parties du corps sur les HRTF grâce à des modélisations de la morphologie d'un auditeur, les travaux de Katz [Katz (1998)] ont montré que la BEM pouvait travailler avec des maillages aux formes complexes comme le maillage de la tête d'un auditeur. Des maillages d'une grande précision (valable jusqu'à 57 kHz) ont été réalisés sur la base de scan laser 3D cylindrique d'un ensemble tête + cou. Se limitant à 5,4 kHz comme fréquence maximale de calcul à cause de ressources informatiques limitées, il a mis en évidence par la BEM l'influence du pavillon externe et l'influence de l'impédance des cheveux et de la peau sur les HRTF. La peau peut être facilement assimilée à une surface totalement réfléchissante dans la bande de fréquences audibles, par contre les cheveux, malgré un modèle d'impédance simpliste, ont une influence pouvant atteindre 6 dB (pour l'oreille contralatérale). Il a aussi étudié la réponse d'un modèle de tête sphérique jusqu'à 6 kHz.

Les travaux de Kahana [Kahana (2000)] ont permis une meilleure compréhension des techniques d'éléments finis et de frontière pour la prédiction des HRTF. Il a effectué des comparaisons entre BEM, IBEM, FEM et IFEM. Ainsi, si la méthode FEM est difficilement applicable aux calculs de HRTF en raison de la complexité du maillage volumique, BEM, IBEM et IFEM semblent bien appropriées. La BEM est plus rapide que la IBEM mais cette dernière nécessite moins d'espace mémoire. Il a aussi travaillé sur des formulations permettant la réduction des coûts de calculs. Le principe de réciprocité a été appliqué et validé : une source virtuelle est placée à proximité de la position du canal auditif du modèle et les points de mesures sont des microphones virtuels autour du modèle. Un raffinement local du maillage à proximité de la source virtuelle doit alors être apporté. Une égalisation en champ libre est nécessaire pour se replacer dans les conditions de mesure. L'application du principe de réciprocité permet un gain de temps considérable : un seul calcul donne la réponse pour toutes les positions souhaitées. De plus il a montré qu'il était possible, dans le cas de modèles symétriques (par rapport à un axe ou un plan) de réduire le temps de calcul en n'utilisant qu'une partie du modèle. Grâce à ses méthodes optimisées et aux progrès de l'informatique, il a calculé les réponses de modèles plus complets : tête ellipsoïdale et maillage issu d'un scanner laser d'une tête artificielle comprenant les épaules et le torse. Limité par le stockage mémoire, le modèle complet n'a pu être calculé que jusqu'à 2500 Hz. Kahana a aussi calculé des réponses de pavillon disposé sur un plan infini (quelques mètres de côté) et ce pour des modèles très précis jusqu'à 16000 éléments. Ces réponses sont proches de réponses de pavillon sur une tête humaine mais il a été ainsi mis en évidence la nécessité de travailler avec des maillages haute résolution pour reproduire fidèlement la réponse de pavillons réels. Les

difficultés principales sont dues à la mauvaise reproduction du volume arrière du pavillon et de la forme de la conque.

Des travaux plus récents ont été menés par Algazi et al [Algazi and Duda (2002)]. Leurs motivations concernaient la faisabilité de prédiction de HRTF à partir de modélisation simple de la morphologie de l'auditeur. La tête était modélisée par une sphère et le torse soit par une sphère, soit par une ellipsoïde, le cou n'a pas été pris en compte dans leur modèle baptisé *Snowman*. Les résultats des simulations BEM, effectuées dans la bande de fréquence [500 - 5000] Hz, montrent de très bonnes correspondances avec les mesures sur une réplique physique du modèle, ceci donnant une autre preuve de la validité de la BEM pour la prédiction de HRTF. Une identification des contributions de la tête et du torse est effectuée par l'observation de figures de diffraction. Ces figures sont présentes sur les HRTF de tête humaine et sur les HRTF du *Snowman*. Les auteurs concluent sur la validité de modèles géométriques simples pour la prédiction de HRTF en basses fréquences. Le modèle *Snowman* étant individualisable, cette étude montre aussi que l'on peut adapter des modèles géométriques simples à la morphologie d'un auditeur.

FIG. III.11 – Réplique physique du modèle *Snowman*. La tête est réalisée avec une balle de croquet de 4.15 cm de rayon et le torse avec une boulle de bowling de 10.9 cm de rayon [Algazi and Duda (2002)].

Un modèle paramétrique obtenu par un logiciel de CAO (Conception Assisté par Ordinateur) représentant la tête avec un pavillon, le cou et le haut du torse avait déjà été proposé par Genuit [Genuit (1984)][12] et a été adapté par J. Fels et al [Fels et al. (2004)] pour un travail sur la modélisation des HRTF d'enfants. Ce modèle comporte 34 paramètres dont 5 pour le pavillon (cf. fig.III.12). Fels a d'abord travaillé sur la génération de maillage à partir de 4 photographies numériques. Le sujet porte des points de mesure sur le visage, sous la forme de pastilles, qui décrivent des courbes utilisées ensuite par un

[12]Les informations décrites sur le modèle de Genuit ont été extraites de [Fels et al. (2004)].

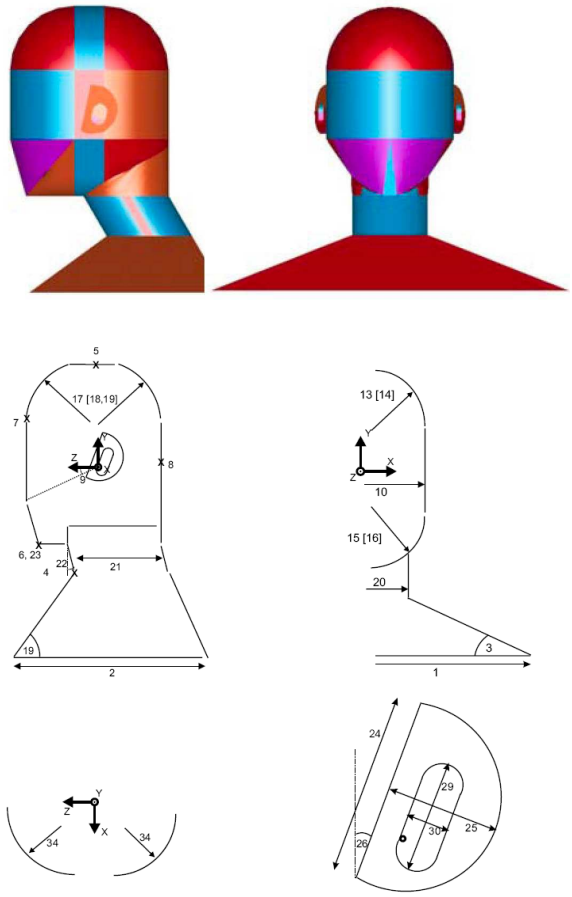


FIG. III.12 – Modèle paramétrique pour le calcul de HRTF [Genuit (1984)].

logiciel de CAO pour définir un maillage. Cette technique permet d'obtenir des maillages très fidèles. Seulement, elle souffre d'un temps de mis en place considérable, de l'utilisation d'un système photographique complexe. La définition d'un maillage adéquat pour la BEM (surface continue et fermée) complique la création du maillage. C'est pourquoi Fels *et al.* (cf. fig.III.13) utilisent le modèle de Genuit. Ils se servent de la formulation réciproque pour abaisser la charge de calcul BEM en définissant l'entrée du canal auditif comme un piston vibrant. Les calculs sont menés jusqu'à 6-8 kHz et ont été validés sur une réplique physique du modèle paramétrique de Genuit.

L'étude réalisée avec la BEM au cours de ces travaux de thèse se situent dans la droite lignée de ceux commencés par Pernaux [Pernaux (2003)]. Il a calculé des HRTF de modèles géométriques simples, partant de la sphère avec oreilles centrées à un modèle complet tête sphérique- cou cylindrique - demi torse ellipsoïdal jusqu'à 3 kHz en passant par l'ellipsoïde avec oreilles décalées. Son travail s'est porté sur les performances d'individualisation de ses modèles autant pour l'ITD que pour les HRTF. Ses travaux ont donné lieu à un brevet sur l'individualisation des HRTF. A partir d'une photographie de face et une autre de profil, les paramètres du modèle complet sont estimés et les HRTF du modèle sont calculées pour être utilisées pour un rendu binaural personnalisé.

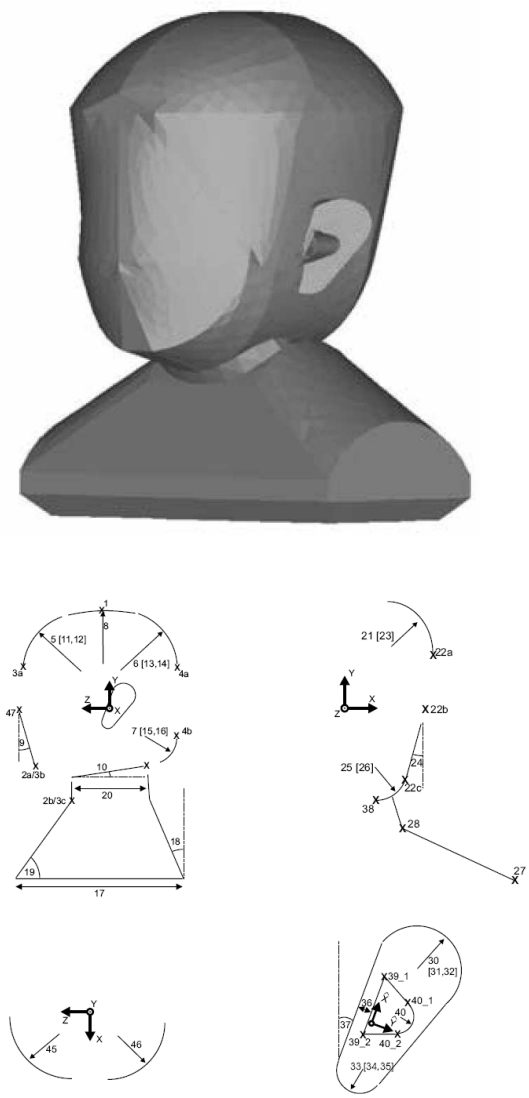


FIG. III.13 – Modèle CAO pour le calcul de HRTF d'enfants [Fels et al. (2004)].

2.11 Application de la BEM à des géométries simplifiées de morphologie pour le calcul de HRTF

L'utilisation de modèles géométriques simples permet une étude systématique des influences de leurs paramètres. Ces modèles décrivent la morphologie de l'auditeur de manière grossière, une sphère pour une tête par exemple, mais sont capables de rendre compte de nombreux détails présents dans les HRTF de personnes humaines. Etant donné les limites matérielles liées à la BEM, les modèles ne seront évalués que jusqu'à la fréquence de 5000 hz. De plus, comme les premières contributions du pavillon se situent autour de 3 kHz [Algazi and Duda (2002); Kahana (2000)], voire autour de 1.5 kHz pour les sources frontales [Katz (2001a)], le calcul pour des modèles sans pavillon pour des fréquences au-delà de 4 kHz entraînerait de grandes différences lors d'une comparaison avec des mesures. Cependant il a été montré de manière empirique que l'on peut addi-

tionner les réponses d'éléments séparés pour reconstituer une réponse complète [Algazi et al. (2001d)]. Il est donc possible de calculer la réponse d'un modèle sans pavillon et de lui adjoindre la réponse d'un pavillon bafflé. Cependant, cette composition fait appel à des paramètres estimés empiriquement tels des angles de rotation pour faire correspondre les éléments entre eux. Cette approche n'a donc pas été retenue.

2.11.1 Motivation des travaux

Le but principal de cette étude est de connaître les performances d'individualisation de modèles géométriques simples. Pour ce faire, une première partie sera consacrée à la validation des deux codes de calcul utilisés. Ensuite, les différentes géométries utilisées seront comparées aux mesures réalisées sur un sujet pour estimer les contributions des différentes parties du corps. Enfin une évaluation des performances d'individualisation d'un modèle comprenant une partie pour la tête, pour le cou et pour le torse sera effectuée.

2.11.2 Validation des codes de calculs

La formulation analytique de la diffraction d'une onde plane par une sphère étant disponible, la validation des codes de calcul se fera donc pour un modèle de tête sphérique. Le modèle présentant un nombre infini d'axes de symétrie passant par le centre, la validation ne se fera que sur le plan horizontal. La pression à la surface d'une sphère pour une élévation nulle et excitée par une onde plane est [Morse and Ingrad (1968)] :

$$p_{tot}(\theta) = p_0 \exp(j\omega t) \sum_{m=0}^{\infty} \frac{j^{m-1}(2m+1)(ka)^{-2} P_m(\cos\theta)}{h_m^{-\prime}(ka)} \tag{III.18}$$

avec p_0, ω et θ respectivement l'amplitude, la pulsation et l'angle d'incidence de l'onde incidente, $k = \frac{\omega}{c}$ le nombre d'onde, a le rayon de la sphère, $P_m(\cos\theta)$ est le polynôme de Legendre d'ordre m et h'_{m2} la dérivée première de la fonction de Hankel de second espèce d'ordre m. Le maillage utilisé pour ce calcul est représenté figure III.16. La zone de raffinement autour de l'axe interaural est d'environ 20 mm de rayon. Le rayon de la sphère est de 118 mm. Ce rayon correspond à la moyenne des trois mesures de la tête effectué sur le sujet $X_1\, X_2\, X_3$ dans [Algazi et al. (2001c)], respectivement demi-hauteur, demi-largeur, et demi-profondeur de la tête. Le maillage est composé de 3572 noeuds pour 7140 éléments. La distance inter-noeuds maximale est de 10 mm, ce qui autorise une fréquence maximale de calcul de $F_{max} = 5000Hz$ à la condition $\lambda/6$.

La figure III.14 représente l'évolution de la pression en dB en fonction de la fréquence pour l'incidence 0° et la figure III.15 pour l'incidence 90° dans le plan horizontal. Pour les deux incidences, les valeurs des résultats des simulations, en prenant en compte un facteur de calibration, sont très proches et l'écart reste inférieur à 1 dB. Les différences observées probablement dues aux différences de placement du microphone. En effet, pour la simulation VNOISE2.0, la formulation réciproque est utilisée : la source sonore est donc placée à quelques millimètres de la sphère, alors que le point de mesure pour la formule analytique se situe sur la sphère. Comme l'écart est de l'ordre du décibel, il peut être conclu que les codes de calcul sont validés.

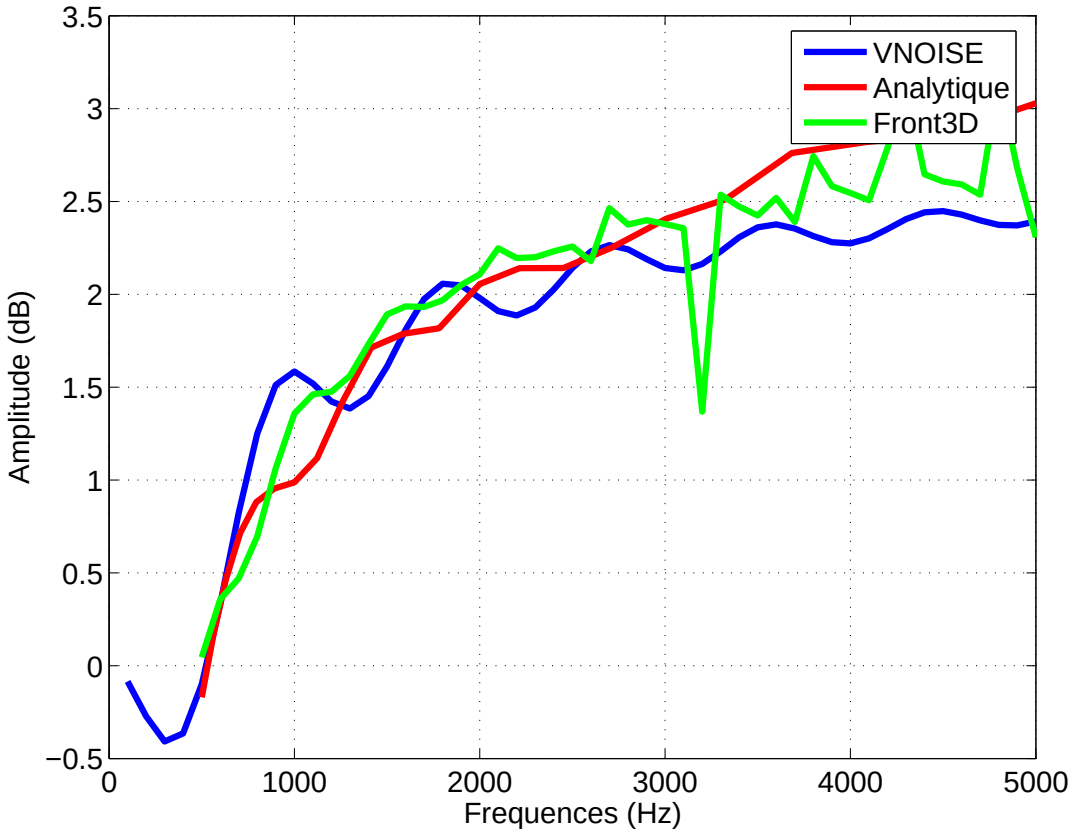


FIG. III.14 – Validation des codes de calculs pour la position ($\theta = 0°$; $\phi = 0°$).

2.11.3 Modèle de tête sphérique

Le modèle de tête sphérique a été le premier à être utilisé avec la technique BEM pour simuler des HRTF [Weinrich (1984)]. On peut définir un modèle sphérique avec un seul paramètre : le rayon. D'autres travaux ont ensuite apportés le décalage des oreilles par rapport au centre de la sphère [Kahana (2000); Pernaux (2003)]. Le modèle sphérique calculé pour cette étude correspond à un modèle individualisé. Le rayon est obtenu de la même façon qu'au paragraphe précédent, c'est-à-dire une moyenne de trois paramètres morphologiques. Le maillage utilisé est le même que celui décrit dans le paragraphe précédent. Le placement de la source sonore, en utilisant la formulation réciproque, est $\theta = 270°$ et ϕ = -22° . La zone grisée à cet endroit correspond au raffinement local du maillage nécessaire en formulation réciproque. Les figures III.16a) et III.16b) montrent la correspondance entre modèle et photographie du sujet.

La modélisation de la tête du sujet par une sphère vise à identifier le volume de la sphère et de la tête. Une sphère de ce type est mieux adaptée à la prédiction de HRTF qu'une sphère ayant comme valeur de diamètre la distance interaurale (Sphère EqVol et IAS dans [Katz (1998)]. Les figures III.17, III.18 et III.19 montrent les HRTF de l'oreille droite pour respectivement les mesures lissées en bandes critiques [Smith (1983)], le modèle sphérique calculé avec VNOISE2.0 et le modèle sphérique calculé avec Front3D. La plage de fréquence est [500 ; 5000] Hz.

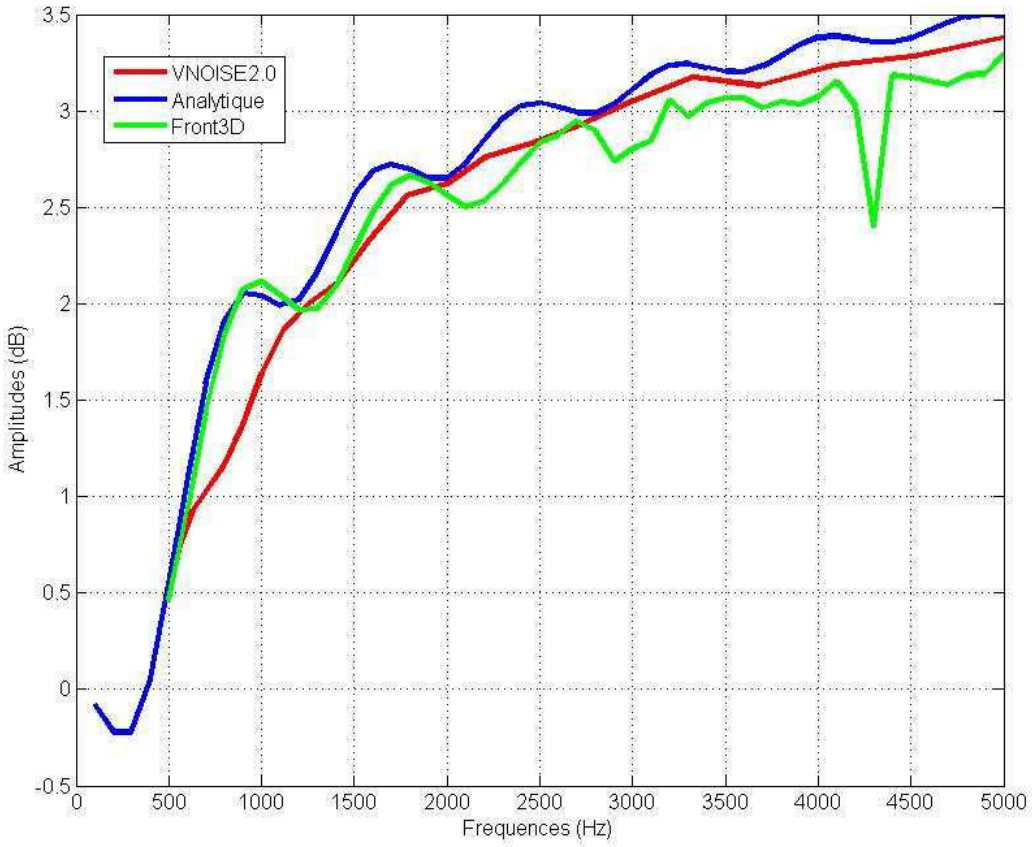
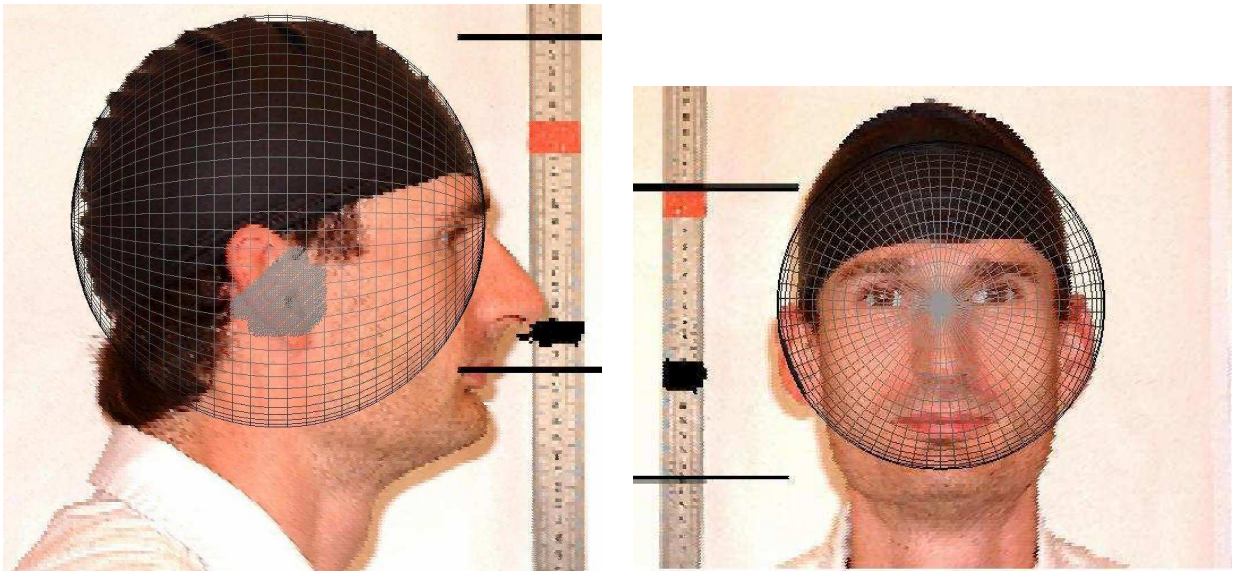


FIG. III.15 – Validation des codes de calculs pour ($\theta = 90°$; $\phi = 0°$).

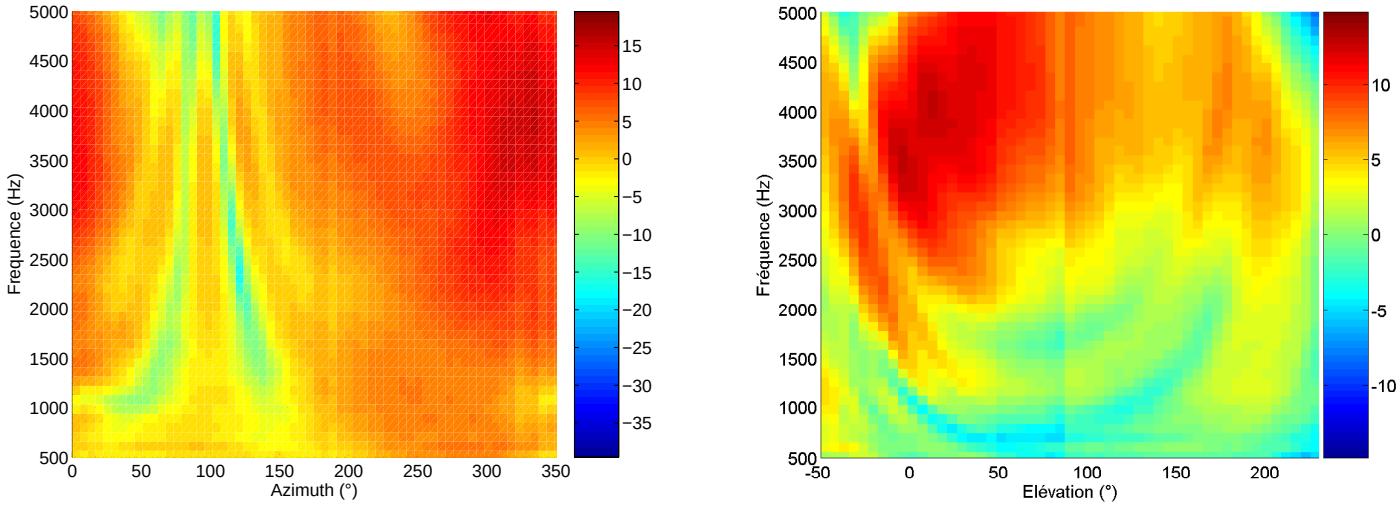
Les figures III.17 a), III.18 a) et III.19 représentent les variations des HRTF sur le plan horizontal (définit par $\phi = 0°$) et les figures notées III.17 b) et III.18 b) les HRTF pour l'élévation variant de -56° à 240° (plan définit par θ= 0°). Bien que représentant une approximation grossière d'une tête humaine, le modèle sphérique contient des figures de diffraction proches de celles présentes sur les mesures notamment dans le plan horizontal.

Plan horizontal La figure III.17 fait apparaître deux parties bien distinctes : la partie ipsilatérale et la partie contralatérale. Les HRTF ipsilatérales sont du côté droit des figures notées *a)* et correspondent aux HRTF recevant le plus d'énergie (rouge) et aux angles d'azimuth inclus dans $\theta = [180°;360°]$, tandis que les HRTF contralatérales correspondent aux angles d'azimuth compris dans [0° ; 180°] et aux HRTF recevant le moins d'énergie (vert/bleu). Ces deux parties sont bien présentes sur la réponse de la sphère individualisée (cf. fig.III.18 et fig. III.19). De plus, la figure de diffraction observée sur la partie contralatérale est aussi reproduite mais avec moins de détails. Cette figure est due à la zone d'ombre créée par la tête où la superposition de l'onde directe et de l'onde diffractée donne naissance à des interférences destructives. La régularité et la symétrie de cette figure de diffraction est due à la symétrie du modèle sphérique. La différence



a) Vue de droite b) Vue de face

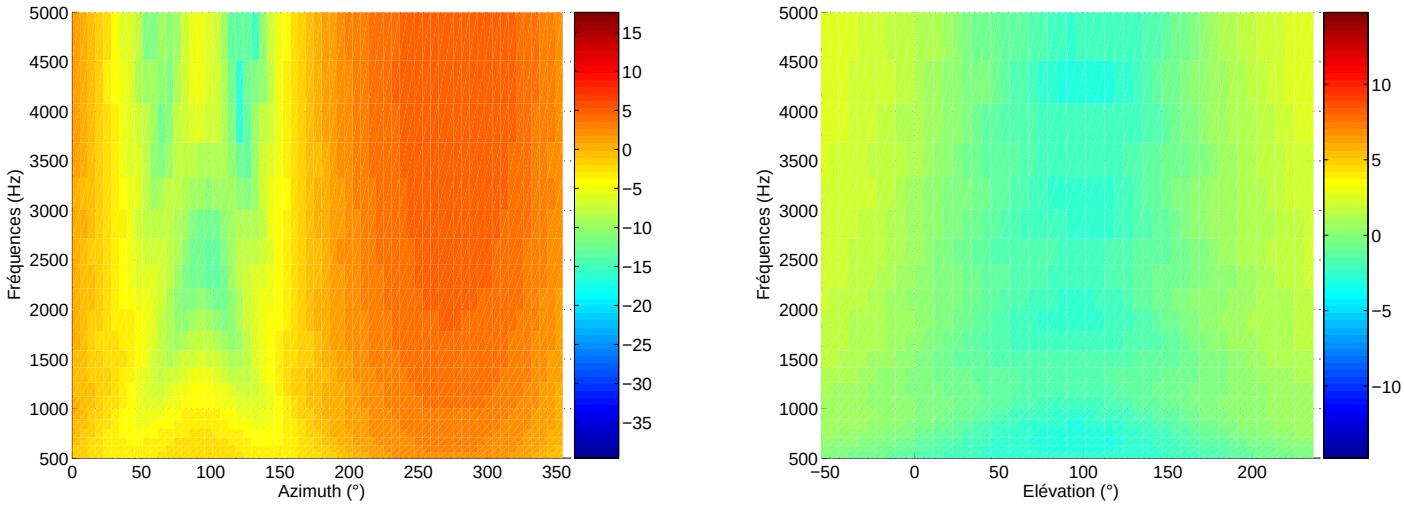
FIG. III.16 – Correspondance entre photographie du sujet et sphère individualisée. a) Vue de droite, b) Vue de face. Le bonnet noir permet une meilleur estimation de la taille réelle de la tête du sujet.



a) b)

FIG. III.17 – Modules des HRTF en dB en fonction de la fréquence. Mesures lissées. a) Plan horizontal, b) Plan vertical. L'amplitude du module est indiquée par la barre colorée sur la droite des figures.

majeure entre HRTF et modèle sphérique est l'amplitude des variations des réponses. Pour les mesures, la plage de variation du module est de [-42 ; 15] dB tandis qu'elle n'est que de [-20 ; 2] dB pour le modèle sphérique. Cette différence majeure est due à l'absence de pavillon sur le modèle sphérique. Le pavillon crée des modulations d'amplitude



a) b)

FIG. III.18 – Modules des HRTF modélisées par une sphère individualisé en dB en fonction de la fréquence. Résultats issus de VNOISE2.0. a) Plan horizontal , b) Plan vertical. L'amplitude du module est indiquée par la barre colorée sur la droite des figures.

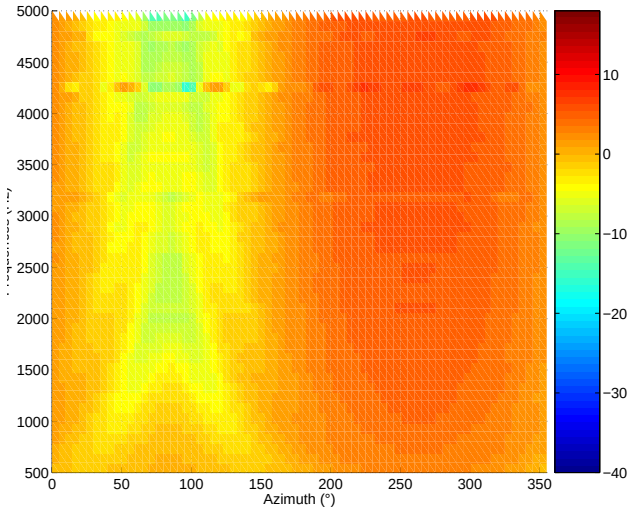


FIG. III.19 – Modules des HRTF modélisées dans la plan horizontal par une sphère individualisé en dB en fonction de la fréquence. Résultats issus de Front3D. L'amplitude du module est indiquée par la barre colorée sur la droite de la figure.

comprises entre [-20 ; 15] dB [Kahana (2000)] et ce à partir de 2-3 kHz. La résonance du pavillon est visible sur la partie ipsilatérale. Le modèle sphérique ainsi que les mesures ont des modules variant peu dans la plage de fréquences [500 ; 1500] Hz avec un niveau autour de 5 dB. Par contre au-delà de 1500 Hz des variations d'amplitude sont visibles sur les mesures alors que pour le modèle sphérique, le module varie peu pour toute la ré-

gion ipsilatérale (fond orange). Ces variations sont dues au torse, aux formes irrégulières d'une tête humaine [Algazi et al. (2001b)] et au pavillon. Les effets du torse sont plus visibles sur le plan vertical.

Plan vertical Les modules des réponses sont référencés par la lettre *b*). Si le module est assez bien reconstruit sur le plan horizontal, la réponse du modèle sphérique sur le plan vertical est très éloignée des mesures. Le module du modèle sphérique varie très peu ([-6 ; 0] dB) alors que les mesures font apparaître des figures de diffraction avec une amplitude comprise dans [-36 ; 13] dB. Seule une petite variation avec l'élévation est observée : l'amplitude est plus importante pour les régions situées en-dessous du niveau de l'axe interaural. Cette variation est due au placement individualisé du point source (qui correspond à l'entrée du canal auditif) à une élévation négative $\phi = -22°$. La position du point de mesure, ou point source[13], est donc importante à reproduire. La figureIII.17b) comporte une zone de forte énergie dans la région $\phi \in [-30° ; 90°]$ correspondant aux résonances du pavillon. Les figures de diffraction en formes d'arches visibles en basses fréquences sont dues au torse [Algazi et al. (2001b)].

Comparaison des codes de calculs Les figures III.18 et III.19 indiquent les résultats des calculs BEM effectués avec VNOISE2.0 et Front3D respectivement. Les résultats sont très similaires autant au niveau des figures de diffraction que de la dynamique du module. Les résultats de VNOISE2.0 et Front3D ont même amplitude de variation de 18 dB. Si les figures présentent des différences, elles sont dues en grande partie à un facteur de calibration qui peut modifier le niveau moyen, et donc la teinte moyenne des figures de HRTF. Cependant les résultats de Front3D font apparaître trois fréquences irrégulières qui font diverger les calculs pour tous les positions : 3200 Hz, 4300 Hz et 4900 Hz (cf. fig 2.11.3). Ces fréquences sont aussi visibles sur la figure III.19 : des traits horizontaux apparaissent aux fréquences irrégulières et sur les figures III.14 et III.14. En effet, au contraire de VNOISE2.0 qui utilise la méthode CHIEF [Schenck (1968)] pour résoudre le problème des fréquences irrégulières, Front3D ne dispose pas d'algorithme particulier pour prendre en compte cet artefact mathématique. On peut donc noter que la méthode CHIEF montre ici de bons résultats. De plus, la différence de temps de calcul entre les deux codes est assez considérable : pour VNOISE2.0, 4 heures pour les 965 positions de la base de HRTF TNO et environ 12 heures pour Front3D et ceci seulement pour 130 positions. Il convient de noter l'apport de la parallèlisation du calcul avec VNOISE2.0, qui pour la présente étude ne s'est faite qu'en utilisant 3 PC, alors que VNOISE2.0 propose d'utiliser jusqu'à 5 PC. Les techniques de répartition de charge de calcul permettent donc un gain de temps considérable en plus d'une capacité de calcul augmentée par l'addition des ressources mémoires des PC. Le calcul en parallèle peut aussi être implémenté avec FRONT3D.

Conclusion sur le modèle sphérique Malgré sa simplicité, le modèle de tête sphérique individualisé apporte une assez bonne prédiction des HRTF du sujet pour le plan horizontal. Il est même reporté que son comportement est très similaire à un modèle de tête

[13]La dénomination est *point de mesure* si on parle des mesures ou d'une formulation direct de la BEM, Front3D par exemple, ou *point source* si l'on parle de formulation indirecte, avec VNOISE2.0 par exemple.

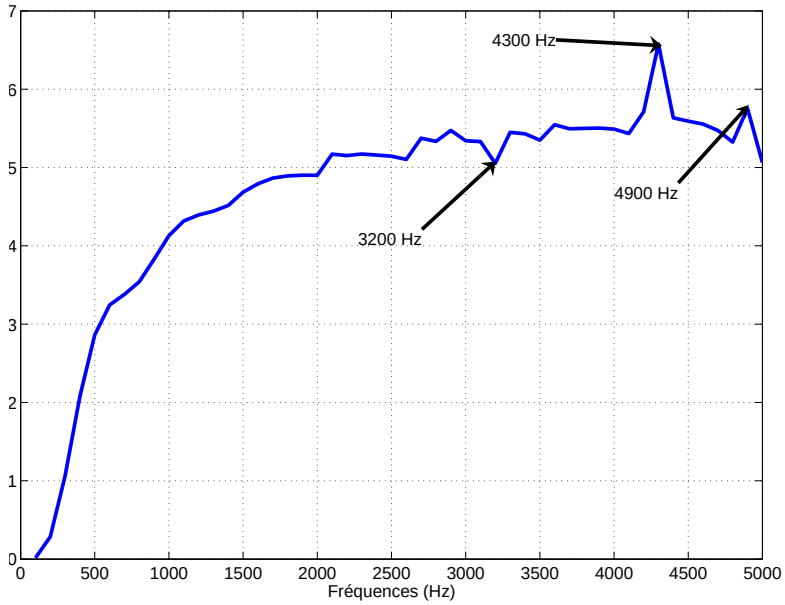


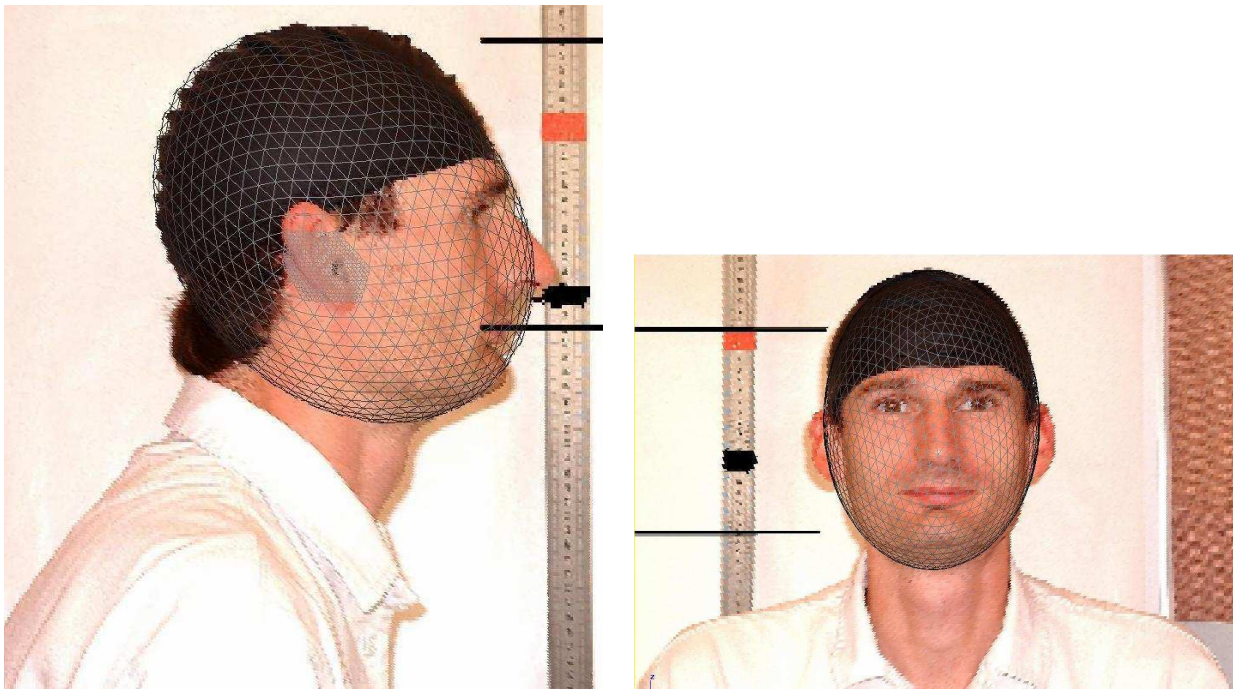
FIG. III.20 – Module de la HRTF d'une sphère calculée avec FRONT3D pour la position ($\theta = 50°$; $\phi = 0°$). Les textes fléchés indiquent les fréquences irrégulières.

sans pavillon jusqu'à 2 kHz [Katz (1998)]. Cependant, la grande symétrie de ce modèle ne permet pas une bonne prédiction pour le plan vertical. Cependant la perception des sources sur le plan horizontal est majoritairement gouvernée par l'ITD. L'apport du modèle sphérique pour une reproduction fidèle de la synthèse binaurale semble alors faible, mais il peut toutefois servir à reconstruire la partie basse fréquence des mesures (les basses fréquences sont difficilement mesurables (cf. § 1.2)).

2.11.4 Modèle de tête ellipsoïdale

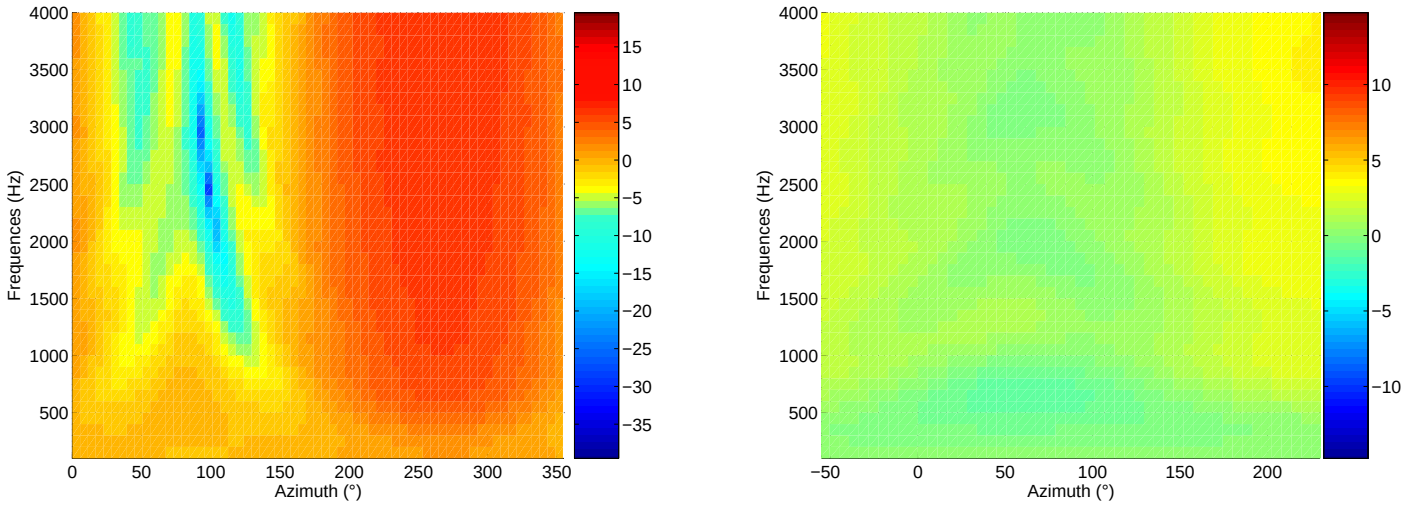
Le modèle de tête ellipsoïdale a déjà été étudié dans [Kahana (2000); Pernaux (2003)]. Les auteurs reportent une meilleure prédiction des HRTF par rapport au modèle de tête sphérique. Le maillage ellipsoïdal utilisé dans la présente étude comporte 1949 noeuds et 3894 éléments. Les figures III.21a) et b) indiquent la correspondance entre le maillage et la morphologie du sujet. Le point source est identifiable par la zone grisée qui correspond au raffinement nécessaire du maillage en formulation réciproque de la BEM.

Sur le plan horizontal, l'apport principal du modèle ellipsoïdal par rapport au modèle sphérique est la figure de diffraction mieux définie sur le côté contralatéral (masquage de la tête) avec des anti-résonances plus accentuées. Globalement la dynamique est mieux respectée : [-27 ; 7] dB. Par contre le modèle ellipsoïdal semble apporter peu de modification par rapport au modèle sphérique sur le plan vertical, malgré l'avantage que peu procurer un tel modèle pour la reconstruction de l'ITD en élévation [Duda et al. (1999)]. La dynamique reste faible : [-1 ; 5] dB.



a) Vue de droite b) Vue de face

FIG. III.21 – Correspondance entre photographie du sujet et ellipsoïde individualisée. a) Vue de droite , b) Vue de face.



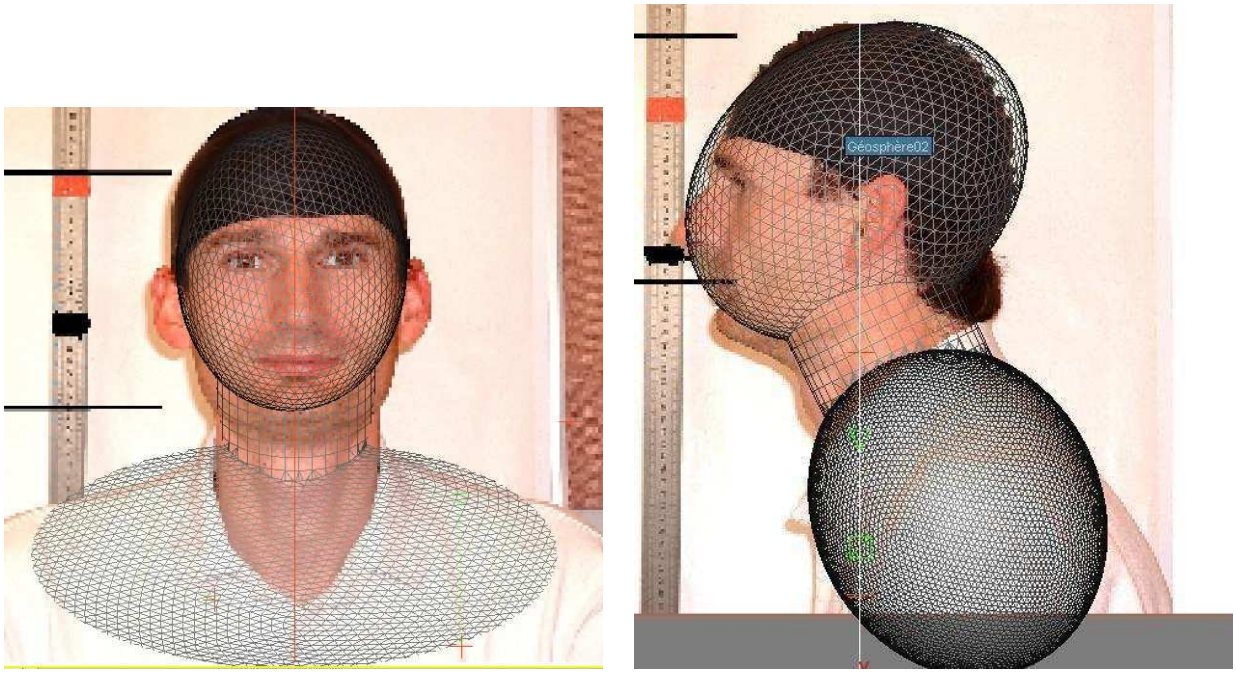
a) b)

FIG. III.22 – Modules des fonctions de transferts du modèle ellipsoïdal individualisé en décibels en fonction de la fréquence pour : a) Plan horizontal , b) Plan vertical

2.11.5 Modèle complet

Le modèle complet calculé pour cette étude est composé d'une tête ellipsoïdale, d'un cou cylindrique à section elliptique et d'un torse ellipsoïdal. La figure III.23 indique la

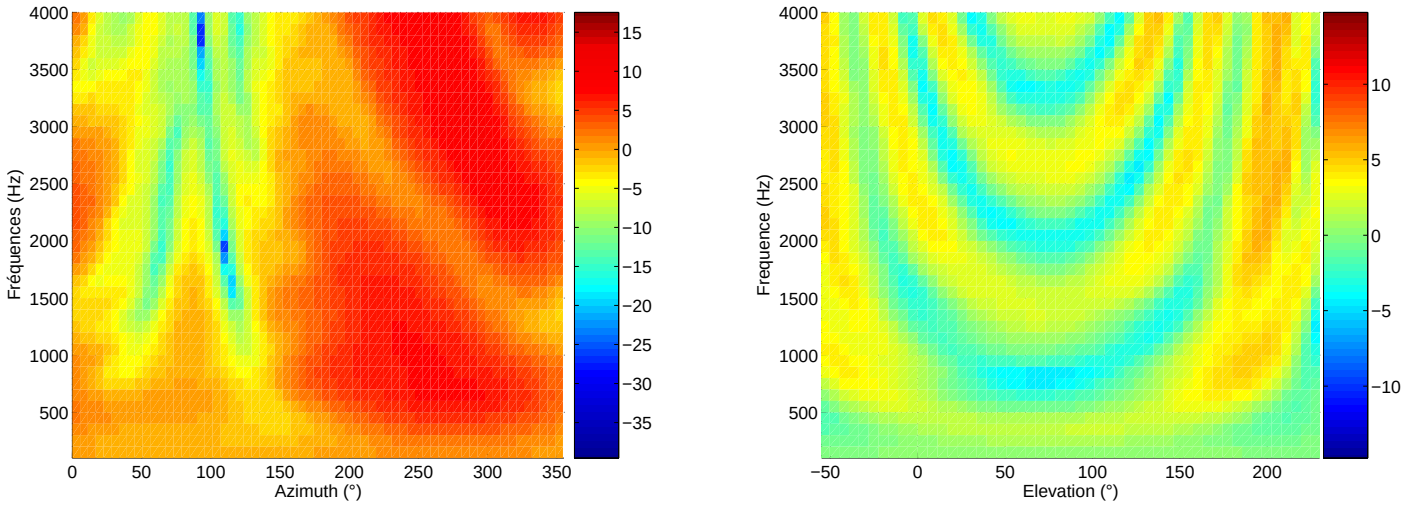
correspondance entre le modèle et les photos du sujet. La taille du torse est un facteur limitant pour les calculs BEM. Ainsi pour le modèle présenté ici, le maillage du torse représente plus de 76 % du nombre total de noeuds (6252/8219) alors que le maillage de la tête contient 1695 noeuds (21%) et celui du cou 272 (3%). Le calcul réalisé sur 3 PC en parallèle a pris moins d'un jour.



a) Vue de face b) Vue de gauche

FIG. III.23 – Modèle complet individualisé : a) Vue de face , b) Vue de gauche.

Les résultats sont reproduits sur la figure III.24. Pour le **plan horizontal**, la présence du cou fait ré-apparaître la figure de diffraction liée à la tête sur le côté ipsitralatéral et qui était occultée dans le modèle précédent. Le modèle complet apporte aussi une asymétrie sur cette figure de diffraction, asymétrie introduite par l'anisotropie d'une ellipsoïde. La dynamique est bien reproduite et vaut 36 dB. Globalement, la correspondance avec des mesures effectuées sur un mannequin KEMAR sans pavillon (fig. 13 et 14 de [Algazi and Duda (2002)]) est très bonne. Le coté ipsilatéral, comporte désormais une figure de diffraction sour la forme de deux creux. Ce dernier motif est toutefois atténué sur les mesures car le pavillon introduit alors des modifications considérables. Pour le **plan vertical**, le modèle complet reproduit la plupart des motifs présents sur les mesures. Comme le modèle précédent, les figures de diffraction se traduisent par des modulations en arches. Ces modulations dépendent de l'élévation. Cependant, le caractère symétrique de ces modulations pourrait empêcher une différenciation avant/arrière. La différence principale avec les mesures est l'absence de pavillon dans les modèles tête + cou + torse. Ceci est cohérent avec le fait que le pavillon apporte les indices majeurs pour la perception de l'élévation. Cependant, et jusqu'à 2000 Hz, c'est le torse qui donne la perception de l'élévation.



a) b)

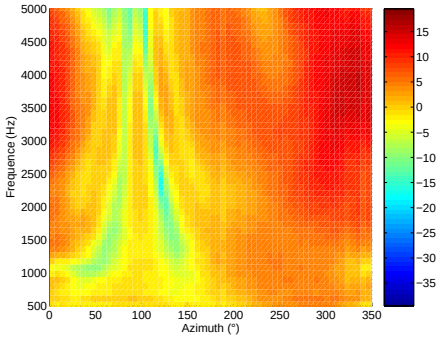
FIG. III.24 – Modules des fonctions de transferts du modèle complet individualisé en décibels en fonction de la fréquence pour : a) Plan horizontal , b) Plan vertical

3 CONCLUSION

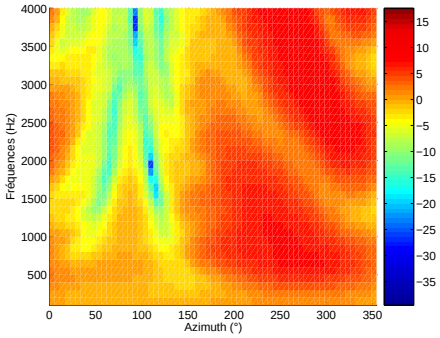
3.1 Travaux présents et futurs

La technique BEM a été utilisée pour la modélisation des HRTF. Par le biais de mesures simples, à partir de photographies du sujet par exemple, un modèle géométrique simple comportant un nombre limité de paramètres est estimé et le maillage correspondant a été réalisé. Ce maillage, représentation surfacique du volume créé, a été introduit dans les codes de calculs BEM mis à disposition. Le modèle le plus évolué comporte 14 paramètres et traduit une modélisation de la tête, du cou et du torse. Un maillage de ce modèle contient environ 8000 noeuds. Ce modèle apporte des fonctions de transfert très proches de HRTF mesurées et presque identiques à des mesures effectuées sur un KEMAR sans pavillon [Algazi and Duda (2002)]. Ce modèle a permis l'identification des rôles joués par la tête, le cou et le torse dans les HRTF. Ainsi, en accord avec [Algazi et al. (2001b); Algazi and Duda (2002)], le torse permet la perception de l'élévation pour des sources basses fréquences (F < 3000 Hz). La taille du torse étant très variable d'une personne à l'autre, il est attendu que l'individualisation du modèle complet soit bénéfique pour la prédiction des HRTF. La figure III.25 représente les différentes acquisitions réalisées au cours de l'étude (plan horizontal).

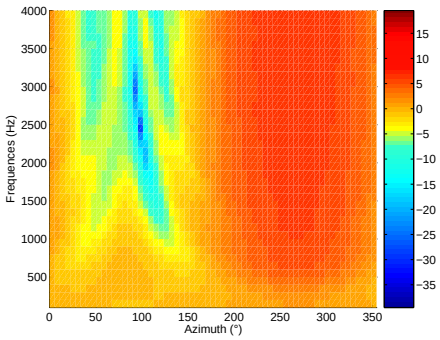
Une démarche est proposée ici pour tester les capacités d'individualisation du modèle complet. Il faut commencer par photographier les 8 sujets de la base TNO, de face et de profil et réaliser les 8 modèles complets. Auparavant, il pourra être nécessaire d'évaluer la reproductibilité de la construction du modèle. Ensuite, il convient de calculer les 8 modèles et d'estimer la variance inter-modèle sur les fonctions de transfert calculée. Cette variance en fonction de la fréquence et des coordonnées d'espace doit alors être comparée à la variance inter-individuelle des mesures des sujets filtrées à 4000 hz et lissées.



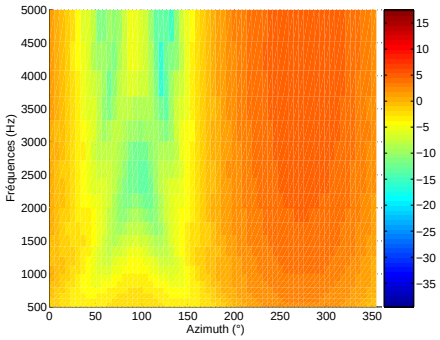
a) Mesures



b) Modèle complet



c) Modèle ellipsoïdal



d) Modèle sphérique

Fig. III.25 – Comparaison des HRTF sur le plan horizontal obtenues pour différentes acquisitions. De haut en bas : mesures, modèle complet, tête ellipsoïdale et tête sphérique.

3.2 La BEM est une méthode basse fréquence

La BEM est une méthode extrêmement coûteuse en ressources mémoire et en temps de calcul. La formulation de Front3D, qui calcule les valeurs de pression sur les éléments, ne permet pas des calculs sur des maillages supérieurs à 5000 noeuds, et donc de travailler sur le modèle complet. Les limitations de VNOISE2.0 sont moins importantes. En théorie, VNOISE2.0 n'a pas de limite du fait de la parallèlisation du calcul. En pratique, il faut éviter un calcul en RAM-disk car les temps de calculs sont alors prohibitifs. Ce qui au mieux, en considérant 5 PC avec 2 Go de RAM dédiés au calcul, donnerait un maillage de 35000 noeuds. Seulement, Kahana [Kahana (2000)] a montré que pour reproduire fidèlement une réponse d'un pavillon seul, un maillage très fin est nécessaire. Donc, les limites de VNOISE2.0 sont presque atteintes. Toutefois, ces remarques dépendent largement du système informatique dans lesquels sont installés les logiciels de calcul. De meilleures performances sont réalisables avec des systèmes dédiés aux calculs. De plus, la parallèlisation de calcul peut être implémentée dans FRONT3D.

Une première remarque est qu'une utilisation en temps réel de la BEM pour une large bande de fréquences semble délicate. De plus, la réponse d'un pavillon est un système interférentiel : le moindre changement dans la géométrie peut apporter des figures de diffractions et la question de la viabilité de la modélisation du pavillon par BEM est posée. Par contre, il a été montré que le modèle complet permet une bonne reproduction des HRTF et des écoutes informelles [Algazi and Duda (2002); Pernaux (2003)] ont indiqué l'amélioration apportée par la présence du torse. Ainsi, la BEM peut apporter des réponses aux très basses fréquences, là où les mesures sont difficilement réalisables (cf. § 1.2). Les modèles calculés à partir de la BEM peuvent donc servir à apporter l'information manquante en basses fréquences des HRTF et/ou à représenter la partie basse fréquence d'une HRTF hybride dont le reste du spectre serait estimé par une autre méthode.

BIBLIOGRAPHIE

Algazi, V., Duda, R., and Thompson, D. (2001a). The cipic hrtf database.

Algazi, V. R., Avendano, C., and Duda, R. O. (2001b). Elevation localization and head-related transfer function analysis at low frequencies. *J. Acoust. Soc. of Am.*, 109(3).

Algazi, V. R., Avendano, C., and Duda, R. O. (2001c). Estimation of a spherical-head model from anthropometry. *J. Acoust. Soc. of Am.*, 49 :472–478.

Algazi, V. R., Avendano, C., and Thompson, D. (1999). Dependence of subjects and measurement position in binaural signal acquisition. *J. Audio Engin. Soc.*, 47(11) :937–947.

Algazi, V. R. and Duda, R. O. (2002). Approximating the head-related transfer function using simple geometric models of the head and torso. *J. Acoust. Soc. of Am.*, 112(5) :2053–2064.

Algazi, V. R., Duda, R. O., Morison, R. P., and Thompson, D. M. (2001d). Structural composition and decomposition of hrtfs. New Paltz, New-York. IEEE Trans. Speech and Audio Proc.

Bronkhorst, A. W. (1995). Localization of real and virtual sound sources. *J. Acoust. Soc. of Am.*, 98(5) :2542–2553.

Bruneau, M. (1998). *Manuel d'acoustique fondamentale*. Hermes.

Brungart, D. S. and Rabonowitz, W. M. (1998). Auditory localization of nearby sources. *J. Acoust. Soc. of Am.*, 106 :1465–1479.

Burton, A. J. (1973). The solution of helmholtz equation in exterior domains using integral equations. Technical report, National Physics Laboratory.

Chateau, N. (1996). *Localisation de sources sonores multiples dans l'hémisphère supérieur*. PhD thesis, Université de la méditerranée Aix Marseille II, laboratoire de mécanique et d'acoustique.

Cheng, C. I. and Wakefield, G. H. (2001). Introduction to head-related transfer functions (hrtfs) : representations of hrtfs in time, frequency and space. *J. Audio Engin. Soc.*, 49(4) :231–249.

Ciscowski, R. D. and Brebbia, C. A. (1991). *Boundary Element methods in acoustics*. Computational Mechanics Publication & Elsevier Applied Science, Southampton.

Duda, R. O., Avendano, C., and Algazi, V. (1999). An adaptable ellipsoidal head model for the interaural time difference. In *Proc. IEEE International Conference on Acoustics Speech and Signal Processing*, volume II, pages 965–968. ICASSP'99.

Duraiswami, R., Zotkin, D. N., and Gumerov, N. A. (2004). Interpolation and range extrapolation of hrtf. volume IV, pages 45–48.

Fels, J., Buthmann, P., and Vörlander, M. (2004). Head-related transfer functions of children. *Acta Acoustica united with Acoustica*, 90 :918–927.

Francescantonio, P. D. (2003). *Vnoise theorotical manual*. http : www.sts-soft.com.

Genuit, K. (1984). *Ein modell zur beschreibung von aussenohrübertragungseigenschaften*. PhD thesis, RWTH Aachen.

Kahana, Y. (2000). *Numerical modelling of the head related transfer function*. PhD thesis, University of Southampton.

Katz, B. F. G. (1998). *Measurement and calculation of individual head-related transfer function using a boundary element model incuding the measurement and effect of skin and hair impedance*. PhD thesis, Pennsylvania State University.

Katz, B. F. G. (2001a). Boundary element method calculation of individual head-related transfer function. i. rigid sphere calculation. *J. Acoust. Soc. of Am.*, 110(5).

Katz, B. F. G. (2001b). Boundary element method calculation of individual head-related transfer function. ii. impedance effects and comparison to real measurements. *J. Acoust. Soc. of Am.*, 110(5) :2449–2455.

Kulkarni, A. and Colburn, H. S. (2000). Variability in the characterisation of the headphone transfer-function. *J. Acoust. Soc. of Am.*, 107(2) :1071–1074.

Hammershøï, D. and Møller, H. (2002). Methods for binaural recording and reproduction. *Acta Acoustica united with Acoustica*, 88 :303–311.

Møller, H. (1992). Fundamentals of binaural technology. *Applied Acoustics*, 36(5) :171–218.

Møller, H., rensen, M. F. S., i, D. H., and Jensen, C. B. (1995). Head related transfer functions of human subjects. *J. Audio Engin. Soc.*, 43(5) :300–321.

Miller, J. D. (2001a). Hrtf error analysis using spectral power ratios.

Miller, J. D. (2001b). Modelling iteraural time difference assuming à spherical head.

Morse, P. and Ingrad, K. U. (1968). *Theorical acoustics*. Mc-Graw-Hill.

Pernaux, J. (2003). *Spatialisation du son par les tehcniques binaurales : application aux services de télécommunication*. PhD thesis, Institut National de Polytechnique de Grenoble.

Schenck, H. A. (1968). Improved integral formulation for acoustic radiation problems. *J. Acoust. Soc. of Am.*, 44(1) :58–41.

Smith, J. O. (1983). *Techniques for Digital Filter Design and System Identification with Application to the Violin*. PhD thesis, Stanford University.

Vandernoot, G. Sytem and protocol. http://recherche.ircam.fr/equipes/salles/listen/system/protocol.html.

Weinrich, S. (1984). Sound field caculations around the human head. Technical report, The acoustic laboratory, Technical university of Denmark.

Wightman, F. L. and Kistler, D. J. (1989). Headphone simulation of free-field listening. i : Stimulus synthesis. *J. Acoust. Soc. of Am.*, 85(2) :858–867.

Zotkin, D. N., Duraiswami, R., and Gumerov, E. G. N. A. (2004). Fast head related transfer function measurment via reciprocity. Technical report CS-4620 and UMIACS-20004-62, Universitt of Maryland, Computer Science and UMIACS.

IV

Modélisation de HRTF par réseaux de neurones

INTRODUCTION

Le chapitre précédent a présenté deux techniques d'acquisition de HRTF : la mesure et le calcul par éléments de frontières. La mesure, seule technique qui permet l'obtention de toutes les caractéristiques individuelles des HRTF, reste une technique coûteuse en temps et en matériel. La BEM permet l'acquisition reproductible de HRTF et peut être utilisée en complément des mesures. Cependant, la BEM nécessite des ressources informatiques considérables qui semblent la cantonner à l'acquisition en basses fréquences des HRTF. Le présent chapitre s'intéresse à une alternative à ces deux méthodes d'acquisition. L'approche envisagée utilise les techniques d'apprentissage statistique et notamment les réseaux de neurones.

Le principe général de cette approche est de créer un modèle mathématique permettant de calculer les HRTF d'un individu à partir d'un ensemble de paramètres qui lui sont propres. Les paramètres d'entrée du modèle peuvent être par exemple des données anthropométriques ou tout autres données contenant des informations à partir desquelles les HRTF individuelles peuvent être reconstruites. Les paramètres d'entrées peuvent aussi contenir des HRTF mesurées.

Ainsi, cette approche de modélisation de HRTF est à la frontière entre deux problématiques largement étudiées de la synthèse binaurale : l'interpolation et la prédiction de HRTF. L'interpolation consiste à créer des HRTF à des positions non mesurées à

partir de HRTF mesurées : il s'agit d'une interpolation spatiale des HRTF. Pour la prédiction, on considère ici les modèles qui permettent la création de HRTF à partir de paramètres d'entrée ne contenant pas de HRTF mesurées. L'approche considérée pour les travaux présentés dans ce chapitre a aussi un lien avec les techniques de réduction de données comme l'analyse en composante principale (ACP) ou l'analyse en composante indépendante (ACI) qui ont été largement appliquées aux bases de données des HRTF.

Ce chapitre est divisé en trois parties. La première partie est consacrée à la présentation des méthodes classiques de modélisation de HRTF. Elle rappelle les principes gouvernant ces méthodes et les travaux les plus importants qui les ont utilisées. La deuxième partie présente les techniques d'apprentissage par réseaux de neurones et leurs applications en synthèse binaurale. La troisième partie rapporte le travail effectué selon l'approche considérée grâce à des techniques statistiques et neuronales. Ce travail est présenté sous la forme d'une étude de faisabilité quant à la réduction du nombre de mesures pour la prédiction de tout un ensemble de HRTF. La question du nombre et de la position des points de mesures est abordée grâce à une classification des HRTF et la prédiction des autres HRTF est réalisée grâce à un apprentissage des données par un réseau de neurones artificiels (RNA). La conclusion de ce chapitre propose une comparaison entre les techniques classiques de réduction de données et l'approche abordée pour le travail de thèse.

Ces travaux ont fait l'objet de deux stages co-encadrés au sein de France Telecom R&D. Pour plus de détails, le lecteur pourra se reporter aux rapports de stage [Choqueuse (2004); Vovor (2005)]. Ces travaux ont fait l'objet de deux dépôts de brevets (numéro FR 05 00218 et INPI 05 10995).

1 MÉTHODES CLASSIQUES DE MODÉLISATION DE HRTF

Cette partie présente l'application de méthodes classiques de réduction de données à la synthèse binaurale : l'interpolation et la décomposition linéaire de HRTF. L'interpolation de HRTF est présentée succintement tant cette technique a été et est encore un sujet d'étude. Des références sont données sur le sujet. Une plus grande partie est consacrée à la décomposition linéaire des HRTF car cette méthode permet notamment un gain d'implémentation de la synthèse binaurale.

1.1 Interpolation de HRTF

L'interpolation de HRTF est considérée ici sous l'angle des techniques de réduction de données : un ensemble réduit de HRTF permet la création d'un ensemble complet par interpolation. De nombreuses méthodes existent pour interpoler une HRTF à une position non mesurée. L'interpolation peut servir à recréer une HRTF manquante, éviter le stockage coûteux des HRTF en termes de ressources mémoires sur les micro-processeurs ou encore obtenir une répartition continue des HRTF pour la spatialisation dynamique. Le choix d'une technique d'interpolation se fait en deux étapes. D'abord il faut déterminer si l'interpolation s'effectue de manière globale, sur des fonctions de base, ou de manière locale, par exemple entre une combinaison de positions mesurées proches au sens géodésique. Ensuite, il faut choisir les variables d'interpolation.

Dans le cas d'une interpolation locale linéaire entre des HRIR, les coefficients de la

HRIR peuvent être obtenus de la manière suivante :

$$\widehat{H}(n) = \sum_i \frac{\alpha_i H_i(n)}{\alpha_i} \tag{IV.1}$$

avec H_i l'ensemble des plus proches voisins de la position à interpoler.

L'interpolation linéaire des coefficients des HRIR ou des HRTF (méthode validée perceptivement dans [Langendjik et Bronkhorst (2000)]), donne de très bons résultats en comparaison des méthodes plus complexes comme les réseaux de neurones à fonction d'activation non-linéaires [Nishino et al. (1996)] ou l'interpolation de type spline [Nishino et al. (2000); Matusmoto et al. (2004)]. Ces méthodes sont aussi appliquées à des représentations des mesures : modèles RII [Jenison (1995)] ou coefficients spatiaux issus d'analyse en composante principale [Chen et al. (1995); Jin et al. (2002)].

1.2 Décomposition linéaire des HRTF

Les HRTF peuvent être rassemblées sous la forme d'une matrice dont les colonnes forment des fonctions de directivité dépendant de la fréquence et les lignes des filtres de spatialisation en fonction de la position. Ainsi, les techniques de réduction matricielle issues de l'algèbre linéaire peuvent être appliquées aux bases de données de HRTF. Ces techniques consistent à trouver une représentation compacte des données. Une réduction du nombre de filtres, ou du nombre de fonctions spatiales est obtenue. Ces techniques ont été largement étudiées et sont utilisées dans certains moteurs de spatialisation sonore. Si, par exemple, 10 filtres de reconstruction sont calculés, alors il suffit d'estimer les 10 coefficients à appliquer aux filtres de reconstruction pour obtenir une HRTF reconstruite proche de la HRTF mesurée. Un gain d'implémentation est ainsi obtenu : le nombre de filtres reste constant quel que soit le nombre de sources à simuler, la synthèse d'une source supplémentaire consiste alors à l'ajout de 2n gains. Dans l'optique d'une réduction de la procédure de mesures, les techniques de décomposition offrent la possibilité d'une procédure simplifiée d'individualisation de la synthèse binaurale. L'individualisation revient alors à estimer les coefficients à appliquer aux filtres de reconstruction. Pour que cette approche soit viable pour n'importe quel auditeur, il faut que l'identification des filtres de reconstruction s'effectue sur une base de HRTF la plus *universelle* possible (cf. § III.1.3.4) pour obtenir par exemple des filtres de reconstruction *universels*.

La décomposition linéaire des HRTF représentées sous la forme d'une matrice H de taille $N \times M$, où N est le nombre de positions et M le nombre de points fréquentiels, consiste à trouver la matrice C de fonctions spatiales, de dimension $N \times r$, ainsi que la matrice L des filtres de reconstruction, de dimension $r \times M$, de telle sorte que H puisse être exprimée par le produit de C et de L [Larcher (2001)] :

$$H \approx C.L \tag{IV.2}$$

L'erreur de reconstruction E_r est alors évaluée par :

$$E_r = \|H - C.L\| \tag{IV.3}$$

Un exemple d'un telle décomposition est donné par la figure IV.1. Cette décomposition peut être obtenue par différentes méthodes de l'algèbre linéaire :

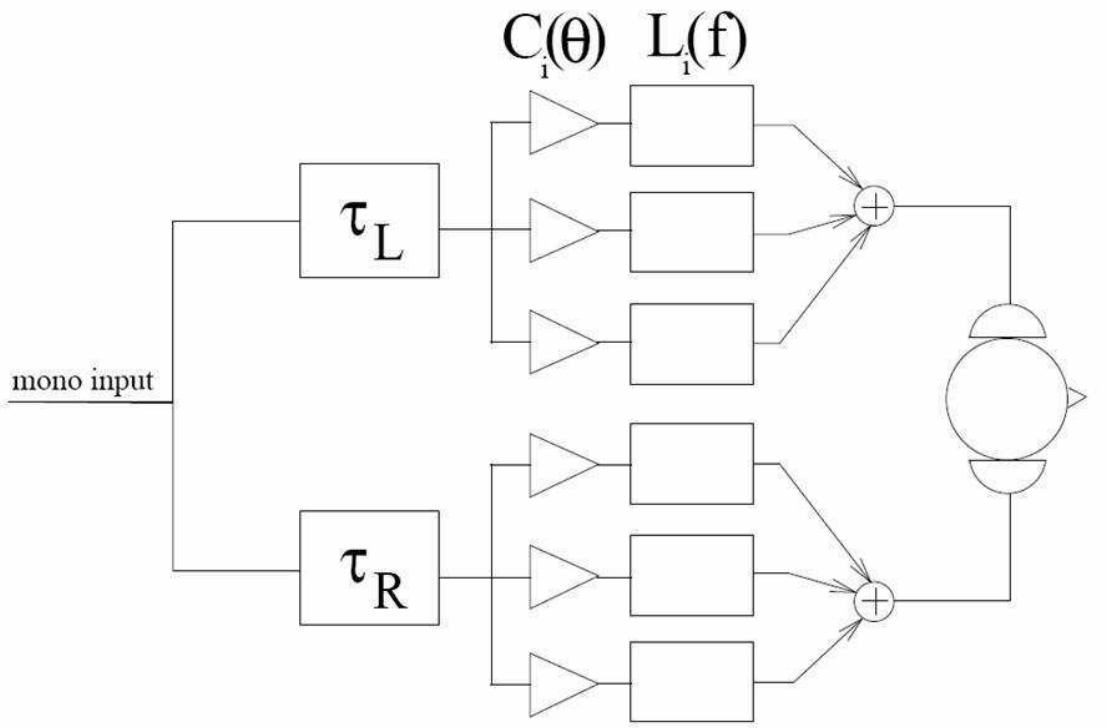


FIG. IV.1 – Implémentation multi-canale de la synthèse binaurale : décomposition linéaire des $HRTF_{min}$.

- Optimisation des fonctions spatiales pour des filtres de reconstruction fixés. Les fonctions spatiales sont obtenues par projection orthogonale des HRTF sur les filtres de reconstruction.
- Optimisation des filtres de reconstruction pour des fonctions spatiales fixées (par exemple les harmoniques sphériques). Les filtres de reconstruction sont obtenus par projection orthogonale des HRTF sur les fonctions spatiales.
- Optimisation commune des fonctions spatiales et des filtres de reconstruction : une ACP, ou ACI, est conduite sur les HRTF. Une telle décomposition assure une minimisation de l'erreur de reconstruction E_r (cf. eq IV.3).

1.2.1 Décomposition en valeurs propres et valeurs singulières

La méthode de décomposition de l'algèbre linéaire la plus répandue est la décomposition d'une matrice en valeurs propres. Les valeurs propres λ_k d'une matrice carrée A sont les solutions de $\det(A - \lambda_k Id) = 0$, où Id représente la matrice identité, et les vecteurs propres X_k associés à λ_k sont tels que $A \cdot X_k = \lambda_k X_k$. Cette méthode permet d'écrire une matrice carrée A sous la forme d'un produit de trois matrices dont une est diagonale :

$$A = PDP^{-1} \tag{IV.4}$$

avec P matrice de passage composée des vecteurs propres de A et D matrice diagonale composée des valeurs propres de A. Comme la matrice des HRTF n'est pas carrée, cette technique n'est pas applicable et la décomposition en valeur singulière (SVD en anglais pour Singular Value Decomposition) est utilisée. Dans le contexte de l'algèbre linéaire, les valeurs singulières d'une matrice $A \in \mathbb{C}_{p,f}$ sont les racines carrées des valeurs propres de AA^T, où A^T représente la matrice trans-conjuguée de A. La SVD est donnée par :

$$A = U^T DV \tag{IV.5}$$

avec $U \in \mathbb{C}_{p,r}$ et $V \in \mathbb{C}_{r,f}$ matrices unitaires, c'est-à-dire $U^T U = Id$ et $V^T V = Id$, et $D \in \mathbb{C}_{r,r}$ matrice diagonale composée des valeurs singulières de A rangées par valeurs absolues décroissantes. Les colonnes de U sont les vecteurs propres de AA^T. Cette technique a été appliquée aux HRIR droites et gauches dans [Grantham et al. (2005)] pour réduire la complexité de filtres transauraux. Une réduction de la taille des filtres RIF de 512 à 90 échantillons n'entraîne pas de dégradation des performances de localisation. Dans [Kahana (2000)], la technique SVD a été utilisée pour mettre en évidence le lien entre la base de décomposition des HRTF et la base des fonctions propres de l'équation d'onde en coordonnées sphériques (cf. équation III.4) c'est-à-dire les harmoniques sphériques. Le lien a d'abord été montré analytiquement sur une sphère puis étendu à des géométries complexes tel des moules de pavillons. La SVD est aussi utilisée, notamment par [Larcher (2001)], pour la technique de décomposition des HRTF nommée *subset selection*. Cette technique applique les fonctions spatiales issues d'une SVD à des HRTF sur lesquelles est appliquée une décomposition QR[1] avec pivot de colonne. Cette décomposition permet une hiérarchisation des fonctions spatiales selon leur norme. Cet ordonnancement des positions est utilisé pour sélectionner les HRTF qui constitueront les filtres de reconstruction. Ces HRTF particulières peuvent aussi indiquer des positions privilégiées pour la mesure des HRTF.

1.2.2 Analyse en Composante Principale (ACP)

La technique d'analyse en composante principale, ou développement de Kahunen-Loeve, permet d'exprimer une matrice dans une base où son expression devient plus compacte. Ceci est réalisé en réduisant l'information partagée par les variables de départ, les fonctions de directivité par exemple, et en répartissant l'erreur résiduelle sur les nouvelles variables de dimension réduite [Larcher (2001)]. Les nouvelles variables sont orthogonales et sont ordonnées en terme de pourcentage de variance apportée par chaque composante. Ainsi, la première composante principale contient les informations relatives à la variance maximale, la deuxième contient les informations relatives à la variance suivante. Le processus est répété jusqu'à l'obtention de la n$^{\text{me}}$ et dernière composante principale. Les pertes d'informations diminuent d'une étape à la suivante. Les différentes opérations de l'ACP sont :

- calcul de la matrice de covariance
- calcul des valeurs et vecteurs propres de la matrice de covariance

Les résultats de l'ACP sont les n premières valeurs propres non nulles de la matrice de covariance. On peut ne garder que les variables les plus représentatives de la variance

[1] Une décomposition QR consiste à décomposer une matrice un en produit dune matrice orthogonale et d'une matrice diagonale supérieure.

de départ. La sélection est alors effectuée sur la base de la minimisation de l'erreur de reconstruction (cf. équation IV.3) au sens des moindres carrés.

Les premières applications de l'ACP se sont réalisées pour la décomposition du spectre d'amplitude des HRTF [Martens (1987); Middlebrooks et Green (1992); Kistler et Wightman (1992)]. Ainsi, Kistler et al [Kistler et Wightman (1992)] effectuent une ACP sur 5300 modules en décibel de HRTF à phase minimale égalisées en champ diffus (10 sujets $\times$ 265 directions $\times$ 2 oreilles) pour en extraire 5 composantes principales, ici 5 filtres de bases, représentant 90% de la variance (la première composante contient entre 70 et 80 % de la variance). Un test d'écoute mené sur 5 sujets montre une bonne correspondance perceptive entre HRTF mesurées et HRTF reconstruites à partir des 5 composantes principales. Dans [Chen et al. (1995)] une ACP est conduite sur des HRTF à phase mixte d'un mannequin KEMAR (2188 HRTF) et d'un chat (1816 HRTF). Seize composantes principales fréquentielles donnent une erreur de reconstruction inférieure à 1%. Les fonctions spatiales extraites sont lissées pour améliorer l'interpolation pour des positions non mesurées. Les auteurs de [Wu et al. (1997)] effectuent une ACP sur des HRIR dont il a été extrait l'ITD et l'ILD. Sur les mêmes données que [Chen et al. (1995)], c'est-à-dire 1816 HRTF de chat, 20 composantes principales donnent 99,9 % de la variance.

Les liens entre composantes principales de HRTF et paramètres morphologiques peuvent conduire à une individualisation de la synthèse binaurale à moindre coût. Ainsi dans [Jin et al. (2000a)] une comparaison des performances de localisation de HRTF réduites par ACP est effectuée pour une base de 36 sujets. Les résultats montrent qu'une reconstruction des HRTF contenant 60 % de la variance (7 composantes) des HRTF de départ ne dégrade pas les performances des sujets. Une ACP est aussi conduite sur des paramètres morphologiques des sujets. Ces 20 paramètres sont mesurés par le relevé de leurs coordonnées spatiales. Une régression linéaire est ensuite effectuée sur les composantes principales morphologiques pour prédire les 7 composantes principales fréquentielles. L'évaluation perceptive de cette dernière étape n'est pas donnée dans l'article, mais cette approche a fait l'objet d'un brevet (numéro US 0138107).

1.2.3 Analyse en Composante Indépendante (ACI)

L'analyse en composante indépendante permet une représentation compacte d'une matrice en déterminant une base de projection où les vecteurs de base sont indépendants. Cette nouvelle base est obtenue par la maximisation d'une *fonction de contraste* qui détermine le niveau d'indépendance des variables. L'ACI revêt un intérêt particulier pour la synthèse binaurale grâce au lien entre indépendance des variables statistiques et compacité de support [Larcher (2001)]. Ainsi, l'ACI peut être menée sur les HRIR pour obtenir une base de réponses impulsionnelles à support compact [Dudouet et Martin (1998)], ou sur les HRTF soit pour calculer une base de filtres de reconstruction à supports fréquentiels disjoints [Emerit et Martin (1995)], soit pour la définition de fonctions spatiales discrètes [Larcher (2001)]. Cette dernière approche est particulièrement intéressante pour le problème de la réduction du nombre de points de mesures des HRTF. Les fonctions spatiales issues de l'ACI indiquent des directions privilégiées pour les points de mesure des HRTF.

Le paragraphe suivant présente une méthode qui est à la frontière entre interpolation et réduction de données. Depuis une vingtaine d'années, les techniques d'analyse de grandes bases de données, ou *data mining* en anglais, ont connu un essor considérable.

Capable de déterminer des groupes de variables à caractéristiques communes ou encore de mettre en évidence des relations de très haut niveau entre groupes de variables, ces techniques sont appliquées dans des domaines aussi variés que les transactions boursières ou la pharmacologie. Ces techniques sont aussi utilisées en synthèse binaurale pour l'interpolation de HRTF ou pour la prédiction de paramètres de spatialisation tels les coefficients à appliquer à des filtres de reconstruction.

2 LES RÉSEAUX DE NEURONES ARTIFICIELS

2.1 Principe

2.1.1 Historique

La théorie des réseaux de neurones artificiels (RNA) repose sur l'observation des neurones biologiques constituant le cerveau. Le cerveau humain contient entre 10^{10} et 10^{12} neurones et chaque neurone est connecté à environ 1000 autres neurones. Le neurone effectue une intégration spatiale et temporelle des signaux électriques en provenance d'autres neurones via les dendrites (cf. fig.IV.2).

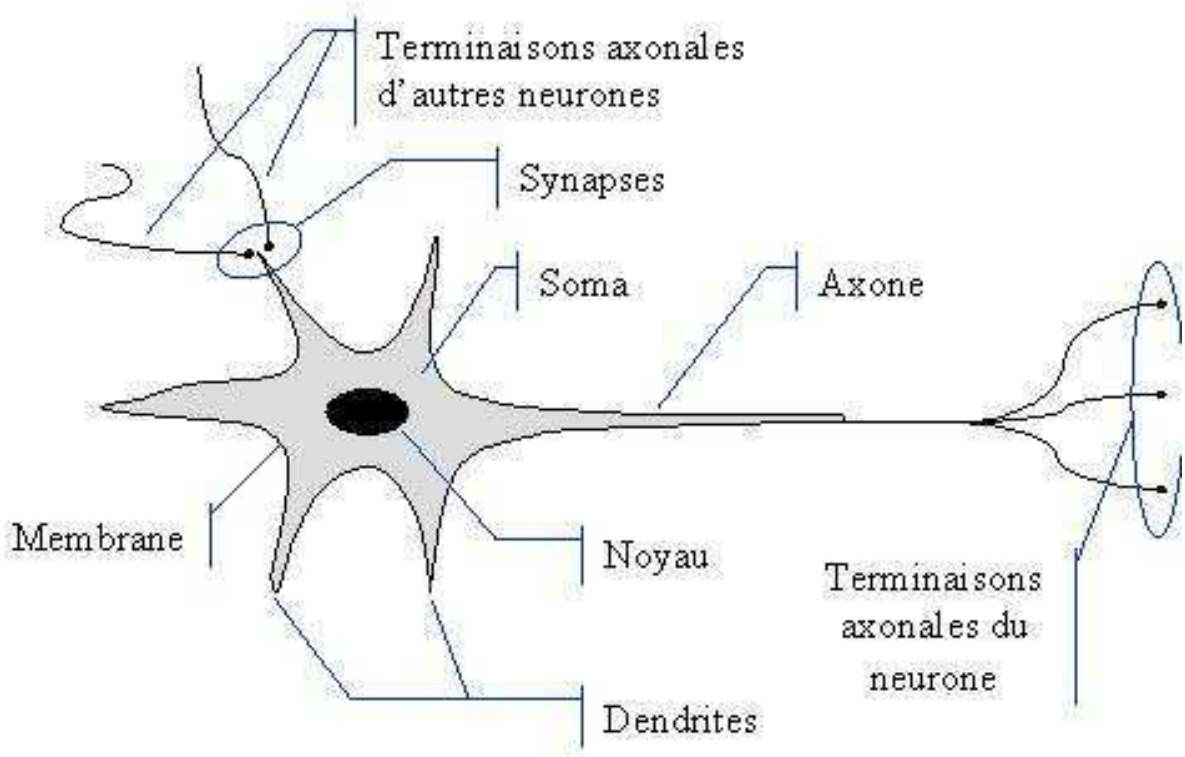


FIG. IV.2 – Représentation schématique d'un neurone biologique.

De façon générale, le début des réseaux de neurones artificiels est situé en 1943 avec

les travaux de McCulloch et Pitts qui montrent qu'un réseau de neurones discret, sans contrainte de topologie, peut représenter n'importe quelle fonction booléenne et donc émuler un ordinateur. Le *neurone formel* (cf. fig. IV.3) de McCulloch et Pitts est un opérateur binaire effectuant la somme pondérée des valeurs d'entrée par des coefficients synaptiques à valeurs réelles. Si cette somme pondérée dépasse un certain seuil θ alors le neurone est actif. Un *neurone formel* est constitué des éléments suivants :

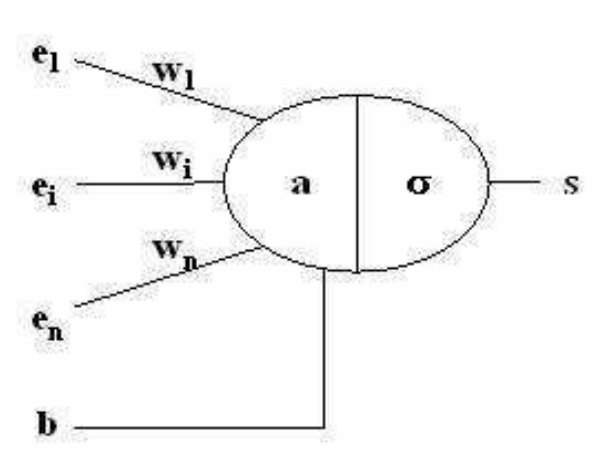


Fig. IV.3 – Neurone formel de McCulloch et Pitts

- **e** est le vecteur d'entrée du neurone, il représente soit l'ensemble des signaux en provenance de l'environnement, soit les sorties d'autres neurones.
- **w** est un vecteur dont le rôle est de simuler les pondérations synaptiques à l'aide de poids associés respectivement à chaque entrée du neurone.
- **b** est le biais du neurone.
- **a** est une variable qui représente le potentiel du neurone. Cet état interne du neurone (activité) est défini par une somme des entrées pondérées par le vecteur de poids
- **s** est une variable qui représente la sortie du neurone envoyée vers d'autres unités. Elle est calculée en appliquant une fonction mathématique à l'activité du neurone.
- σ est la fonction d'activation (ou de transfert) qui est appliquée à l'activité. Il peut s'agir par exemple de la fonction signe ou de la fonction de Heaviside.

Il s'agit d'une simplification grossière du neurone biologique. Cependant la généralisation de ce modèle va permettre des architectures de réseaux de neurones complexes et de nouveaux algorithmes pour travailler sur ces modèles.

Ainsi, en 1958 Rosenblatt propose le premier algorithme d'apprentissage, qui permet d'ajuster les paramètres d'un neurone. En 1969, Minsky et Papert publient le livre *Perceptrons* dans lequel ils utilisent une argumentation mathématique pour démontrer les limitations des réseaux de neurones à une seule couche. En 1982, Hopfield propose des réseaux de neurones associatifs. En 1986, Rumelhart, Hinton et Williams publient, l'algorithme de la *rétropropagation de l'erreur* qui permet d'optimiser les paramètres d'un réseau de neurones à plusieurs couches. La recherche sur les réseaux de neurones connaît alors un essor fulgurant et les applications commerciales de ce succès académique suivent au cours des années 90.

Les applications sont nombreuses et partagent toutes un point commun essentiel à l'utilité des réseaux de neurones : les processus pour lesquels on désire émettre des prédictions comportent de nombreuses variables explicatives et surtout, il existe la possibilité de dépendances non-linéaires de haut niveau entre ces variables qui, si elles sont découvertes et exploitées, peuvent servir à l'amélioration de la prédiction du processus. L'avantage fondamental des réseaux de neurones par rapport aux modèles statistiques traditionnels, comme les décompositions linéaires, réside dans le fait qu'ils permettent d'automatiser la découverte des dépendances les plus importantes du point de vue de la prédiction du processus [APSTAT Technologies (2002)].

2.1.2 Le perceptron multi-couche

Un RNA est un réseau d'unités élémentaires interconnectées à fonctions d'activation linéaires ou non. Ces unités se décomposent (pour les réseaux multi-couches) en au moins deux sous-ensembles de neurones (cf. fig. IV.4) ou couches [2] : une couche d'entrée, une autre de neurones de sortie et éventuellement une couche de neurones cachés. Il existe de nombreux modèles de réseaux de neurones : les réseaux de Kohonen [Kohonen (1995)], ou carte de Kohonen, les réseaux à fonctions de base radiales [Jenison (1995); Jenison et Fissell (1996)], les perceptrons multicouches [Cun (1987)] (MLP pour Multi Layer Perceptron) ayant des architectures de complexité variable (les différentes unités sont interconnectées aux autres, soit complètement, soit partiellement). La *connaissance* incluse dans le réseau de neurones est *mémorisée* dans les poids et l'architecture du réseau. Chaque couche d'un MPL est entièrement connectée à la suivante et son graphique de connectivité ne possède pas de cycle.

Le MLP est un réseau de neurones avec un algorithme d'apprentissage, c'est-à-dire avec un processus d'adaptation des poids du réseau. Le choix d'un processus d'apprentissage est déterminé par la nature de la tâche à effectuer. Dans l'étude présentée ici, il s'agit d'une tâche d'**approximation** : on cherche à modéliser une application non-linéaire de $\mathbb{R}^m$ vers $\mathbb{R}^n$ à l'aide de couples d'exemples d'entrée-sortie. Il existe deux classes d'apprentissage :

1 **L'apprentissage non-supervisé** : Le réseau va essayer de s'adapter aux régularités des statistiques des données d'entrée. Il est alors considéré que toute l'information nécessaire se trouve dans les données d'entrée et dans la structure topologique qui est imposée aux données d'entrée par l'analyste. L'apprentissage non-supervisé permet des tâches de catégorisation (cf. § 4.2).

2 **L'apprentissage supervisé** : Il s'agit d'un mode d'apprentissage où le réseau est guidé. La méthode classique pour l'apprentissage supervisé consiste à présenter un ensemble d'exemples, c'est-à-dire un ensemble fini de couples de vecteurs (x_i, y_i) avec x_i l'entrée du réseau et y_i la sortie désirée pour cette entrée. La fonction calculée par le réseau est écrite sous une forme paramétrique : $f(x, w)$ ce qui désigne la sortie du réseau quand on lui présente en entrée le vecteur x et qu'il utilise les poids synaptiques contenus dans la matrice w. Suivant un critère d'erreur donné d, une mesure de l'erreur entre la sortie effective du RNA, $f(x, w)$, et la sortie désirée, y_i, est calculée. Le but est alors de trouver w qui minimise l'erreur totale introduite par le réseau. L'erreur totale peut s'exprimer de la manière suivante :

[2] Le terme *couche* désigne un ensemble de poids synaptiques

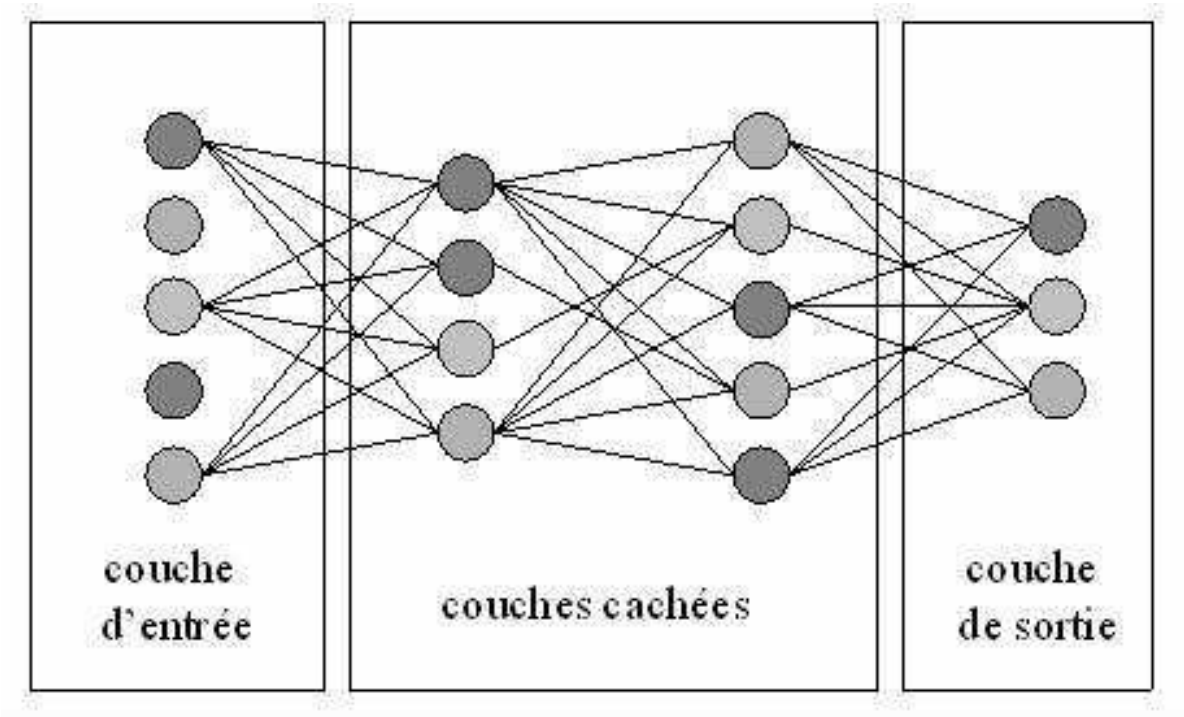


FIG. IV.4 – Graphique de connectivité d'un perceptron multi-couches

$$E_{totale} = \sum_{i}^{n} d(f(xi; w); yi) \tag{IV.6}$$

avec n le nombre total de neurones. Cette erreur est fonction de l'ensemble des poids synaptiques et est appelée *fonction de coût*. Les techniques classiques d'optimisation de fonctions pour trouver son minimum peuvent être utilisées et notamment la technique de *descente du gradient* qui est exposée en annexe D.

2.1.3 Apprentissage et généralisation

Dans un processus d'apprentissage le réseau de neurones est construit en minimisant, par exemple, une fonction de coût sur un ensemble fini d'exemples : *l'ensemble d'apprentissage*. Cependant, le réseau doit pouvoir généraliser la représentation construite par le réseau à d'autres données, y compris celles n'appartenant pas à l'ensemble d'apprentissage. Une manière d'évaluer cette faculté consiste à mesurer les performances du réseau de neurones sur des données non apprises. Il s'agit d'une évaluation de l'erreur de généralisation. Cette erreur est calculée entre les sorties prédites par le RNA et les sorties connues de l'ensemble de généralisation. La différence entre l'erreur d'apprentissage et l'erreur de généralisation représente une mesure de la qualité de l'apprentissage.

Certaines méthodes qui évaluent l'erreur de généralisation sont basées sur la partition de l'ensemble des données en plusieurs sous-ensembles : par exemple, un ensemble utilisé pour l'*apprentissage* et un ensemble de *validation*. L'ensemble de validation est utilisé pour contrôler et mesurer la généralisation du réseau, ou *erreur de test*. Pendant

l'apprentissage l'erreur d'apprentissage décroît, tandis que sur l'ensemble de validation elle commence à diminuer pour atteindre un minimum. A partir de ce point, le réseau apprend par coeur les données de l'ensemble d'apprentissage et l'apprentissage doit être stoppé (cf fig.IV.5).

Différents critères d'arrêt ont été développés :

– quand l'erreur d'apprentissage a atteint un seuil fixé
– après un nombre fixé de cycles d'apprentissage
– quand une estimation de l'erreur de généralisation est minimum.

Si l'apprentissage continue, le RNA apprend par coeur les données de l'ensemble d'apprentissage et n'est plus capable de généraliser. L'erreur de test se met alors à croître. La figure IV.6 illustre ce dilemme.

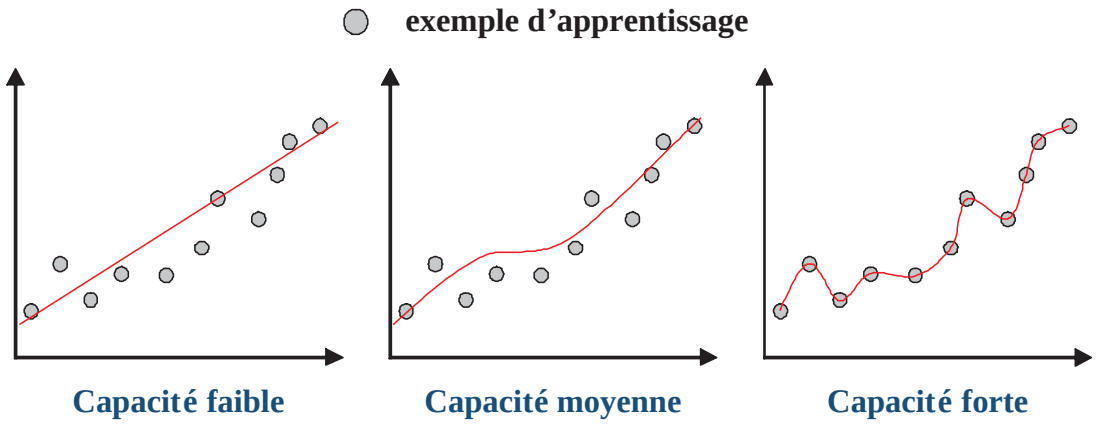


FIG. IV.5 – Représentation de données cibles (points) et de données prédites (ligne) : capacité du réseau et sur-apprentissage. Figure de gauche : modélisation de la tendance générale des données. Figure du milieu : prise en compte de variations fines. Figure de droite : la modélisation a appris par coeur les données [Lemaire (2001)].

Les ensembles d'apprentissage et de validation servent à déterminer l'architecture la plus appropriée : pour différentes architectures (nombre de neurones cachés variable), l'erreur de validation est contrôlée et l'architecture pour laquelle elle est minimale est choisie. Plusieurs techniques qui utilisent cette méthodologie existent. La méthode retenue pour l'apprentissage du réseau est la technique nommée *séparation d'échantillons*.

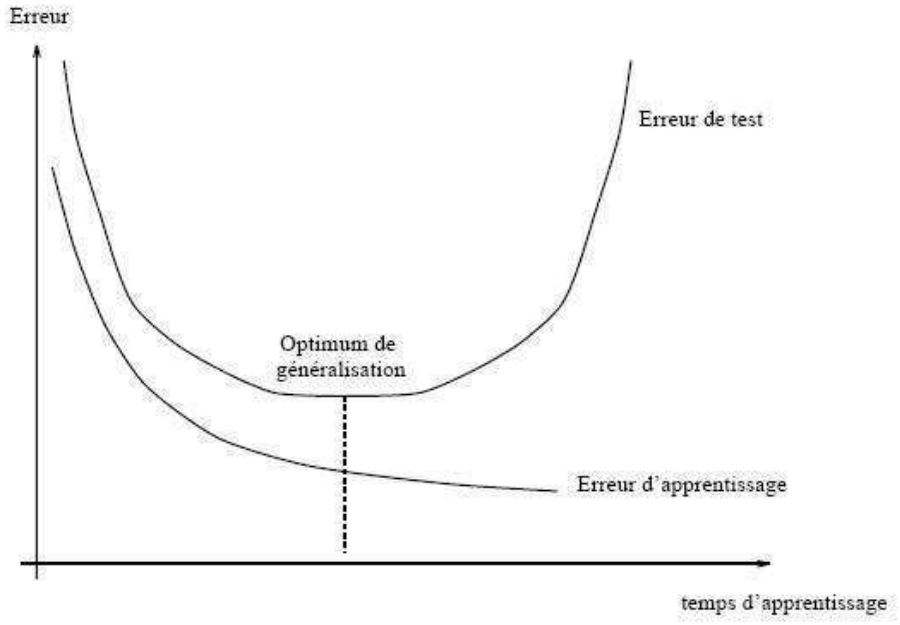


FIG. IV.6 – Evolution des erreurs d'apprentissage et de test en fonction du temps d'apprentissage [Lemaire (1999)].

Elle réserve un troisième ensemble de données appelé *ensemble de test*, pour tester le réseau sur des données qui n'ont jamais été utilisées ni pour l'apprentissage ni pour la validation. La technique de *validation croisée* est utilisée pour déterminer l'architecture du réseau. Une séparation de la base de données en k ensembles est réalisée, k pouvant être la taille de la base. L'apprentissage est effectué k fois avec $k-1$ parties réservées à l'apprentissage et la partie restante servant à la validation et le test.

2.2 Application des réseaux de neurones artificiels aux HRTF

Les travaux antérieurs concernant l'application des réseaux de neurones au contexte de la spatialisation sonore et notamment de la synthèse binaurale ont porté principalement sur deux aspects :

1 La création de modèles de perception pour la localisation auditive : il s'agit de simuler les processus perceptifs et de prédire la direction dans laquelle une source virtuelle sera localisée à partir de signaux présentés aux oreilles d'un auditeur.
2 La prédiction de filtres binauraux : à partir d'un ensemble de paramètres d'entrées, une ou plusieurs HRTF sont créées.

Modèle d'écoute binaurale Dans [Palomäki et al. (2000)], les auteurs ont comparé les performances des cartes de Kohonen et des réseaux des neurones MLP pour la prédiction de localisation de sources sonores dans l'espace. Les entrées des réseaux de neurones sont les sorties d'un modèle de localisation sonore [Pulkki et al. (1998)] qui donne une estimation de l'ITD et de l'ILD en 32 bandes ERB. L'étude conclut que les deux méthodes offrent des performances similaires et que des données psychoacoustiques sont nécessaires

pour entraîner le MLP et adapter le modèle binaural afin que leurs performances soient proches de celles d'un vrai auditeur.

Les modèles d'écoute binaurale peuvent aider à la compréhension de certains mécanismes physiologiques de la perception auditive. Dans [Jin et al. (2000b)], un modèle d'image auditive (AIM pour Auditory Image Model), basé sur le modèle décrit dans [Patterson et Allerhand (1995)], est utilisé avec un MLP qui incorpore une composante temporelle (TDNN pour Time Delay Neural Network) pour prédire la localisation de stimuli filtrés par des HRTF. Un TDNN a été préféré à un MLP car il peut traiter à la fois les données spectrales et temporelles comme le *ferait* le système auditif. Ainsi la tonopie du codage réalisé par le système auditif a pu être incorporée dans un RNA. Le modèle AIM génère un profil d'activité neuronale qui simule la sortie des cellules cilliées internes de l'organe de Corti. Ce profil d'activité neuronale indique la probabilité d'activité du nerf auditif. L'association de l'AIM et d'un TDNN a permis d'atteindre qualitativement les performances humaines pour des stimuli large bande et des stimuli filtrés à la condition que le TDNN soit entraîné sur un large panel de stimuli (différentes largeur de bandes et de différentes fréquences centrales).

Prédiction de HRTF Dans [Jenison et Fissell (1996)] un apprentissage est réalisé pour la prédiction de poids à associer à des composantes principales issues d'une ACP sur 400 modules de HRTF. Les entrées du réseau sont les coordonnées de la HRTF à reconstruire. La particularité de cette étude est la comparaison des performances obtenues pour deux types de fonctions d'activation : les fonctions de base radiale (RBF pour Radial Basis Functions) et les fonctions de base de von Misses (VMBF pour Von Misses Basis Functions). Les VMBF sont l'équivalent d'une fonction gaussienne en coordonnées sphériques. Les réseaux sont entraînés sur 400 HRTF décomposées et testées sur 50 HRTF. L'étude montre un bon comportement du RNA pour la prédiction des poids à associer aux composantes principales et l'avantage des VMBF sur les RBF.

Utilisant à nouveau les VMBF, Jenison [Jenison (1995)] étudie la prédiction par un RNA des coefficients IIR d'un ensemble de HRTF, les entrées étant les coordonnées de la HRTF à reconstruire. Les résultats montrent une très bonne correspondance dans la bande [5-12] kHz. Cependant un écart de 10 dB à 2kHz est observé. La modélisation RII donne un ensemble de 30 pôles communs aux 50 HRTF à phase minimale apprise. L'article conclut alors que les pôles ne dépendent pas de la direction mais dépendraient de l'individu. Ce résultat est cohérent avec l'approche d'individualisation des HRTF utilisée dans [Middlebrooks (1999)].

Dans [Wu et al. (1998)], un apprentissage par RNA, qui utilise comme entrées les coordonnées de la HRTF à reconstruire, est effectué pour la prédiction de fonctions spatiales issues d'une ACP. Le but est d'obtenir une reconstruction de HRIR pour n'importe quelle position de l'espace. 710 HRIR égalisées en champ diffus, 600 pour l'apprentissage et 110 pour la validation, mesurées sur un mannequin KEMAR sont utilisées sous une forme décomposée.

2.3 Présentation des problématiques abordées pour la modélisation de HRTF

Les RNA, qualifiés parfois d'*approximateurs universels*, présentent un intérêt en terme de modélisation. Ils sont capables de déterminer des relations de haut niveau entre les variables d'entrée et les variables de sortie. Le travail mené dans la présente étude vise à déterminer dans quelle mesure et avec quelle fiabilité les RNA sont capables de modéliser des HRTF mesurées. Une première différence avec les autres études de la littérature est l'obtention de HRTF directement *utilisables* et non une modélisation des HRTF (poids à associer à des composantes principales, coefficients d'un filtre RII). Les RNA vont être utilisés pour modéliser la fonction f de $\mathbb{R}^m$ dans $\mathbb{R}^n$ suivante :

$$HRTF = f(\text{paramètres d'entrée}) \tag{IV.7}$$

Un point important de cette modélisation porte sur le choix et la nature des paramètres d'entrée du modèle. Ces paramètres peuvent être, par exemple, la direction de la HRTF cible, ou des paramètres morphologiques de l'individu. Une des originalités de l'approche décrite dans la suite du chapitre est d'appliquer des HRTF mesurées en entrée du modèle. Le but est d'obtenir une procédure simplifiée de mesures qui ne nécessite qu'un nombre réduit de directions à mesurer : le RNA *apprend* et prédit les directions manquantes. Le problème réside alors dans le choix et le nombre des HRTF à appliquer en entrée du modèle. La solution adoptée pour identifier les HRTF à appliquer en entrée consiste à utiliser une méthode de regroupement de variable, ou *clustering* en anglais, qui est également basée sur des RNA nommés *cartes de Kohonen*.

Une question commune à tout problème de modélisation est la définition d'un critère d'erreur pour la construction du modèle et pour l'évaluation de ses performances. Cette question est problématique pour les HRTF à cause de leur nature duale : les HRTF sont à la fois des objets mathématiques, qui traduisent des phénomènes physiques, et des objets perceptifs. Il conviendrait alors de se doter d'un critère qui rende compte de la perception.

La suite du chapitre présente les détails de cette étude. Premièrement, une réflexion sera portée sur la nature du critère d'erreur. Les critères d'erreur classiques sont présentés ainsi que les critères généralement utilisés en synthèse binaurale. Ensuite, la question du nombre et de la position des HRTF mesurées à mettre en entrée de la fonction f (cf. équation IV.7) sera abordée. Deux méthodes sont comparées dont une qui utilise des regroupements statistiques des HRTF. Enfin, un RNA sera construit pour la prédiction de HRTF à partir d'un ensemble réduit de HRTF mesurées. Cette étude évaluera la faisabilité de la prédiction d'abord pour un seul individu d'une base de HRTF, puis pour tous les individus de la base.

3 DE L'IMPORTANCE DU CRITÈRE D'ERREUR

3.1 Introduction

Un critère d'erreur est une mesure de distance entre deux objets. La distance est ici une mesure de la différence entre un objet cible et un objet modélisé : elle sert pour la validation de la construction du modèle et pour la mesure de la qualité de modélisation.

Du point de vue mathématique, une distance d définit une relation stable entre deux éléments x, y d'un ensemble E et jouit des propriétés suivantes :

$$d(x,y) = d(y,x) \quad \text{(IV.8)}$$
$$d(x,y) = 0 \Leftrightarrow x = y \quad \text{(IV.9)}$$
$$\forall x,y,z \in E, \quad d(x,z) \leqslant d(x,y) + d(y,z) \quad \text{(IV.10)}$$

Une mesure de dissimilarité possède les mêmes propriétés sans l'inégalité triangulaire. Les objets particuliers que sont les HRTF peuvent être purement numériques (échantillons d'un signal), acoustiques (réalisation sonore d'un objet numérique) ou perceptif (image mentale créée par un auditeur à la réception d'un objet acoustique). En synthèse binaurale, il s'agit d'évaluer une distance entre deux HRTF (du point de vue signal) ou d'évaluer une distance perceptive entre deux réalisations sonores. Le *ou* est exclusif : le fonctionnement du système auditif n'est pas entièrement connu et donc un critère de distance appliqué entre deux HRTF n'est qu'une représentation d'une distance perceptive.

Pour obtenir une distance perceptive, il peut être envisagé de déterminer un espace perceptif multidimensionnel dans lequel des HRTF pourraient être projetées. Ainsi, les distances entre HRTF dans cet espace multidimensionnel représenteraient les distances perceptives recherchées. Cette approche a été utilisée notamment par McAdams pour la détermination d'un espace perceptif multidimensionnel du timbre musical [McAdams (1994)]. L'espace a été construit sur la base de jugements de dissimilarité entre des paires de sons d'instruments de musique. Les jugements de dissimilarité sont traduits en distances et la matrice des distances permet la projection de l'ensemble des objets sonores dans un espace multi-dimensionnel, selon un modèle mathématique. Dans cet espace, les objets similaires sont proches et les objets dissemblables sont éloignés. L'étape suivante consiste à caractériser, d'un point de vue perceptif, chacune des dimensions de l'espace et à les interpréter d'un point de vue physique, c'est-à-dire à les relier à des critères objectifs. Ainsi, McAdams a relié la première dimension de l'espace des timbres à la *brillance* qui semble être liée à l'enveloppe spectrale. Cette démarche appliquée aux HRTF permettrait à la fois la construction d'un espace perceptif où les HRTF pourraient être projetées, et l'interprétation des distances perceptives entre HRTF par des descripteurs physiques.

Cependant cette approche se heurte à plusieurs difficultés. Premièrement,l'hypothèse que l'objet d'étude peut être décrit selon plusieurs dimensions continues est faite. Or il se peut très bien que les HRTF soient catégorielles. De plus on peut se demander si le jugement de similitude, ou de préférence, à un sens pour les HRTF ? Quand bien même un sujet serait capable de juger une paire de HRTF selon leur similitude, quel type de stimuli (trajectoire ou point, son complexe ou son pur, scène complexe ou événement ponctuel) pourrait permettre d'en juger ? Deuxièmement, après construction de l'espace, il faut donner une signification objective aux axes. Ce travail peut être complexe tant les axes de projection peuvent exprimer des relations non linéaires des variables objectives classiques (niveau de pression, fréquences, sonie, centre de gravité).

Une alternative à la construction d'un espace perceptif des HRTF serait un espace sémantique. Les HRTF seraient alors projetées dans un espace où les relations de proximité traduisent une verbalisation, et donc un sens, similaire. La verbalisation associée aux axes principaux est ensuite identifiée en termes de paramètre objectif. Cette approche a été développée par Osgood sous le nom de *différentiation sémantique*. Les sujets doivent

placer des stimuli sur plusieurs échelles bipôlaires (par exemple axe aïgu ⇔ grave). Chacune des échelles a ses extrémités définies par des attributs verbaux opposés. Ensuite, au moyen d'une analyse factorielle (ACP par exemple), les composantes qui expriment le plus de variances sont extraites pour définir un sous-espace de représentation des HRTF. Cependant, on pressent la difficulté de trouver plusieurs qualificatifs verbaux permettant de décrire la sensation produite par l'écoute d'une scène via des HRTF (qualificatifs verbaux qui pourraient par ailleurs être différents selon les stimuli mis en jeu).

Une autre approche serait alors de répertorier tous les paramètres objectifs des HRTF et de faire une étude systématique sur leur influence perceptive. Cette approche à déjà été abordée, notamment dans [Bronkhorst (1995)], où est décrit l'influence de zones fréquentielles des HRTF. Cependant, une étude systématique ne saurait être exhaustive vu le nombre de paramètres en jeu (coordonnées spatiales, fréquences, amplitude).

Un critère de distance perceptive entre deux HRTF n'existe pas à l'heure actuelle et sa réalisation ne semble pas aisée. Les questions posées sont inhérentes au problème de relation entre paramètres physiques estimables et ressentis perceptifs, et dépassent largement le cadre des travaux de cette thèse. L'alternative proposée est alors d'appliquer des critères objectifs classiques, comme la distance euclidienne. La distance n'est pas alors interprétable en terme de qualité de rendu binaural, mais l'utilisation de critères classiques autorise un premier travail *exploratoire* qui permettra éventuellement la définition d'axes de recherches pour la création d'un critère d'erreur adapté à la prédiction de HRTF. En l'absence d'un critère perceptif, des écoutes doivent toujours être réalisées pour s'assurer de la qualité du rendu sonore. Une première méthode consiste à réaliser une comparaison par paire entre modélisation et mesure.

Dans la suite de cette section, les principales distances classiques sont décrites : elles servent de formulation de base à des critères d'erreur plus adaptés aux spécificités de la synthèse binaurale. Ensuite, les modifications qui peuvent être apportées, à la fois aux HRTF et à la formulation des critères classiques, seront présentés sous le terme de *critère d'erreur pour les filtres binauraux*.

3.2 Critère d'erreur classique

La qualité objective des outils de prédiction ou de modélisation est évaluée grâce à la fonction d'erreur complexe :

$$E(e^{\mathbf{i}\omega t}) = \|H(e^{\mathbf{i}\omega t}) - \widehat{H}(e^{\mathbf{i}\omega t})\| \tag{IV.11}$$

où $H(e^{\mathbf{i}\omega t})$ est la HRTF à reproduire et $\widehat{H}(e^{\mathbf{i}\omega t})$ est la HRTF modélisée. Plusieurs normes existent au sens mathématique. L'objet de ce paragraphe est de présenter les principales normes qui sont utilisées comme critère d'erreur.

3.2.1 Norme L_1

La norme L_1 permet une estimation linéaire de l'erreur et est souvent utilisée en premier lieu. Son expression est la suivante :

$$\begin{aligned} \|E\|_1 &= \|H(e^{\mathbf{i}\omega t}) - \widehat{H}(e^{\mathbf{i}\omega t})\|_1 && \text{(IV.12)} \\ &= \sum_{n=0}^{\infty} |h(n) - \widehat{h}(n)| && \text{(IV.13)} \\ &= \|h(n) - \widehat{h}(n)\|_1 && \text{(IV.14)} \\ & && \text{(IV.15)} \end{aligned}$$

Cette norme présente l'avantage d'être continue par morceaux et donc d'être différentiable par morceaux ce qui autorise l'emploi de techniques de gradient pour l'apprentissage du modèle.

3.2.2 Norme L_2

La norme Euclidienne, notée L_2 et aussi appelée norme de *Frobenius*, est particulièrement bien adaptée à la synthèse binaurale car, grâce à la relation de Parseval, L_2 peut être appliquée sur la composante à phae minimale des HRIR ou des HRTF. Le critère d'erreur associé prend la forme suivante :

$$\begin{aligned} \|E\|_2 &= \|H(e^{\mathbf{i}\omega t}) - \widehat{H}(e^{\mathbf{i}\omega t})\|_2 && \text{(IV.16)} \\ &= \sqrt{\sum_{n=0}^{\infty} (h(n) - \hat{h}(n))^2} && \text{(IV.17)} \\ &= \|h(n) - \widehat{h}(n)\|_2 && \text{(IV.18)} \end{aligned}$$

L'avantage de cette norme est la présence d'un minimum global, car la fonction d'erreur est quadratique. Un véritable avantage de cette norme est qu'elle est différentiable, ce qui autorise l'emploi de techniques de gradient pour les opérations d'optimisation.

3.2.3 Norme de Chebyshev

La norme L_∞ cherche à minimiser le maximum de la fonction d'erreur E :

$$\begin{aligned} \|E\|_\infty &= \|H(e^{\mathbf{i}\omega t}) - \widehat{H}(e^{\mathbf{i}\omega t})\|_\infty && \text{(IV.19)} \\ &= \max_{-\pi<\omega<\pi} |H(e^{\mathbf{i}\omega t}) - \widehat{H}(e^{\mathbf{i}\omega t})| && \text{(IV.20)} \end{aligned}$$

La norme de Chebyshev peut représenter un bon choix pour la prédiction de HRTF car les HRTF comportent de nombreux pics et creux qui sont perceptivement importants à reproduire [Huopaniemi et Smith (1999)]. La résolution logarithmique en amplitude de l'oreille peut être incorporée dans la nome L_∞. Un désavantage à l'utilisation de cette norme est la possibilité d'instabilité et/ou la présence de minima locaux fréquentiels. De plus, elle n'est pas différentiable et peut se révéler délicate à employer en optimisation.

3.3 Critère d'erreur pour les filtres binauraux

De nombreux critères d'erreur existent dans la littérature scientifique consacrée à la synthèse binaurale. La comparaison peut se faire sur de nombreux points. Etant donné le but de l'étude, qui est la prédiction de HRTF, les critères d'erreurs binauraux doivent donner une indication sur l'écart perceptif entre la HRTF à modéliser et la HRTF prédite. Ainsi un critère d'erreur optimal serait un critère binaire : 0 = "l'écart n'est pas perçu " = la prédiction est correcte, 1 = "l'écart est perçu " = la prédiction n'est pas correcte. Seulement comme indiqué en introduction du § 3, la création d'un tel critère sort du domaine de cette étude (de nombreuses expériences sont nécessaires). L'écart de prédiction se fait donc par l'intermédiaire d'une fonction d'erreur qui peut être difficile à interpréter en terme de qualité de prédiction.

Finalement, la comparaison entre des HRTF s'effectue souvent sur l'observation de leur représentation, comme par exemple la comparaison du module des HRTF en fonction de coordonnées spatiales. De plus, les HRTF étant un système interférentiel où le moindre décalage peut engendrer des écarts considérables, une figure permet d'avoir une vue d'ensemble des adéquations entre deux ensembles de HRTF. Ainsi, dans le chapitre III, consacré à l'acquisition de HRTF par BEM, les comparaisons entre les calculs BEM et les mesures se fait grâce à l'affichage des HRTF sur un plan (élévation constante par exemple). L'observation des figures permet entre autre de comparer la dynamique des modules et la reproduction ou non des figures de diffraction qui sont des indices pertinents pour la localisation sonore.

L'objet de ce paragraphe est de présenter les principaux critères utilisés en synthèse binaurale. Ces critères se basent sur des distances mathématiques. Les modifications apportées sont soit sur les données en entrée soit dans le critère lui-même.

3.3.1 Modification de l'échelle des fréquences

Echelle perceptive des fréquences L'utilisation de la transformée de Fourier pour l'analyse des signaux impose l'utilisation d'une échelle fréquentielle à pas constant alors qu'il a été montré que le système auditif est sensible à une échelle des fréquences non-linéaire du type échelle ERB (Equivalent Rectangular Bandwidth), dont l'échelle de Bark [Hartmann (1998)]. La résolution fréquentielle du système auditif est alors traduite en terme de largeur de bandes fréquentielles à l'intérieur desquelles le système auditif extrait une unique valeur, moyenne des contributions relatives de chaque fréquence dans la bande. Plusieurs échelles ont été proposées, les trois principales sont présentées. Pour l'échelle ERB, la largeur de bande est donnée en fonction de la fréquence centrale f_c :

$$\Delta f_{CE} = 24.7(4.37 f_c + 1) \tag{IV.21}$$

L'échelle de Bark utilise la formule suivante :

$$\Delta f_{CB} = 25 + 75\left(1 + 1.4(\frac{f_c}{1000})^2\right)^{0.69} \tag{IV.22}$$

et la largeur des bandes critiques est donnée par :

$$\Delta f_{cc} = 100Hz \quad \text{pour} f < 500Hz \tag{IV.23}$$
$$\Delta f_{cc} = 0.2 \times f \quad \text{pour} f \geq 500Hz \tag{IV.24}$$

Grâce à ces échelles fréquentielles plus proches de la perception qu'une échelle linéaire, des critères d'erreurs prenant en compte la résolution du système auditif peuvent être alors utilisées.

Distorsion de l'échelle des fréquences Afin d'améliorer la modélisation dans une zone particulière de fréquences, une distortion de l'échelle des fréquences peut être appliquée afin, par exemple, de dilater les basses fréquences et de comprimer les hautes fréquences. Cette technique, dite de *warping fréquentiel* est notamment utilisée pour la modélisation des HRTF en filtres RII, ce qui permet de concentrer l'effort de modélisation sur les basses fréquences [Marin (1996)]. Il s'agit d'une transformée bilinéaire opérée dans le plan des z qui rééchantillonne le spectre sur une nouvelle échelle des fréquences :

$$z \rightarrow \frac{z+\lambda}{\lambda z+1} \tag{IV.25}$$

où λ est le paramètre de dilatation ou *warping* en anglais. Pour des valeurs de λ comprises entre 0 et 1, les basses fréquences sont étirées et les hautes fréquences sont compressées (cf. fig. IV.7). Les valeurs de λ correspondant aux échelles ERB et Bark sont données dans [Smith et Abel (1999)]. Le spectre initial $H(e^{j\omega_k})$, avec $k = 0 \ldots N$ est remplacé par :

$$H_\lambda(e^{j\omega_k}) = H\big(\frac{e^{j\omega_k}+\lambda}{\lambda e^{j\omega_k}+1}\big) \tag{IV.26}$$

Les nouveaux vecteurs d'entrée du modèle sont alors les H_λ. Les vecteurs de sortie $\widehat{H}_\lambda$ doivent alors subir la transformée bilinéaire inverse pour revenir à l'échelle de fréquence initiale :

$$z \rightarrow \frac{z-\lambda}{1-\lambda z} \tag{IV.27}$$

3.3.2 Pré-traitement des HRTF

Trois techniques de traitement des HRTF sont présentées. Ces techniques permettent la prise en compte de la résolution du système auditif dans la représentation des signaux.

Lissage des HRTF Le lissage des HRTF consiste à réaliser des moyennes glissantes du module des HRTF sur des tailles de fenêtres fréquentielles variables de la manière suivante :

$$|H_s(f)| = \sqrt{\frac{1}{f_1-f_0}\int_{f_0}^{f_1} |H_s(f)|^2 df} \tag{IV.28}$$

avec $f_1 - f_0$ la largeur de bande autour de f. La taille de la fenêtre peut être définie par une échelle ERB, par bande d'octave entière ou fractionnaire ou encore par les bandes

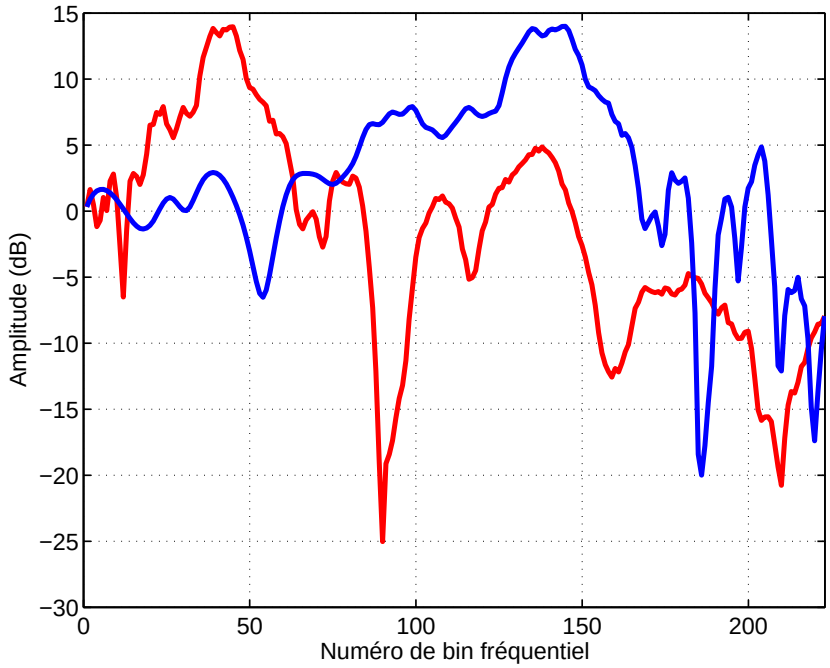


FIG. IV.7 – Ré-arrangement des fréquences par la technique de *warping fréquentiel*. Modules de HRTF correspondant à az = 0° et el = 0°. Rouge : HRTF initiale, Bleue : HRTF modifiée

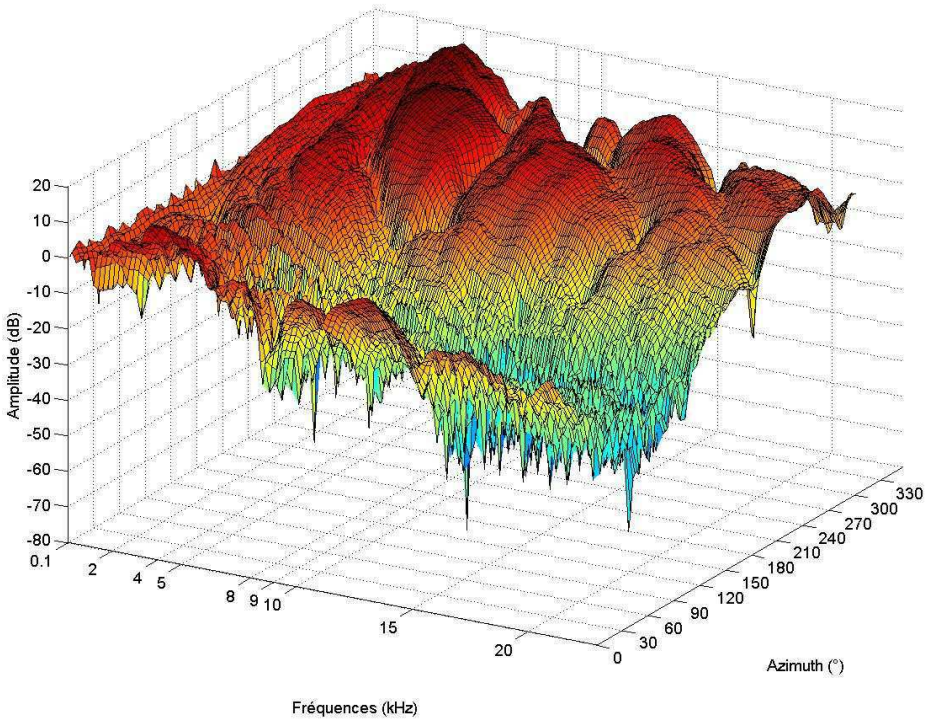
critiques du système auditif. L'effet du lissage des spectres des HRTF est illustré sur la figure IV.8. Le lissage présenté est réalisé en bandes critiques. Ce type de lissage n'est pas audible.

Une autre technique de lissage, utilisée le plus souvent en traitement de la parole, peut être appliquée aux HRTF [Kulkarni et Colburn (1995)]. Cette technique consiste a effectuer une moyenne glissante sur le cepstre de la réponse impulsionnelle à phase minimum. Cependant cette technique ne constitue pas un lissage perceptif car l'échelle de fréquence utilisée est constante.

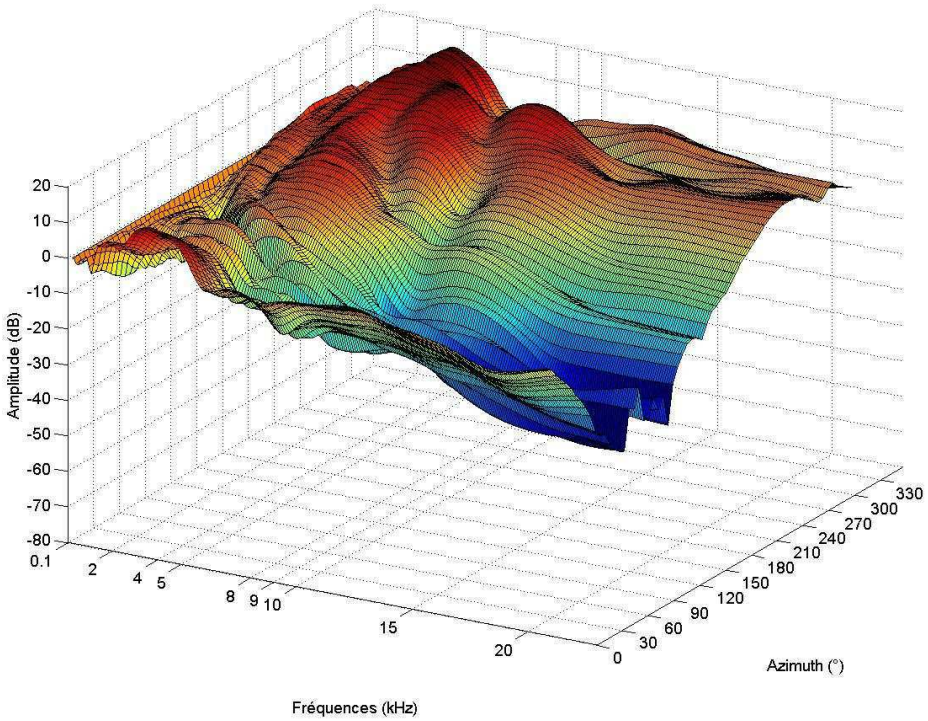
Pondération fréquentielle La résolution fréquentielle de l'oreille peut être utilisée pour pondérer le signal de manière perceptive. Ainsi les poids appliqués peuvent correspondre à l'inverse de la largeur de bande ERB (cf. équation IV.21) ou Bark (cf. équation IV.22) en fonction de la fréquence. Par exemple, pour la norme L_2 entre deux HRTF échantillonnées :

$$E_{pond} = \sqrt{\sum_{n=0}^{\infty} w_n (h(n) - \hat{h}(n))^2} \tag{IV.29}$$

avec $w_n = \frac{1}{\Delta f_n}$ largeur de bande autour de f. La figure IV.9 donne les valeurs des coefficients fréquentiels Bark et ERB dans le cas d'une pondération par l'inverse de la largeur de bande.



a) Mesures non lissées.



b) Mesure lissées en bandes critiques.

FIG. IV.8 – Illustration de l'effet d'un lissage des modules de HRTF. Les HRTF sont présentées pour le plan horizontal de 500 à 5000 Hz. a) Mesures non lissées , b) Mesure lissées en bandes critiques.

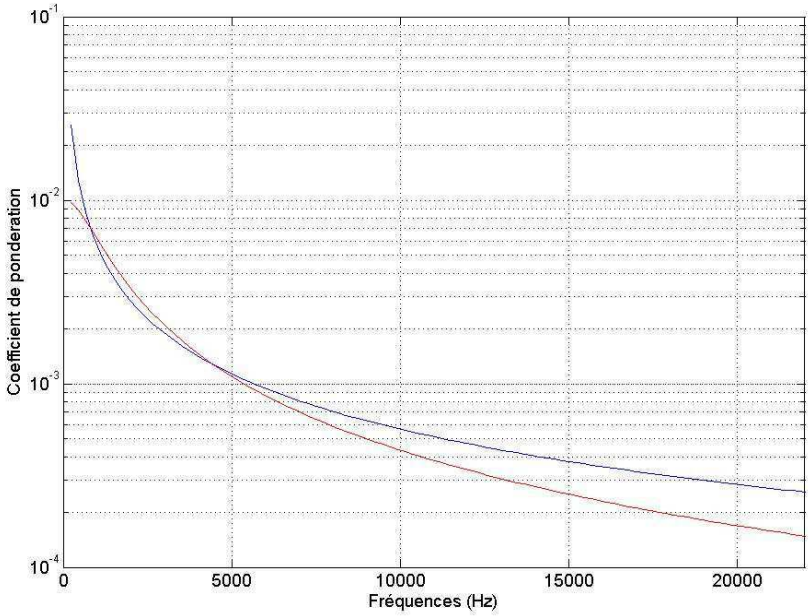


FIG. IV.9 – Coefficients de pondération fréquentielle relatifs à l'inverse de la largeur de bande considérée : échelle de Bark (bleu) et échelle ERB (rouge).

Pondération spatiale Les performances du système auditif de localisation d'une source sonore dépendent fortement de la position de la source. Ainsi, il a été montré que le pouvoir de résolution du système auditif est plus important devant, de l'ordre de 2°-3°, que derrière, (5°) et est plus faible sur les côtés, (7°-10°) [Blauert (1983); Mills (1958); Oldfield et Parker (1984)]. Ainsi il peut être intéressant d'introduire ces dépendances dans les calculs d'erreur. Une première pondération est effectuée lors de l'égalisation des HRTF issues de mesures (cf. § I.4.2) : l'égalisation *champ diffus* ne privilégie aucune direction de l'espace tandis que l'égalisation *champ libre* privilégie la reconstruction pour une certaine direction, typiquement la direction frontale ($\theta = 0°$ et $\phi = 0°$). Il peut être aussi intéressant d'améliorer la prédiction pour des HRTF ayant une faible énergie (principalement les HRTF contralatérales), car les critères d'erreur usuels négligent les erreurs faibles. Une pondération par l'inverse de l'énergie de chaque position remplit ce rôle :

$$w_i = \frac{1}{\sum_{j=1}^{q} |H_{i,j}|^2} \tag{IV.30}$$

avec i indice de position et j indice de fréquence. La reproduction de sources frontales est souvent problématique en synthèse binaurale, surtout dans le cas d'une écoute avec des HRTF non-individuelles. C'est pourquoi, une pondération spécifique privilégiant les positions situées devant l'auditeur peut être apportée. Un exemple d'une telle pondération est donnée dans [Rio et Warusfel (2002)] et passe par la définition d'un vecteur d'écart par rapport à la position frontale :

$$v_i(\lambda) = \frac{(1-\lambda)(\cos(\theta_i)\cos(\phi_i)+1)}{2(1-\lambda\cos(\theta_i)\cos(\phi_i))} \tag{IV.31}$$

où λ est un coefficient déterminant la largeur de la courbe définie par $v_i(\lambda)$. Le vecteur d'écart vaut 1, son maximum, pour $\theta = 0$ et $\phi = 0$ et vaut 0, son minimum pour $\theta = 0$ et $\phi = 180$. Ce vecteur est ensuite utilisé dans la fonction de pondération w_i :

$$w_i(\alpha, \lambda) = 1 + (\alpha - 1)v_i(\lambda) \tag{IV.32}$$

où α désigne le rapport entre poids arrière et poids avant. La figure IV.10 montre sur sa partie gauche les valeurs de v_i pour différentes valeurs de λ et sur sa partie droite la famille de fonction de poids pour $\alpha = 1.5$.

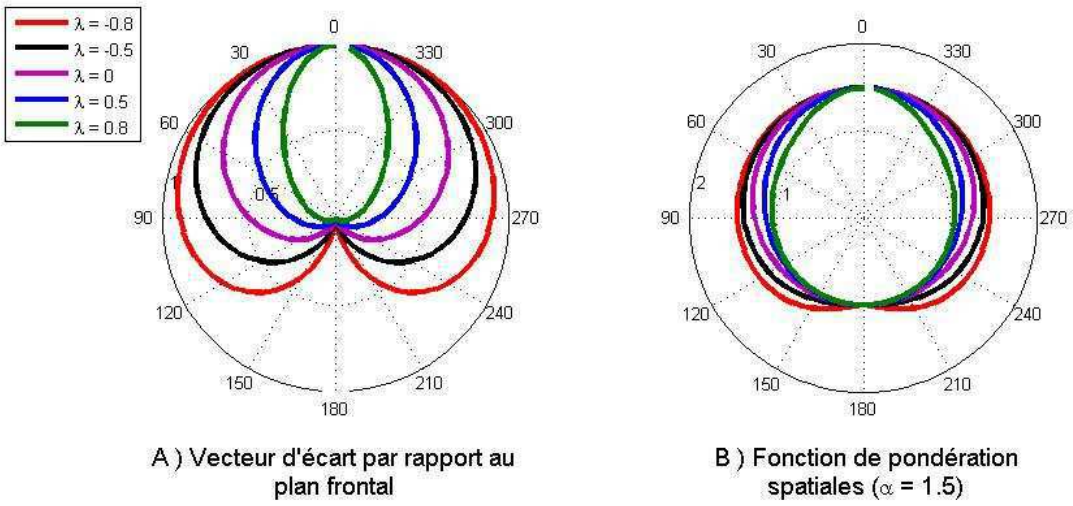


FIG. IV.10 – Diagrammes polaires de pondérations spatiales privilégiant les positions frontales.

3.4 Conclusion

Un large choix de critères d'erreur objectifs est disponible et de nouveaux critères qui tiennent compte de la résolution fréquentielle et angulaire peuvent être créés. L'étude

présentée dans la suite du document est de nature *exploratoire*, car le réseau est entraîné pour prédire des HRTF, et non des paramètres de modélisation de HRTF. Dans ce cas, il est préférable d'utiliser des outils bien maîtrisés. Ainsi, l'erreur basée sur la norme L_2 est utilisée notamment pour ces propriétés de dérivabilité qui autorise l'utilisation d'algorithme d'optimisation tel que la rétropropagation du gradient de l'erreur. La norme L_1 est elle utilisée pour l'évaluation des performances de modélisation. Un axe de recherche pertinent serait la prise en compte du critère d'erreur comme paramètre de modélisation et d'évaluation.

4 CLASSIFICATION DE HRTF : QUELS SONT LES VECTEURS D'ENTRÉE DU MODÈLE ?

Le but de l'étude présentée dans ce chapitre est de déterminer la faisabilité d'une réduction du nombre de HRTF dans une procédure de mesure et de prédire les autres au moyen d'un RNA. La technique neuronale est vue ici comme une technique de réduction de données tout comme les techniques de décomposition linéaire des HRTF (cf. § 1.2. L'apport du RNA est que la décomposition est non linéaire ce qui lui autorise l'apprentissage de relations de haut niveau entre les HRTF en entrée du modèle (les HRTF mesurées) et HRTF en sortie (les HRTF à des positions non mesurées).

L'étude est séparée en deux parties : choix du nombre et de la position des points de mesure et prédiction des autres mesures. L'idée est de réaliser un regroupement des HRTF par similarité et d'élire des *centres*, ou *représentants*, au sein de chaque groupe : les centres désignent alors les positions des HRTF à soumettre en entrée du modèle de prédiction. Deux types de regroupement sont effectués : un regroupement sur les modules des HRTF, désigné *méthode statistique* par la suite, et un regroupement sur les coordonnées des positions des HRTF, labellisé *méthode géométrique*.

Cette section présente en premier lieu le principe des méthodes utilisées pour la classification des données. Trois méthodes de classification sont abordées : l'algorithme des *k-moyennes*, la *classification hiérarchique ascendante* (CHA) et les *cartes de Kohonen*. Ensuite le travail de sélection des HRTF *représentatives*, c'est-à-dire les HRTF à appliquer en entrée du RNA, est présenté.

4.1 Méthodes de Classification

4.1.1 Introduction

Le but des méthodes de clustering (de l'anglais *cluster* = groupe, amas) est de regrouper entre elles des variables *semblables*. Il est alors possible de remplacer les groupes ainsi déterminés par un représentant. Un exemple de clustering est l'algorithme des k-moyennes. La grande difficulté de cet algorithme est sa non-reproductibilité et le choix difficile d'un nombre *optimal* de centres. Les techniques de classification hiérarchique ascendante (CHA) sont aussi des techniques de clustering, même s'il est plus difficile de trouver un représentant pour les classes (les classes ne sont pas nécessairement convexes et la moyenne de la classe peut être en dehors de la classe par exemple). Ces techniques sont difficiles à appliquer directement sur les données (trop d'individus). Souvent, ces techniques de classification hiérarchique, dont le but est plus de montrer la structure des

données, sont appliquées après une application de l'algorithme des k-moyennes avec k grand, ce qui permet d'éviter d'avoir à choisir au préalable le nombre de centres.

Dans tous les cas, le défaut de ces techniques est que la description des données est encore faite dans un espace de grande dimension (même si c'est avec nettement moins de variables) et qu'il reste à comparer ces variables entre elles pour prendre vraiment connaissance des données. La classification hiérarchique ascendante répond en partie à ce problème en regroupant entre eux des groupes proches mais elle n'y répond qu'en partie car elle n'apprend rien par exemple sur les relations de plus ou moins grande proximité entre des groupes classés dans des branches différentes de l'arbre.

La succession de deux algorithmes de classification, telle qu'un premier regroupement effectué par la méthode des k-moyennes et un deuxième avec la méthode CHA sur les groupes issus de la première méthode, a beaucoup d'avantages car elle corrige dans une large mesure les défauts des deux algorithmes :

- choix d'un nombre important de classes pour les k-moyennes
- la sensibilité aux conditions initiales devient anecdotique puisque la CHA va faire de "grands" regroupements (la structure "ultra-fine" de ces regroupements n'est d'aucun intérêt)
- la CHA travaille avec peu de variables : elle est donc viable
- la méthode des k-moyennes a réglé en grande partie le problème des points aberrants et isolés ce qui améliore la classification de la CHA.

La méthode statistique de sélection de HRTF représentatives utilise une succession de deux méthodes : une carte de kohonen pour obtenir un premier regroupement des HRTF suivie d'un CHA pour réduire le nombre de groupes.

4.1.2 Algorithme des k-moyennes

Le clustering par l'algorithme des *k-moyennes* permet le regroupement de données en k classes, k étant fixé au départ [Gray (1984)]. Les quatre étapes de cet algorithme sont les suivantes :

1 définir k vecteurs comme étant les *centres*,
2 attribuer chaque vecteur au centre dont il est le plus proche, au sens de la distance euclidienne,
3 calculer les coordonnées des nouveaux centres comme la moyenne des vecteurs qui leurs sont attribués,
4 tant que l'étape 3 modifie les centres, aller à l'étape 2

4.1.3 Classification Hiérarchique Ascendante

Le principe des algorithmes de classification hiérarchique ascendante est simple :

- Initialisation : chaque élément de l'espace de départ constitue une classe. Une *distance* D est calculée entre toutes les classes.
- Tant que nombre de classes > 1
  - regrouper les deux classes les plus proches au sens de la *distance* D,
  - calcul des *distances* entre la nouvelle classe et les autres.

Dans l'étude présentée ici, cette méthode est appliquée à une partition de l'ensemble de données issue de la carte de Kohonen des HRTF. Le critère de distance inter-classe utilisé est le critère de Ward qui mesure l'inertie intra-classe. L'utilisation de ce critère

permet l'obtention de classes convexes. La valeur du critère est fixée selon un compromis entre le nombre de classes et la perte de variance intra-classe. La figure IV.11 indique l'évolution du critère de Ward en fonction du nombre de classes. La valeur du critère de Ward normalisé est choisie égale à 0,5.

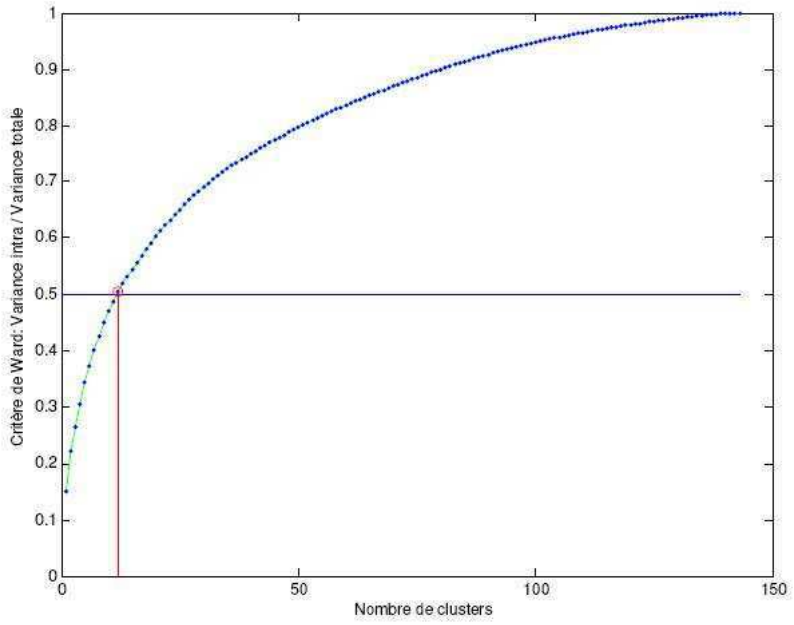


FIG. IV.11 – Evolution du critère de Ward normalisé en fonction du nombre de classes. La valeur du critère choisie correspond ici à 13 clusters.

4.1.4 Clustering de HRTF

Le clustering peut servir de plusieurs manières en synthèse binaurale, selon que l'on cherche à prédire les HRTF [Lemaire et al. (2005)], interpoler les HRTF [Fahn et Lo (2003)] ou encore travailler sur des modèles d'écoute binaurale [Palomäki et al. (2000)].

Dans une étude sur le clustering des HRTF d'un mannequin KEMAR dans le plan horizontal, Fahn [Fahn et Lo (2003)] utilisent la représentation cepstrale des HRTF modélisées par des filtres IIR en entrée d'un algorithme LBG. Cet algorithme permet de regrouper les 72 HRTF du plan horizontal en 12 clusters. 12 HRTF représentatives sont alors élues et l'étude montre que les HRTF restantes peuvent être correctement interpolées par une interpolation linéaire entre les HRTF représentatives les plus proches.

4.2 Cartes de Kohonen

4.2.1 Présentation

Une carte de Kohonen est composée d'un ensemble de k points de l'espace (espace dans lequel reposent les données) liés entre eux par des relations de voisinage. L'ensemble de ces relations de voisinage constitue la topologie de la carte. La topologie définit des

relations de voisinage entre points de la carte et ces relations de voisinage n'ont a priori aucun caractère métrique : des *voisins* au sens de la topologie peuvent être très éloignés dans l'espace des données (la carte en deux dimensions ne représente pas bien l'organisation des données) et de même des points très proches de la carte peuvent ne pas être voisins (cas rare où la carte doit être séparée en deux pour une meilleure représentation de la topologie des données). La notion de voisinage dans l'espace des données est liée au calcul de distance entre deux éléments de l'espace des données. Dans notre étude il s'agit de la distance inter-spectre de la norme L_2. Le processus d'apprentissage de la carte de Kohonen consiste à faire correspondre le voisinage topologique à une proximité métrique et peut être résumé par les 4 étapes suivantes, une fois la topologie choisie et initialisée :

1 Un vecteur est tiré au hasard dans l'espace des données,
2 Le point de la carte le plus proche de ce vecteur est déterminé (le *point gagnant* ou BMU en anglais pour Best Matching Unit),
3 Le BMU est rapproché du vecteur,
4 Les voisins du BMU sont aussi rapprochés du vecteur (cf. fig.IV.12).

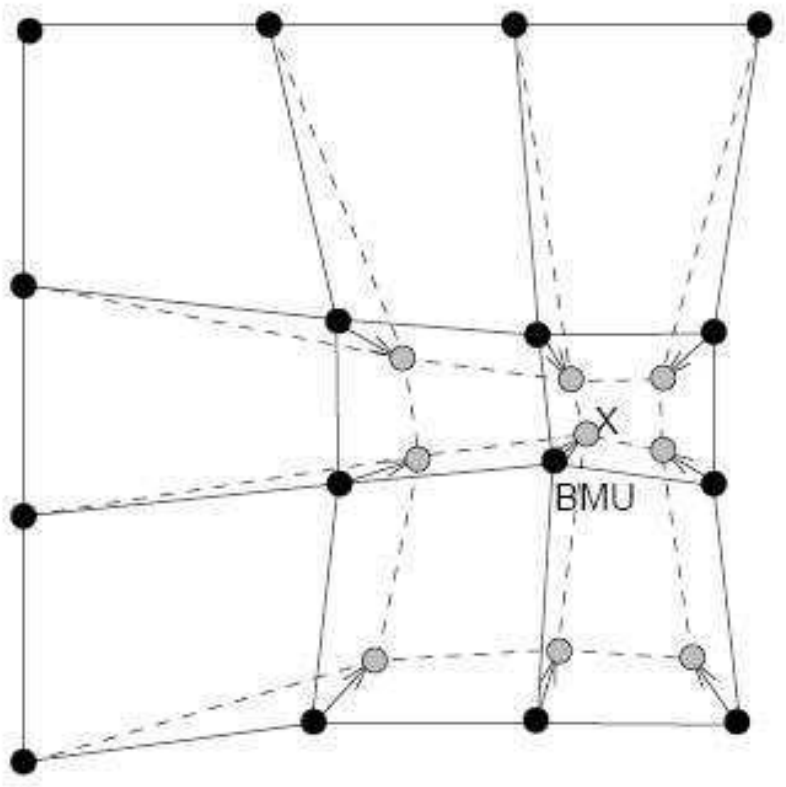


Fig. IV.12 – Processus d'adaptation du BMU et des ses voisins à la présentation d'un échantillon **x**. Les lignes continues correspondent à la situation antérieure et les lignes en pointillé à la nouvelle situation.

Cet apprentissage fait en sorte que le voisinage topologique vienne coïncider avec la proximité métrique : comme le BMU *traîne* ses voisins avec lui quand il est attiré par un individu, on comprend que des points voisins au sens de la topologie ont tendance à

aller du même côté. Une carte de Kohonen de dimension 2 peut être interprétée comme des points représentatifs d'une sous-variété de dimension 2 qui *passe* le mieux possible (au sens de la métrique) à travers les données. Comme une méthode factorielle, ACP ou ACI par exemple, la carte de Kohonen effectue une diminution du nombre de dimensions pour la représentation des données ; cette diminution est même drastique, la dimension de représentation étant en général 2 mais à la différence des méthodes factorielles classiques, cette représentation en dimension 2 n'est pas un plan mais une sous-variété de dimension 2 ce qui permet d'épouser beaucoup mieux la distribution des données tout en restant lisible comme une carte. Les cartes de Kohonen sont donc à la fois une méthode de clustering des données (les k points de l'espace des données) et une technique de réduction de la dimensionnalité des données (la représentation en deux dimensions des données).

Si on visualise sur la carte la valeur de la projection d'une variable pour chaque point représentatif, on a une image de la répartition de la variable sur la carte. Ces images peuvent être considérées comme les vecteurs représentatifs des variables et ces vecteurs peuvent à leur tour servir à la construction d'une carte de Kohonen, la carte des variables, qui peut à son tour servir de base à l'étude des relations des variables entre elles (clustering des variables en particulier).

4.2.2 Construction d'une carte de Kohonen

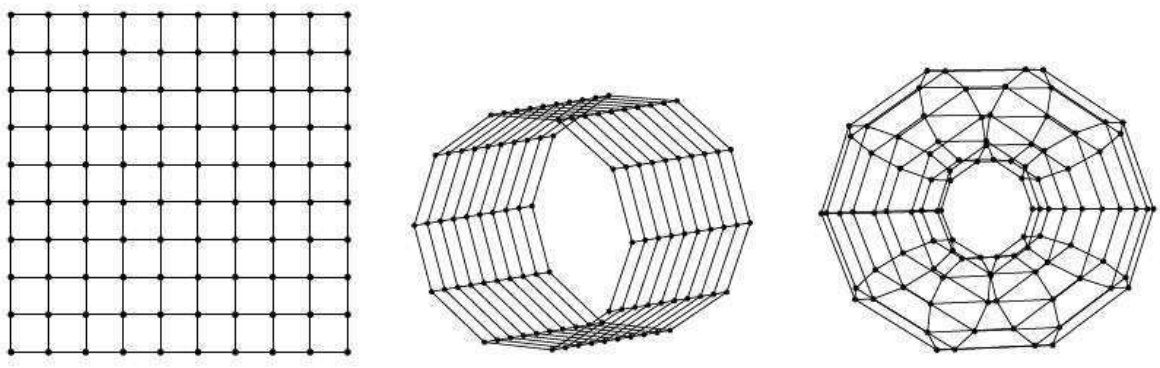
Les cartes de Kohonen, ou cartes auto-organisatrices, sont une disposition de neurones artificiels[3], les n points de l'espace de départ, selon une grille régulière. Le nombre de neurones est fixé a priori et peut atteindre plusieurs milliers. Chaque neurone de la carte est représenté par un vecteur de poids $\mathbf{m} = [m_1, ..., m_d]$, ou prototype, de dimension d égale à la dimension des vecteurs d'entrée. Le prototype représente une forme moyenne des vecteurs contenus dans le neurone et est composé des attributs caractéristiques des vecteurs du neurone. La topologie de la carte est définie à l'aide de deux facteurs : la forme globale de la carte (cf. fig.IV.13) et la structure locale du treillis(cf. fig.IV.14).

L'algorithme d'apprentissage est proche d'une méthode de quantification vectorielle telle que l'algorithme des *k-moyennes* (cf. § 4.1.2), dont le résultat est le prototype pour chaque neurone. La différence avec un algorithme des *k-moyennes* est que le voisinage du prototype s'étire en direction d'un nouvel échantillon (cf. fig.IV.12).

Deux types de fonctionnement sont distingués. Dans un premier temps, l'ensemble des vecteurs devant être appris est présenté au réseau et les vecteurs de poids sont mis à jour de manière à approximer les vecteurs d'entrée (algorithme d'apprentissage global). Les paramètres de la carte sont adaptés au fur et à mesure pour qu'elle se stabilise de plus en plus : c'est l'étape d'initialisation de la carte. La deuxième étape est la phase d'utilisation proprement dite. Dans ce cas, on présente un motif particulier et c'est le neurone dont le vecteur de poids minimise la distance avec le vecteur d'entrée qui réagit (algorithme d'apprentissage séquentiel).

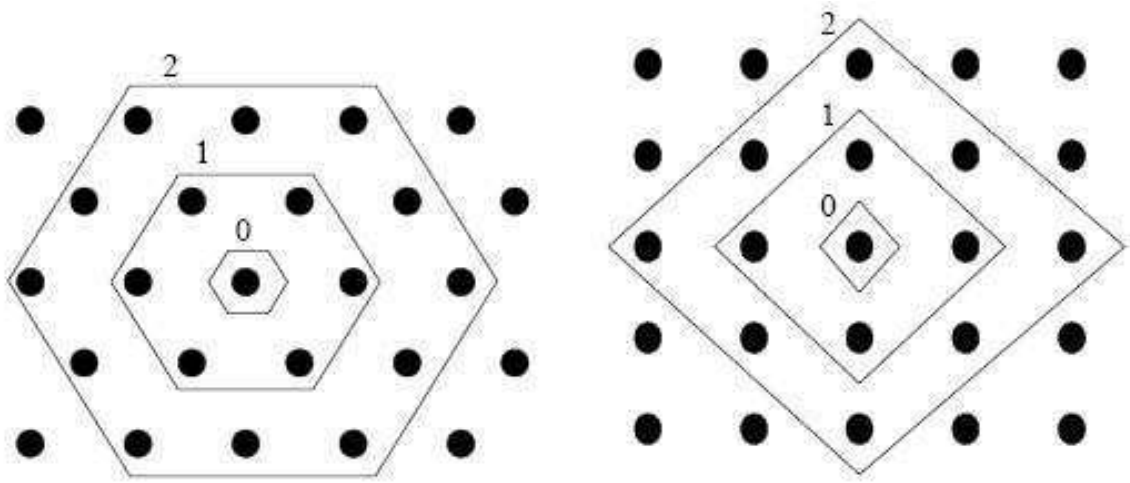
Algorithme d'apprentissage global L'algorithme d'apprentissage global est itératif. Les données en entrée sont la totalité des données partagées en n groupes de neurones : un vecteur d'entrée appartient au groupe de neurones qui minimise la distance entre le vecteur d'entrée et le groupe de neurones. Après cette étape de partage des données, les

[3]Pour l'étude présentée ici, un neurone regroupe plusieurs modules de HRTF



A) Forme par défaut. B) Forme cylindrique. C) Forme toroïdale.

FIG. IV.13 – Différentes formes de carte de Kohonen. A) Forme par défaut, et deux formes adaptées à des données circulaires B) Cylindrique et C) Toroïdale [Vesanto et al. (2000)].



A) Treillis hexagonal. B) Treillis rectangulaire.

FIG. IV.14 – Différent treillis avec leurs voisinages 0, 1 et 2. A) Treillis hexagonal B) Treillis rectangulaire. Le polygone le plus au centre correspond au voisinage 0, le second au voisinage 1 et le plus grand au voisinage 3 [Vesanto et al. (2000)].

prototypes sont calculés :

$$\mathbf{m_i}(t+1) = \frac{\sum_{j=1}^{n} h_{ic}(t)\mathbf{x}_j}{\sum_{j=1}^{n} h_{ic}(t)} \tag{IV.33}$$

avec $c = \arg\min_k \|\mathbf{x_j} - \mathbf{m_k}\|$ est l'index du BMU correspondant au vecteur $\mathbf{x_j}$, et le n le nombre de groupes de neurones. Les nouveaux prototypes sont des moyennes des vecteurs d'entrée pondérés par la valeur de la fonction de voisinage $h_{ic}(t)$ prise pour le BMU, c'est-à-dire à l'index c. L'apprentissage s'arrête quand les prototypes ne sont plus

modifiés. La fonction de voisinage est une fonction décroissante de l'itération et de la distance entre le neurone i et le neurone contenant le BMU (neurone c). Elle décrit la zone d'influence d'un vecteur d'entrée sur la carte (cf. fig. IV.16).

Algorithme d'apprentissage séquentiel Les cartes de Kohonen sont construites de manière itérative. A chaque pas d'apprentissage, un vecteur $\mathbf{x}$ des données de départ est choisi aléatoirement et les distances entre lui et tous les prototypes sont calculées. Le prototype qui minimise la distance avec $\mathbf{x}$, le BMU, correspond à :

$$\|\mathbf{x} - \mathbf{m_c}\| = \min_i \|\mathbf{x} - \mathbf{m_i}\| \tag{IV.34}$$

avec $\|.\|$ critère de distance, généralement la distance euclidienne ou moindre carrée, $\mathbf{m_i}$ le prototype associé au neurone i et $\mathbf{m_c}$ le BMU. Après avoir élu le BMU, la topographie de la carte est modifiée ainsi que les valeurs des prototypes. La position des prototypes est modifiée de telle sorte que le BMU et ses voisins se retrouvent plus près du vecteur d'entrée et ce dans l'espace de départ et sur la carte. La modification topologique induite par ce processus est schématisée sur la figure IV.12. La règle d'adaptation d'un prototype est donnée par :

$$\mathbf{m_i}(t+1) = \mathbf{m_i}(t) + \alpha(t) h_{ci}(t)[\mathbf{x}(t) - \mathbf{m_i}(t)] \tag{IV.35}$$

où i est l'indice du neurone, t le temps d'apprentissage, $\mathbf{x}(t)$ un vecteur d'entrée choisi aléatoirement au temps t, $h_{ci}(t)$ est la fonction de voisinage autour du BMU $\mathbf{m_c}$ et $\alpha(t)$ le taux d'apprentissage au temps t (cf. fig. IV.15). Le processus d'apprentissage est généralement réalisé en deux phases. Dans la première, des taux d'apprentissage α_0 et des rayons de voisinage σ_0 élevés sont utilisés. Ceci permet d'accorder approximativement la carte à l'espace des données. Ensuite, dans la deuxième phase, des taux d'apprentissage et des rayons de voisinage relativement petits sont utilisés ce qui permet un accord fin de la carte. Dans la présente étude, un taux d'apprentissage fixe et petit (0,001 typiquement) et une fonction de voisinage *fenêtre* décroissant linéairement jusqu'à un voisinage de taille 1 (cf. fig.IV.14), pour conserver le lien avec les premiers voisins, sont utilisés.

4.3 Application des cartes de Kohonen à la classification des HRTF

Le but de cette étude est de connaître le nombre et la localisation des HRTF représentatives pour une base de données. Deux méthodes sont comparées : une méthode statistique et une méthode géométrique. La méthode statistique est une méthode de sélection de variables : clustering des modules des HRTF par carte de Kohonen suivi d'une CHA (Clustering Hierarchique Ascendant) pour réduire le nombre de groupes définis par la carte de Kohonen. La méthode géométrique est une sélection des coordonnées spatiales des HRTF uniformément réparties sur la sphère des points de mesures. La figure IV.17 illustre les méthodes employées et la figure IV.18 indique les résultats des deux méthodes pour le cas de 4 représentants. L'approche statistique pour le regroupement des HRTF permet d'appréhender la présence de redondance entre les HRTF et ainsi appliquer une réduction des données qui assimile les HRTF d'un groupe à la HRTF représentative. L'erreur introduite par cette réduction de donnée est appelée *erreur de quantification*. Cette erreur permet d'estimer la *qualité* du regroupement.

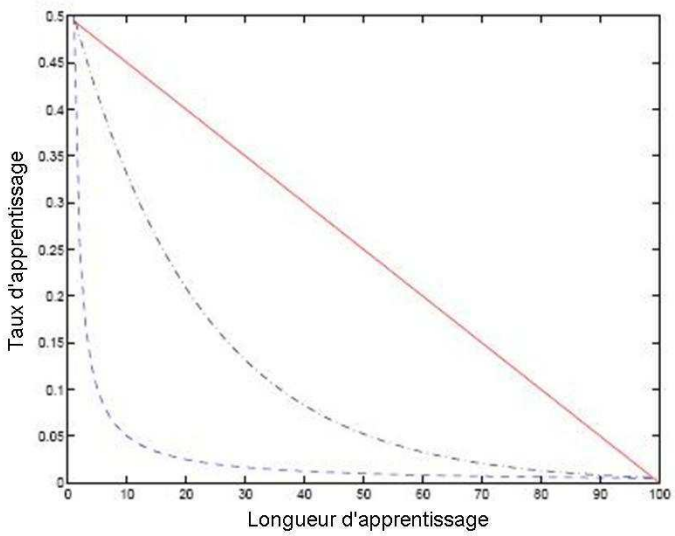


FIG. IV.15 – Différents taux d'apprentissage : *linéaire* (ligne continue) $\alpha(t) = \alpha_0(1 - \frac{t}{T})$, *puissance* (ligne interrompue) $\alpha(t) = \alpha_0(\frac{0.005}{\alpha_0})^{\frac{t}{T}}$ et *inverse* (ligne en pointillés) $\alpha(t) = \frac{\alpha_0}{(1+\frac{100t}{T})}$. T représente la longueur de l'apprentissage et α_0 le taux initial d'apprentissage [Vesanto et al. (2000)].

4.3.1 Nature des données

Les HRIR disponibles dans la base de données CIPIC sont utilisées pour cette étude [Algazi et al. (2001)]. Les HRIR y sont répertoriées pour chaque oreille des individus et pour les différentes positions de mesures déterminées par les angles θ et ϕ, respectivement azimut et élévation en coordonnées sphériques-interaurales (cf. § III.1.3.2). La base comporte 44 individus et 2500 HRIR par individus (1250 positions * 2 oreilles). Le travail s'est effectué sur les modules des HRTF, c'est-à-dire sur les 100 composantes des fréquences positives de la transformée de Fourier des HRIR de l'oreille droite (1250*100 échantillons fréquentiels). Une des règles de travail avec les RNA est d'apporter toutes les informations disponibles a priori. Cela *aide* le réseau dans sa tâche : des relations de haut niveau et de nature très complexe peuvent apparaître entre les entrées et les sorties. Cependant il se peut aussi que le réseau donne de meilleurs résultats avec des données brutes non transformées. C'est pourquoi l'influence de différents types de vecteurs d'entrée est étudiée. Principalement deux types d'égalisation des HRTF existent (cf. § I.4.2) et sont appliqués aux vecteurs d'entrées : l'égalisation champ libre et l'égalisation champ diffus. De plus, afin de tenir compte de la résolution fréquentielle du système auditif, une pondération fréquentielle de type ERB est apportée aux données (cf. § 3.3.2). C'est pourquoi la première partie de l'étude compare les performances d'apprentissage du réseau en utilisant 4 types de vecteurs d'entrées qui sont définis dans le tableau IV.1.

Qui plus est, l'oreille est un capteur sensible à des variations logarithmiques de la pression acoustique, toutes les études présentées au tableau IV.1 sont effectuées sur le logarithme des modules de HRTF. Enfin, pour réduire la dynamique des vecteurs d'entrée,

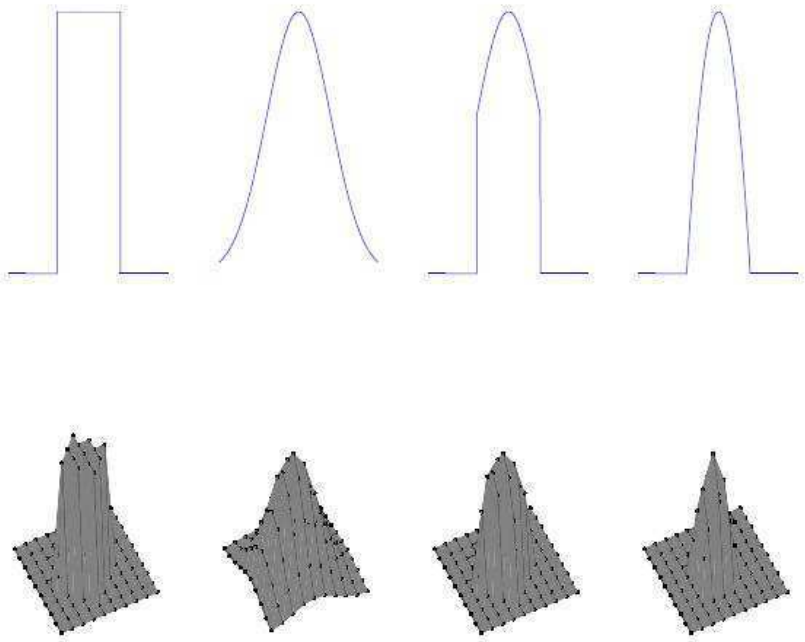


FIG. IV.16 – Différentes fonctions de voisinage. De gauche à droite : *fenêtre* $h_{ci}(t) = \mathbf{L}(\sigma_t - d_{ci})$, *gaussienne* $h_{ci}(t) = exp(\frac{-d_{ci}^2}{2\sigma_t^2})$, *gausienne coupée* $h_{ci}(t) = exp(\frac{-d_{ci}^2}{2\sigma_t^2})\mathbf{L}(\sigma_t - d_{ci})$ et *ep* $\max 0, 1 - (\sigma_t - d_{ci})^2$, avec σ_t le rayon de voisinage au temps t, $d_{ci} = \|\mathbf{r}_c - \mathbf{r}_i\|$ la distance entre le neurone c et le neurone i and $\mathbf{L}(x)$ est la function de Heaviside : $\mathbf{L}(x) = 0$ si $x < 0$ et $\mathbf{L}(x) = 1$ si $x > 0$. Les figures situées sur la rangée supérieure représentent les fonctions de voisinage pour une carte à une dimension et les figures sur la rangée inférieure pour des cartes en deux dimensions [Vesanto et al. (2000)].

TAB. IV.1 – Nature des vecteurs d'entrées.

Pondération fréquentielle	Egalisation	Identification
non	champs libre	étude 1
ERB	champs libre	étude 2
non	champs diffus	étude 3
ERB	champs diffus	étude 4

les variations ont été limitées au seuil de -80 dB. Ainsi les vecteurs d'entrée sont de la forme :

$$V_e = \begin{pmatrix} 20.\sqrt{\alpha_1}.\log_{10}(\max(HRTF_{\theta,\phi}(f1), 0.0001)) \\ 20.\sqrt{\alpha_2}.\log_{10}(\max(HRTF_{\theta,\phi}(f2), 0.0001)) \\ \vdots \\ 20.\sqrt{\alpha_{100}}.\log_{10}(\max(HRTF_{\theta,\phi}(f100), 0.0001)) \end{pmatrix} \quad \text{(IV.36)}$$

avec $\alpha_i = 1$ dans le cas des études 1 et 3 et $\alpha_i = (24.7(4.37 f_i + 1))^{-1}$ pour les études 2 et 4.

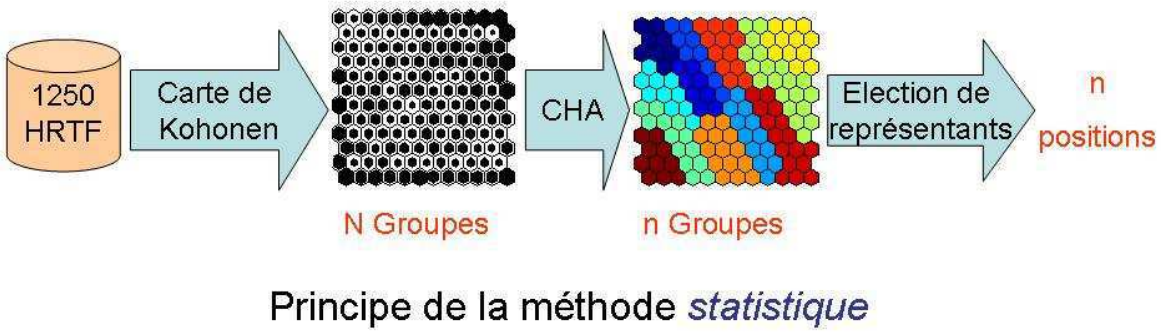


FIG. IV.17 – Schéma illustrant les différentes étapes des deux méthodes de sélection des réprésentants.

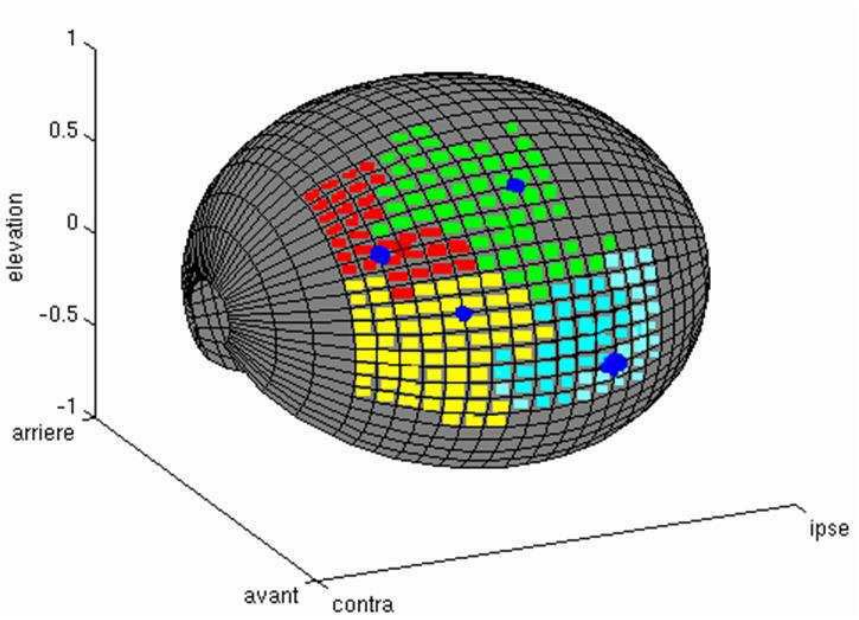
4.3.2 Clustering par carte de Kohonen

La carte de Kohonen des HRTF est entraînée avec la distance euclidienne entre deux vecteurs d'entrée :

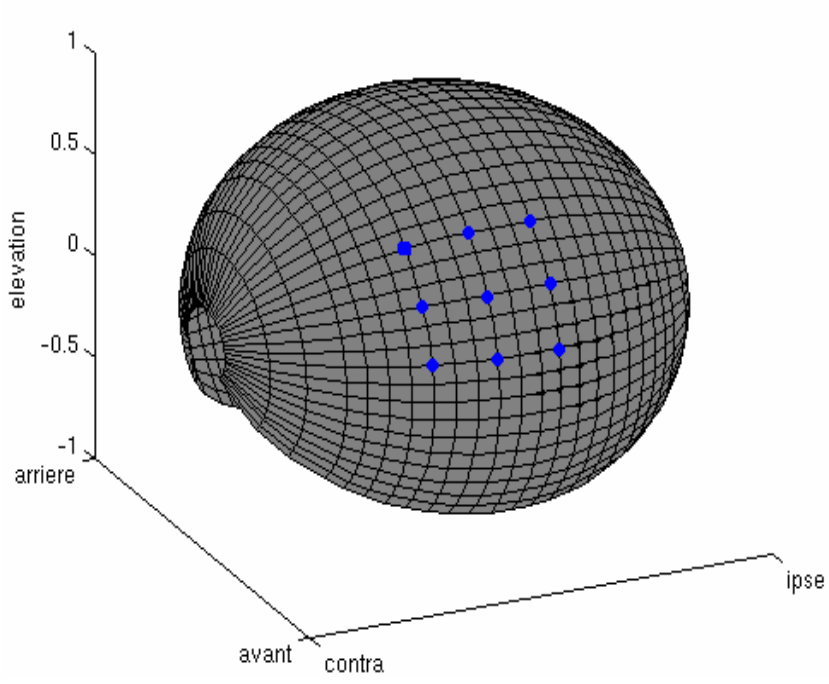
$$d(HRTF_{\lambda,\theta,\phi}, HRTF_{\lambda',\theta',\phi'}) = \frac{1}{100}\sum_{n=1}^{100}\left(20.\sqrt{\alpha_1}.log_{10}\left(\frac{\max(abs(HRTF_{\theta,\phi}(f_n)), 0.0001)}{\max(abs(HRTF_{\lambda',\theta',\phi'}(f_n)), 0.0001)}\right)\right)^2 \quad \text{(IV.37)}$$

avec λ indice relatif à l'individu. Tous les calculs de carte de Kohonen sont réalisés grâce à la boîte à outils *SOM toolbox* [Vesanto et al. (2000)]. La topologie de la carte est choisie en deux dimensions avec un voisinage hexagonal et est composée de 144 neurones (12×12). Cette topologie est déterminée empiriquement : il n'existe pas de méthode d'optimisation d'une topologie de carte pour un problème donné. L'algorithme de création de la carte est schématisé sur la figure IV.19. La figure IV.20 montre les résultats obtenus pour la carte de Kohonen des 1250 HRTF d'un individu la base CIPIC.

Le regroupement des HRTF donné par la carte de la figure IV.20 est maintenant analysé en étudiant les différents paramètres, ou variables, associés aux HRTF. La projection



Méthode de sélection statistique



Méthode de sélection géométrique

FIG. IV.18 – Les deux méthodes de sélection des HRTF représentatives. Figure supérieure : méthode statistique (les représentants sont les centres des clusters), Figure inférieure : méthode géométrique (les représentants sont uniformément répartis sur la sphère).

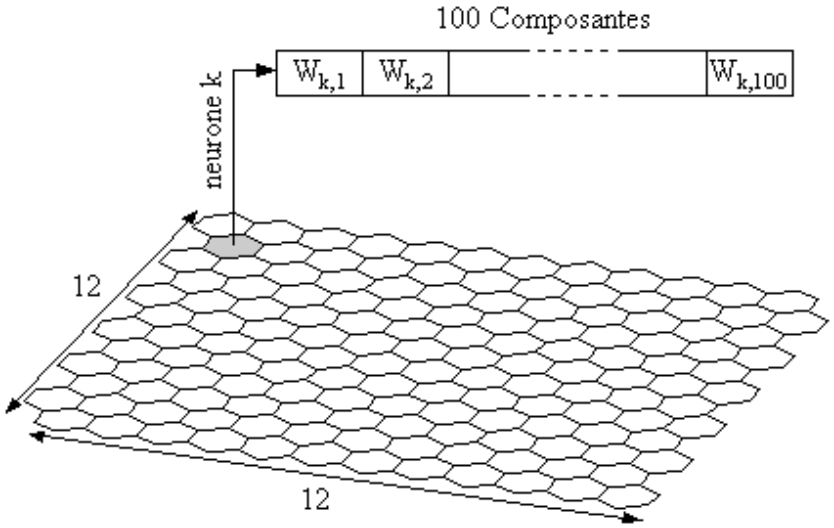


FIG. IV.19 – Topologie de la carte de Kohonen 12*12.

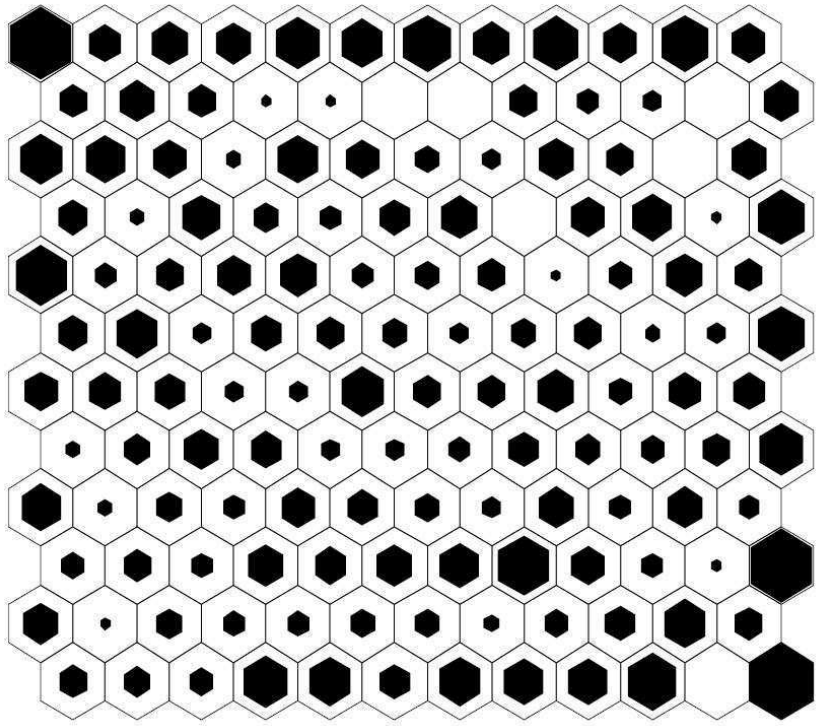


FIG. IV.20 – Carte de Kohonen des 1250 HRTF d'un individu la base CIPIC. La taille des losanges noirs est proportionnelle au nombre de HRTF contenu dans le neurone.

de variables sur la carte permet d'avoir une vue d'ensemble de la répartition des HRTF en fonction des variables projetées. Ainsi, la figure IV.21 montre l'évolution géographique de la moyenne dans chaque neurone des azimuts des HRTF (figure de droite) et de la moyenne des élévations (figure de gauche).

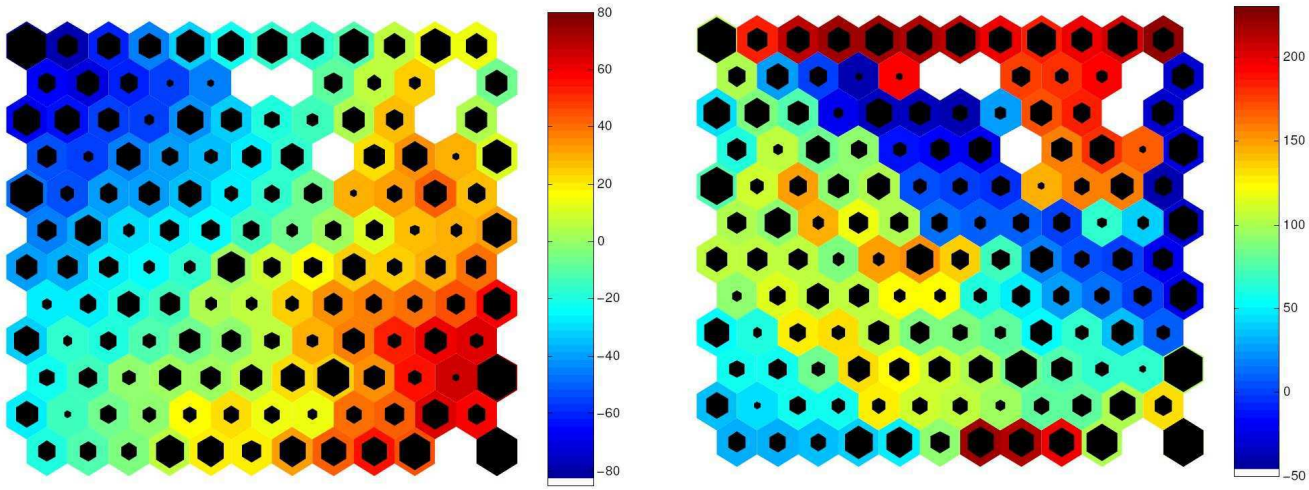


FIG. IV.21 – Carte de Kohonen des 1250 HRTF d'un individu de la base CIPIC. Figure de droite : moyenne des azimuts des HRTF de chaque neurone, Figure de gauche : moyenne des élévations des HRTF de chaque neurone.

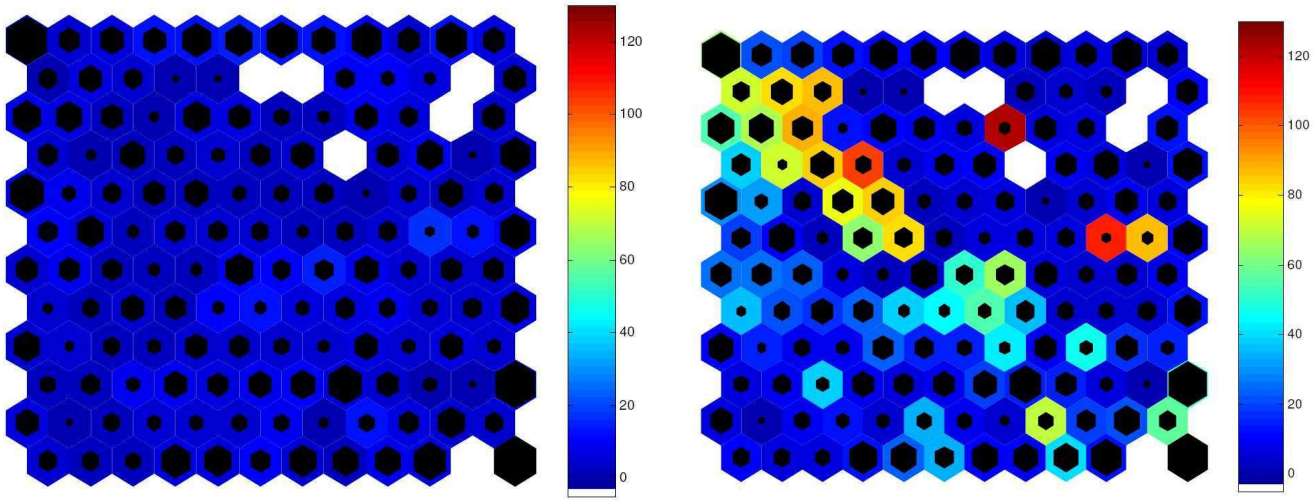


FIG. IV.22 – Carte de Kohonen des 1250 HRTF d'un individu de la base CIPIC. Figure de droite : écart-type des azimuts des HRTF de chaque neurone, Figure de gauche : écart-type des élévations des HRTF de chaque neurone.

La figure des azimuts moyens (cf. fig. IV.21 droite) fait apparaître un bonne homogénéité du regroupement en fonction de l'azimut : les couleurs transitent graduellement du clair au sombre quand la carte est parcourue du sud-ouest au nord-est. Cela signifie que l'azimut est un facteur pertinent pour le regroupement statistique des HRTF sur la base de leur proximité fréquentielle. La projection de l'écart-type des azimuts sur la carte, figure IV.22, indique aussi une bonne homogénéité intra-groupe : les HRTF proches spatialement sont regroupées autour du même prototype. L'écart-type reste inférieur à 10°. Par contre la projection de la moyenne de l'élévation (cf. fig. IV.21 gauche), fait

apparaître une forte hétérogénéité inter-groupe. Des élévations spatialement distantes sont regroupées dans des neurones voisins : les couleurs de proche en proche varient brutalement. La projection de l'écart-type (cf. fig. IV.22 gauche), confirme une forte hétérogénéité intra-groupe pour l'élévation. Ainsi certains neurones regroupent des HRTF appartenant à des hémisphères différents : de fortes valeurs de l'écart-type (niveau représenté en rouge) sont observées. Ce problème est bien illustré sur la figure IV.25 qui représente 13 clusters construits à partir des 144 neurones de la carte de Kohonen.

Etant donné que la carte de Kohonen présente de fortes hétérogénéités, les données sont séparées selon leur appartenance à un hémisphère. Deux cartes de Kohonen sont alors construites : une pour les données relatives aux HRTF de l'hémisphère avant et une pour les données de l'hémisphère arrière. La topologie reste identique à la première carte. La projection des moyennes et des variances des données d'azimut donne des résultats similaires à ceux présentés en figures IV.21 droite et IV.22 droite. Afin d'illustrer l'amélioration en terme d'homogénéité des nouvelles cartes, seules les projections des moyennes de l'élévation sont indiquées sur la figure IV.23, figure de droite pour l'hémisphère avant et figure de gauche pour l'hémisphère arrière, ainsi que les projections des écart-types sur la figure IV.24. La séparation en deux hémisphères permet donc de réduire les hétérogénéités intra et inter-groupe : sur la carte des moyennes, les couleurs évoluent graduellement et l'écart-type est largement réduit par rapport à la carte globale (cf. fig. IV.22 droite). Cette configuration de carte sera conservée dans la suite car la topologie est ici adaptée aux variations spatiales des HRTF.

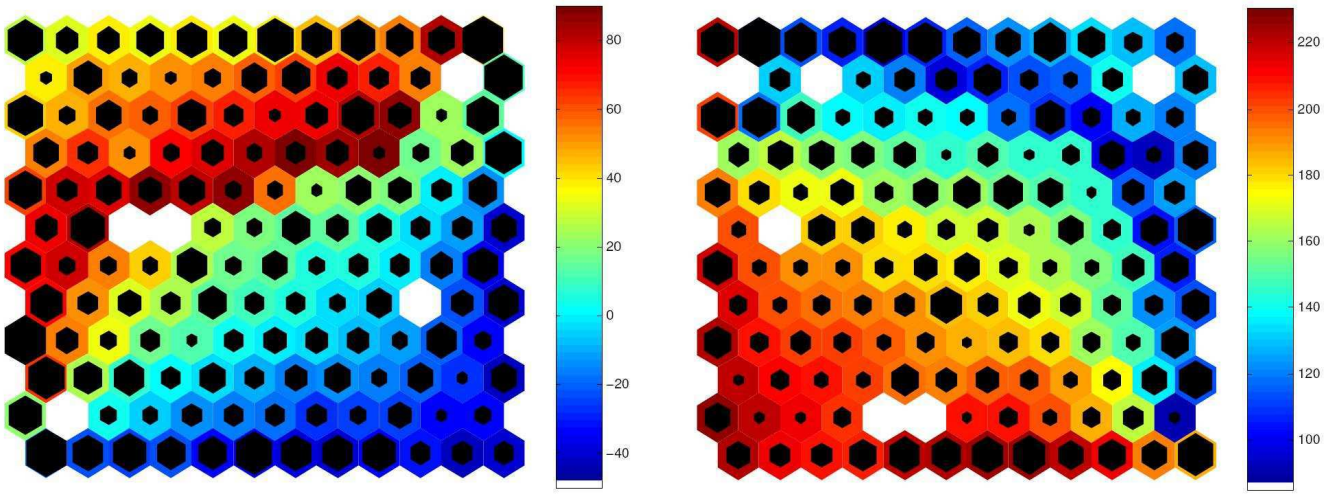


Fig. IV.23 – Carte de Kohonen des 1250 HRTF d'un individu de la base CIPIC. Séparation en deux cartes de Kohonen : projection de l'élévation moyenne des HRTF dans chaque neurone. Figure de droite : hémisphère avant, Figure de gauche, hémisphère arrière.

4.3.3 Regroupement des neurones de la carte de Kohonen par CHA

Une CHA (cf. § 4.1.3) est réalisée sur les prototypes pour déterminer des groupes de prototypes. Cette opération permet de réduire encore les dimensions du problème.

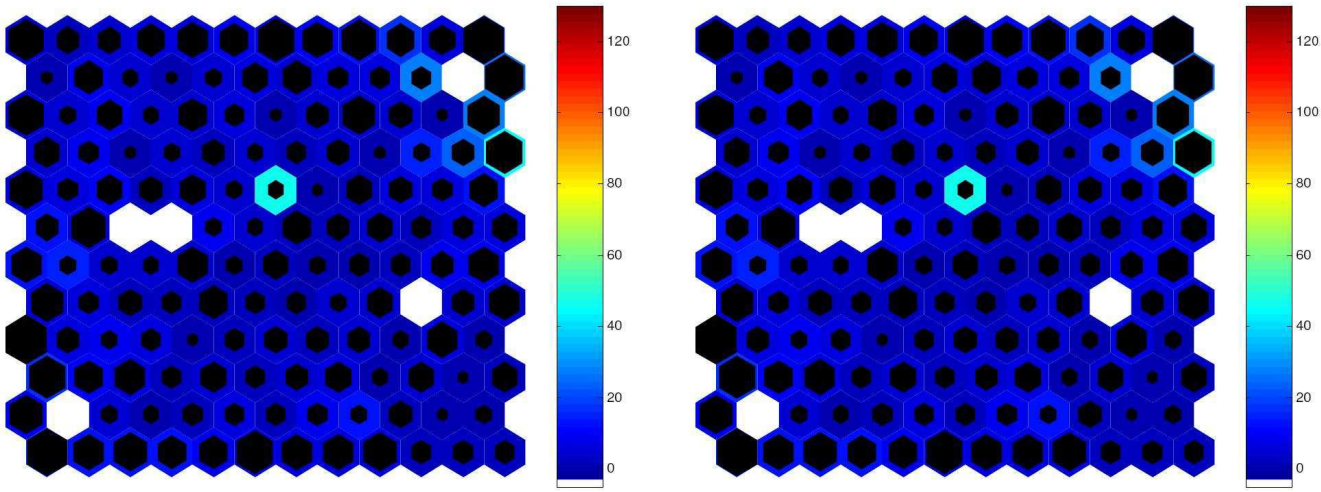


FIG. IV.24 – Carte de Kohonen des 1250 HRTF d'un individu de la base CIPIC. Séparation en deux cartes de Kohonen : projection de l'écart-type en élévation des HRTF dans chaque neurone. Figure de droite : hémisphère avant, Figure de gauche, hémisphère arrière.

Le choix d'un critère de Ward égal à 0.5 permet de regrouper les neurones en évitant des variations intra-classe élevées. Les 144 neurones de chaque carte sont alors regroupés en 13 clusters (26 pour la carte entière). Le fait que chaque carte se regroupe en 13 classes est une coïncidence car les données sont indépendantes. Les figures IV.26 et IV.27 présentent les clusters obtenus par CHA pour les deux hémisphères et la figure IV.25 indique les clusters de la carte de Kohonen entière. La comparaison des clusters obtenus avec et sans séparation de la carte permet de mieux appréhender l'intérêt d'un telle séparation. Globalement, il est observé sur les figures IV.26 et IV.27 que les clusters varient plus en azimut qu'en élévation : pour passer de -80° à 80° en azimut 6 à 7 clusters sont traversés alors que pour passer de -50° à 90° en élévation seulement 4 clusters sont parcourus. La séparation de la carte de Kohonen évite l'apparition de clusters présents dans les deux hémisphères. Par exemple, le cluster représenté par la couleur bleue foncée sur la figure IV.25, est présent sur l'hémisphère avant et sur l'hémisphère arrière. Une telle inhomogénéité montre que la topologie de la carte ne traduit pas correctement les variations spatiales des HRTF. L'étape de CHA fait aussi apparaître un regroupement intéressant des HRTF. Certains clusters regroupent les HRTF sur des plans d'azimuts constants. Ces plans coïncident avec des cônes de confusion liées aux indices interauraux tels que l'ITD et l'ILD du modèle de tête sphérique (cf. § I.2.3.2). Une des hypothèses avancées dans la littérature pour comprendre les mécanismes de discrimination auditive de sources sonores localisées sur les cônes de confusion, est que ces HRTF ne partagent pas les mêmes indices spectraux. Les résultats présentés ici ne représentent que les HRTF de l'oreille droite, mais il peut être supposé que le même traitement sur les HRTF de l'oreille gauche donnerait globalement des résultats similaires. L'observation de la figure IV.25 permet donc de nuancer cette hypothèse : des HRTF réparties sur un même cône de confusion partagent les mêmes caractéristiques spectrales moyennes. Ceci signifierait que le système auditif se base sur des différences spectrales interaurales fines telles que

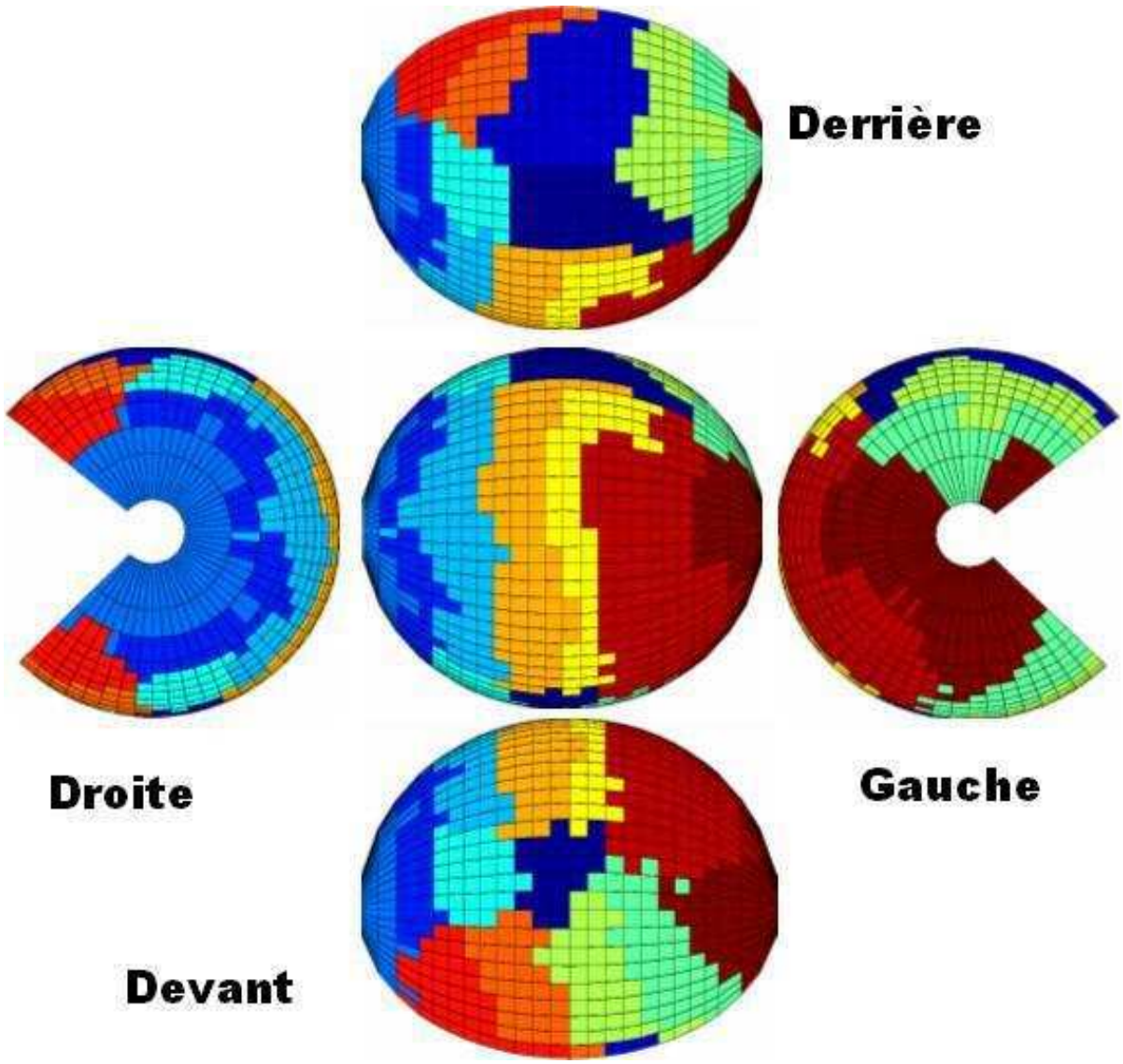


FIG. IV.25 – Projection en 3 dimensions des clusters obtenus par CHA sur la carte de Kohonen de l'étude 4. Une même couleur indique des HRTF appartenant au même cluster, c'est-à-dire partageant les mêmes caractéristiques spectrales moyennes. De haut en bas et de gauche à droite : Vue de derrière, Vue de doite, Vue de dessus, Vue de gauche et Vue de face.

celles introduites par des petits mouvements de tête.

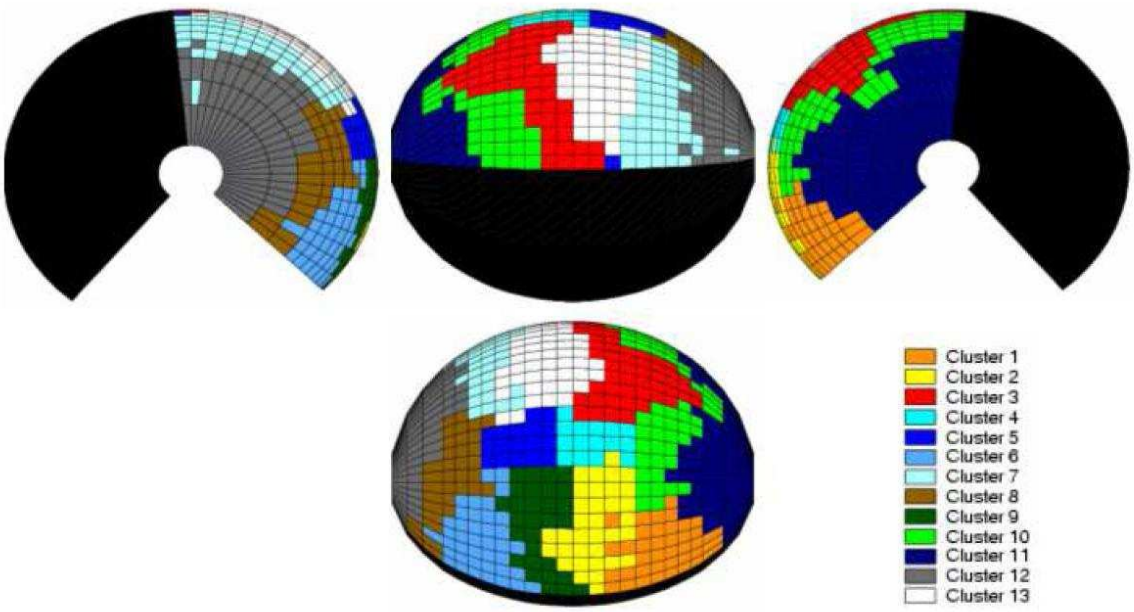


FIG. IV.26 – Cluster obtenus par CHA sur les prototypes de l'hémisphère avant.

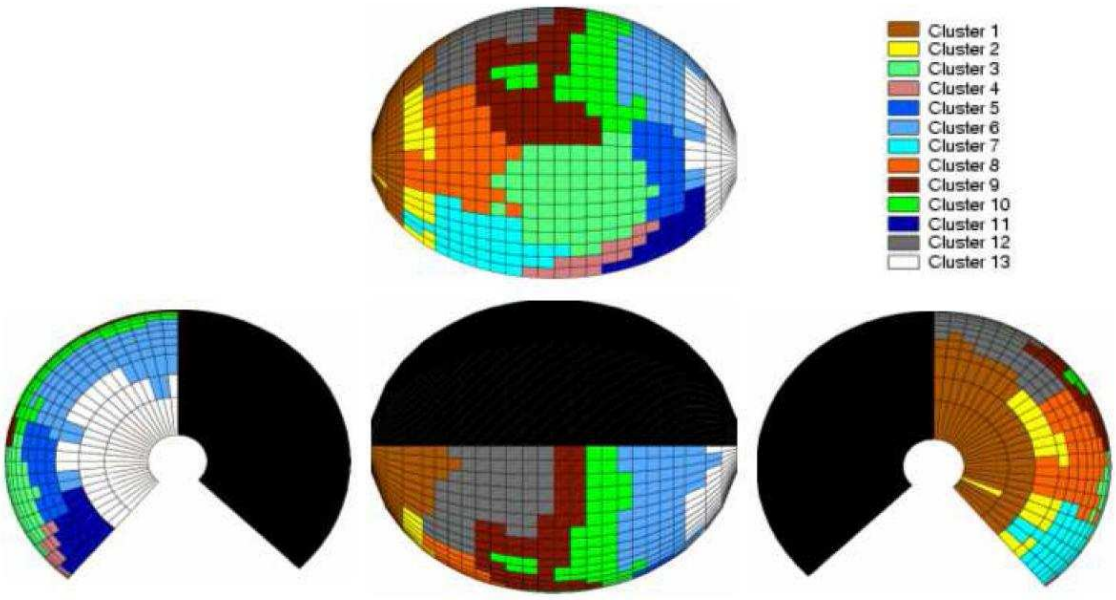


FIG. IV.27 – Cluster obtenus par CHA sur les prototypes de l'hémisphère arrière.

4.3.4 Sélection des représentants

L'étape de CHA permet donc d'obtenir 26 clusters de HRTF. Les HRTF représentatives sont sélectionnées avec l'aide des clusters. Elles sont les *parangons* des clusters : elles minimisent la somme des distances avec les autres HRTF du même cluster. La distance considérée est ici la distance euclidienne sur chaque cluster C_q :

$$R_q = \min_{n \in C_q} \Big(\sum_{j=1}^{N_q} \sum_{i=1}^{100} (V_n(f_i) - V_j(f_i))^2 \Big) \tag{IV.38}$$

où R_q désigne la HRTF représentative, V_i une HRTF du cluster C_q et N_q le nombre de HRTF dans le cluster. La position des représentants sur la sphère des mesures est indiquée en figure IV.28.

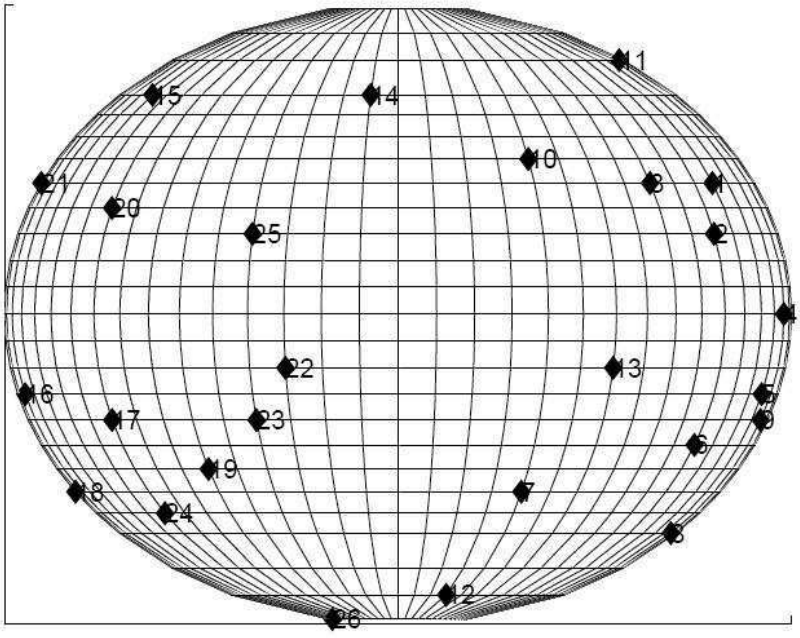


Fig. IV.28 – Position des HRTF représentatives sur la sphère des mesures. Vue de dessus.

4.3.5 Méthode de regroupement géométrique

M représentants uniformément répartis sur la sphère des mesures sont élus selon la méthode suivante :

- conversion en coordonnées cartésiennes des positions des HRTF
- regroupement des données spatiales des HRTF en k-clusters, (k=M), par l'algorithme des k-moyennes
- élection de k positions de HRTF représentatives (cf. équation IV.38).

L'échantillonnage spatial de la base CIPIC n'est pas uniforme. Grâce à l'algorithme des k-moyennes sur les coordonnées spatiales des HRTF, la partition obtenue des posi-

Tab. IV.2 – Erreur de quantification pour un individu.

Numéro de l'étude	Nombre de cluster obtenus	Erreur moyenne	Erreur minimale	Erreur médiane	Erreur maximale
1	26	3.40	1.68	3.43	5.17
2	26	3.36	1.63	3.38	5.15
3	27	3.41	1.63	3.42	5.28
4	28	3.36	1.69	3.35	5.23

tions rend maximum la distance entre les représentants ce qui assure une bonne uniformité de l'échantillonnage.

4.3.6 Influence des vecteurs d'entrée pour la sélection des HRTF représentatives

L'étude de l'influence des représentations des vecteurs d'entrée (cf. tableau IV.1) ainsi que la comparaison des méthodes de sélection des HRTF représentatives, statistique ou géométrique, se fait grâce au calcul de l'*erreur de quantification*. C'est un calcul de distance entre la HRTF représentative $\widehat{H}_{\lambda,\theta,\phi}$ et les autres HRTF au sein d'un même cluster $H_{\lambda,\theta,\phi}$. Cela permet de mesurer l'erreur introduite si la HRTF représentative est utilisée au lieu de la vraie HRTF. L'erreur de quantification est donnée par l'équation IV.39.

$$\mathbf{E}_{\lambda,\theta,\phi} = \frac{1}{100}\sum_{n=1}^{100}\left|\widehat{H}_{\lambda,\theta,\phi}(f_n) - H_{\lambda,\theta,\phi}(f_n)\right| \tag{IV.39}$$

$E_{\lambda,\theta,\phi}$ est ensuite moyennée sur toutes les positions, pour l'étude sur un seul individu et sur tous les individus pour l'étude sur tous les individus. Pour chaque étude, des couples (θ, ϕ) décrivant les positions des HRTF représentatives sont obtenues à partir de la méthode décrite précédemment dont les principales étapes sont rappelées ici :

- Séparation des données en deux hémisphères
- Création de deux cartes de Kohonen
- CHA utilisant un critère de Ward de 0.5
- Election des HRTF représentatives

Le tableau IV.2 présente la valeur moyenne pour chaque étude calculée sur l'ensemble de test d'un seul individu. Les différentes études montrent des résultats très similaires. L'étude 2, c'est-à-dire pour des vecteurs d'entrées dont les modules sont égalisés en champ libre et pondérés par des coefficients ERB (cf. § 3.3.2), est retenue car elle offre un nombre minimal de représentants et une erreur de quantification minimale.

4.3.7 Evaluation de la méthode statistique de sélection des HRTF représentatives

Erreur de quantification pour un individu La figure IV.29 montre l'évolution de l'erreur de quantification moyenne en fonction du nombre de représentants pour la méthode statistique (SOMs) et pour la méthode géométrique (Uniform). Comme attendu, l'er-

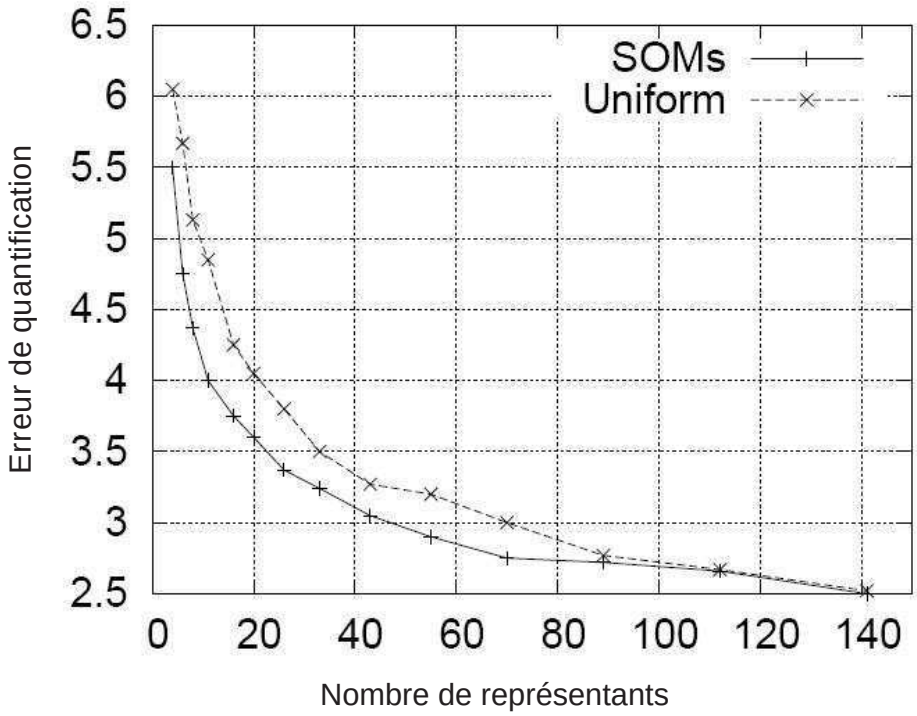


FIG. IV.29 – Erreur de quantification moyenne pour un individu en fonction du nombre de représentants pour la méthode statistique (SOMs) et pour la méthode géométrique (Uniform).

reur de quantification diminue avec le nombre de représentants. Pour un individu, la méthode statistique atteint la valeur de $E_{quant} = 3$ pour 45 représentants alors qu'il faut 70 représentants uniformément répartis pour atteindre ce niveau d'erreur. Globalement la méthode statistique apporte de meilleurs résultats. Cependant les erreurs entre les deux méthodes restent très proches bien que les deux méthodes soient très éloignées en terme d'implémentation : la méthode géométrique n'a besoin d'aucune mesure sur l'individu, tandis que la méthode statistique requiert un nombre important de mesures pour constituer un ensemble suffisant à l'apprentissage de la carte de Kohonen.

Erreur de quantification pour tous les individus L'évaluation de la méthode de sélection des HRTF représentatives est maintenant conduite sur tous les individus de l'ensemble de test (11 individus). L'erreur de quantification est calculée avec les positions des HRTF représentatives déterminées sur un individu et appliquées à tous les autres. La figure IV.31 montre l'évolution de l'erreur de quantification en fonction du nombre de clusters. Les tendances sont inversées par rapport à la figure IV.29 : la méthode géométrique présente une erreur globalement plus faible. Le niveau de $E_{quant} = 3,5$ pour la méthode statistique est cette fois-ci atteint pour 90 représentants.

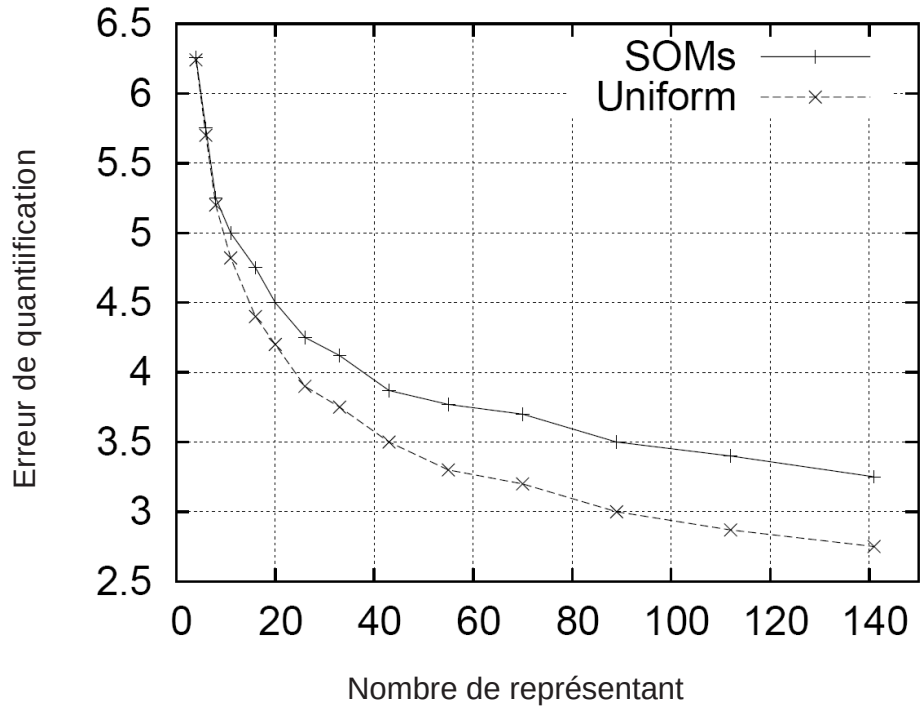


FIG. IV.30 – Erreur de quantification moyenne pour tous les individus de la base en fonction du nombre de représentants pour la méthode statistique (SOMs) pour un individu et pour la méthode géométrique (Uniform).

4.4 Discussion

Les erreurs de quantification entre les deux méthodes sont reportées dans le tableau IV.3. L'erreur de quantification est plus faible pour la méthode de sélection statistique dans le cas ou les représentants sont déterminés sur un individu et appliqués à ce même individu. Par contre, la méthode géométrique donne une erreur de quantification plus faible quand les positions des HRTF représentantes déterminées sur un individu sont appliquées à tous les individus. Le manque de généralisation de la méthode statistique amène plusieurs questions. D'abord, le choix de l'individu sur lequel est appliquée la méthode n'est peut-être pas le *bon*. Un *bon* individu serait alors celui dont les mesures sont une bonne représentation de la moyenne des HRTF. A ce titre, les HRTF d'une tête artificielle pourraient être une solution. Cependant rien ne garantit la pertinence de l'utilisation d'une telle modélisation de la morphologie d'un individu. En effet, les différences inter-individuelles des HRTF sont surtout situées en hautes fréquences, zone fréquentielle gouvernée par la diffraction de l'onde sonore sur le pavillon des sujets. Or la définition d'une forme moyenne d'un pavillon artificiel semble extrêmement délicate vue la complexité et la variance des formes de vrais pavillons. Pour tenter de répondre à la question du *bon* individu, il faudrait mener la même étude comparative sur tous les individus de la base de données. Cette étude n'a pas été menée pour des raisons de temps. Ensuite, la différence entre les deux méthodes sur tous les individus peut être interprétée en termes d'individualisation de la synthèse binaurale. La méthode statistique apporte des informations individuelles du sujet sur lequel elle est conduite et ces informations ne

conviennent pas aux autres individus. Les positions des HRTF représentatives doivent donc être individualisées.

Tab. IV.3 – Comparaison des erreurs de quantification pour les deux méthodes utilisées dans le cas de 26 représentants.

Nombre d'individus	HRTF représentatives	Erreur moyenne	Erreur minimale	Erreur médiane	Erreur maximale
1	SOM	3.40	1.68	3.43	5.17
	Uniform	3.79	1.66	3.72	6.95
Ensemble de test	SOM	4.31	1.95	4.21	6.98
	Uniform	3.93	1.81	3.86	7.06

5 MODÉLISATION DE HRTF PAR RÉSEAU DE NEURONES MLP

Une fois les HRTF représentatives sélectionnées, elle sont présentées à un MLP (Multi Layer Perceptron). Le MLP utilisé est composé de trois couches : une couche d'entrée, une couche cachée composée de 50 neurones et 100 neurones dans la couche de sortie. La fonction d'activation est la fonction tangente hyperbolique. Le réseau est entraîné par la méthode de rétropropagation de l'erreur quadratique

5.1 Ensembles d'apprentissage statistique

L'étude décrite ici est de nature exploratoire. Il est alors approprié de diviser la base de données en trois ensembles :

1. un ensemble d'apprentissage servant à ajuster les paramètres du modèle,
2. un ensemble de validation qui permet de contrôler la généralisation lors de l'apprentissage,
3. un ensemble de test pour mesurer les capacités de modélisation sur des données non apprises.

Ces ensembles représentent respectivement (50 %), (25 %) et (25 %) du nombre de vecteurs total de la base de donnée utilisée pour l'étude.

Concernant l'étude sur un individu, ces trois ensembles correspondent à une partition des 1250 modules des HRTF de l'oreille droite d'un seul individu. En indexant les HRTF de 1 à 1250, par plan azimuth constant et par élévation croissante, l'ensemble d'apprentissage est composée des HRTF {1, 5, 9,..} et {2, 6, 10,..}, l'ensemble de validation des HRTF {3, 7, 11,..} et l'ensemble de validation des HRTF {4, 8, 12,..}.

Pour l'étude sur toute la base de données CIPIC, les trois ensembles sont des groupes d'individus. L'indexation concerne ici le numéro du sujet de la base et l'ensemble d'apprentissage, l'ensemble de validation, l'ensemble de validation regroupent respectivement les individus 1 à 22, 23 à 33 et 34 à 44.

Les données présentées pour l'apprentissage du MLP sont centrées et réduites[4]. Cette opération évite au MLP d'apprendre la moyenne des vecteurs, fonction qu'il apprend en premier. La structure du réseau est déterminé par la méthode de *cross-validation* (cf. § 2.1.3).

5.2 Apprentissage sur un seul individu

Le MLP est utilisé pour prédire les HRTF d'un individu à partir des 26 représentants élus pour le même individu. Les vecteurs d'entrée sont composés d'une HRTF représentative statistique $HRTF_{ri}$ et des coordonnées (θ_0, ϕ_0) de la HRTF à prédire. La HRTF représentative du vecteur d'entrée est le représentant du cluster auquel appartient la HRTF à prédire. Le vecteur de sortie est $HRTF_{\lambda,\theta_0,\phi_0}$. Le neurone apprend la fonction qui dépend de l'individu λ, f_λ de $\mathbb{R}_{102} \longrightarrow \mathbb{R}_{100}$ suivante :

$$HRTF_{\lambda,\theta_0,\phi_0} = f_\lambda(HRTF_{ri}, \theta_0, \phi_0) \tag{IV.40}$$

Afin d'évaluer les performances du MLP créé, un apprentissage en régression linéaire utilisant un MLP avec un seul neurone dans la couche cachée et qui contient une fonction d'activation linéaire est conduit avec les mêmes données. Les résultats en terme d'erreur de modélisation sur l'ensemble de test sont reportés dans le tableau IV.4.

Tab. IV.4 – Comparaison des erreurs de modélisation pour une régression linéaire et pour le MLP.

Type de régression	Erreur moyenne	Erreur minimale	Erreur médiane	Erreur maximale
Linéaire	3.59	2.29	3.60	5.04
Non linéaire	1.88	0.92	1.71	3.24

L'erreur de modélisation du MLP non-linéaire est très faible. L'erreur est plus importante pour les positions contra-latérales et pour les hautes fréquences. Une réduction considérable de l'erreur de modélisation est apportée par l'apprentissage non-linéaire (47%) par rapport à une régression linéaire. De plus, il y a bien apprentissage du RNA car l'erreur de modélisation est plus faible que l'erreur de quantification. La réduction de l'erreur grâce à l'apprentissage est de 45%. Qui plus est, une méthode de sélection de variable décrite dans [Lemaire et Clérot (2004)] a mis en évidence que le vecteur d'entrée peut se résumer aux simples coordonnées de la HRTF désirée. Les résultats montrent alors une erreur de modélisation de 1.81.

5.3 Apprentissage sur tous les individus

Le MLP est utilisé pour prédire les HRTF de n'importe quel individu, de l'ensemble de test, à partir de ses propres HRTF représentatives élues par la méthode géométrique (cf. § 4.3.5). La fonction que le réseau apprend est alors la suivante :

[4]La transformation suivante est opérée : $x \rightarrow \frac{x-m}{\sigma}$ où m est la moyenne de x et σ son écart-type

$$HRTF_{\lambda,\theta_0,\phi_0} = f(HRTF_{ri_\lambda}, \theta_0, \phi_0) \tag{IV.41}$$

où $HRTF_{ri_\lambda}$ est le plus proche représentant de $HRTF_\lambda(\theta_0, \phi_0)$ la HRTF à prédire. L'erreur de quantification ainsi que l'erreur de modélisation sont reportées en figure IV.31 en fonction du nombre de HRTF représentatives. Etant donné que l'étape de sélection des représentants n'est effectuée ici que sur un individu, le nombre de 26 réprésentants n'est plus statistiquement correct et il est alors intéressant de comparer l'évolution des erreurs en fonction du nombre de représentants utilisés.

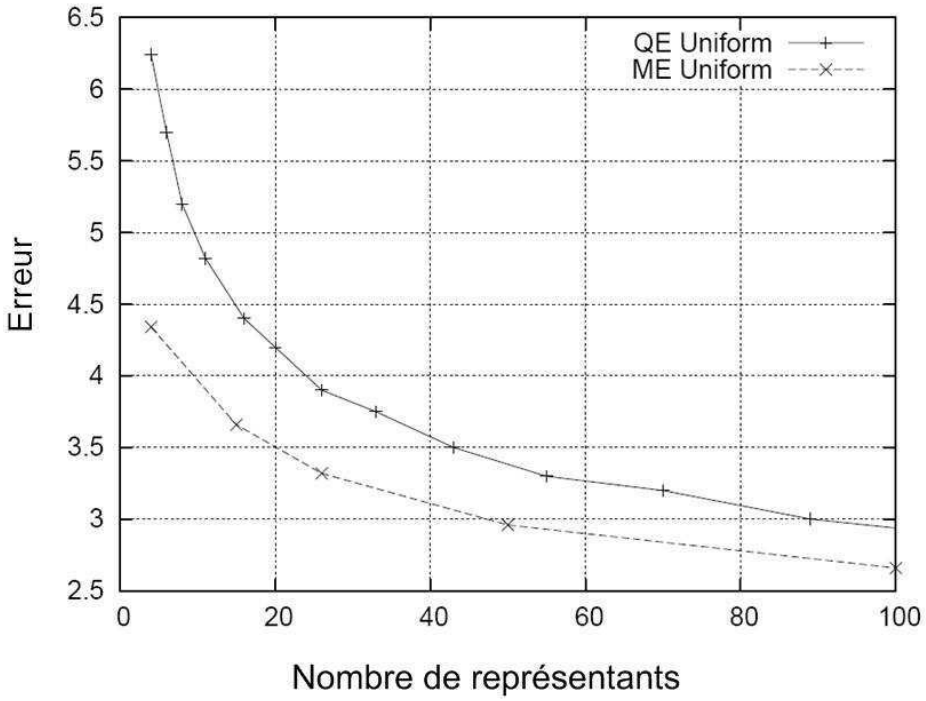


Fig. IV.31 – Erreur de quantification (QE) et erreur de modélisation (ME) pour les individus de l'ensemble de test en fonction du nombre de représentants géométriques.

L'erreur de modélisation est globalement inférieure à l'erreur de quantification et le niveau de $E_{quant} = 3$ est atteint avec 50 représentants pour le MLP et 90 pour la quantification. Ce résultat prouve que le réseau généralise et s'adapte à la présentation d'un nouveau sujet. La fonction apprise contient les informations pour traduire les dépendances entre des HRTF uniformément réparties sur la sphère des mesures et les autres HRTF. Le nombre de 50 mesures devant être réalisées est très encourageant et permet d'envisager des sessions de mesures réduites.

La figure IV.32 permet de faire la comparaison visuelle entre des HRTF prédites et des HRTF mesurées (figures du haut). Les HRTF d'un individu de l'ensemble de test sont représentées dans un plan d'azimut 0°, figures de droites, et pour un plan d'élévation 0°, figures de gauche, pour un entraînement du réseau avec 100 représentants ($E_{mod} = 2.66$) et pour 50 représentants ($E_{mod} = 2.95$). Les lignes noires correspondent aux emplacements des HRTF représentatives.

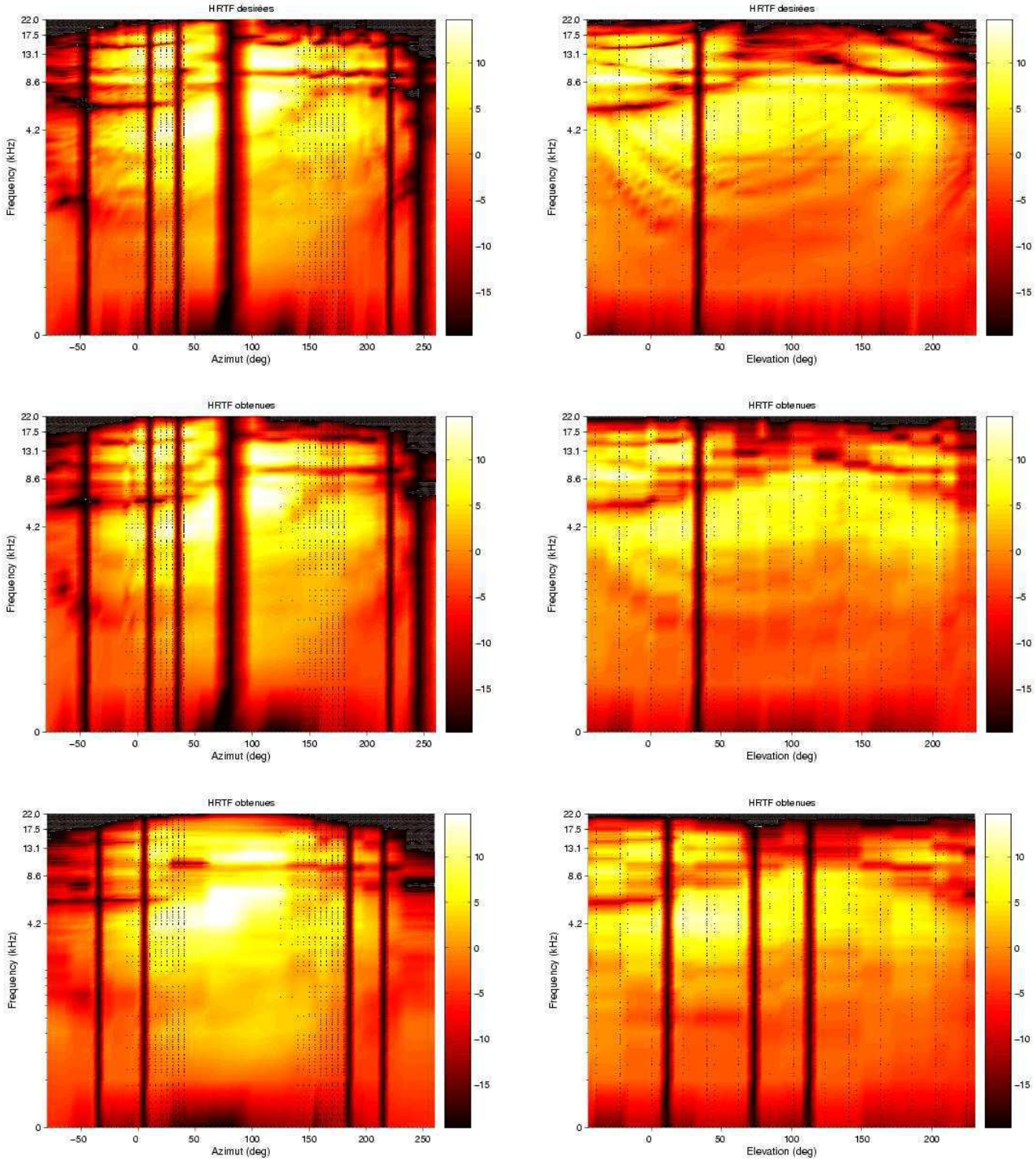


Fig. IV.32 – HRTF pour l'individu 4 de l'ensemble de test. Le niveau des HRTF est indiqué grâce à un code couleur calibré sur la barre colorée située à droite des figures. Figures de gauche : plan horizontal, Figure de droite : plan vertical. Figure du haut : mesures, Figures du milieu : modélisation avec 100 représentants, Figures du bas : modélisation avec 50 représentants. Les lignes verticales noires sont les positions des représentants non affichées.

Les figures du haut et du milieu sont globalement très similaires et la dynamique est correctement reproduite. Les figures de diffraction dues au torse et à la tête sont très bien reproduites (plan vertical $\phi \in$ [-50°;50°] et $f \in [2kHz - 4kHz]$). Le plus remarquable est la prédiction de la diffraction induite par le pavillon qui reste fidèle même en hautes fréquences ($f > 6kHz$). Cette observation indique que le réseau est capable d'individualiser à un fort niveau : les variations inter-individuelles sont plus fortes en hautes fréquences. Les figures du bas permettent de se rendre compte des dégradations introduites avec seulement 50 représentants. Les résonances et anti-résonances du pavillon sont légèrement moins nettes. La figure de diffraction du torse reste cependant bien décrite.

Des écoutes informelles avec 100 représentants indiquent que les HRTF modélisées sont très proches perceptivement des HRTF mesurées. La trajectoire décrivant le plan d'élévation 0° est particulièrement bien reproduite.

6 CONCLUSION

L'étude menée a permis de valider l'utilisation d'outils statistiques tels que les réseaux de neurones et les cartes de Kohonen pour la classification et la modélisation de HRTF. Premièrement, la sélection des représentants par les cartes de Kohonen pour un individu a montré de bons résultats comparés à la méthode géométrique. Les clusters créés par la CHA apportent des informations sur des zones de similitudes des HRTF notamment pour les zones mises en jeu dans le problème perceptif des confusions avant / arrière. Le fait que les représentants statistiquement élus n'offrent pas une bonne généralisation, indique qu'il faut appliquer cette technique à toute la base de données pour sélectionner des HRTF représentatives *universelles*. Il serait alors intéressant de comparer les positions représentatives de toute la base avec les directions indiquées par la méthode de type ACI ou *subset selection*. Deuxièmement, le réseau est capable de traduire des dépendances non-linéaires de haut niveau entre les données et reproduit des détails très fins des HRTF individuelles. Les écoutes informelles sont prometteuses. Des tests psychoacoustiques doivent être réalisés pour évaluer perceptivement les performances du MLP, notamment en fonction du nombre de représentants.

La capacité de généralisation de la fonction apprise peut être mesurée en utilisant la fonction sur d'autres mesures issues d'une autre base de données de HRTF. Des performances équivalentes devraient être obtenues en supposant que les différences entre les systèmes de mesure sont négligeables. Trouver des différences importantes reviendrait à mettre indirectement en évidence une forte contribution du dispositif aux mesures. Une capacité de généralisation plus importante sera obtenue avec des bases de données contenant plus de sujets. Une solution, qui évite une campagne de mesures de HRTF, serait de réunir des bases existantes. Cependant, l'union des bases pose un problème pour la création d'une base unique d'apprentissage du fait notamment des répartitions spatiales différentes des points de mesures.

L'évaluation de l'apport des techniques neuronales peut se faire en comparaison des performances des méthodes classiques. Dans les méthodes présentées en début de chapitre, seules les techniques ACI et *subset selection* pourraient être comparées. En effet, l'entrée du réseau de neurone est composée de HRTF mesurées. Les techniques ACI et *subset selection* peuvent aboutir à des fonctions spatiales très directives qui engendrent

l'espace de représentation des HRTF. Ces directions ainsi isolées indiquent alors des positions de mesures privilégiées. En revanche, la technique ACP utilise des composantes principales qui sont des formes *moyennes* génératrices du sous-espace de représentation. Une modélisation des HRTF d'un individu avec l'ACP nécessite la définition d'un protocole de mesure pour estimer les poids à associer aux composantes principales.

BIBLIOGRAPHIE

V. Algazi, R. Duda, et D. Thompson. The cipic hrtf database. 2001.

APSTAT Technologies. Introduction aux réseaux de neurones. www.apstat.com, 2002.

J. Blauert. *Spatial Hearing*. The mit press edition, 1983.

A. W. Bronkhorst. Localization of real and virtual sound sources. *J. Acoust. Soc. of Am.*, 98(5) :2542–2553, 1995.

J. Chen, B. D. Van Veen, et K. E. Hecox. A spatial feature extraction and regularization model for the head related transfer function. *J. Acoust. Soc. of Am.*, 97(1) :439–452, January 1995.

V. Choqueuse. Utilisation d'outils statistiques pour l'individualisation des hrtf. Rapport de stage ingénieur, Université de Technologie de Troie, 2004.

Y. Le Cun. *Modèles connexionnistes de l'apprentissage*. Thèse de Doctorat, Université Paris 6, 1987.

E. Dudouet et J. Martin. Analyses multidimensionnelles des hrtf pour la spatialisation de sources sonores. Technical report, CSTB, 1998.

M. Emerit et J. Martin. Head-related transfer functions and high-order statistics. 1995.

C.S. Fahn et Y.C. Lo. On the clustering of head-related transfer functions used for 3-d sound localisation. *Journal of Information Science en Engineering*, 19 :141–157, 2003.

D. W. Grantham, J. A. Whillhite, K. D. Frampton, et D. H. Ashmead. Reduced order modeling of head related impulse responses for virtual acoustics displays. *J. Acoust. Soc. of Am.*, 117(5) :3116–3152, May 2005.

R. M. Gray. Vector quantization. *IEEE ASSP Magzine*, pages 4–29, April 1984.

W. M. Hartmann. *Signal, sound and sensation*. Springer, 1998.

J. Huopaniemi et J. O. Smith. Spectral and time-domain preprocessing and the choice of modelling error criteria for binaural digital filters. avril 1999.

R. L. Jenison. A spherical basis function neural network for pole-zero modeling of head-related transfer functions. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, October 1995.

R. L. Jenison et K. Fissell. A spherical basis function neural network for pole-zero modeling auditory space. *Neural computation*, 8 :115–128, 1996.

C. Jin, P. Leong, J. Leung, A. Corderoy, et S. Carlile. Enabling individualized virtual auditory space using morphological measurements. pages 235–238. Proceedings of the First IEEE Pacific-Rim Conference on Multimedia (2000 International Symposium on Multimedia Information Processing), 2000a.

C. Jin, M. Schenkel, et S. Carlile. Neural system identification model of human sound localization. *J. Acoust. Soc. of Am.*, 108(3) :1215–1235, September 2000b.

C. Jin, T. Tan, J. Leung, A. van Schaik, et S. Carlile. Human interface support for information technology for mars research missions. pages 1261–1266, 2002. http ://www.marssociety.org.au/amec2002/.

Y. Kahana. *Numerical modelling of the head related transfer function.* Thèse de Doctorat, University of Southampton, December 2000.

D. J. Kistler et F. L. Wightman. A model of head-related transfert functions based on principal components analysis and minimum-phase reconstruction. *J. Acoust. Soc. of Am.*, 91(3) :1637–1647, March 1992.

T. Kohonen. *Self-organizing maps*, volume 30 of *Springer series in information sciences.* Springer, 1995.

A. Kulkarni et H. S. Colburn. Efficient finite-impluse-response filter of the head related transfer function. *J. Acoust. Soc. of Am.*, 5 :3278, may 1995.

E. H. A. Langendjik et A. W. Bronkhorst. Fidelity of three-dimensional-sound reproduction using a virtual auditory display. *J. Acoust. Soc. of Am.*, 107(1) :528–537, 2000.

V. Larcher. *Techniques de spatialisation des sons pour la réalité virtuelle.* Thèse de Doctorat, Université Paris VI, Mai 2001.

V. Lemaire. *Une nouvelle fonction de coût régularisante dans les réseaux de neurones artificles : application à l'estimation des temps de blocage dans un noeuds ATM.* Thèse de Doctorat, Université Paris VI, Septembre 1999.

V. Lemaire. Les réseaux de neurones. Cours présenté à l'Ecole National de Statistique et d'Analyse de l'Information, Février 2001.

V. Lemaire et F. Clérot. An input variable importance definition based on empirical data probability and ist use in variable selection. pages 1375–1380, Budapest 2004.

V. Lemaire, F. Clérot, S. Busson, R. Nicol, et V. Choqueuse. Individualized hrtfs from few measurements : a statistical learning approach. August 2005.

M. Marin. *Etude de la localisation en restitution du son pour la téléconférence de haute qualité.* Thèse de Doctorat, Université du Maine, 1996.

W. Martens. Principal componant analysis and resynthesis of spectral cures to perceived direction. pages 274–281. ICMC, 1987.

M. Matusmoto, S. Yamanaka, et M. Tohyama. Effect of arrival correction on the accuracy of binaural impulse response interpolation. *J. Audio Engin. Soc.*, 52(1/2) :56–61, January/February 2004.

S. McAdams. *Audition : physiologie, perception et cognition.* Presses Universitaires de France, 1994. pp 283–344.

J. Middlebrooks. Individual differences in external-ear transfer function reduced by scaling in frequency. *J. Acoust. Soc. of Am.*, 106(3) :1480–1492, 1999.

J. Middlebrooks et D. Green. Observations on a principal componante analysis of head related transfer function. *J. Acoust. Soc. of Am.*, 92(1) :597–592, 1992.

A. W. Mills. On the minimum audible angle. *J. Acoust. Soc. of Am.*, 30 :237–248, April 1958.

T. Nishino, M. Ikeda, K. Takeda, et F. Itakura. Interpolating head related transfer functions. October 2000.

T. Nishino, S. Mase, S. Kajita, K. Takeda, et F. Itakura. Interpolating hrtf for auditory virtual reality. pages 1261–1266, 1996.

S. R. Oldfield et S. P. A. Parker. Acuity of sound localisation : a topography of auditory space. i. normal hearing. *Perception*, 13 :581–600, July 1984.

K. Palomäki, V. Pulkki, et M. Karjalainen. Neural network approach to analyse spatial sound. 2000.

R. D. Patterson et M. H. Allerhand. Time-domain modeling of periphal auditory processing : a modular architecture and a softxar paltform. *J. Acoust. Soc. of Am.*, 98 (6) :3435–3444, 1995.

V. Pulkki, M. Karjalainen, et J. Huopaniemi. Analysing virtual sound source attributes using a binaural model. *AES* 114$^{\text{th}}$ *Convention*, 1998.

E. Rio et O. Warusfel. Optimizations of multi-channel binaural formats based on statistical analysis. Seville, 2002. FORUM ACUSTICUM SEVILLA.

J. O. Smith et J. S. Abel. Bark and erb bilinear transforms. In *IEEE Trans. on speech and audio processing*. IEE, November 1999.

J. Vesanto, J. Himberg, E. Alhoniemi, et J. Parhankangas. Som toolbox for matlab 5. Technical report, Helsinki Uninersity of Technology, 2000.

P. Vovor. Utilisation d'outils statistiques pour l'individualisation des hrtf. Rapport de stage master 2, Université Paris IV, Master ATIAM, 2005.

Z. Wu, F. H. Y. Chan, et Lam F. K. A time domain binaural model based on spatial feature extrcation for the head related transfer function. *J. Acoust. Soc. of Am.*, 102 (4) :2211–2218, October 1997.

Z. Wu, T. Weng, W. Wang, T. F. Lao, F. H. Y. Chan, et F. K. Lam. Neural network model of binaural hearing based on spatial feature extraction of the head related transfer function. volume 20, 1998.

V

Conclusion

1 RÉSULTATS

L'individualisation de la synthèse binaurale a été abordée sur les deux principaux indices qui gouvernent la localisation auditive : la différence interaurale de temps (ITD) et les indices spectraux contenus dans les fonctions de transfert binaurales (HRTF). Cette dichotomie permet une évaluation des capacités d'individualisation de l'implémentation la plus commune de la synthèse binaurale qui utilise un retard pur associé à un filtre à phase minimale ($\{ITD \bigoplus HRTF_{min}\}$).

La première étude sur l'individualisation de l'ITD a permis, grâce à un protocole expérimental spécifique pour l'estimation psychoacoustique de l'ITD, de valider l'implémentation choisie et ce même pour les positions où les $HRTF$ sont mal décrites par un filtre à phase minimale. La comparaison entre l'ITD psychoacoustique et l'ITD donnée par les méthodes classiques de calcul a permis de dégager les méthodes les plus aptes à reproduire des variations individuelles de l'ITD. La contribution principale de ce travail se situe dans une nouvelle formulation de l'ITD issue d'une modélisation sphérique de la tête d'un auditeur. Cette formulation permet une meilleure reproduction de l'ITD pour toutes les positions et l'ajustement des paramètres de cette formule autorise un niveau supérieur d'individualisation de l'ITD. En effet, la formule FDO (Formule de Déplacement des Oreilles) permet la reproduction des variations individuelles de l'ITD sur les cônes de confusion. La deuxième étude expérimentale sur l'ITD a montré que ces variations devaient être reproduites pour une meilleure perception des sources en élévation. Les résultats de cette deuxième étude, qui a complété des données psychoacoustiques, a

permis d'établir des paramètres moyens de la formule FDO qui semblent convenir à un grand nombre de personnes.

Le travail d'individualisation sur les indices spectraux des HRTF s'est basé sur l'hypothèse communément faite qu'une synthèse binaurale optimale est obtenue avec des HRTF propres à chaque individu, c'est-à-dire acquises par la mesure. Une alternative à la mesure de HRTF, qui est une tâche délicate et qui représente le principal obstacle à l'émergence d'une synthèse binaurale pour le grand public, est l'acquisition de HRTF par le calcul BEM. Différentes modélisations de la morphologie de l'auditeur sont présentées et leur capacité à reproduire les motifs spectraux présents dans les HRTF a été évaluée. L'analyse de différentes modélisations de l'auditeur par des formes simples (sphère, ellipsoïde) permet en outre une meilleure compréhension de l'influence des différentes parties du corps sur les HRTF. Ainsi l'ajout d'une modélisation du torse apporte des indices spectraux pour la perception en élévation des sources basses fréquences. La reproduction correcte des basses fréquences est primordiale pour la perception des sons environnementaux et n'est pas assurée par la mesure de HRTF. Bien que l'utilisation de la BEM pour l'acquisition de HRTF reste encore limitée, principalement à cause des ressources informatiques requises, la méthode revêt un intérêt particulier pour la prédiction en basses fréquences des HRTF.

Enfin, une alternative à l'acquisition de HRTF par BEM et par la mesure a été proposée. Cette approche combine différentes méthodes d'apprentissage statistique pour d'une part déterminer un nombre réduit de points de mesures *optimaux* et d'autre part construire un modèle qui prédit un ensemble de HRTF à partir des points de mesures *optimaux*. Le choix des points de mesures *optimaux* est réalisé par des méthodes de classification, dont les cartes de Kohonen. Le modèle de prédiction construit avec un réseau de neurones permet la reproduction de motifs spectraux fins responsables du caractère individuel des HRTF. Ces résultats suggèrent la définition d'un protocole simplifié de mesure de HRTF : un nombre réduit de mesures est effectué et le modèle mathématique prédit les autres HRTF aux positions non mesurées.

2 PERSPECTIVES

Validation perceptive du modèle de prédiction de HRTF Des écoutes informelles ont été réalisées pour l'évaluation perceptive du modèle de prédiction. Elles indiquent que les HRTF modélisées sont très proches perceptivement des HRTF mesurées. Il reste à mener une étude subjective complète des HRTF modélisées. Il faudrait notamment établir l'influence du nombre de HRTF en entrée du modèle sur la qualité de prédiction. En appliquant ce modèle pour des HRTF d'une base de données qui n'a pas servi pour la construction du modèle, sa capacité de généralisation pourra être évaluée. L'analyse des erreurs perceptives introduites servira pour l'optimisation d'un critère d'erreur utilisé pour la construction du modèle, afin de prendre en compte à la fois les spécificités des techniques neuronales et la résolution du système auditif.

Création d'un espace multidimensionnel de représentation perceptive des HRTF Les écoutes doivent toujours avoir lieu pour évaluer perceptivement un modèle car un critère d'erreur perceptif sur la base de paramètres objectifs n'existe pas à l'heure actuelle. La projection des HRTF dans un espace de représentation perceptif, où deux HRTF proches

perceptivement, sont proches dans l'espace de projection, pourrait aider à la définition d'un tel critère. Une des difficultés rencontrées pour la construction d'un espace perceptif est l'étape de verbalisation des dimensions. Une approche psycholinguistique, notamment avec la méthode de catégorisaton prototypique, pourrait aider à dégager des descripteurs sémantiques liés aux HRTF.

Remise en cause de l'hypothèse de départ L'étude sur l'individualisation des HRTF est basée sur l'hypothèse qu'une synthèse binaurale de qualité est obtenue avec des HRTF individuelles. Un test réalisé à France Telecom R&D a mis en évidence des éléments en contradiction avec cette hypothèse. Ce test a consisté à présenter aux sujets des HRTF de différents individus et à leur demander leur HRTF *préférée*. Ce test a mis en évidence une importante non reproductibilité des résultats. De plus, certains sujets ont préféré des HRTF qui n'étaient pas les leurs. Par ailleurs, des travaux reportés dans la littérature ont montré que des HRTF modélisées pouvaient offrir de meilleures performances de localisation que les HRTF mesurées. Ainsi, Cheng montre que des sources synthétisées avec des HRTF interpolées sont mieux localisées que des sources synthétisées avec des HRTF issues de mesure [Cheng (2001)]. Qui plus est, des études ont mis en évidence un phénomène de plasticité de la localisation sonore, à long terme [Hofman et al. (1998)], mais aussi à court terme [Blum et al. (2004)]. Une démarche complémentaire entre modélisation et plasticité pourrait représenter une alternative à la recherche systématique de HRTF individuelles.

BIBLIOGRAPHIE

Blum, A., Katz, B. F. G., and Warusfel, O. (2004). Eliciting adaptation to non-individual hrtf spectral cues with multi-modal training.

Cheng, C. I. (2001). *Visualization, measurement, and interpolation of Head Related Transfer Function (HRTF's) with applications in electroacoustic music.* PhD thesis, University of Michigan.

Hofman, P. M., van Riswick, J. G. A., and van Opstal, A. J. (1998). relearning sound localisation with new ears. *Nature Neuroscience*, 1(5) :417–421.

Bibliographie

Algazi, V., Duda, R., and Thompson, D. (2001a). The cipic hrtf database.

Algazi, V. R., Avendano, C., and Duda, R. O. (2001b). Elevation localization and head-related transfer function analysis at low frequencies. *J. Acoust. Soc. of Am.*, 109(3).

Algazi, V. R., Avendano, C., and Duda, R. O. (2001c). Estimation of a spherical-head model from anthropometry. *J. Acoust. Soc. of Am.*, 49 :472–478.

Algazi, V. R., Avendano, C., and Thompson, D. (1999). Dependence of subjects and measurement position in binaural signal acquisition. *J. Audio Engin. Soc.*, 47(11) :937–947.

Algazi, V. R. and Duda, R. O. (2002). Approximating the head-related transfer function using simple geometric models of the head and torso. *J. Acoust. Soc. of Am.*, 112(5) :2053–2064.

Algazi, V. R., Duda, R. O., Morison, R. P., and Thompson, D. M. (2001d). Structural composition and decomposition of hrtfs. New Paltz, New-York. IEEE Trans. Speech and Audio Proc.

APSTAT Technologies (2002). Introduction aux réseaux de neurones. www.apstat.com.

Asano, F., Suzuki, Y., and Sone, T. (1990). Role of spectral cues in median plane localization. *J. Acoust. Soc. of Am.*, 88 :159–168.

Avendano, C., Duda, R. O., and Algazi, R. (1999). Modelling the contralateral hrtf.

Baumgarte, F. and Faller, C. (2003). Binaural cue coding-part i : psychoacoustic fundamentals and design principles. *IEEE Transactions on Speech and Audio Processing*, 11(6) :509–519.

Best, V., van Schaik, A., and Carlile, S. (2004). Separation of concurrent broadband soun sources by human listeners. *J. Acoust. Soc. of Am.*, 115(1) :324–336.

Blauert, J. (1983). *Spatial Hearing*. The mit press edition.

Blommer, M. A. and Wakefield, G. H. (1997). Pole-zero approximations for head-related transfer functions using a logarithmic error criterion. *IEEE Transactions on speech and audio processing*, 5(3) :278–287.

Blum, A., Katz, B. F. G., and Warusfel, O. (2004). Eliciting adaptation to non-individual hrtf spectral cues with multi-modal training.

Braasch, J. and Hartung, K. (2002). Localization in the presence of a distracter and reverberation in the frontal horizontal plane. i. psychoacoustical data. *Acta Acoustica united with Acoustica*, 88 :942–955.

Bronkhorst, A. W. (1995). Localization of real and virtual sound sources. *J. Acoust. Soc. of Am.*, 98(5) :2542–2553.

Brown, C. P. and Duda, R. O. (1998). A structural model for binaural sound synthethis. volume 6. IEEE Trans. Speech and Audio Proc.

Bruneau, M. (1998). *Manuel d'acoustique fondamentale*. Hermes.

Brungart, D. S. and Rabonowitz, W. M. (1998). Auditory localization of nearby sources. *J. Acoust. Soc. of Am.*, 106 :1465–1479.

Burton, A. J. (1973). The solution of helmholtz equation in exterior domains using integral equations. Technical report, National Physics Laboratory.

Busson, S., Nicol, R., and Katz, B. F. G. (2005a). Subjective investigations of the interaural time difference in the horizontal plane. *Presented at the* 118$^{\text{th}}$ *AES Convention, Barcelona, Spain*. Convention Paper 6324.

Busson, S., Nicol, R., and Warusfel, O. (2004). Influence of the ears canal location on spherical head model for the individualised interaural time difference. *Proceedings of CFA / DAGA Joint Meeting, Strasbourg, France*.

Busson, S., Nicol, R., Warusfel, O., and Gros, L. (2005b). Just noticeable difference of the interaural time difference on cones of confusion. to be summited to J. Acoust. Soc. of Am.

Chateau, N. (1996). *Localisation de sources sonores multiples dans l'hémisphère supérieur*. PhD thesis, Université de la méditerranée Aix Marseille II, laboratoire de mécanique et d'acoustique.

Chen, J., Veen, B. D. V., and Hecox, K. E. (1995). A spatial feature extraction and regularization model for the head related transfer function. *J. Acoust. Soc. of Am.*, 97(1) :439–452.

Cheng, C. I. (2001). *Visualization, measurement, and interpolation of Head Related Transfer Function (HRTF's) with applications in electroacoustic music*. PhD thesis, University of Michigan.

Cheng, C. I. and Wakefield, G. H. (2001). Introduction to head-related transfer functions (hrtfs) : representations of hrtfs in time, frequency and space. *J. Audio Engin. Soc.*, 49(4) :231–249.

Choqueuse, V. (2004). Utilisation d'outils statistiques pour l'individualisation des hrtf. Rapport de stage ingénieur, Université de Technologie de Troie.

Ciscowski, R. D. and Brebbia, C. A. (1991). *Boundary Element methods in acoustics*. Computational Mechanics Publication & Elsevier Applied Science, Southampton.

Cun, Y. L. (1987). *Modèles connexionnistes de l'apprentissage*. PhD thesis, Université Paris 6.

Daniel, J. (2000). *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.

Divenyi, P. L. and Oliver, S. K. (1989). Resolution os steady-state sound in simulated auditory space. *J. Acoust. Soc. of Am.*, 85(5) :2042–2052.

Domnitz, R. H. (1973). The interaural time jnd as a simultaneous function of interaural time and interaural amplitude. *J. Acoust. Soc. of Am.*, 53(6) :1549–1552.

Domnitz, R. H. and Colburn, H. S. (1977). Lateral position and interaural discrimination. *J. Acoust. Soc. of Am.*, 61(5) :1586–1598.

Duda, R. O., Avendano, C., and Algazi, V. (1999). An adaptable ellipsoidal head model for the interaural time difference. In *Proc. IEEE International Conference on Acoustics Speech and Signal Processing*, volume II, pages 965–968. ICASSP'99.

Dudouet, E. and Martin, J. (1998). Analyses multidimensionnelles des hrtf pour la spatialisation de sources sonores. Technical report, CSTB.

Duraiswami, R., Zotkin, D. N., and Gumerov, N. A. (2004). Interpolation and range extrapolation of hrtf. volume IV, pages 45–48.

Emerit, M. and Martin, J. (1995). Head-related transfer functions and high-order statistics.

Ericson, M. and McKinley, R. L. (1989). Auditory localization cue synthesis and human performance. pages 718–725, Dayton, OH. NAECON 89.

Fahn, C. and Lo, Y. (2003). On the clustering of head-related transfer functions used for 3-d sound localisation. *Journal of Information Science en Engineering*, 19 :141–157.

Faller, C. and Baumgarte, F. (2002). Binaural cue coding : a novel and efficient representation of spatial audio. volume 2, pages 1841–1844.

Faller, C. and Baumgarte, F. (2003). Binaural cue coding-part ii : Schemes and applications. *IEEE Transactions on Speech and Audio Processing*, 11(6) :520–531.

Fels, J., Buthmann, P., and Vörlander, M. (2004). Head-related transfer functions of children. *Acta Acoustica united with Acoustica*, 90 :918–927.

Francescantonio, P. D. (2003). *Vnoise theorotical manual*. http : www.sts-soft.com.

Genuit, K. (1984). *Ein modell zur beschreibung von aussenohrübertragungseigenschaften*. PhD thesis, RWTH Aachen.

Grantham, D. W., Hornsby, B. W. Y., and Erpenbeck, E. A. (2003). Auditory spatial resolution in horizontal, vertical, and diagonal planes. *J. Acoust. Soc. of Am.*, 114(2) :3030–3038.

Grantham, D. W., Whillhite, J. A., Frampton, K. D., and Ashmead, D. H. (2005). Reduced order modeling of head related impulse responses for virtual acoustics displays. *J. Acoust. Soc. of Am.*, 117(5) :3116–3152.

Gray, R. M. (1984). Vector quantization. *IEEE ASSP Magzine*, pages 4–29.

Hafter, E. R. and Maio, J. D. (1975). Difference thresholds for interaural delay. *J. Acoust. Soc. of Am.*, 57(1) :181–187.

Han, H. L. (1994). Measuring a dummy head in search of pinna cues. *J. Acoust. Soc. of Am.*, 42 :15–37.

Hartmann, W. M. (1998). *Signal, sound and sensation*. Springer.

Hasegawa, H., Kasuga, M., Matsumoto, S., and Koike, A. (2000). Simple relaization of sound localization using hrtf approximated by iir filter. *IEICE Transactions fundamentals*, E83-A(6) :973–978.

Henning, G. B. (1974). Detectability of interaural delay in high-frequency complex waveforms. *J. Acoust. Soc. of Am.*, 55(1) :84–90.

Hershkowitz, R. M. and Durlach, N. I. (1969). Interaural time and amplitude jnds for a 500-hz tone. *J. Acoust. Soc. of Am.*, 46(6, Part 2) :1464–1467.

Hofman, P. M., van Riswick, J. G. A., and van Opstal, A. J. (1998). relearning sound localisation with new ears. *Nature Neuroscience*, 1(5) :417–421.

Huopaniemi, J. and Smith, J. O. (1999). Spectral and time-domain preprocessing and the choice of modelling error criteria for binaural digital filters.

Jenison, R. L. (1995). A spherical basis function neural network for pole-zero modeling of head-related transfer functions. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics.

Jenison, R. L. and Fissell, K. (1996). A spherical basis function neural network for pole-zero modeling auditory space. *Neural computation*, 8 :115–128.

Jin, C., Leong, P., Leung, J., Corderoy, A., and Carlile, S. (2000a). Enabling individualized virtual auditory space using morphological measurements. pages 235–238. Proceedings of the First IEEE Pacific-Rim Conference on Multimedia (2000 International Symposium on Multimedia Information Processing).

Jin, C., Schenkel, M., and Carlile, S. (2000b). Neural system identification model of human sound localization. *J. Acoust. Soc. of Am.*, 108(3) :1215–1235.

Jin, C., Tan, T., Leung, J., van Schaik, A., and Carlile, S. (2002). Human interface support for information technology for mars research missions. pages 1261–1266. http ://www.marssociety.org.au/amec2002/.

Jin, G., Corderoy, A., Carlile, S., and van Shaik, A. (2004). Contrasting monaural and interaural spectral cues for human sound localization. *J. Acoust. Soc. of Am.*, 115(6) :3124–3141.

Kahana, Y. (2000). *Numerical modelling of the head related transfer function*. PhD thesis, University of Southampton.

Katz, B. F. G. (1998). *Measurement and calculation of individual head-related transfer function using a boundary element model incuding the measurement and effect of skin and hair impedance*. PhD thesis, Pennsylvania State University.

Katz, B. F. G. (2001a). Boundary element method calculation of individual head-related transfer function. i. rigid sphere calculation. *J. Acoust. Soc. of Am.*, 110(5).

Katz, B. F. G. (2001b). Boundary element method calculation of individual head-related transfer function. ii. impedance effects and comparison to real measurements. *J. Acoust. Soc. of Am.*, 110(5) :2449–2455.

Kistler, D. J. and Wightman, F. L. (1992). A model of head-related transfert functions based on principal components analysis and minimum-phase reconstruction. *J. Acoust. Soc. of Am.*, 91(3) :1637–1647.

Klump, R. G. and Eady, H. R. (1956). Lateral position and interaural discrimination. *J. Acoust. Soc. of Am.*, 28(5).

Kohonen, T. (1995). *Self-organizing maps*, volume 30 of *Springer series in information sciences*. Springer.

Kollmeier, B., Gilkey, R. H., and Sieben, U. K. (1988). Adaptative staircase techniques in psychoacoustics : A comparison of human data and a mathematical model. *J. Acoust. Soc. of Am.*, 83(5) :1852–1862.

Kuhn, G. F. (1977). Model for the interaural time differences in the azimuthal plane. *J. Acoust. Soc. of Am.*, 62(1).

Kulkarni, A. and Colburn, H. S. (1995). Efficient finite-impluse-response filter of the head related transfer function. *J. Acoust. Soc. of Am.*, 5 :3278.

Kulkarni, A. and Colburn, H. S. (2000). Variability in the characterisation of the headphone transfer-function. *J. Acoust. Soc. of Am.*, 107(2) :1071–1074.

Kulkarni, A. and Colburn, H. S. (2004). Infinite-impulse-response models of the head-related transfer function. *J. Acoust. Soc. of Am.*, 115(4) :1714–1728.

Kulkarni, A., Isabelle, S., and Colburn, H. (1999). Sensitivity of human subjects to head-related transfer function phase spectra. *J. Acoust. Soc. of Am.*, 105(5) :2821–2840.

Laakso, T. I., Välimäki, V., Karjailanen, M., and Laine, K. U. (1996). Splitting the unit delay : tools for fractional delay filter design. *IEEE Signal Processing Magazine*, 12(1) :30–60.

Langendijk, E. H. A. and Bronkhorst, A. W. (2002). Contribution of spectral cues to human sound localization. *J. Acoust. Soc. of Am.*, 112(4) :1583.

Langendjik, E. H. A. and Bronkhorst, A. W. (2000). Fidelity of three-dimensional-sound reproduction using a virtual auditory display. *J. Acoust. Soc. of Am.*, 107(1) :528–537.

Larcher, V. (2001). *Techniques de spatialisation des sons pour la réalité virtuelle.* PhD thesis, Université Paris VI.

Larcher, V. and Jot, J. M. (1999). Techniques d'interpolations de filtres audionumériques : Application à la reproduction spatiale des sons sur écouteurs. Marseilles. Présenté au Congès Français d'Acoustique.

Lemaire, V. (1999). *Une nouvelle fonction de coût régularisante dans les réseaux de neurones artificles : application à l'estimation des temps de blocage dans un noeuds ATM.* PhD thesis, Université Paris VI.

Lemaire, V. (2001). Les réseaux de neurones. Cours présenté à l'Ecole National de Statistique et d'Analyse de l'Information.

Lemaire, V. and Clérot, F. (2004). An input variable importance definition based on empirical data probability and ist use in variable selection. pages 1375–1380.

Lemaire, V., Clérot, F., Busson, S., Nicol, R., and Choqueuse, V. (2005). Individualized hrtfs from few measurements : a statistical learning approach.

Levitt, H. (1970). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. of Am.*, 49 :467–477.

Litovsky, R. Y., Hawley, M. L., and Fligor, B. J. (2000). Failure to unlearn the precedence effect. *J. Acoust. Soc. of Am.*, 108(5) :2345–2352.

Lord Rayleigh (1876). Our perception of the direction of sound. *Nature*, XIV :32–33.

Lord Rayleigh (1907). On our perception of sound direction. *Phylosophy Magazine*, 13 :214–232.

Mackensen, P. (2004). *Auditive localization. Head movements, an additional cue in localization.* PhD thesis, Technischen Universität Berlin.

Macpherson, E. A. and Middlebrooks, J. C. (2002). Listener weighting of cues for lateral angle : The duplex theory of sound localization revisited. *J. Acoust. Soc. of Am.*, 111(5) :2219–2236.

Marin, M. (1996). *Etude de la localisation en restitution du son pour la téléconférence de haute qualité.* PhD thesis, Université du Maine.

Martens, W. (1987). Principal componant analysis and resynthesis of spectral cures to perceived direction. pages 274–281. ICMC.

Matusmoto, M., Yamanaka, S., and Tohyama, M. (2004). Effect of arrival correction on the accuracy of binaural impulse response interpolation. *J. Audio Engin. Soc.*, 52(1/2) :56–61.

Hammershøï, D. and Møller, H. (2002). Methods for binaural recording and reproduction. *Acta Acoustica united with Acoustica*, 88 :303–311.

Møller, H. (1992). Fundamentals of binaural technology. *Applied Acoustics*, 36(5) :171–218.

Møller, H., rensen, M. F. S., i, D. H., and Jensen, C. B. (1995). Head related transfer functions of human subjects. *J. Audio Engin. Soc.*, 43(5) :300–321.

McAdams, S. (1994). *Audition : physiologie, perception et cognition.* Presses Universitaires de France. pp 283–344.

McFadden, D. (1981). The problem of different interaural time differences at different frequencies. *J. Acoust. Soc. of Am.*, 69(6) :1586–1598. Letters to the Editor.

McFadden, D. and Pasanen, E. G. (1976). Lateralization at high frequnecies based on interaural time differences. *J. Acoust. Soc. of Am.*, 59(3) :634–639.

Mercier, D. (1993). *Le livre de techniques du son. Tome III.* Fréquences, paris edition. 458 p.

Middlebrooks, J. (1999). Individual differences in external-ear transfer function reduced by scaling in frequency. *J. Acoust. Soc. of Am.*, 106(3) :1480–1492.

Middlebrooks, J. and Green, D. (1992). Observations on a principal componante analysis of head related transfer function. *J. Acoust. Soc. of Am.*, 92(1) :597–592.

Middlebrooks, J., Makous, J. C., and Green, D. M. (1989). Directional sensitivity of sound-pressure levels in the human ear canal. *J. Acoust. Soc. of Am.*, 86 :89–107.

Middlebrooks, J. C. (1992). Narrow-band sound localization related external ears acoustics. *J. Acoust. Soc. of Am.*, 92 :2607–2624.

Miller, J. D. (2001a). Hrtf error analysis using spectral power ratios.

Miller, J. D. (2001b). Modelling iteraural time difference assuming à spherical head.

Mills, A. W. (1958). On the minimum audible angle. *J. Acoust. Soc. of Am.*, 30 :237–248.

Minaar, P., Plosgsties, J., Olesen, S. K., Christensen, F., and Moller, H. (2000). The interaural time difference in binaural synthesis. *J. Acoust. Soc. of Am.*, 92 :207.

Minnaar, P., Christensen, F., Moller, H., Olesen, S. K., and Plogsties, J. (1999). Audibility of all-pass components in binaural synthesis. Munich. AES 106$^{\text{th}}$ Convention.

Morse, P. and Ingrad, K. U. (1968). *Theorical acoustics.* Mc-Graw-Hill.

Moushegian, G. and Jeffress, L. A. (1959). Role of the interaural time and intensity differences in the lateralization of low-frequency tones. *J. Acoust. Soc. of Am.*, 31(11) :1441–1445.

Nishino, T., Ikeda, M., Takeda, K., and Itakura, F. (2000). Interpolating head related transfer functions.

Nishino, T., Mase, S., Kajita, S., Takeda, K., and Itakura, F. (1996). Interpolating hrtf for auditory virtual reality. pages 1261–1266.

Oldfield, S. R. and Parker, S. P. A. (1984). Acuity of sound localisation : a topography of auditory space. i. normal hearing. *Perception*, 13 :581–600.

Oppenheim, A. V. and Schafer, R. W. (1989). *Discrete-Time Signal Processing*. Prentice Hall, Englewood Cliffs, New Jersey.

Palomäki, K., Pulkki, V., and Karjalainen, M. (2000). Neural network approach to analyse spatial sound.

Patterson, R. D. and Allerhand, M. H. (1995). Time-domain modeling of periphal auditory processing : a modular architecture and a softxar paltform. *J. Acoust. Soc. of Am.*, 98(6) :3435–3444.

Pernaux, J. (2003). *Spatialisation du son par les tehcniques binaurales : application aux services de télécommunication*. PhD thesis, Institut National de Polytechnique de Grenoble.

Pernaux, J.-M., Emerit, M., and Nicol, R. (2003). Perceptual evaluation of binaural sound synthesis : the problem of reporting localization judgments. *Presented at the* 114$^{\text{th}}$ *AES Convention, Amsterdam, The Netherlands*. Convention paper.

Perrot, D. R. (1984). Concurrent minimum audible angle : a re-examination of the concept of auditory spatial acuity. *J. Acoust. Soc. of Am.*, 75(4) :1201–1206.

Perrot, D. R. and Saberi, K. (1990). Minimum audible angle thresholds for sources varying in both elevation in azimuth. *J. Acoust. Soc. of Am.*, 87(4) :1728–1731.

Plogsties, J., Minnaar, P., Olesen, S. K., Christensen, F., and Moller, H. (2000). Audibility of all-pass components in head-related transfer function. Paris. AES 108$^{\text{th}}$ Convention.

Pulkki, V., Karjalainen, M., and Huopaniemi, J. (1998). Analysing virtual sound source attributes using a binaural model. *AES* 114$^{\text{th}}$ *Convention*.

Rio, E. and Warusfel, O. (2002). Optimizations of multi-channel binaural formats based on statistical analysis. Seville. FORUM ACUSTICUM SEVILLA.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning internal representations by error propagation. *In Parallel Distribued Processing : Explorations in the Microstructures of Cognition*, 1 :318–362.

Sandvad, J. and Hammershøi, D. (1994). Comparison of fir and iir filter representation of hrirs. pages 1–16. preprint 3862.

Saviojaa, L., Huopaniemi, J., Lokki, T., and Vaanen, R. (1999). Creating interactive virtual acoustic environnements. *J. Acoust. Soc. of Am.*, 47(9) :675–705.

Schenck, H. A. (1968). Improved integral formulation for acoustic radiation problems. *J. Acoust. Soc. of Am.*, 44(1) :58–41.

Shaw, E. A. G. (1982). *Localization of sound : theory and applications.* Amphora Press. chapitre External ear response and sound localization.

Shin-Cunningham, B. G., Santelli, S., and Kopco, N. (2000). Tori of confusion : binaural localization cues for sources within reach of a listener. *J. Acoust. Soc. of Am.*, 107(3).

Smith, J. O. (1983). *Techniques for Digital Filter Design and System Identification with Application to the Violin.* PhD thesis, Stanford University.

Smith, J. O. and Abel, J. S. (1999). Bark and erb bilinear transforms. In *IEEE Trans. on speech and audio processing.* IEE.

Tollin, D. J. and B., H. G. (1998). Some aspects of the lateralization of echoed sound in man. i. the classical interaural-delay based precedence effect. *J. Acoust. Soc. of Am.*, 104 :3030–3038.

Trahiotis, C., Bernstein, L. R., Buell, T. N., and Spektor, Z. (1990). On the use of adaptative procedures in binaural experiments. *J. Acoust. Soc. of Am.*, 87(3) :1359–1361.

Vandernoot, G. Sytem and protocol. http://recherche.ircam.fr/equipes/salles/listen/system/protocol.html.

Vesanto, J., Himberg, J., Alhoniemi, E., and Parhankangas, J. (2000). Som toolbox for matlab 5. Technical report, Helsinki Uninersity of Technology.

von Békésy, G. (1960). *Experiments in hearing.* Acoustical Society of America, mcgraw-hill book company edition.

Vovor, P. (2005). Utilisation d'outils statistiques pour l'individualisation des hrtf. Rapport de stage master 2, Université Paris IV, Master ATIAM.

Weinrich, S. (1984). Sound field caculations around the human head. Technical report, The acoustic laboratory, Technical university of Denmark.

Wightman, F. L. and Kistler, D. J. (1989a). Headphone simulation of free-field listening. i : Stimulus synthesis. *J. Acoust. Soc. of Am.*, 85(2) :858–867.

Wightman, F. L. and Kistler, D. J. (1989b). Headphone simulation of free-field listening. ii : Psychophysical validation. *J. Acoust. Soc. of Am.*, 85(2) :868–878.

Wightman, F. L. and Kistler, D. J. (1997). Monaural sound localization revisited. *J. Acoust. Soc. of Am.*, 101(2) :1050–1063.

Wightman, F. L. and Kistler, D. L. (1999). Resolution of front–back ambiguity in spatial hearing by listener and source movement. *J. Acoust. Soc. of Am.*, 105(5) :2841–2853.

Woodworth, R. S. and Schloesberg, G. (1962). *Experimental psychology.* New-York. pp 349-361.

Wu, Z., Chan, F. H. Y., and K., L. F. (1997). A time domain binaural model based on spatial feature extrcation for the head related transfer function. *J. Acoust. Soc. of Am.*, 102(4) :2211–2218.

Wu, Z., Weng, T., Wang, W., Lao, T. F., Chan, F. H. Y., and Lam, F. K. (1998). Neural network model of binaural hearing based on spatial feature extraction of the head related transfer function. volume 20.

Yost, W. A., Turner, R., and Bergert, B. (1974). Comparison among four psychophysical procedures used in lateralization. *Perception & Psychophysics*, 15(3) :483–487.

Zotkin, D. N., Duraiswami, R., and Gumerov, E. G. N. A. (2004). Fast head related transfer function measurment via reciprocity. Technical report CS-4620 and UMIACS-20004-62, Universitt of Maryland, Computer Science and UMIACS.

Zurek, P. M. (1985). Spectral dominance in sensitivity to interaural delay for broadband stimuli. *Presented at the* 110^{th} *meeting of the Acoustical Society of America.*

Annexes

A

Expérimentation sur l'implémentation des filtres binauraux

Le test comporte deux taches :

- écoute des différentes estimations des HRIR qui conduisent à l'implémentation *Spat˜* et qui sont succeptibles d'introduire des dégradations perceptibles.
- comparaison du type A-B-X.

Premièrement, il est demandé aux sujets d'écouter les différentes étapes de l'implémentation des HRIR. Les sujet cliquent sur des boutons pour écouter un son spatialisé aux azimuts et élévations choisies au préalable (cf. fig.A.1). Sur les boutons, apparaissent le nom de la modification des HRIR par rapport aux HRIR originales (HRIR mesurées à l'IRCAM et égalisées champ diffus). Il est demandé au sujet de bien écouter toutes les étapes et d'exprimer leurs impressions sur les éventuels défauts perçus. Ensuite, la deuxième tâche s'effectue en aveugle. Trois sons sont diffusés l'un après l'autre, le troisième étant une répétition d'un des deux premiers. Il est alors demandé au sujet d'identifier à quel son correspond le troisième. Si le sujet donne la bonne réponse, il lui est demandé sur quels critères se base son analyse.

Le stimulus correspond à un bruit blanc gaussian d'une durée de 600 ms et échantillonné à 44100 Hz, avec une attaque et une décroissance en cosinus d'une durée de 5 ms. Le test utilise une interface graphique écrite sous Matlab. Cette interface est composée,

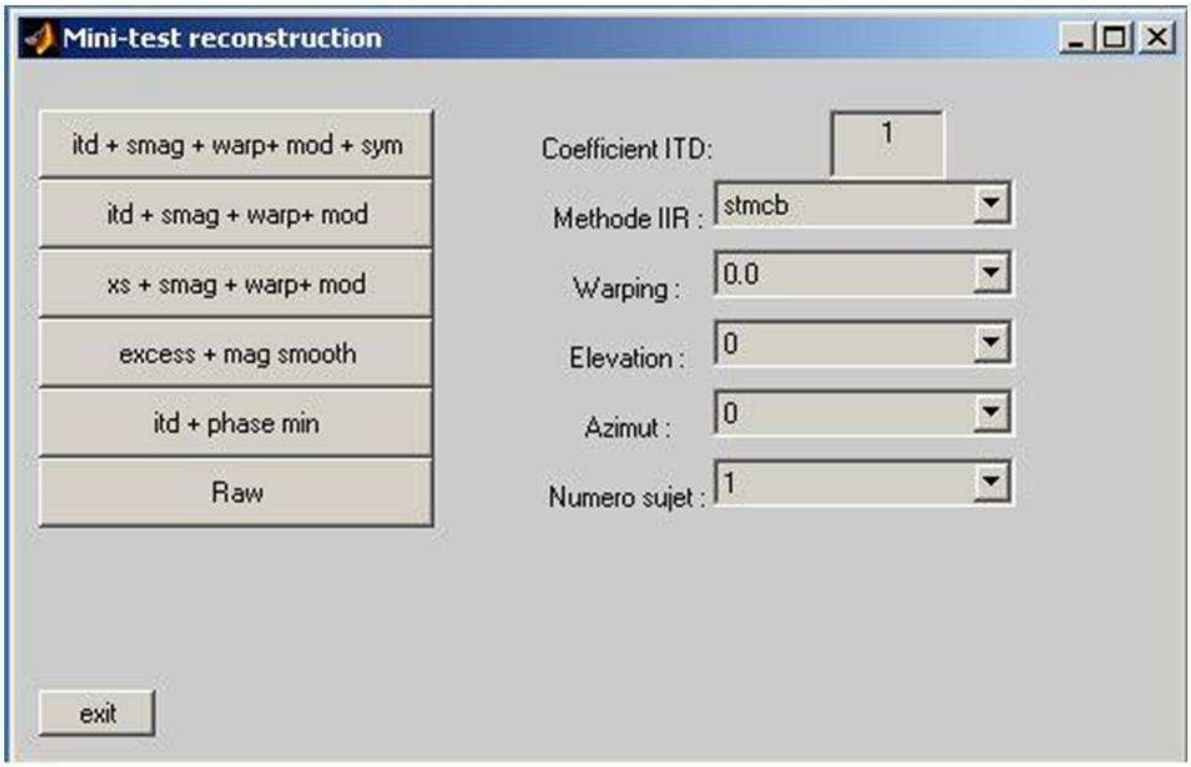


FIG. A.1 – Interface d'évaluation des différentes étapes de l'implémentation des HRTF présentes dans *Spat˜*.

sur la gauche d'une colonne de six boutons pour les différentes étapes de l'implémentation et sur la droite de cinq menus déroulants pour les paramètres et une case d'édition d'un coefficient multiplicateur de l'ITD (cf. fig.A.1). Les six boutons de gauche permettent d'écouter la convolution du son avec les HRIR :

- compensées en champ diffus (*Raw*).
- phase minimum avec estimation de l'ITD par la méthode MaxIACC.
- phase mixte avec lissage du module.
- phase mixte, lissage, warping fréquentiel[1] et modélisation RII.
- ITD, lissage, warping fréquentiel et modélisation RII.
- ITD, lissage, warping fréquentiel, modélisation RII et symétrisation des deux oreilles[2] (=implémentation *Spat˜*).

Les paramètres réglables sont :

- l'azimut par pas de 15°, le choix "plan" permettait la synthèse d'une trajectoire sur toutes les azimuts.

[1] *warping fréquentiel* désigne la transformation homothétique des HRTF selon l'axe des fréquences cf. § .

[2] Une moyenne est effectuée entre les HRTF de l'oreille droite et celles de l'oreille gauche pour obtenir un jeu commun de HRTF pour les deux oreilles

– l'élévation, seulement 0° et 30°.
– le coefficient du warping fréquentiel, 0, 0.4 et 0.7.
– la méthode de modélisation RII d'orde 12, Yule-Walker et Steiglitz-Mc Bride.
– un coefficient multiplicatif de la valeur de l'ITD.

Huits sujets ont passé le test. Tous sont experts en acoustique.

B

Formule de Déplacement des Oreilles

Démonstration de l'équivalence entre la formule FDO et la formule de Woodworth (cf. eq.B.1) pour le calcul de l'ITD dans le plan horizontal. La modèlisation physique de l'auditeur est alors une sphère de rayon a et les oreilles sont diamétralement opposées. L'incidence de l'onde qui arrive aux oreilles de l'auditeur est repérée par l'angle θ (cf. fig.B.1). La direction et le sens de l'onde incidente et l'orientation des oreilles par rapport au centre de la sphère sont repérés par l'intermédiaire de vecteurs unitaires. Le vecteur unitaire de l'onde incidente, $\overrightarrow{U}_{inc}$, est orienté vers le centre de la sphère et les vecteurs unitaires des oreilles, respectivement $\overrightarrow{U}_d$ pour l'oreille droite et $\overrightarrow{U}_g$ pour l'oreille gauche, sont orientés vers l'extérieur de la sphère. Les distances d'arc sont calculées avec des produits scalaires entre les vecteurs d'ondes et les vecteurs d'oreille. L'angle θ varie entre 0 et 2π. La démonstration se restreint au cas où l'ITD est positif, c'est-à-dire pour $\theta \in [0 - \pi]$.

$$ITD_{sphere}(\theta) = \frac{a}{c}(\sin(\theta) + \theta) \tag{B.1}$$

Quatres cas sont distingués selon que l'onde incidente éclaire ou pas les oreilles (cf. fig.II.20). Le cas 3 correspond à la plage de variation choisie.

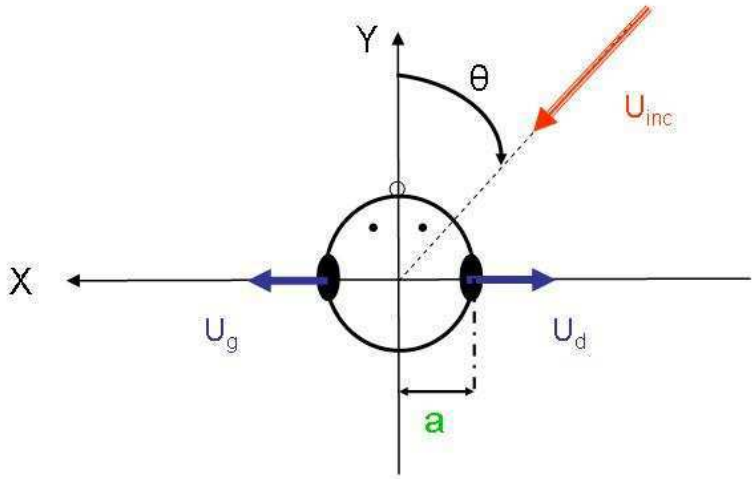


FIG. B.1 – Convention d'orientation des angles. $\overrightarrow{U}_{inc}$, $\overrightarrow{U}_{d}$ et $\overrightarrow{U}_{g}$ représentent respectivement, les orientations de l'onde incidente, de l'oreille droite et de l'oreille gauche.

$$
\begin{aligned}
ITD &= -\frac{R}{C} * (\frac{\pi}{2} - \arccos(\overrightarrow{U}_{inc} \cdot \overrightarrow{U}_{d}) - \overrightarrow{U}_{inc} \cdot \overrightarrow{U}_{g}) && \text{(B.2)} \\
&= -\frac{R}{C} * (\frac{\pi}{2} - \arccos(\cos(\theta + \frac{\pi}{2})) - \cos(\frac{\pi}{2} - \theta)) && \text{(B.3)} \\
&= -\frac{R}{C} * (\frac{\pi}{2} - \theta - \frac{\pi}{2} - \sin(\theta)) && \text{(B.4)} \\
&= \frac{R}{C} * (\theta + \sin(\theta)) && \text{(B.5)} \\
& && \text{(B.6)}
\end{aligned}
$$

C

fdo.m

FDO.M

```
function ITD = itd_sphere_dec(Inc,G,D,R)

%   ITD_ SPHERE_DEC calcul de l'ITD Hautes Fréquences (> 1500 Hz) pour une
%   sphère avec oreilles décalées
%   Paramètres d'entrée :
%     Inc = structure relative à l'onde incidente
%     G   = structure relative à la position de l'oreille gauche
%     D   = structure relative à la position de l'oreille droite
%     R   = Rayon de la sphère (en m)
%     chaque structure de paramètres doit contenir les champs :
%       elev_v et azim_v
%  example : ITD = itd_sphere_dec(Inc,G,D,R)
%
% S. Busson, R. Nicol 09/2003

c = 344;   % célérite du son

u_inc = zeros(1,3); u_gau = zeros(1,3); u_dro = zeros(1,3);

[u_gau_x,u_gau_y,u_gau_z] = sph2cart(G.azim_v,G.elev_v,R);
```

```
[u_dro_x,u_dro_y,u_dro_z] = sph2cart(D.azim_v,D.elev_v,R);

u_gau = [u_gau_x,u_gau_y,u_gau_z]; u_gau = u_gau./norm(u_gau);
u_dro = [u_dro_x,u_dro_y,u_dro_z]; u_dro = u_dro./norm(u_dro);

for i = 1:length(Inc.azim_v)

    %Vecteurs unitaires en coordonnées cartésiennes

    [u_inc_x,u_inc_y,u_inc_z] = sph2cart(Inc.azim_v(i),Inc.elev_v(i),R);

    u_inc = [u_inc_x,u_inc_y,u_inc_z];
    u_inc = u_inc./norm(u_inc);

    %Les trois cas à distinguer selon le signe des produits scalaires

    scal_1 = u_inc*u_gau' ;
    scal_2 = u_inc*u_dro' ;


    % 1) l'onde incidente voie les deux oreilles
    if scal_1 <= 0 & scal_2 <= 0

        ITD(i) = -R/c.*(scal_2-scal_1);


    % 2) l'onde incidente voie une des deux oreilles
    elseif scal_1 <= 0 & scal_2 > 0;

        ITD(i) = -R/c.*(pi/2 - acos(scal_2) - scal_1 );

    elseif scal_2 <= 0 & scal_1 > 0;

        ITD(i) = R/c.*(pi/2 - acos(scal_1) - scal_2 );

     % 3) l'onde incidente ne voie aucune des oreilles
    elseif scal_1 > 0 & scal_2 > 0

        ITD(i) = -R/c.*( - acos(scal_2) + acos(scal_1) );

    end
end
```

D

Méthode d'apprentissage par descente de gradient

L'apprentissage d'un RNA par correction d'erreur consiste à minimiser la fonction de coût. Pour ce faire, l'algorithme le plus utilisé est l'algorithme de *descente de gradient*. Le calcul du gradient se fait en utilisant l'algorithme de la rétro-propagation de l'erreur [Rumelhart et al. (1986)]. Cette algorithme consiste en 4 étapes appliquées à chaque pas d'apprentissage du RNA (cf. fig.D.1) :

1. proposer des couples entrée - sortie voulues
2. calculer l'erreur totale (cf. eq IV.6)
3. calculer la contribution de chaque neurone à l'erreur totale, c'est-à-dire calculer les dérivées partielles de l'erreur totale par rapport à chaque neurone, ou encore calculer le gradient de l'erreur par rapport à w.
4. ajuster les coefficients synaptiques dans le sens inverse du gradient de l'erreur.

L'étape 3 est la plus délicate à réaliser et impose la continuité et la dérivabilité des fonctions d'activations. La démonstration qui suit fait l'hypothèse d'une fonction d'activation de type sigmoïde à seuil nul (cf. fig.D.2) :

$$\sigma(x) = \frac{e^{kx}}{1 + e^{kx}} = \frac{1}{1 - e^{kx}} \tag{D.1}$$

avec $k \geq 1$. Dans la suite $k = 1$.Cette fonction est une approximation indéfiniment dérivable de la fonction à seuil de Heaviside, d'autant meilleure que k est grand. La

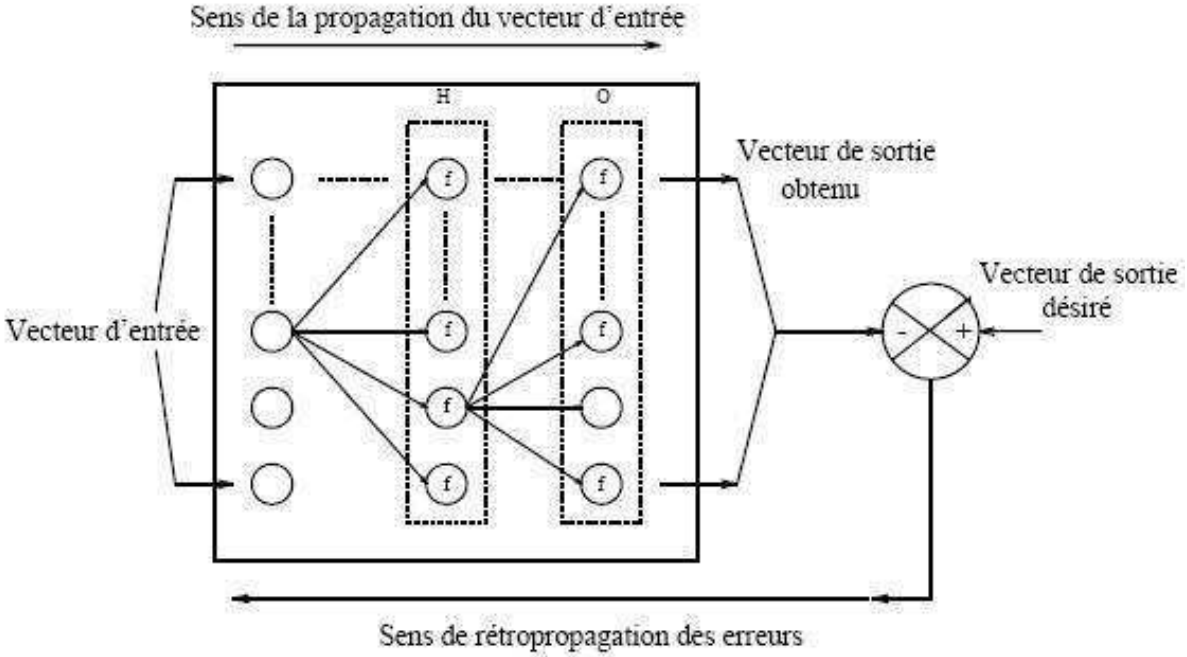


FIG. D.1 – Représentation schématique de la rétropropagation de l'erreur

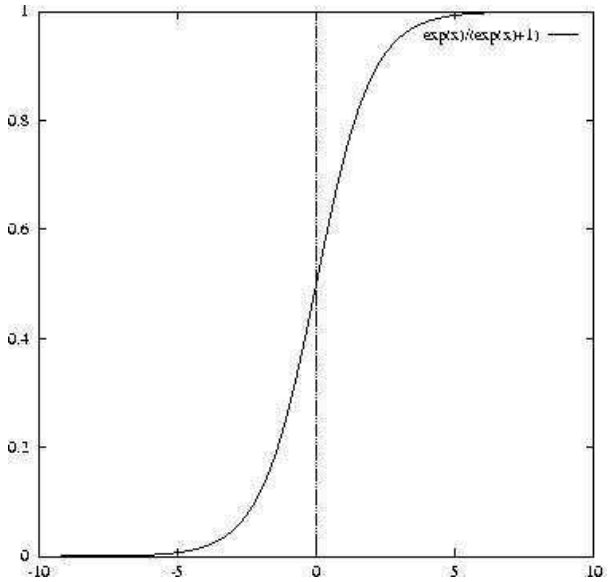


FIG. D.2 – Fonction sigmoide

dérivée de la fonction σ est simple à calculer :

$$\sigma'(x) = \frac{e^x}{(1+e^x)^2} = \sigma(x)(1-\sigma(x)) \tag{D.2}$$

Dans souci de lisibilité, le calcul ne s'effectue que sur un exemple et non sur la totalité de l'ensemble des exemples. Il est essentiel que ce calcul soit simple, puisqu'il doit être effectué autant de fois qu'il y a de neurones, à chaque étape de modification des poids synaptiques. Un réseau comprenant $q+1$ couches numérotées de 0 à q est considéré et les notation suivantes sont utilisées (cf. fig.D.3) :

- n_L le nombre de cellules de la couche L
- $y_i^{(L)}$ l'entrée de la i^{eme} cellule de la couche L pour $L \geq 1$.
- $z_i^{(L)}$ l'état de la i^{eme} cellule de la couche L. L'état coïncide avec l'entrée si $L = 0$ et avec la sortie si $L = q$.
- $\alpha_{ij}^{(L)}$ le coefficient synaptique entre la j^{eme} cellule de la couche $L-1$ et la i^{eme} cellule de la couche L
- $t_j^{(q)}$ la sortie attendue de la j^{eme} cellule de la couche de sortie

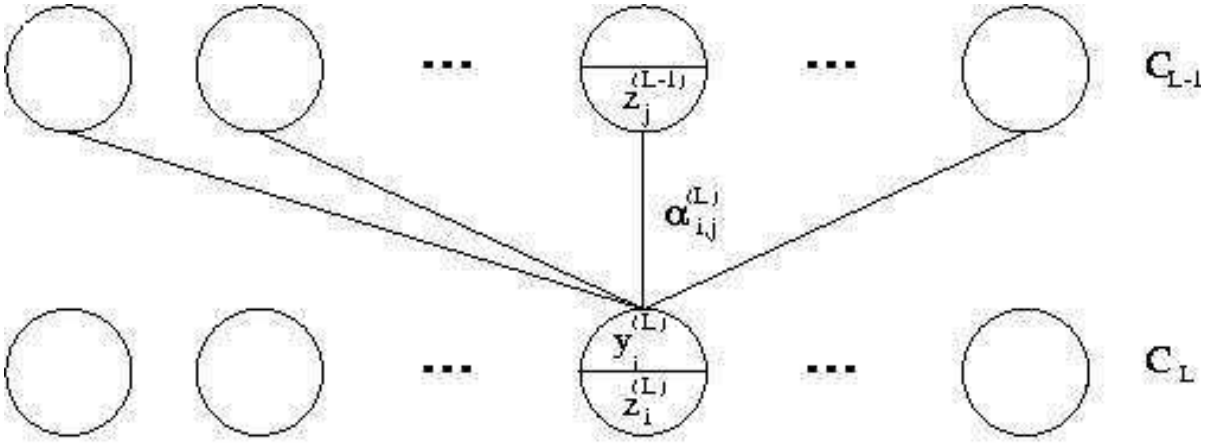


Fig. D.3 – Représentation d'une couche cachée

La sortie d'une couche L est donnée par :

$$z_i^{(L)} = \sigma(y_i^{(L)}) \quad \text{et} \quad y_i^{(L)} = \sum_j \alpha_{ij}^{(L)} z_j^{(L-1)} \tag{D.3}$$

L'erreur du réseau pour une entrée $s = (y_i^{(0)}, ..., y_{n_0}^{(0)})$ est égale à l'erreur quadratique moyenne :

$$E(s) = \frac{1}{2} \sum_j (z_j^{(q)} - t_j^{(q)})^2 \tag{D.4}$$

La rétropropagation de l'erreur consiste à calculer récursivement les quantités $\frac{\partial E}{\partial y_j^{(L)}}$. Pour $L = q$:

$$\frac{\partial E}{\partial y_j^{(q)}} = \frac{\partial E}{\partial z_j^{(q)}} \frac{\partial z_j^{(q)}}{\partial y_j^{(q)}} = z_j^{(q)}(1 - z_j^{(q)}) \frac{\partial E}{\partial z_j^{(q)}} = z_j^{(q)}(1 - z_j^{(q)})(z_j^{(q)} - t_j^{(q)}) \tag{D.5}$$

Pour L<q, on a :

$$\frac{\partial E}{\partial y_j^{(L)}} = \frac{\partial E}{\partial z_j^{(L)}} \frac{\partial z_j^{(L)}}{\partial y_j^{(L)}} = z_j^{(L)}(1 - z_j^{(L)}) \frac{\partial E}{\partial z_j^{(L)}} \tag{D.6}$$

et

$$\frac{\partial E}{\partial z_j^{(L)}} = \sum_k \alpha_{kj}^{(L+1)} \frac{\partial E}{\partial y_k^{(L+1)}} \tag{D.7}$$

Les quantités $\frac{\partial E}{\partial y_j^{(L)}}$ étant connues, la contribution à l'erreur des coefficients $\alpha_{ji}^{(L)}$ peut être calculée :

$$\frac{\partial E}{\partial \alpha_{ji}^{(L)}} = \frac{\partial E}{\partial y_j^{(L)}} \frac{\partial y_j^{(L)}}{\partial \alpha_{ji}^{(L)}} = z_j^{(L-1)} \frac{\partial E}{\partial y_j^{(L)}} \tag{D.8}$$

Par suite on obtient les valeurs des modifications des poids du perceptron au coefficient ϵ près :

$$\Delta \alpha_{ji}^{(L)} = -\epsilon \frac{\partial E}{\partial \alpha_{ji}^{(L)}} \tag{D.9}$$

ϵ est un coefficient réel positif qui représente le pas de déplacement en direction de la pente de la fonction de coût. Dans la recherche d'un mimimum global, il se peut que l'algorithme présenté converge vers un minimum local. C'est pourquoi un terme d'inertie peut-être rajouté à l'équation eq.D.9 :

$$\Delta \alpha_{ji}^{(L)} = -\epsilon \frac{\partial E}{\partial \alpha_{ji}^{(L)}} + \beta \Delta \alpha_{ji} \tag{D.10}$$

E

Articles

Influence of the ears canals location on spherical head model for the individualized interaural time difference

Sylvain Busson[1], Rozenn Nicol[1], Olivier Warusfel[2]

[1] *France Telecom R&D, 22300 Lannion,France, Email: sylvain.busson@rd.francetelecom.com, rozenn.nicol@rd.francetelecom.com*

[2] *Institut de Recherche et de Coordination Acoustique Musicale, 75004 Paris, France, Email: olivier.warusfel@ircam.fr*

Introduction

Human sound localization is governed by interaural and monaural cues which are embedded by the head related transfert function (HRTF). Measuring HRTF is a time-consuming and expensive task. Furthermore, using non-individual HRTF results in localization blur, intracranial perception and front-back confusion. To avoid measurement of HRTF, physical models which take the transformations of the sound field, generated by the source, by the listeners head, torso and pinnae into account are used. The most simple and widely used model is the spherical head model with centered ears canals. This model provides quite good approximation of human HRTF and analytical formulations of localization cues. The most salient localization cue is the interaural time difference (ITD) mainly for the lateralization of sound sources. Many formulations of the ITD for a sphere are available. Associated with the Woodworth formula [1]:

$$ITD = \frac{a}{c}(\theta + \sin\theta) \quad (1)$$

where a is the sphere radius, θ is the azimuth angle and c the speed of sound, the Algazi radius well predicts CIPIC subjects ITD [2] [3]. Comparison between constant ITD contours of a sphere (cones of confusion) and a human head shows deviation up to 18% of the maximum value [4]. These deviations are caused by the non spherical shape of the head and the location of the entrance of the ears canals with respect to the center of the head.

In this paper, a new formula is proposed to take the ear displacement into account in the calculation of the spherical head model ITD. As expected, the displacement of the ear canals causes the ITD to vary along a cone of confusion. A psychoacoustic experiment is conducted in order to study to what extent the difference between measured and simulated ITD is audible.

Ear Displacement Formula

Validation

A simple spherical head model ITD formula which takes the ears displacement with respect to the center of the head into account, does not exist. The ear displacement formula (EDF) considers the inner product between both unit right and left ear vectors (Ur, Ul), corresponding to the ears' location, and the unit incident vector of the sound ($Uinc$). Four cases have to be distinguished as the incident plane reach the ears before or after head masking. For example, the formulation in the case of both ear being shadowed is given below :

$$ITD = -\frac{R}{C} * (\arccos(\overrightarrow{Uinc} \cdot \overrightarrow{Ul}) + \arccos(\overrightarrow{Uinc} \cdot \overrightarrow{Ur})) \quad (2)$$

EDF with no ears' displacement is comparable with elevation dependant Larcher and Jot formula [5]. Figure 1 shows the exact matching of the EDF.

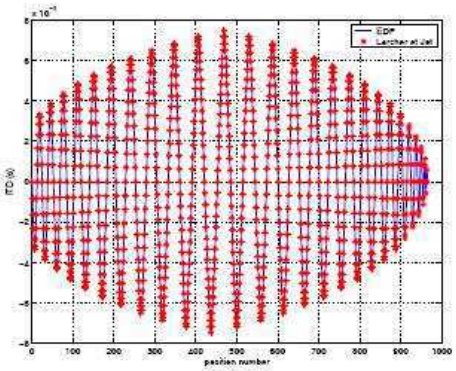


Figure 1: Comparison between EDF and Larcher and Jot Formula

ITD Elevation Dependent Variation

The ear displacement is given by the angular shift to the central position. For exemple, a 10° azimuth and 5° elevation displacement of the ear canal means that the right and left ear locations are 280° azimuth and 5° elevation and 80° azimuth and 5° elevation respectively. Figure 2 shows the ratio between ITD with ears displacement and ITD with ears centered versus elevation for the −60° cone of confusion and several azimuthal ears displacements. The EDF variations compared with constant spherical head ITD are in the same range of the values observed on the ITD estimated from measurement on human subjects [4]. The main effects of ears displacement are : a decrease of the ITD mean value on a cone of confusion and increase of variations and peak values with ears displacement. A least squares minimization procedure could be run to set the optimal ears' displacement giving the best matching. Some deviations remain anyway, probably due to the non-spherical shape of the human head.

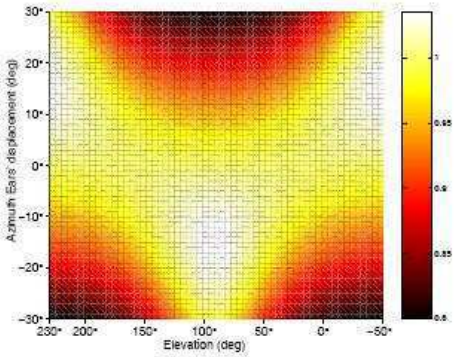


Figure 2: Ratio between ITD with and without ears displacement on the −60° cone of confusion versus elevation with azimuthal ears displacement as parameter

Psychoacoustics Experiment

Subjects and Task

In order to examine to what extent the ITD variations along a cone of confusion are audible, an informal psychoacoustic experiment was led with the FTRD subjects [6]. 8 subjects with normal hearing, 5 males and 3 females from 25 to 43 years were tested. The experiment consisted in listening to 40 pairs of sound through headphones. The subjects were asked to determine if they perceive audible differences between the two sounds without a forced choice protocol. Subjects could listen to the sounds as many times as they wanted before giving their answer. They were further asked to verbalize the differences, if any.

Stimuli

8 sounds locations along 5 different cones of confusion were tested. Each sound of a pair was a white noise convolved with the subject's own HRTFs (except for the subject 8 who listened with another subject HRTFs) over-sampled at 96000 Khz (in order to have 10,4 μs as temporal precision). Implementation of HRTF was made with minimum-phase FIR filter and a pure delay corresponding to the ITD. For one sound, the pure delay was the ITD estimated from the HRTF as in [2] and for the other it was the ITD computed from the spherical-head model with ears centered.

Results

Table 1 indicates that subjects can be divided into two groups : those who perceive none or few differences (Ga Group) and those who perceive differences for half of the presented stimuli (Gb group). Gb contains only experienced subject in listening test.

	S1	S2	S3	S4	S5	S6	S7	S8
Differences	2.5	67.5	50	60	5	47.5	0	10

Table 1: Percentage of discriminate differences

Each Gb subject reported that the audible differences are subtle and required many listenings for the same pair to be discriminate. The verbalization analysis shows that the audible differences source are localization shift and spectral modification. The second difference is quite unexpected because each pair of stimuli was based on the same FIR filters. These results show the ambiguity for the subjects to verbalize a difference in localization. The auditory system can analyze the difference both as a location displacement and a spectral shift. The analysis of the Gb results (see Figure 3) are consistent with those of the former studies in terms of noticeable differences as a function of base line ITD [8] [7]. Moreover, no logical links between ITD differences, or spatial location, and audible differences have been found.

Conclusion

Thanks to EDF, it has been shown that the displacement of the entrance of the ear canals with respect to the center of the head for the spherical head model causes the ITD to vary with elevation. The results of the psychoacoustic experiment highlights differences between subjects' answers. Further detailed and longer listening tests with a forced choice protocol still have to be conducted to know sharply if ITD variation along a cone of confusion have to be reproduced to obtain a satisfying binaural synthesis with the spherical head model.

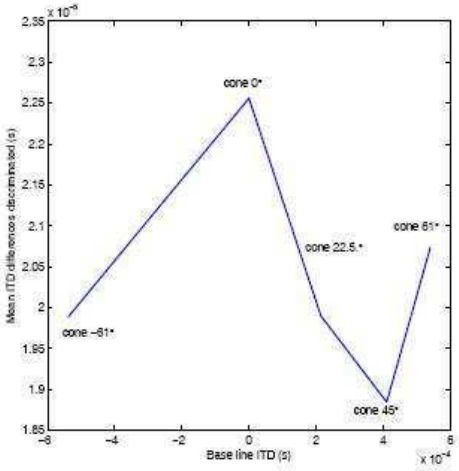


Figure 3: Mean ITD differences discriminated versus constant ITD

References

[1] Woodworth R. S., H. Schloesberg, *Experimental Psychology*, Holt, Rinehard and Winston, NY, pp. (348-361), (1962).

[2] Algazi V. R., C. Avendano and R. O. Duda, "Estimation of a spherical Head-Model from anthropometry", *J. Audio. Eng. Soc.*, Vol. **49**, No. **3**, March (2001).

[3] Algazi V. R., R. O. Duda, D. M. Thompson, "The CIPIC HRTF database", *IEEE Workshop on Applications of Signal Processing*, NY, October (2001).

[4] Duda R. O. , C. Avendano and V. R. Algazi, "An adaptable ellipsoïdal Head-Model for the interaural time difference", *IEEE international conference on Acoustic, Speech and Signal Processing*, Phoenix AZ, pp. (965-968), (1999).

[5] Larcher V. and J. M. Jot, "Techniques d'interpolation de filtres audionumériques : Application à la reproduction spatial des sons sur écouteurs", *Presented at the Congrès Français d'Acoustique*, Marseille, France, April (1999).

[6] Pernaux J. M., "Spatialisation du son par les techniques binaurales : Application aux services de télécommunications", *Ph-D of the Institut National Polytechnique de Grenoble*, May (2003).

[7] Moushegian G. and L. A. Jeffress, "Role of Interaural Time and Intensity Differences in the Lateralization of Low-Frequency Tones", *J. Acoust. Soc. of Am.*, Vol. **31**, No. **11**, November (1959).

[8] Domnitz R., "The interaural time jnd as a simultaneous function of interaural time and interaural amplitude", *J. Acoust. Soc. of Am.*, Vol. **53**, No. **6**, February (1973).

Audio Engineering Society

Convention Paper

Presented at the 118th Convention
2005 May 28–31 Barcelona, Spain

Subjective investigations of the interaural time difference in the horizontal plane

Sylvain Busson[1], Rozenn Nicol[1], and Brian F. G. Katz[2]

[1] *France Telecom R&D, TECH/SSTP, Lannion, France*

[2] *LIMSI, Université d'Orsay, Paris, France*

Correspondence should be addressed to Sylvain Busson (sylvain.busson@francetelecom.com)

ABSTRACT

The Interaural Time Difference (ITD) is the primary cue for sound source localization. Although the ITD has been intensively studied, there are still many open questions. This paper focuses on the ITD in relation to minimum phase filter modeling for binaural synthesis. Comparison of the ITD estimate methods highlights discrepancies for locations near the interaural axis. A subjective experiment was performed to estimate the ITD value to use with a given pair of minimum phase filters. The task of the subject consists in adjusting the ITD of the test sound until its perceived location matches the target sound. The psychoacoustic value of the ITD is compared with computational estimates from Head Related Transfer Function (HRTF).

1. INTRODUCTION

There are several definitions of the ITD, depending on the context. For instance, one definition of the ITD is psychoacoustics : the ITD is one of the localization cue used by the auditory system to identify the location of a sound source [1]. In this context, the ITD is the difference between the arrival time of the sound waves reaching the left and the right ear of a listener. However it should be noticed that this definition is not so clear. In particular, it is not obvious to measure the arrival time of the sound wave reaching one ear. Therefore, the ITD is not obtained directly from measurement, but estimated by signal processing from the measured HRTF, the Head Related Transfer Functions, A great amount of estimate methods exists and the question of which is the most suitable is important.

In the context of binaural synthesis, another definition of the ITD is proposed. Binaural synthesis is based on the HRTF which are the acoustic transfer functions measured between a sound source at a given location and the listener's eardrums. Bin-

aural synthesis consists in rendering 3D audio scene by convolving sounds by binaural filters which corresponds to the HRTF associated to the desired location. Binaural synthesis requires the design of binaural filters according to the HRTF specifications. One common implementation of the binaural filters consists in modeling the HRTF by a minimum phase filter (which reproduces the magnitude component of the HRTF) and a pure delay (which reproduces the phase component of the HRTF, under the assumption that the excess phase is linear) [2] [3]. In this context, the ITD is referred to the difference between the time delay associated with the left and right binaural filters. This is the definition of ITD from the point of view of binaural synthesis implementation. At first sight, this definition seems less ambiguous than the psychoacoustic however there are different methods to design the binaural filters and to compute the pure delay from the HRTF so there are different plausible values for the ITD associated with the same pair of HRTF.

This implementation is strongly dependant on the assumption that the HRTF can be approximated by a minimum phase filter. This is not true for all directions. Particularly for contralateral locations of the sound source (i.e. in the head shadow, where the direct sound is absent and only the diffracted waves remain) the HRTF may not be approximated by a minimum phase filter [4]. It also occurs when a strong reflection is induced, for instance by the torso [5]. If the minimum phase approximation is not valid, what about the validity of the pure delay? Furthermore, it should be also noticed that the excess phase of the HRTF is decomposed into a linear component and an all-pass component. Typically, the ITD is computed from the linear component and the all-pass component is omitted. But it has been shown that the absence of the all-pass component is audible in some cases e.g. for contralateral HRTF [6]. However Plogsties & al suggest that for these HRTF the all-pass component may be replaced by an additional pure delay. This result leads to two conclusions. First, the HRTF model composed of a minimum phase filter and a pure delay is validated. Secondly, to improve this model the ITD value must be slightly increased in some cases,so as the all-pass component has to be taken into account .

These two examples show that the ITD is still not fully understood and that there are many issues to investigate. This paper focuses on the ITD regarding to the implementation definition from a psychoacoustic point of view. Whereas various signal processing methods provide us several plausible values of ITD, we are looking for the psychoacoustic value of ITD, i.e. the value which offers the best perceptive matching between the measured HRTF and the model based on a minimum phase filter and a pure delay. It is expected that this psychoacoustic estimate of ITD will include both the linear component and the all-pass component of the HRTF excess phase.

As an introduction to ITD estimate issues, the following section raises the problem of quantifying the ITD for common binaural synthesis implementation. The various methods of computing the ITD will be briefly discussed. The experiment and its methodology is described in the second section. The last two sections comment and analyze the results of the various experiments.

2. HOW TO QUANTIFY THE ITD ?

As the ITD is the primary psychoacoustic cue for sound localization, it was intensively studied from Rayleigh [7] up to now and ITD estimation methods are numerous. The purpose of this section is not to describe in details all the available methods, but to give an overview of the two main categories of ITD computation , which are :

- Physical modeling of human morphology
- Signal processing method

2.1. Physical modeling

Physical modeling of human morphology is the first method used to get an insight into the spatial variations of the ITD The method is based on elementary geometrical modeling of the human head. Associated with physical considerations on the wave propagation, mathematical formulas are developed. The simplest model is only based on the ear spacing. More complex models allow to specify the ear location in respect with the center of the head. Instead of a spherical head, an ellipsoidal shape has also been proposed [8]. The B.E.M. (Boundary Element Method) brings another opportunity for physical modeling of the ITD but this approach has not

been thoroughly investigated [9]. The main advantage of physical modeling is its robustness.

The Woodworth's formula is based on a spherical head and is commonly used in commercial application. This formula gives a good approximation of the ITD in the zero elevation plane [10]:

$$ITD = \frac{a}{c}(\theta + \sin\theta) \quad (1)$$

where a is the sphere radius, θ is the azimuth angle (in the vertical-polar coordinates) and c the speed of sound. The ITD obtained particulary well matches the measured ITD[1] on a dummy head for high frequency range (f > 1.5 kHz) [11]. This formula has been extended to other elevation plane [12] [13]. Further improvement was provided by Algazi et al. who added individual scaling by estimating an individual radius from anthropometry [14]. This latter method has been used by the authors and will be named the *individualized Woodworth's* method in subsequent sections.

2.2. Signal processing

The signal processing based methods require HRTF measurements. They are all concerned with typical estimation errors (robustness and confidence). These methods can be divided into three groups, but it should be kept in mind that each group includes various competing formulas, which are not reported here.

Considering the ITD according to its physical definition, the *threshold* method determines the time when the Hear Related Impulse Response[2] (HRIR) reaches a percentage of its maximum value. The ITD is then the difference between the time estimated time for the right HRIR and for the left HRIR. Although this method seems to be as close as possible to the ITD definition, the values obtained are strongly dependent on the percentage value chosen to estimate the right and left delays. Furthermore difficulties arise for lateral positions because of the low signal to noise ratio for the contralateral HRIR. The 50 % value has been chosen which corresponds to a -3 dB threshold to avoid these difficulties.

[1] The ITD was mesured with an oscilloscope from phase shift through two microphones-amplifier channels

[2] The Head Related Impulse Response is the inverse Fourier Transform of the HRTF

The computation of the maximum of the InterAural Cross Correlation (IACC) function is based on the assumption that the central auditory system uses the correlation between left and right signal reaching the ears to achieve source identification and localization. This technique was first proposed by Whightman & al [15]. The ITD is estimated by determining the time when the result of the correlation between the right and the left HRIR reaches its maximum. It should be noticed that the method used by the authors does not consider the impulse response itself, but its envelope, in order to improve the estimate. In the following, this method will be referred to as the *IACCe* method. Although this approach is rather robust, it may suffer from the lack of coherence between the contralateral and the ipsilateral HRIR, which leads to misestimated values. This issue occurs mainly for source positions in the vicinity of the interaural axis.

The last group of signal processing methods comes from a linear fitting applied to the right and left excess phase component of the HRTF The right and left delay is derived from the slope of each fitting and the ITD is obtained by computing the difference between the two delays. This method, the *linear phase* one, is greatly dependent on the frequency range chosen for the linear fitting. As in [16], the frequency range used in the present study is [1000-5000 Hz].

Another phase method was recently introduced by Minnaar & al [17]. The ITD is derived from the group delay of the excess phase component of the HRTF The ITD value is computed as the interaural group delay difference evaluated at 0 Hz.

Fig.1 illustrates the ITD estimates of three methods: the *threshold* method, the *IACCE* method and the *linear phase* method. This figure depicts the average ITD for all subjects as a function of azimuth angle in degrees. Whereas the various methods agree in the median plane, noticeable departures rise as the source position approaches the interaural axis.

3. DESCRIPTION OF THE EXPERIMENT

3.1. Goal

The purpose of the experiment was to estimate the psychoacoustic value of the ITD. Psychoacoustic value is considered to be the ITD value which gives

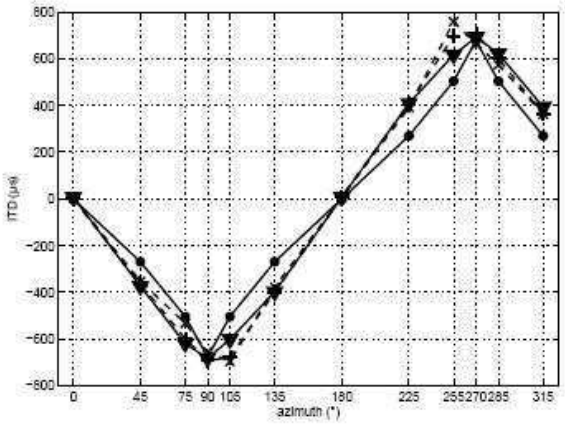


Fig. 1: Comparison between the various estimate methods of ITD All the values plotted are averaged for subjects participating in the control experiment. The estimate methods are: the *individualized Woodworth's* method [14] (solid line with circles), the *threshold* method based on 50 % maximum (solid line with triangles), the *linear phase* method (dotted line with crosses) and the *IACCE* method (dotted line with plus marks).

the best matching (in terms of sound localization) between the measured HRIR and its model composed of a F.I.R. minimum phase filter and a pure delay. Subjects listen to 2 sequential sounds: a target sound followed by a test sound. The target sound is obtained by convolving the sound stimulus by the raw HRIR, which is the measured HRIR after diffuse field equalization. The test sound results from the processing by the model, including minimum phase filter and pure delay. The task of subjects consists in adjusting the test sound delay until its perceived location matches the target. The time delay is varied by steps, which are an integer number of samples. The sampling frequency is $F_s = 44100$ Hz, which equates to an intersample time of $T_s = 23\ \mu s$. It should be noted that the HRIR were individualized, i.e. each subject listened to their own measured HRIR.

3.2. Control experiment

Preliminary experiments suggested that the matching task was difficult, especially for lateral positions. Therefore, it was decided to carry out a control experiment with a modified protocol, in order to estimate the reliability for the matching task. For the control condition, both the target sound and the test sound are processed by the HRIR model (i.e. minimum phase filter and pure delay). Then, the ITD value expected is the value applied to the target sound. Thus, the control experiment provides information about subject precision error for the ITD matching test.

3.3. Sound stimulus

The stimulus used was a gaussian noise burst of 20-ms duration low pass filtered [0-3000 Hz]. A 5 ms raised-cosine ramp was applied to the onset and offset. This frequency range was chosen to maximize the detectability of ITD which is a low frequency cue and to minimize the effect of interaural level difference (ILD) which is a high frequency cue.

3.4. Location of the virtual sound sources

The sound locations tested were restricted to the horizontal plane. Twelve azimuths were considered : 0°, 45°, 75°, 90°, 105°, 135°, 180°, 235°, 255°, 270°, 285°, 315°. Each location was repeated five times. The various locations were randomly presented.

3.5. Experiment protocol

The experimental protocol followed the principle of a modified adaptative method [18]. The starting pure delay was the ITD estimated by the *IACCe* method at the corresponding azimuth. In order to maximize the detection, a random shift is added to the starting value. Then, the ITD was varied as a function of the subject's response, according to an up-down method. A series of steps in one direction only is defined as a single run. For the first run, the ITD step is equal to five samples ($T_s = 23\ \mu s$). For the last run, it is one sample. The subject adjusts the ITD value by pressing one of two keys on the P.C. keyboard : one key makes the ITD increase and the other makes the ITD decrease.

Because of a potential lack of coherence while using an over-estimated ITD, the ITD value is restricted to the range: $[1.5 \times \tau_{min}$ - $1.5 \times \tau_{max}]$, where τ_{min} and τ_{max} are respectively the minimum and maximum ITD value estimated by the $I.A.C.C$ method. If the subject reaches the maximum number of runs

(which is fixed to six runs), the mean of the minimum and maximum values obtained for the second run is considered as the judgment's value [18]. The subject has also the possibility to validate a value if he(she) perceives no difference between the target and the test sound. Before beginning the experiment, the subject had a four-trial training session.

3.6. Subjects

21 subjects (6 females and 15 males) participated in the two experiments : 10 for the control experiment and 11 for the test experiment. Except for 2 subjects, each subject participated in only one experiment: test or control. The HRTF of the subjects belonged to two different databases: France Telecom R&D one (6 subjects) [9] and IRCAM one (15 subjects) [19]. An A.N.O.V.A. carried on the whole results did not show an effect of the HRTF database.

3.7. Experimental apparatus

Subjects listened through AKG K240 headphones and were located in an anechoic room with an ambient background noise level of 20 dBA. Convolutions, graphical interface and sound playing are processed with Matlab by a 700 Mhz Unix computer. The signal feeds a RME ADI-8 Pro sound card connected to a P2075 Yamaha amplifier. The listening level was 78 dBA for both sounds. Subjects had the possibility to have breaks as much as desired, but were reminded to have breaks every twenty judgments. The whole test lasts about 1 hour with a total of sixty judgments. For each trial, subjects heard four sounds : the target, the test sound a 500 ms pause and e repetitionof the two sounds. The test sound was always played after the target sound.

4. CONTROL EXPERIMENT

4.1. The question of the feasibility of the ITD matching task

Preliminary experiments showed that the ITD matching task was difficult to perform mainly for extreme lateral locations (i.e. for azimuth angle near $\theta = 90°$ and $270°$). There are several hypothesis to explain this finding. First, it is known that the contralateral HRTF is "weakly" minimum-phase[3] [4]. It may be illusory to try to match a raw HRIR with a model composed of minimum phase filter and ITD because of the lack of coherence. Indeed, it has already been pointed out that the HRIR model based on minimum-phase filter and delay fails for extreme locations [3]. Secondly, studies on the localization of real sources in free field report that, even for real sources, subjects have difficulties in localizing sound sources in this region [20] [21]. For these two reasons, we may wonder whether the ITD matching task is feasible for any location.

For the control condition, both the target sound and the test sound were processed by the HRIR model (i.e. minimum phase filter and pure delay). The expected ITD value was the delay value designed for the target sound. Subjects responses are compared to the defined target value. Besides, the standard deviation gives information about the judgment accuracy. Another question, that will be answered by the control experiment, is whether or not the ITD does exist for any locations, and even for extreme locations. If the subjects will succeed in reporting a reliable ITD for any locations in the horizontal plane, it will be a strong clue of the existence of the ITD. During the experiment, subjects responses were checked to ensure the ITD values reported by the subject stayed "realistic", i.e. the values should not exceed a maximum value defined as 150 % of the ITD target value. Whenever this case occurs, it is noticed and counted as an inconsistent answer. The number of inconsistent judgements is an additional clue for assessing the feasibility of the matching task.

4.2. ITD estimate

The subjects answers for the control experiment are shown in (see Fig.2). The ITD values reported by each subject match with the expected value. Indeed, the linear regression obtained for the data has a slope of 0.954. However, it should be noted that for lateral locations, subjects judgment were slightly less than ITD values.

The mean ITD reported by subjects and the standard deviation as a function of the azimuth angle is shown in Fig.3. The standard deviation depends on the azimuth angle. It is the lowest in the median plane and it increases with the lateralization of the virtual sound source. It is the highest for locations near to the interaural axis. This result suggests that the ITD matching task is harder for these locations.

[3] However, it should be mentioned that there are other locations, for which the HRIR is not minimum-phase, and yet the minimum-phase filter modeling is broadly validated and used.

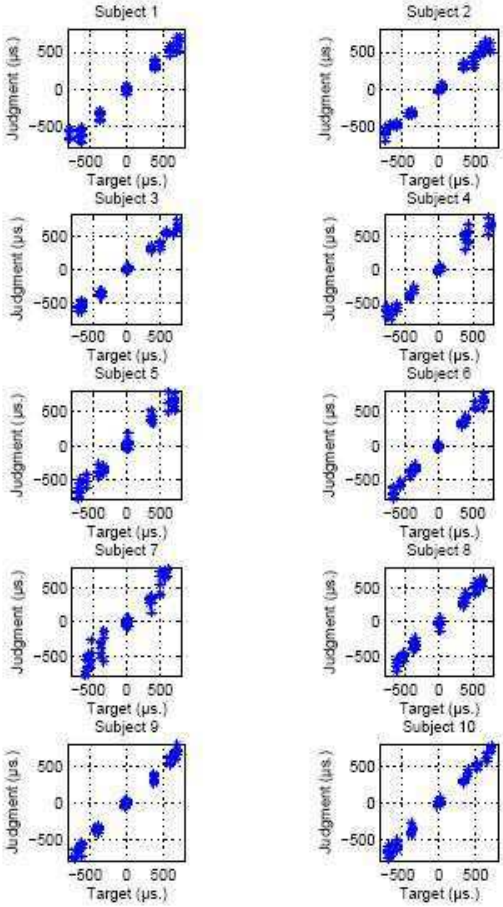


Fig. 2: Control experiment: For each subject, the subject's answer is plotted in function of the the target value.

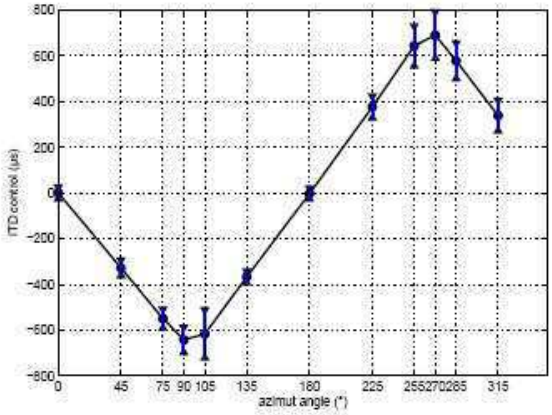


Fig. 3: Control experiment: Mean and standard deviation of the ITD estimate in function of the azimuth angle. The standard deviation is depicted by vertical bars.

4.3. Inconsistent judgments

The number of inconsistent answers for the whole control experiment were 1 % of the total implying that the test is reliable. Fig.4 depicts the percentage of inconsistent judgments as a function of azimuth. The number of inconsistent answers depends on azimuth. There were no inconsistent answers in the median plane, whereas the number of inconsistent answers is greatest near the interaural axis, which confirms that the ITD estimation for lateral locations is not so easy and somewhat problematical. The percentage of inconsistent judgments never exceeded 6 %. It was be noted that the inconsistency analysis results were not symmetrical about the median plane. It is higher (6 %) for the right hemisphere than for the left one (2 %). The subject's judgements can be considered as consistent for all azimuth angles.

4.4. Estimate error

The estimate error is defined as:

$$E(\theta) = \frac{1}{N}\sum_{i=1}^{N} |ITD(\theta,i) - \widehat{ITD}(\theta,i)| \qquad (2)$$

where the ITD value reported by the i^{th} subject is referred to as $\widehat{ITD}(\theta,i)$ and the ITD target value as $ITD(\theta,i)$ (θ is azimuth and N is the number of

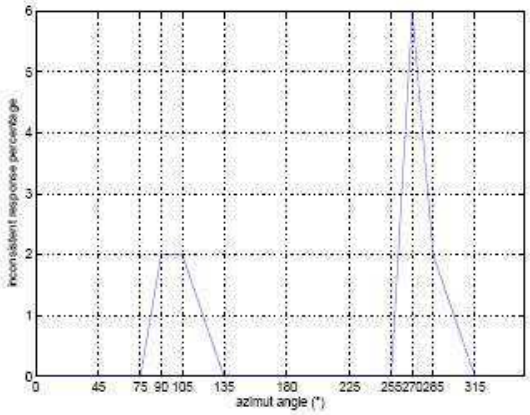


Fig. 4: Control experiment: Percentage of inconsistent judgments in function of the azimuth angle.

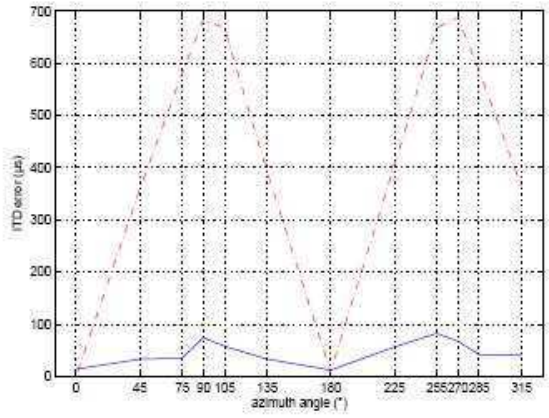


Fig. 5: Control experiment: Estimate error of the ITD (solid line) in function of azimuth. The ITD target (absolute value) is also plotted (- -) as a reference.

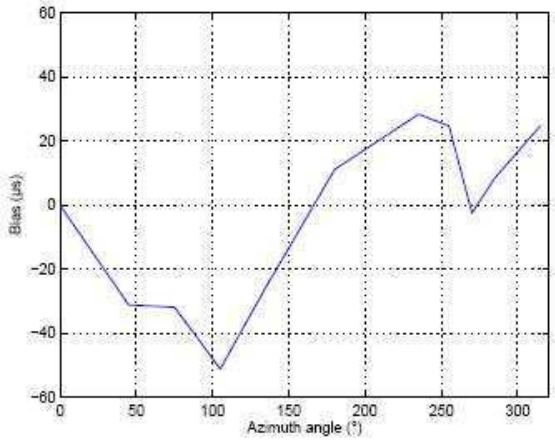


Fig. 6: Control experiment: Bias error of the ITD (—) as a function of azimuth.

subjects, i.e. N = 10). Fig.5 depicts the estimate error as a function of azimuth. The error strongly depends on azimuth. It is the lowest for locations in front of the listener (14 μs for $\theta = 0°$ and 12 μs for $\theta = 180°$) and the highest for lateral locations (74 μs for the left hemisphere and 82 μs for the right hemisphere). In the experiment, the ITD can be varied only by steps of one sample, which here is equal to 23 μs. Therefore, it can be concluded that in the median plane the judgment error is dominated by the experimental procedure. The estimation error for locations behind the listener is quite the same as for locations in front.

The estimate bias is also computed as :

$$B(\theta) = \frac{1}{N}\sum_{i=1}^{N} ITD(\theta) - \frac{1}{N}\sum_{i=1}^{N} \widehat{ITD}(\theta, i) \qquad (3)$$

Fig.6 describes the bias as a function of azimuth. As the error, the bias strongly depends on the azimuth. The bias is almost null in the median plane (0 μs for $\theta = 0°$ and 11 μs for $\theta = 180°$). It is the higher for lateral locations (-51 μs for the left hemisphere and 28 μs for the right hemisphere). It is again noted that the bias is not symmetrical about the median plane, i.e. the bias in the left hemisphere is higher than in the right hemisphere. The bias confirms that the psychoacoustics ITD value is less than the measured ITD value as shown before in Fig.2.

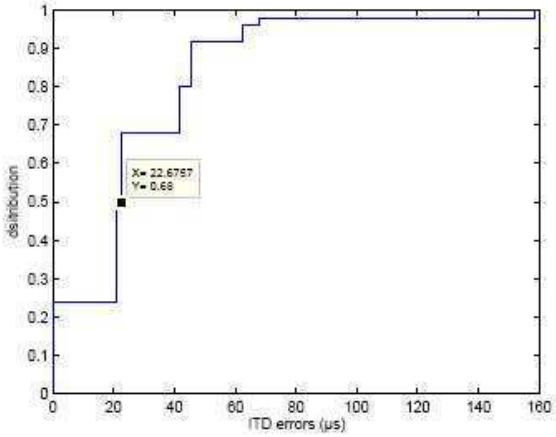


Fig. 7: Control experiment: Empirical cumulative distribution function for azimuth angle $\theta = 0°$.

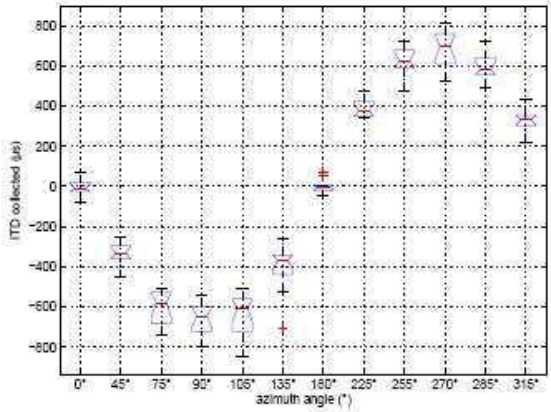


Fig. 8: ITD in μs for the test condition as a function of azimuth. The lower and upper lines of the "box" are the 25^{th} and 75^{th} percentiles of the sample. The line in the middle of the box is the sample median. The "whiskers" are lines extending above and below the box. The plus sign at the top of the plot is an indication of an outlier in the data. The notches in the box are a graphic confidence interval of the sample median.

In order to estimate the deviation of the subject answers, the empirical cumulative distribution function describing the repartition of the estimate error is computed from the experiment data. The ITD error value corresponding to a 50 % threshold of repartition is then derived from this function and compared with the thresholds obtained by [21] or [22]. Fig.7 shows the cumulative empirical distribution function and the estimate of 50 % threshold for azimuth $\theta = 0°$. For this position, the 50% threshold is approximately 23 μs. Mills reported a 50 % threshold of 10 μs for the frontal position. Despite great differences between the Mills experiment and the present work, thresholds are in good agreement.

4.5. Conclusion

We are able to make several conclusions from this study. Firstly, the protocol used was viable : subjects achieved the matching task in the control condition. Secondly, variations of the matching task error are in agreement with those reported in previous studies. The error is azimuth dependant and error is larger for lateral positions than for median positions. Thirdly, the minimum 50 % threshold estimated for the frontal position reaches the precision governed by the sampling rate ($\Delta ITDmin = 23\ \mu s$). The results of the control experiment ensure the validity of the experiment protocol.

5. RESULTS

Before analyzing the results, the consistency of the subject's judgments is checked. If the last value exceeds the range of valid ITD variation defined in Sec.3.5, this value is replaced by the mean of the other values obtained for the same azimuth angle. Only 17 over 660 judgments were thus corrected, which corresponds to 2.6 % of the total number of judgments.

An ANOVA carried on the data for the whole experiment (merging the results of the control and the test experiment) concludes that there is no effect of the experimental condition ($F(1, 176) = 0.3, p = 0.605$). Furthermore, a Tukey test shows that the difference between the two conditions is not significant. In other words, whether homogeneous stimulus (HRIR model vs HRIR model) or heterogeneous stimulus (raw HRIR vs HRIR model) are used, the matching task is successfully performed by the subjects.

An ANOVA was then performed on the results of the test experiment. As expected, the effect of azimuth is greatly significant ($F(11, 120) = 511, 6, p < 0.001$). Fig.8 depicts the statistical variations of sub-

jects' answers versus azimuth. First, it should be highlighted that a consistent value of ITD is obtained for all azimuth angles and that the dispersion is reasonable. Therefore, it can be concluded that the model (minimum phase filter and delay) is valid, even for lateral locations. Secondly, it can be seen that the closer to the interaural axis the source location is, the wider the interquartile is. However, it should be noticed that, for locations near the left ear, the ITD judgments exhibit more dispersion at 75° and 105° than for location at 90°.

It is also interesting to remark that the psychoacoustic ITD is neither symmetrical with respect to the interaural axis (for instance, the mean value for 75° is -608 μs, whereas the mean value for 105° is -647 μs), nor symmetrical with respect to the median plane (for example, the mean value for 90° is -657 μs whereas the mean value for 270° is -683 μs). This asymmetry is an additional criteria for the assessment of the estimate methods of the ITD. Indeed, the estimate methods which are able to reproduce the asymmetrical behavior can be considered as better estimate than those who don't. Indeed, it should be pointed out that the *threshold* and *IACCe* methods exhibit asymmetrical values (see Fig.1.

6. DISCUSSION

Fig.9 depicts the ITD values giving the best matching between a raw HRIR and its model composed of a minimum phase filter and a delay. For comparison, the ITD estimates which have been previously described, are also plotted. All the data plotted here are the mean value of the subject's answers and the subject's ITD estimates. The various estimate methods reproduce quite well the ITD variations with azimuth, except for the *individualized Woodworth's* method, which underestimates the psychoacoustic ITD. In [17], the Woodworth's model is reported as a good ITD estimate, at least as a model which fits remarkably well the interaural group delay difference estimate proposed by Minnaar et al. At first sight, this result is in contradiction with ours, but it should be kept in mind that Minnaar et al. compare two estimate methods, whereas our study assesses the Woodworth's model in comparison with the psychoacoustic ITD.

Fig.9 shows that the *threshold* and *IACCE* agree well with the psychoacoustic value, despite some de-

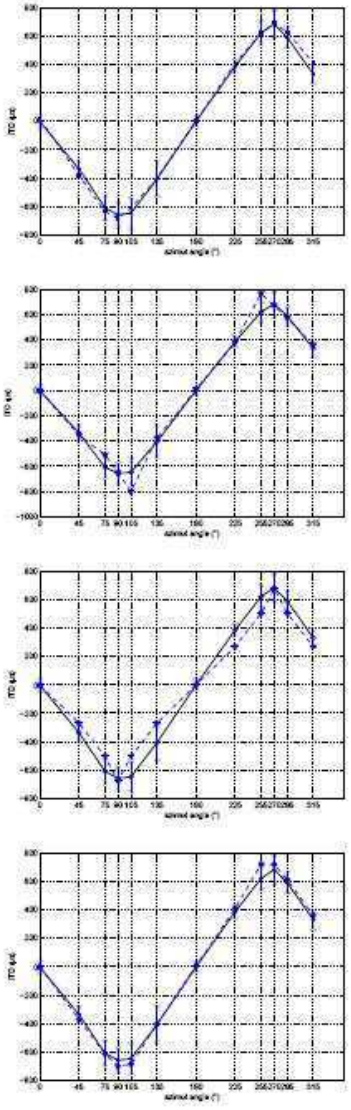


Fig. 9: Psychoacoustic responses versus estimation methods. Data plotted are means of subjects answers and subjects ITD estimations. Subjective ITD values versus azimuth are plotted with ± standard deviation bars. Methods from top to bottom are : threshold values, linear phase estimate, Woodworth ITD, and IACCE.

	Linear regression	Threshold	IACCE	Woodworth
EC1 (μs)	74	56	56	87
EC2 (%)	13.8	14.9	6.7	13.6

Table 1: Error Criterion.

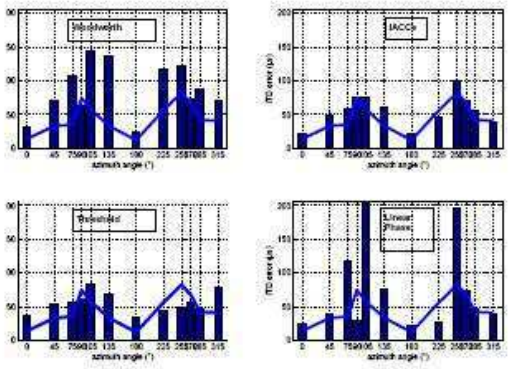


Fig. 10: Mean absolute error obtained for various methods of computing the ITD as a function of azimuth angle. Mean absolute error for the control condition is plotted on each bar chart as a reference.

viations for lateral positions for the *IACCE* method. On the contrary, the *linear phase* method departs from the psychoacoustic value quite more strongly for lateral locations. It should also be pointed out that both the *IACCE* and the *linear phase* methods depict maxima shift from the interaural axis positions ($\theta = 90°$ and $\theta = 180°$). This shift can be related to the listener's morphology. In [8], Duda & al explain that the deviations between the measured ITD and the Woodworth's model can be reduced by shifting the ear canal with respect to the interaural axis. Indeed, if the ear canal location is modified, the ITD maxima is shifted. Thus, for a location slightly toward the back hemisphere, the ITD maxima is pushed towards back hemisphere. As some asymmetrical values are observed in Sec.5 for the psychoacoustic ITD, we can wonder whether the estimate methods should be assessed on the basis of this criterium. In Sec.5, it has been remarked that, unlike the *threshold* method, the *linear phase* and the *IACCE* methods exhibits asymmetrical value.

To go deeply into the comparison of the prediction methods, an error criterion is defined as:

$$EC(\theta) = \frac{1}{N}\sum_{i=1}^{N} |ITD_{psych}(\theta,i) - \widehat{ITD}(\theta,i)| \quad (4)$$

where ITD_{psych} refers to the psychoacoustic value of the ITD and $\widehat{ITD}$ is the value estimated by one of the computational methods (i.e. the *individualized Woodworth's* method, the *threshold* method, the *IACCE* method or the *linear phase* method). This criterion is plotted as a function of azimuth in Fig.10. According to this criterion, the *threshold* method seems the closest to the psychoacoustic value. The mean error is computed as:

$$EC1 = \frac{1}{N \times M}\sum_{j=1}^{M}\sum_{i=1}^{N} |ITD_{psych}(\theta_j,i) - \widehat{ITD}(\theta_j,i)| \quad (5)$$

The results for the various methods are listed in Tab.1. The EC1 criterion is in good agreement with the observations from Fig.9. The mean absolute error between the psychoacoustic ITD and the computational estimate is higher for the *linear phase* and the *individualized Woodworth's* estimates than for the *threshold* and the *IACCE* estimates. In Fig.10, it is clearly seen that, unlike the other estimates, the *individualized Woodworth's* estimate leads to errors, which are distributed over all azimuths. Quite surprisingly, the error of the *linear phase* estimate is of the same order of the control condition error, excepted for lateral locations, which exhibit absolute errors as high as 200 μs. Thus, the poor results of the *linear phase* estimate pointed out by the EC1 criterion is mainly due to the failure of this method for lateral positions. Indeed, excess phase component of the contralateral HRTF for these locations are weakly linear phase. That's why a linear regression can lead to inconsistent delay estimate. In the

light of Fig.10, the *threshold* and the *IACCE* methods seem to provide the best estimate.

As EC1 and Fig.10 refer to mean values, which may flatten individual disparities, another criterion, EC2, is proposed in order to assess the various ITD estimates. Instead of considering strictly the mean value of the psychoacoustic estimate (ITD_{mean}), the standard deviation (SD) is taken into account, so as to define a range $[ITD_{mean} - SD, ITD_{mean} + SD]$. For each individual and each location, it is checked whether the value estimated by the computational methods belongs to this range. The criterion EC2 corresponds to the percentage of estimate values (for 660 responses collected), which do not belong to the range $[ITD_{mean} - SD, ITD_{mean} + SD]$. The criterion EC2 can be interpreted as the percentage of noticeable differences between the psychoacoustic ITD and the estimated one. Results are shown in Tab.1. Unlike EC1, the EC2 criterion leaves no doubt about the ITD estimate assessment. For the *IACCE* estimate, EC2 is less than one half of the EC2 values obtained for the other estimates. Therefore, it is concluded that the *IACCE* method provides the best estimate, not only in terms of the mean error, but also in terms of individualization.

7. CONCLUSION

Many conclusions can be drawn from the results of this study. First, it is shown that, despite many uncertainties about the validity of the modeling, the HRIR model composed of a minimum-phase filter and a pure delay works well for any locations in the horizontal plane. During the experiment, subjects succeed in matching raw HRIR with their phase minimum model. Secondly, a consistent value for the ITD is found for all azimuths, which proves the psychoacoustic validity of the ITD These results agree with those obtained by Plogsties [6]. The last goal of the experiment was to assess the values estimated by various computational methods of the ITD, in comparison with the psychoacoustic value. The results exhibit noticeable differences between the methods and the *IACCE* method is identified as the best estimate from a psychoacoustic point of view.

ACKNOWLEDGMENTS

The authors are very grateful to O. Warusfel, D. Presnitzer, E. Rio, G. Vandernoot, L. Gros and N. Bertet for their precious help. They would like also to thank all the subjects who participated in the experiment at IRCAM and at France Telecom R& D.

8. REFERENCES

[1] J. Blauert, "Spatial Hearing," The MIT Press, 1983.

[2] A. Kulkarni, S.K. Isabelle, H.S. Colburn, "On the minimum-phase approximation of head-related transfer functions", IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, 1995, pp. 84–87

[3] A. Kulkarni, S.K. Isabelle, H.S. Colburn, "Sensitivity of human subjects to head-related transfer function phase spectra", J.A.S.A., Vol. 105(5), pp. 2821-2840, (1999)

[4] C. Avendano, R.O. Duda, R. Algazi, "Modeling the contralateral HRTF", A.E.S. 16^{th} International Conference on Spatial Sound Reproduction, (1999 Apr.)

[5] V.R. Algazi, R.O. Duda, D.M. Thompson "The use of head-and-torso models for improved spatial sound synthesis", AES 113^{th} Convention, October (2002).

[6] J. Plogsties, S.K. Olesen, P. Minnaar, F. Christensen, H. Møller, "Audibility of all-pass components in head-related transfer functions", presented at the 108^{th} Audio Engineering Society Convention, convention preprint 5133 (2000 Feb.)

[7] Lord Rayleigh, "On our perception of sound dircetion," Philos. Mag. Vol. 13, pp. 214-232 (1907).

[8] R.O. Duda , C. Avendano and V.R. Algazi, "An adaptable ellipsodal Head-Model for the interaural time difference", *IEEE international conference on Acoustic, Speech and Signal Processing*, Phoenix Az, pp. (965-968), (1999).

[9] J.-M. Pernaux, "Spatialisation du son par les techniques binaurales : Application aux services de tlcommunications", *Ph-D of the Institut National Polytechnique de Grenoble*, May (2003).

[10] R.S. Woodworth, H. Schloesberg, *Experimental Psychology*, Holt, Rinehard and Winston, NY, pp. (348-361), (1962).

[11] G.F. Khun, "Model for the Interaural Time Difference in the Azimuthal Plane", *J. Acoust. Soc. of Am.*, Vol. **62**, No. **1**, pp. (157-167), July (1977).

[12] V. Larcher and J.-M. Jot, "Techniques d'interpolation de filtres audionumériques : Application à la reproduction spatiale des sons sur écouteurs", *Presented at the Congrès Français d'Acoustique*, Marseille, France, April (1999).

[13] L. Savioja, J. Huopaniemi, T. Lokki, R. Väänänen, "Creating interactive virtual acoustic environments", J.A.E.S., Vol. 47, No. 9, pp. 675-705, September (1999).

[14] V.R. Algazi, C. Avendano, R.O. Duda, "Estimation of a spherical Head-Model from anthropometry", *J. Audio. Eng. Soc.*, Vol. **49**, No. **3**, March (2001).

[15] D.J. Kistler, F.L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction", J.A.S.A., Vol. 91, pp. 1637-1647, (1992).

[16] J.-M. Jot, V. Larcher, O. Warusfel, "Digital signal processing issues in the context of binaural and transaural stereophony", 98th Convention of the Audio Engineering Society (Paris), preprint 3980, 1995.

[17] P. Minnaar, J. Plogsties, S. K. Olesen, F. Christensen, H. Moller, "The interaural time difference in binaural synthesis", presented at the 108^{th} Audio Engineering Society Convention, convention preprint 5133 (2000 Feb.)

[18] H. Levitt, "Transformed up-down methods in psychoacoustics", J. Acoust. Soc. Am, vol 49, pp 467-477 (1970 Feb.)

[19] Listen Project - Information Society Technologies Program - IST-1999-20646 :http://listen.gnd.de/ index.html http://recherche.ircam.fr/equipes/salles/listen/index.html

[20] J. Braasch, K. Hartung, "Localization in the presence of a distracter and reverberation in the frontal horizontal plane. I. Psychoacoustical data", Acta Acustica, Vol. 88, pp. 942-955, (2002)

[21] A.W. Mills, "On the minimum audible angle", J. Acoust. Soc. Am, vol 30, pp. 237–248 (1958 Apr.).

[22] R. Domnitz, "The interaural time jnd as a simultaneous function of interaural time and interaural amplitude", *J. Acoust. Soc. of Am.*, Vol. **53**, No. **6**, February (1973).

Just noticeable difference of the interaural time difference along cones of confusion

Sylvain Busson, Rozenn Nicol
Interfaces Sonores Innovantes, Laboratoire Speech and Sound Technologies and Processing, France Telecom R&D, 2 avenue Pierre Marzin, 22300 Lannion, France.

Olivier Warusfel
Institut de Recherche et Coordination Acoustique/Musique, 1 place Igor Stravinsky, 75004 Paris, France.

Laetitia Gros
Perception Multi-Sensorielle Cognition et Modélisation, Laboratoire Qualité et Valeur Perçue, France Telecom R&D, 2 avenue Pierre Marzin, 22300 Lannion, France.

ASA-PACS numbers : 43.66.Pn, 43.66.Mk

Suggested running title : JND of ITD in sagittal planes

ABSTRACT

Just noticeable differences (JND) of the interaural time difference (ITD) are measured for locations describing four cones of confusion. Wide band stimuli are played back to subjects using minimum phase individual and non-individual Head Related Transfer Function (HRTF). Two psychoacoustic protocols are compared : an adaptive 2 down-1 up 3AFC procedure and a 2AFC procedure. The lowest JND ([20 ; 40] μs) are obtained with the 2AFC procedure which exhibits thresholds close to those reported in previous studies in free field and in dichotic experiments. JND are larger for locations near to the interaural axis than in the front and are elevation independent. In the light of the JND, it is examined to what level of accuracy the ITD variation along a cone of confusion has to be modelled in the context of binaural synthesis. The ITD variation is not audible in the median plane and for low azimuth angles, whereas it may be audible near to the interaural axis. Two ITD models are perceptually assessed. The question of ITD individualization is also investigated. All these issues are analyzed by considering a whole HRTF database.

ASA-PACS numbers : 43.66.Pn, 43.66.Mk

I. INTRODUCTION

Head Related Transfer Functions (HRTF) embody all the necessary cues to render natural spatial sound through headphones. HRTF are determined by the individual morphology and the use of non-individualized HRTF to spatialize sound sources can lead to audible artefacts like front-back confusion or intracranial localization (Middlebrooks, 1999). A common implementation of HRTF is based on the decomposition into a minimum phase component and an excess phase component. It has been shown that the excess phase component can be replaced by a pure delay without audible artefacts (Kistler and Wightmann, 1992). The difference between the right and left pure delay gives the Interaural Time Difference (ITD) which is the primary cue for sound source localization (Rayleigh, 1906 ; Blauert 1974). Instead of estimating ITD from HRTF measurement, virtual auditory display (VAD) may also use mathematical formula to associate an ITD value with a location in space. The most often used formula is the Woodworth's formula (Woodworth and Schloesberg, 1962). The ITD (in seconds) is computed in the horizontal plane for a Spherical Head Model (SHM) with the entries of the ear canals diametrically placed :

$$ITD = \frac{a}{c}\left[\sin(\theta) + \theta\right] \tag{E.1}$$

where a is the head radius (m), c is the speed of sound (m/s) andθ is the azimuth of the source (rad).

The original formula is based on a head radius corresponding to an anthropometric mean (a = 8.75 cm). Algazi et al. have proposed to apply the Woodworth's formula with an individualized head radius which is derived experimentally from morphological data (Algazi, 2001 a). In addition, some studies have suggested improving this formula in order to account for the variation of the ITD with the elevation angle that are observed on subjects (see Fig. 1) (Larcher and Jot, 1999 ; Savioja, 1999). However, SHM fails for sound source locations along a cone of confusion. A cone of confusion is defined as the locus of all points for which the differences between distances to the left and right ear are the same, considering the simplest head model composed of only two points as two ears (see Fig. 2). This locus is described by a hyperboloid centred on the interaural axis, which may be approximated by the shell of a cone in the far field (Blauert, 1983 ; Shin-Cunningham *et al.,* 2000). The intersection of the cone of confusion with a sphere centred on the head are concentric circles distributed along the interaural axis (see Fig. 3, on the right). In interaural-polar coordinates, these circles are defined by a constant azimuth angle. The ITD given by the Woodworth's formula does not depend on the elevation angle and therefore is constant along a cone of confusion. On the contrary, the ITD estimated from individual HRTF measurement varies along a cone of confusion, as can be seen in Fig. 1. Near to the interaural axis, these variations can reach 18 % of the maximum value of the ITD, which corresponds to a shift of almost 15° in azimuth. These variations are mainly due to the shift of the ears according to the head centre and to the non-sphericity of the head (Duda *et al.*, 1999).

The present paper describes a subjective experiment which aims at determining whether the ITD variation along a cone of confusion is audible or not. Subjects are asked to detect variations of ITD for locations along various cones of confusion. For each location, the varied ITD (ITD_{var}) is judged in comparison with a reference ITD (ITD_{ref})which is defined as the ITD estimated from the measured HRTF of the subject according to the

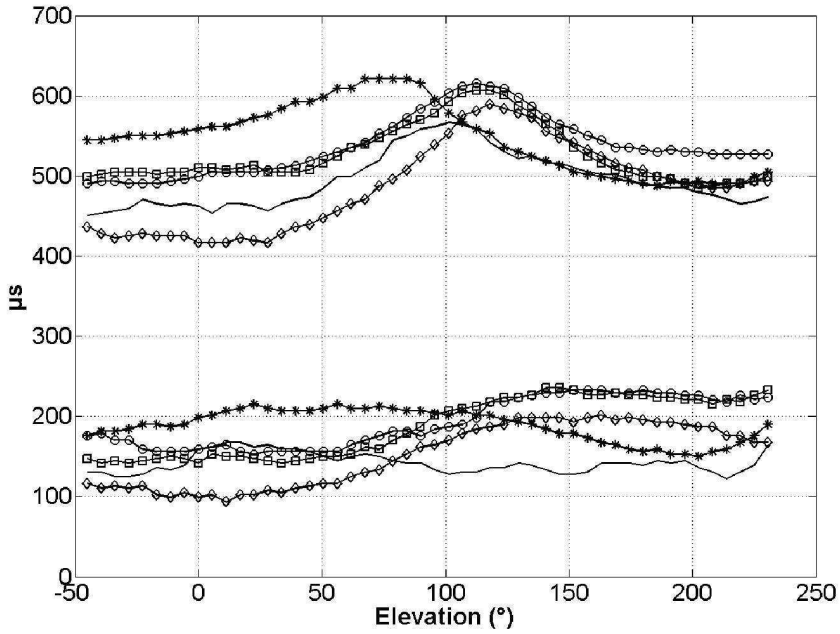


FIG. E.1 – Variation of the ITD (μs) of 5 subjects, taken from the CIPIC database, along two different cones of confusion. Lower curves : 20° azimuth cone, Upper curves : 65° azimuth cone. Azimuth and elevation are considered according to interaural-polar coordinates.

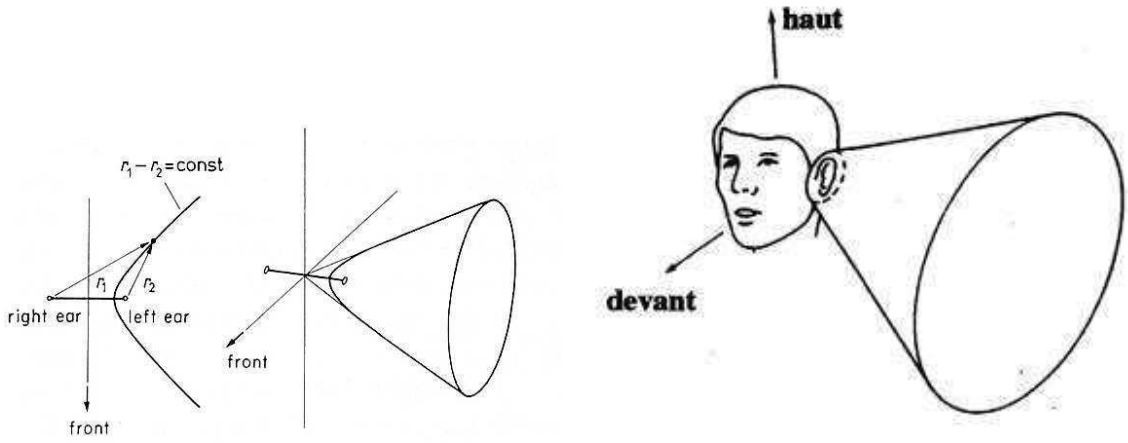


FIG. E.2 – Definition of a cone of confusion – For a simplified model of head [Blauert 1983] (on the left) and for a typical head (on the right).

method described in (Algazi et al., 2001 a)[1]. The Just Noticeable Difference (JND) of ITD is thus collected for various elevation along a set of cones of confusion. With this experiment, it is intended to assess to what extent the SHM is relevant for ITD modelling.

[1]The ITD estimation consists in using a threshold estimate method in the high frequency range in combination with a spherical harmonic expansion.

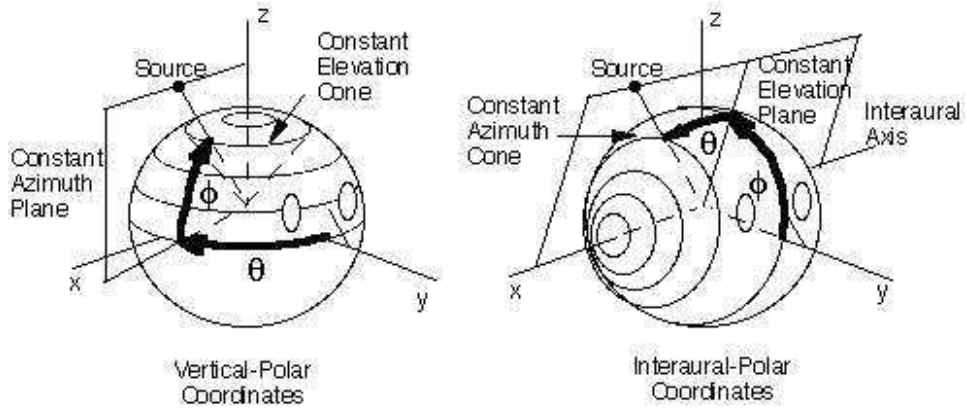


FIG. E.3 – The two coordinate systems commonly used in binaural synthesis – Vertical-polar coordinates (on the left) and interaural-polar coordinates (on the right). Taken from [Pernaux 2003].

If the results lead to the conclusion that the ITD variation is audible, the SHM must be discarded and replaced by a new model which accounts for this ITD variation. Many studies have given insight into the JND of ITD, but most of them consider either the free-field condition or the dichotic listening, which provides poor information about the JND of ITD for binaural synthesis, where spectral cues may disturb the temporal cues. The state of the art concerning the studies about the JND of ITD will be briefly recalled in part 2. Then the experiment will be described in part 3 and the JND values obtained for the various locations are presented in part 4. In the last part, the audibility of ITD variation along the cones of confusion is examined in the light of the JND of ITD. The requirement for individualized ITD variation is also investigated. Two ITD modelling are assessed : the SHM with individualized head radius (Algazi, 2001 a) and the Ear Displacement Formula (EDF) (Busson *et al.*, 2004), which allows shifting the ears with respect to the symmetry axis of the head in order to account for ITD variation along the cones of confusion.

II. PREVIOUS WORK

The JND of the ITD as well as the spatial resolution of the auditory system has been extensively studied. It can be useful to distinguish the studies considering the free-field sound localization (Mills, 1958 ; Blauert, 1983 ; Oldfield and Parker, 1984 ; Perrot and Saberi, 1990) from those concerning the stereophonic rendering over headphones (Klump and Eady, 1956, Hershkowitz and Durlach, 1969 ; Domnitz, 1973 ; Tolling and Henning, 1998 ; Litovsky *et al.*, 2000] and those focussing on binaural synthesis, where the stimulus is filtered by HRTF (Ericsson and McKinley, 1990 ; Grantham *et al.*, 2003 ; Best *et al.,* 2004). Free-field and stereophonic headphone listening give JND values of the same order. It is also observed that the JND of the ITD increases with the azimuth angle, i.e. the JND value is the lowest in the median plan. Some studies have even reported the impossibility of obtaining a JND value for highly lateralized positions, i.e. near to the interaural axis (Mills, 1958).

In (Klump and Eady, 1956) the authors have evaluated the JND of the ITD for various types of stimuli. The subject listens to two sequential sounds and he (she) is asked to

determine which one he (she) localizes more to the left. The results highlight how the JND depends on the stimulus for noises of different bandwidth, for a single click or trains of clicks, for pure tones with frequencies going from 90 Hz to 3200 Hz and for different ITD_{ref}. The lowest JND, equal to 9 μs, is obtained for ITD_{ref}= 0 μs and for a noise with bandwidth [100 - 1700 Hz]. In the study on the Minimum Audible Angle (MAA), Mills (1958) describes a free field experiment where the subjects judge the relative position of two sequentially presented stimuli. Although free field listening conditions also contain an interaural level difference (ILD) which provides an additional localization cue, this study reports a minimum JND[2] of 10 μs for a pure tone of 750 Hz. In (Domnitz and Colburn, 1977) the study is focussed on the combined effect of the ITD and the ILD on the JND of the ITD. The author uses a 500 Hz - 300 ms tone burst. The method of two Alternatives Forced Choice (2AFC) with feedback of correct answer is applied. The lowest threshold obtained is 10 μs. This threshold is also confirmed by Zurek (1985) for broadband stimuli, by Henning (1974) for a harmonic sound modulated in amplitude. This value agrees also with the threshold obtained by Hafter and Maio (1975) for clicks of [100 - 2000 Hz] frequency range (2AFC procedure with correct answer feedback). Even with non-individualized binaural synthesis, a 16 μs JND is obtained in (Grantham et al., 2003). The experiment presented in (Best, 2004) gives brief answer in the case of concurrent sources where an average JND of 50 μs is extracted. The authors explain that it may be more difficult to extract from a concurrent presentation of stimuli than from a sequential one. All these results are listed in Table I.

Tab. E.1 – JND value of the ITD in the median plane according to various experiments.

Reference	**Listening condition**	**JND value of the ITD in the frontal position (az = 0° el =0°)**
[Mills 1958]	free field	10 μs
[Klump & Eady 1956]	stereophonic headphones	9 μs
[Domnitz 1977]	stereophonic headphones	10 μs
[Grantham & al. 2003]	binaural synthesis	16 μs
[Best & al 2004]	Individualized binaural synthesis. Concurrent presentation	50 μs

Fig. 1 illustrates the ITD as a function of the elevation angle for the 20° and 65° cone of confusion. Several questions arise from Fig. 1 :

– Is the variation of ITD along a cone of confusion audible or not ?

[2]The relation between MAA and JND is explained in [Mills 1958]. It is based on HRTF measurement with an artificial head.

– If the variation of ITD along a cone of confusion is not audible, is the ITD given by a spherical head (Algazi *et al.*, 2001 a) sufficient from a psychoacoustic point of view ?
– Are the individual differences of the ITD along a cone of confusion audible or not ?

In the present paper it is intended to investigate the auditory discrimination of ITD along the cones of confusion for binaural synthesis. The objective is to adjust and optimize the accuracy of ITD modelling or rendering in VAD by taking into account the properties of auditory perception. Our experiment aims at evaluating the JND of the ITD for sequential listening, which is relevant for the simulation of spatialized trajectories. In addition, as preliminary experiments show deviations in collected data from one protocol to another, psychoacoustic procedures for JND evaluation will be compared.

III. EXPERIMENT

The choice of the experimental protocol is a critical issue. Preliminary experiments were carried on in order to compare various protocols. These experiments focused on the JND value of the ITD without any additional localization cue such as spectral cues or ILD. The JND values collected by the various protocols showed a great variance, which suggests that the experiment protocol is another experimental factor which must be taken into account. The two protocols, which are considered in the present study, correspond respectively to the ones that provided the highest (protocol 1) and the lowest (protocol 2) threshold estimates during the preliminary experiments.

A. Stimuli

Each sound stimulus is computed using the convolution of a 400 ms Gaussian noise burst (10 ms cosine-squared onset-offset ramp) and the appropriate delayed minimum phase Head Related Impulse Response (HRIR) corresponding to the tested location. The two different stimuli are the stimulus delayed by ITD_{ref} (ITD estimated by the method described in (Algazi *et al.*, 2001 a)) and the stimulus delayed by a modified ITD value : $\mathrm{ITD}_{var} = \mathrm{ITD}_{ref} + \Delta\mathrm{ITD}$. Fine varying of ITD is obtained by oversampling, which is preferred to fractional delay in order to avoid audible artefacts (Laakso et al, 1996). The HRIR is oversampled at a sampling rate of 96 kHz, which allows varying the ITD in steps of about 10 μs ($\Delta\mathrm{ITD}_{min}$= 10,41 μs).

B. Location of the virtual sound source

The JND value of the ITD will be evaluated as a function of azimuth and elevation angles. It is expected that the JND value increases with the azimuth of the cone and for high elevation angles. The 965 measurement locations of the HRTF database used in the experiment cover the elevation angle range [-56° ; 90°] and the azimuth range [0° ; 355°] in vertical-polar coordinates (see coordinates (see Fig. 3)). The azimuth sampling is constant for a given constant elevation plane and varies with elevation angle (Pernaux, 2003). This spatial sampling is not well suited for cones of confusion defined as constant azimuth planes in interaural-polar coordinates (Fig. 3). Therefore, the locations selected to represent a given cone do not exactly match the cone of confusion, but deviate slightly (by a few degrees) from the ideal trajectory. The error introduced is of the same order as the error due to subject's head movement during measurements (Algazi *et al.*, 1999). Furthermore, as the task is to judge ITD_{var} in comparison with ITD_{ref}, this deviation is not disturbing. Fig. 4 shows the distribution of the measurement locations and of the selected locations for the listening test. For convenience, the interaural-polar coordinates will be used in the next parts and the tested locations distributed along a cone of

confusion will be referred to as the azimuth angle of the cone. confusion are considered corresponding respectively to 0°, 22°, 61° and -61° azimuth angle. They are referred to as 0° cone, 22° reasons as previously, it was not possible to have the same elevation angle for each cone. The 26 locations selected are listed in Table II. Fig. 5 illustrates the axis references for a better understanding of the azimuth and elevation angle.

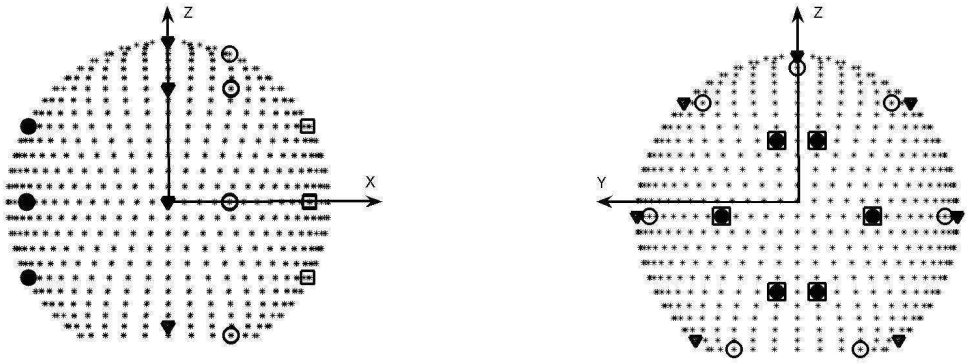


Fig. E.4 – Locations of the measured HRTF (∗) and of the selected HRTF corresponding to the cones of confusion considered in the experiment – Front view (left) and side view (right). Tested locations for the 0° cone are depicted with triangles, for the 22° cone with empty circles, for the 61° cone with empty squares and for the -61° cone with plain circles.

Tab. E.2 – Azimuth and elevation angle of the tested locations in vertical-polar coordinates.

Azimuth Cone	**low elevation**	**eye-level front**	**high elevation front**	**zenith**	**high elevation behind**	**eye-level behind**	**low elevation behind**
0°	-50.5°	0°	45°	90°	135°	180°	230.5°
22°	-56°	0°	45°	67.5°	135°	180°	236°
61°	-28°	0°	28°	-	118°	180°	208°
-61°	-28°	0°	28°	-	118°	180°	208°

C. Protocols

Protocol 1 is an adaptive method based on a 2down-1up three Alternative Forced Choice (3AFC) with correct answer feedback (Levitt, 1971). Each trial is composed of three stimuli : two are identical, one is different. The subject is asked to identify which of the three stimuli is different. The test is stopped either when the maximum number of reversals is reached, which is fixed to six reversals in our experiment, or when

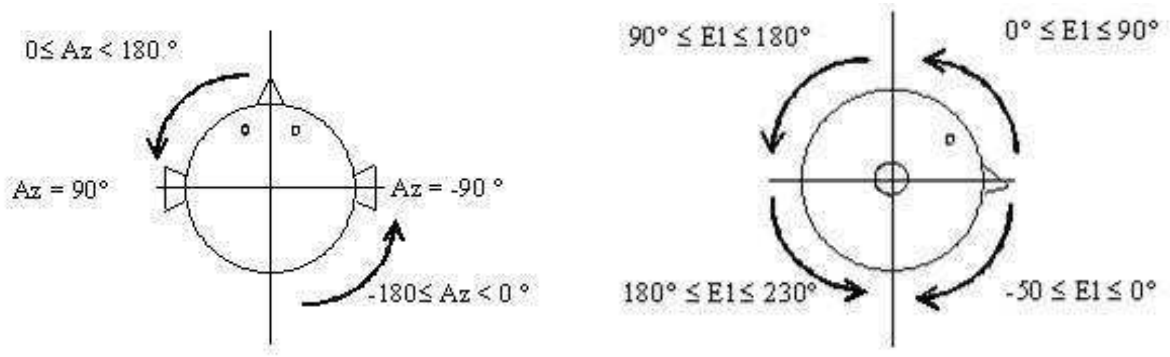


FIG. E.5 – Axis reference for the azimuth and elevation angle used for the description of the location of the virtual sound source.

the minimum difference (ΔITD_{min}) is identified by the subject (two consecutive correct answers). The first step is equal to four ΔITD_{min} and decreases by step of ΔITD_{min} for each new run. The JND value is computed as the midpoint of every second run. This estimate of the JND value corresponds to 70.7 % of right answers (Levitt, 1971). The subject is trained along four independent tests. He (she) can make a pause whenever he (she) wants. Each ITD difference is tested three times. There are two conditions for each position depending on the sign of ΔITD. Indeed, it is reported that the JND value decreases when the ITD variation interacts in the same way[3] as the ILD cue (Domnitz and Colburn, 1977). Therefore, for protocol 1, the subject listens to 156 trials (26 locations x 2 ITD variations x 3 repetitions).

Protocol 2 is a non-adaptive process based on a two Alternative Forced Choice (2AFC) method. Each trial is composed of two stimuli : one with the reference ITD, the other with the modified ITD. A set of ten values of ITD variation from zero sample (i.e. : $\text{ITD}_{var} = \text{ITD}_{ref}$) to nine samples is considered. The subject is asked to identify which of the stimuli is on the left. Each pair of the same stimuli is presented 30 times. The JND value is estimated as the minimum ITD difference for which the subject's answers are 75% right. According to the χ^2-test, the 75% level is significantly different from the chance level (50%) for 30 repetitions. For protocol 2, only the cones 0°, 20° and 61° are considered since the symmetry is studied in trials (20 locations x 10 ITD thresholds x 30 repetitions), which locations x 10 ITD thresholds x 30 repetitions), which are divided into two blocks.

In the two protocols, a correct answer feedback is given to the subject after each trial and all the experimental conditions are randomly presented.

D. Subjects

Five subjects, two men and three women, take part in the experiment using protocol 1. It is checked before the experiment that they have normal hearing. Subjects 1 to 4 listen to binaural rendering with their own HRTF, while the HRTF of another individual are used for subject 5. Only subjects 4 and 5 take part in the experiment with protocol 2.

[3]For example, for a location of 60° azimuth, a negative ΔITD moves the sound localization towards the centre, which counteracts the ILD cue which indicates a location on the right, whereas a positive ΔITD enhances the lateralization of the virtual sound source.

D. Equipment

The subject is seated in an anechoic room with a cut-off frequency of 45 Hz. The stimuli as well as the visual interface is computed with MATLAB®. The virtual sound sources are rendered via a TERRATEC EWS-88 MT® sound card (sampling rate : 96 kHz). The subject listens over a SENNHEISER HD650® headphone via a MCA Audio U4410® headphone amplifier. The sound level is 75 dBA measured on an artificial head B&K HATS 4128C for the frontal position, i.e. azimuth angle 0° and elevation angle 0°.

IV. RESULTS

The first objective of the experiment is to estimate the JND value of ITD in the context of binaural synthesis. Three experimental factors are considered : the cone azimuth, the elevation angle along a cone, and the way of ITD variation (i.e. increasing or decreasing ITD). The second objective of the experiment is to investigate to what extent and how the JND value of ITD is influenced by these experimental factors. The results obtained according to experimental protocol 1 (cf. Sec. III) will first be described. Then the results of protocol 1 and 2 will be compared.

An ANalysis Of VAriance (ANOVA) conducted on the individual JND's considering the experimental factors cone, elevation and way of variation, shows that the only significant effect is the cone azimuth ($F(3,192) = 4.44$, $p < 0.05$). Thus, neither the elevation angle along a cone ($F(5,192) = 0.51$, $p = 0.76$), nor the way of ITD variation ($F(1,192) = 0.01$, $p = 0.93$) has a significant effect. It is striking that the JND value of ITD does not depend on elevation, which is also suggested by the study of Oldfied and Parker (1984). Previous work has highlighted that the ITD and the ILD may interact as localization cues (Domnitz and Colburn, 1977 ; Blauert, 1983). Nevertheless, in the present experiment, the JND value of ITD does not depend on whether the ITD variation counteracts or not the ILD.

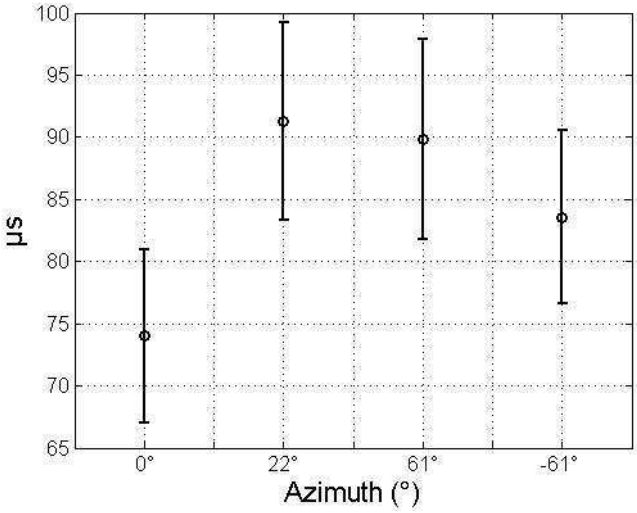


FIG. E.6 – Mean JND value of ITD (μs) in function of the cone azimuth (°) – Mean for all the elevation angles, the 5 subjects and the 2 ways of ITD variation, the corresponding 95% confidence interval is also plotted as an error bar.

Fig. 6 depicts the mean JND value of ITD as a function of the cone azimuth. For

each cone azimuth, the mean value pulls together the data from all elevation angles and the two ways of ITD variation, since the ANOVA points out that these factors have no significant effect. The confidence interval is almost the same for all the cone azimuths. It varies between ± 7 and ± 8 μs, which is quite low. Moreover, it should be reminded that, in this experiment, the ITD is varied in step of one sample. The confidence interval is thus very close to the sample period (10.4 μs for a sample rate of 96 kHz). It can be noticed that the JND value of ITD is the lowest (74 μs) in the median plane (i.e. cone 0°). The JND values of the other cones do not significantly differ from each other (according to a Honestly Significant Difference Tukey Test). In addition, there is no significant difference between the JND value of ITD for the cone 61° and for the cone -61°, which shows the symmetry of JND with respect to the median plane. The JND values of ITD and their corresponding confidence intervals are given in Table III. It is striking that the JND values obtained in our experiment are strongly higher than the thresholds given by previous work which reports JND values as low as 10 μs in the median plane (see Sec. II).

Tab. E.3 – Mean JND value of ITD (protocol 1) and 95% confidence interval (μs)

Subject / **Cone**		**1**	**2**	**3**	**4**	**5**	**Mean**
0°	**Mean**	62	71	114	55	67	74
	95% Confidence Interval	± 8.6	± 17.4	± 22.1	± 8.1	± 10.1	± 7
22°	**Mean**	88	81	138	73	76	91
	95% Confidence Interval	± 10.1	± 22.5	± 22.2	±9.8	±11.3	± 7.9
61°	**Mean**	83	89	103	79	95	90
	95% Confidence Interval	± 15.9	± 18.7	±19.9	± 13.4	±20.7	± 8
-61°	**Mean**	76	83	118	63	77	84
	95% Confidence Interval	± 11.8	± 20.8	±14.5	± 7.7	±14.4	± 7

To go into the details of individual trends, the mean thresholds obtained by each individual are plotted for the 5 subjects in Fig. 7. Numerical values are also listed in Table III. Noticeable differences between the subjects are remarked as reported in (Domnitz, 1973 ; Best *et al.*, 2004). Most of the subjects show JND value within the range [60 ; 100 μs], except for subject 3 who reports JND greater than 100 μs. It should be highlighted that subject 1, 4 and 5 are experts in 3D audio, whereas subjects 2 and 3 are not, which may explain the high thresholds obtained by subject 3. Moreover, contrary to the other four subjects, subject 5 did not use his own HRTF, but the HRTF of another individual. It is noteworthy that, despite non-individualized HRTF, subject 5 shows results which are very close to subject 1 and 4.

Fig. 8 compares the JND value of ITD obtained by subject 4 and subject 5 according to the two experimental protocols, for the three common cone azimuths 0°, 22° and 61°.

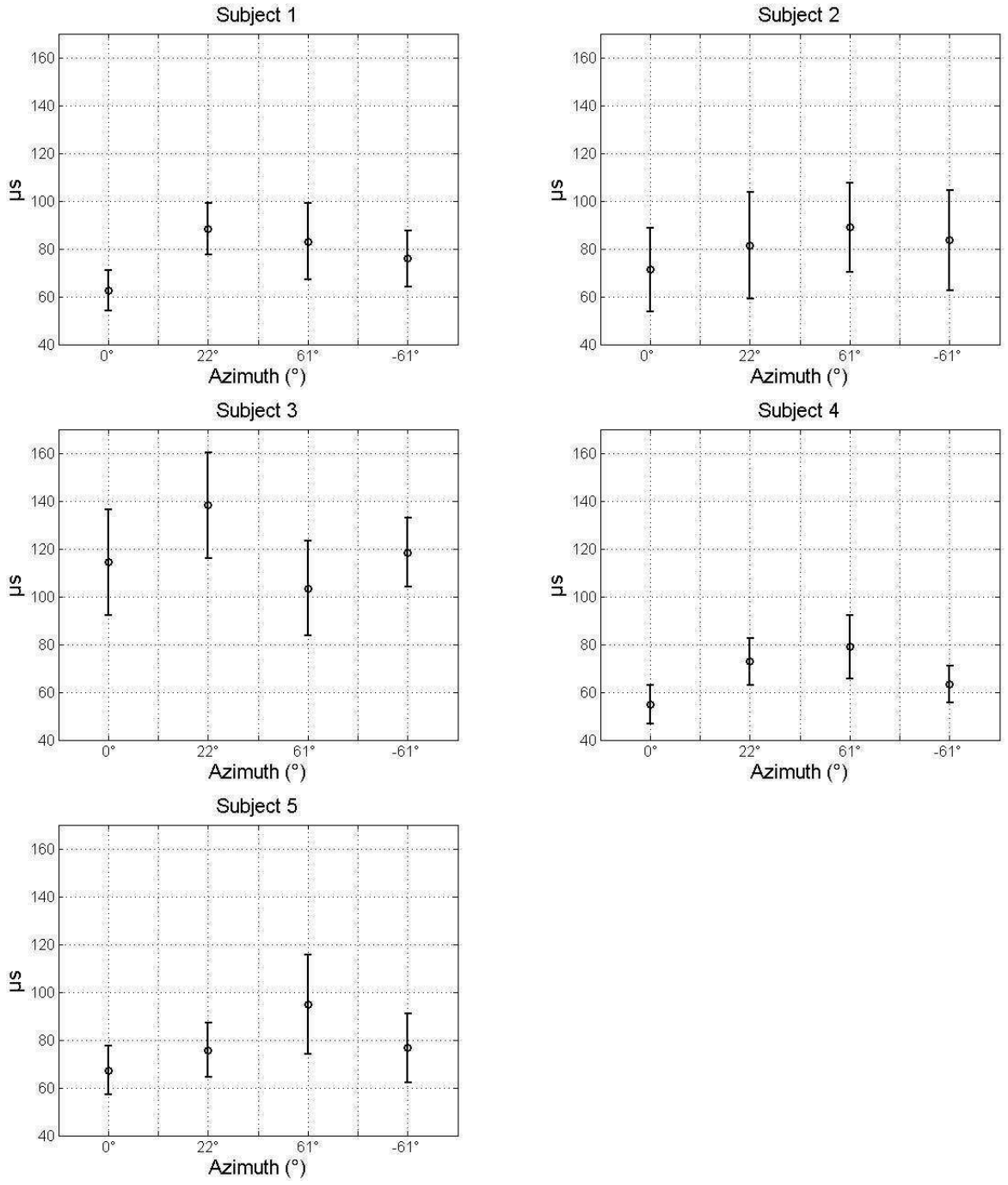


FIG. E.7 – Mean JND value of ITD (μs) in function of the cone azimuth (°) for the 5 subjects – Mean for all the elevation angles and the 2 ways of ITD variation, the corresponding 95% confidence interval is also plotted as an error bar.

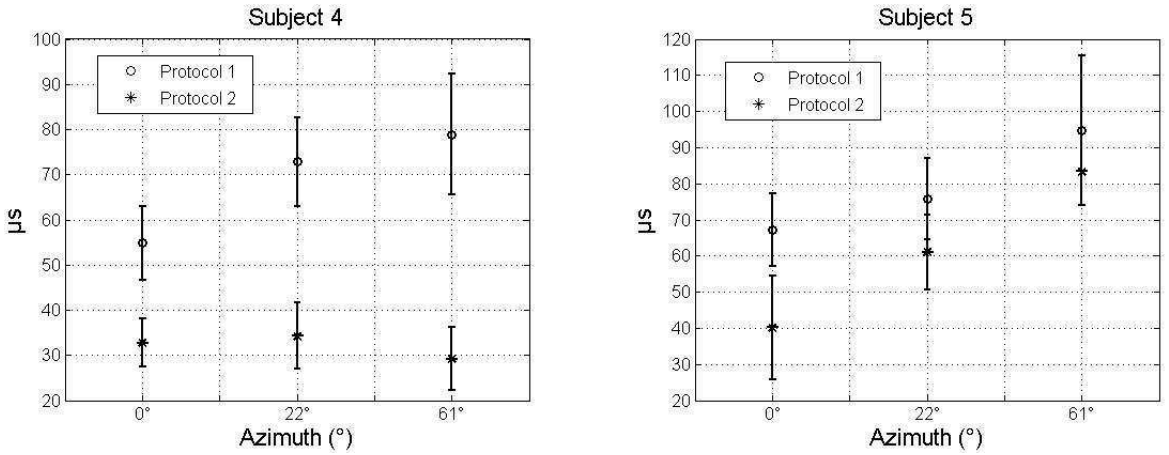


FIG. E.8 – Comparison between protocol 1 and 2 for subject 4 (on the left) and 5 (on the right) - Mean JND value of ITD (μs) in function of the cone azimuth (°) – Mean for all the elevation angles and the 2 ways of ITD variation, the corresponding 95% confidence interval is also plotted as an error bar.

The JND values for all locations according to the two protocols are detailed in Fig. 9. As expected, the JND values of ITD provided by protocol 2 are protocol 2 are globally lower than by protocol 1. Particularly subject 4 exhibits JND two times lower than in protocol 1, which leads to JND values within the range [20 – 40 μs]. These thresholds are closer to the values reported by previous work (see Sec. II). It should also be noticed that for subject 4 the JND from protocol 2 does not depend on the cone azimuth. The comparison between subject 4 and 5 shows discrepancies. Firstly, the decrease of JND for subject 5 is less obvious. Secondly, the JND increases with the cone azimuth for subject 5 in the same way as in protocol 1. Besides, subject 5 exhibits poor judgment for the 61° cone : his discrimination of the ITD variation is not level. Some studies report that JND are difficult to obtain for highly lateralized locations (Mills 1958, Grantham *et al.*, 2003). Subject 5 uses non-individual HRTF, which may lead to front / back confusions (Middlebrooks, 1999). These confusions could make the task more difficult. In the next sections only threshold for subject 4 are considered when dealing with results from protocol 2.

The previous results point out strong discrepancies between the 2 protocols. A thorough investigation of the two experimental procedures is beyond the scope of the present paper. The following is only an attempt to analyze and understand the differences between protocol 1 and 2. Protocol 1 is an adaptive procedure, unlike protocol 2. The advantages of an adaptive procedure are manifold (Trahiotis *et al.*, 1990). First, it is not required to know in advance the range of threshold values, since the experiment process will automatically adapt in order to reach these values. On the contrary, in protocol 2, a set of arbitrary values is tested, which implies preliminary experiments in order to determine the range of threshold values. Second, in protocol 2 the threshold is derived from the psychometric curve, which requires a high number of trials, whereas protocol 1 is based on a quick and reliable estimate of the threshold (Levitt, 1971). The consequence is that protocol 1 leads to a shorter experiment than protocol 2. In our case, protocol 1 lasts 2 hours, whereas protocol 2 lasts up to 7 hours.

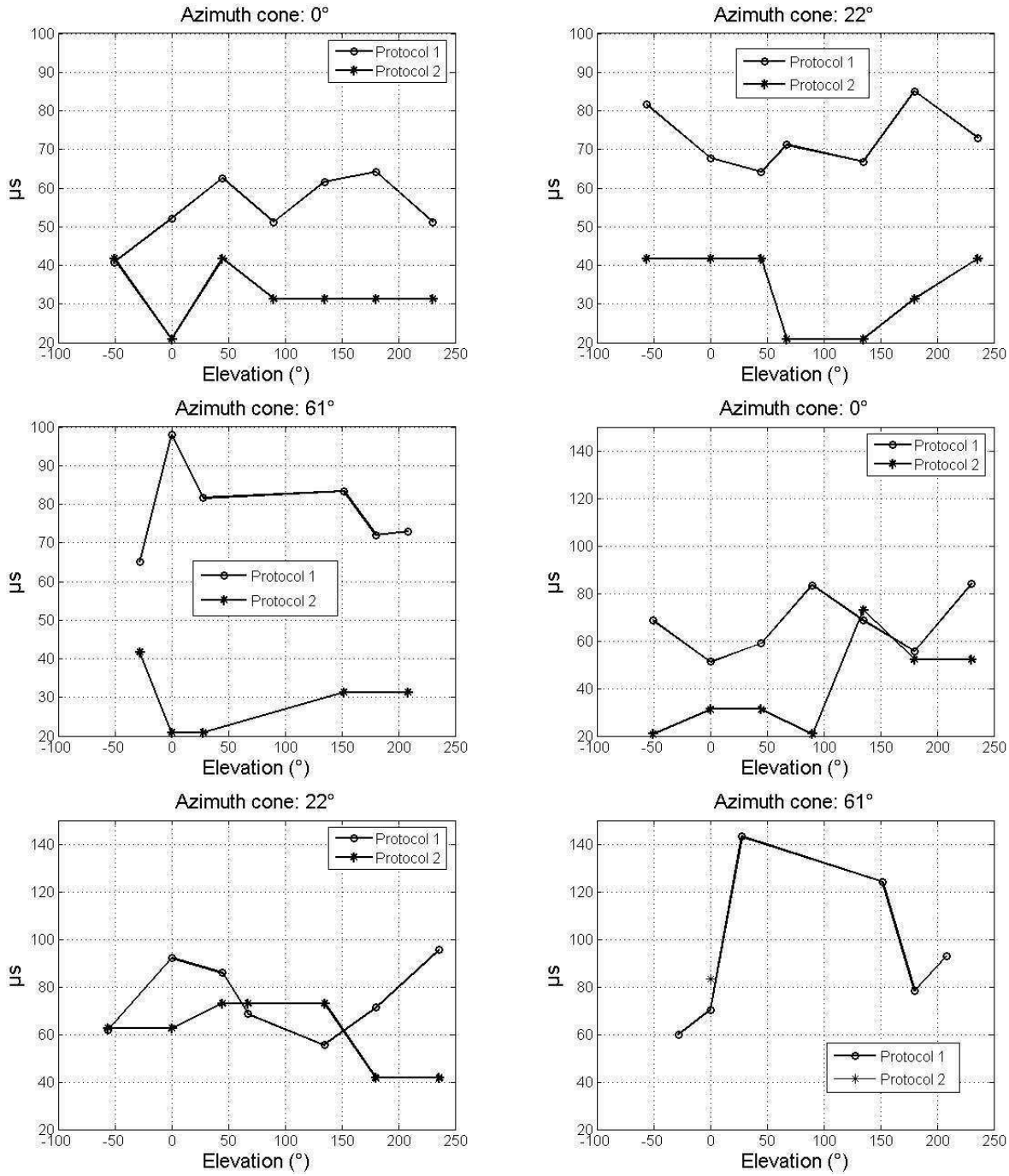


FIG. E.9 – Comparison between protocol 1 and 2 for subject 4 (on the left) and 5 (on the right) - JND value of ITD (μs) for the 20 locations.

However the adaptive procedure may have one drawback. Indeed, it is often reported that thousands of trials are needed to perform discrimination task in a binaural listening test (Trahiotis *et al.*, 1990). In an adaptive procedure without intensive training, the subject will listen to small differences only if he (she) gives a continuous series of correct answers. Thus the subject may not listen to little differences enough to learn the subtle

cues to answer. This could be a reason for higher thresholds in protocol 1. In addition, the authors have made some observations during the experiments which are reported here. Protocol 1 is more sensitive to a lack of attention of the subject than protocol 2. Indeed, a mistake of the subject affects the remaining runs of the trial in protocol 1, whereas it has only a local effect in protocol 2. Another fact is that the subject's task may be more complex in protocol 1 (identifying the dissimilar stimulus among three, without any possibility to replay the sequence) than in protocol 2 (identifying the stimulus which is most on left). A learning effect was also noticed for protocol 2 : because of the high number of trials, the subject learns to sharpen his (her) auditory perception. Indeed, for protocol 2, the experiment was divided into two blocks and it was observed that the JND values of the first part were higher than those of the second part. This learning effect suggests that protocol 2 leads to the lowest threshold that could be expected. Protocol 1 seems to be closer to average performance. For psychoacoustic knowledge or specific task as fighter pilot training, the lowest threshold is required, but for binaural synthesis dedicated to public application, as museum audio visit, the average performance should be sufficient.

V. Discussion

The JND value defines an ITD blur, i.e. the range in which individual ITD values cannot be discriminated from one to another. From the knowledge of this ITD blur, the audibility of individual variation of ITD along cones of confusion will be examined. It is intended to determine to what degree of accuracy the ITD should be modelled in binaural synthesis for VAD. To that respect, all the fine details of ITD variation, which are not perceived according to the JND blur, may be considered as perceptually useless and can be omitted, which may lead to simplified and cost efficient modelling of ITD. Two ITD models will be assessed here : the SHM with individualized head radius (Algazi, 2001 a) and the Ear Displacement Formula (EDF) (Busson *et al.*, 2004), which allows shifting the ears with respect to the symmetry axis of the head in order to account for ITD variation along the cones of confusion. Another advantage of these ITD modellings is that they both provide an easy way to individualize ITD without HRTF measurement by using only anthropometric data. This question of ITD individualization will also be investigated in the light of the JND blur. All these issues will be analysed by considering the CIPIC HRTF database (Algazi, 2001 b), which gives a high number of subjects.

– **Audibility of ITD variation**

The ITD data, which will be used in the following study, are taken from the CIPIC HRTF database. The analysis is focussed on three cones of confusion corresponding to the azimuth angles 0°, 20° and 65°. The data obtained for 43[4] subjects are considered. The first question aims at assessing whether the ITD variation along a cone of confusion is audible or not. In order to quantify the variability of ITD along a cone of confusion, the following criterion of "ITD deviation[5]" $e_i(\theta, \varphi)$ is defined :

$$e_i(\theta, \varphi) = \frac{ITD_i(\theta, \varphi) - ITD_i(\theta, 0)}{ITD_{i_{\max}}} \tag{E.2}$$

where $ITD_i(\theta, \varphi)$ denotes the individualized ITD of azimuth angle θ and elevation angle

[4]The CIPIC HRTF database comprises 45 subjects, but a careful examination of the ITD estimated for each subject shows that the results given subjects are somewhat anomalous. These subjects are discarded from the present study.

[5]However it should be noticed that this criterion of ITD deviation is not a standard deviation.

φ for the ith subject. $ITD_{i_{\max}}$is the maximum value reached by the ITD of the ith subject along the cone of azimuth $\theta = 80°$. This criterion quantifies the deviation of ITD around a reference value which is arbitrarily chosen as the chosen as the ITD at elevation $\varphi = 0°$. The criterion is normalized by the maximum $ITD_{i_{\max}}$of the subject in non-dimensional value which can be compared and merged with the results of the other subjects. For direct comparison with the ITD deviation criterion, the JND value is also normalized by the mean value of the maximum $ITD_{i_{\max}}$of all the subjects :

$$JND_{norm}(\theta) = \frac{JND(\theta)}{\frac{1}{N}\sum_{i=1}^{N} ITD_{i_{\max}}} \tag{E.3}$$

where N is the total number of subjects.

The ITD deviation criterion is computed for each subject and for each location. For one cone of azimuth, a total number of 2150 data (43 subjects x 50 elevations) is collected. Instead of conventional statistical tools (i.e. mean and standard deviation for instance), it is preferred to analyze this data set in terms of probability distribution, because it gives thorough insight into the distribution of the values obtained by the criterion among all the individuals. The empirical cumulative distribution function is thus estimated for the three cones 0°, 20° and (see Fig. 10). This function denoted by F(e) gives the probability that the criterion is less than or equal to the value e. It is examined then where the JND value intersects the cumulative distribution function. The corresponding probability is interpreted as the probability that the ITD variation is not audible. The probability obtained for the three cones and according to the JND of the two protocols is listed in Table IV. For JND1 (Protocol 1), the ITD variation is little audible. For 0° and 20°, the probability of non-audibility is higher than 0.9. than 0.9. However, for 65°, the probability decreases to 0.84. the probability strongly decreases (near to 0.5 for 0°) according to JND2 (Protocol 2). For 65°, the probability is as low as 0.27. Concerning the audibility of ITD variation, it can be concluded that near to the median plane it is very little audible. it is more audible for lateral locations and it may be audible in applications requiring a high degree of accuracy or for expert listeners.

Tab. E.4 – Audibility of ITD variation along cones of confusion

	Probablility that : e $\leq$ JND (normalized)	
Protocol	**1**	**2**
Cone		
0°	0.93	0.47
20°	0.99	0.67
65°	0.84	0.27

– **Individualization of ITD variation**

There are two issues concerning the audibility of ITD variation. First, does the accuracy of the auditory system allow one individual to perceive the ITD variation along a cone of confusion ? This question was examined in the previous paragraph. It is shown that the ITD variation is partly audible. However it should be reminded that the ITD variation along a cone of confusion depends on the individual (see Fig. 1). It is now

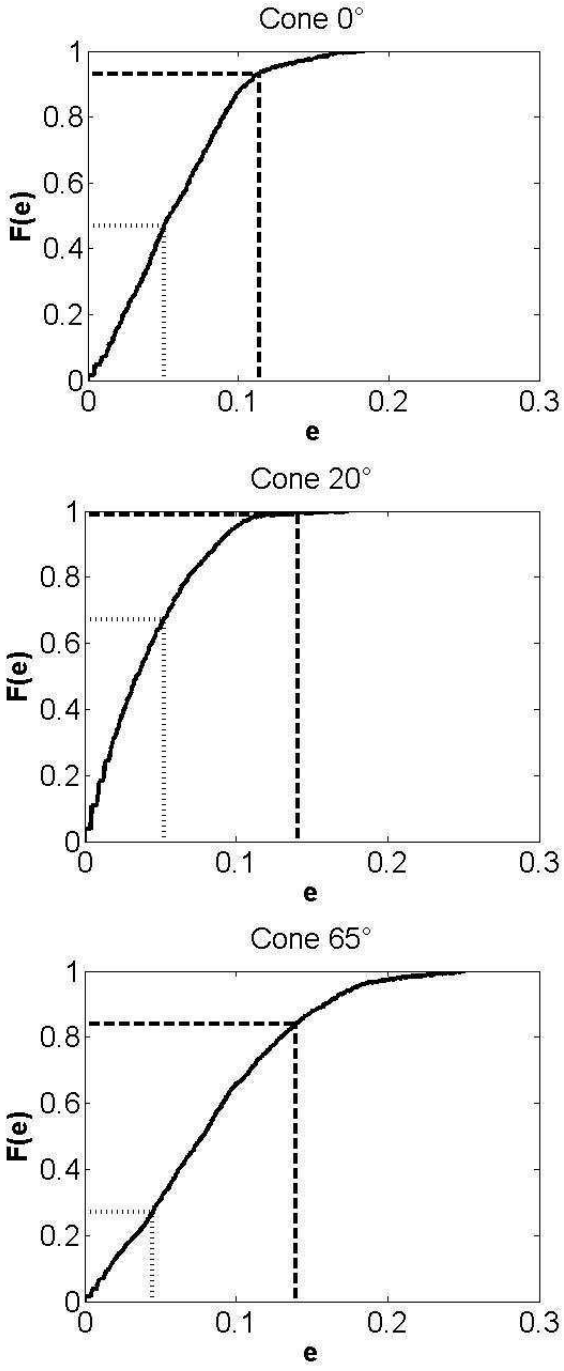


FIG. E.10 – Audibility of ITD variation : Cumulative distribution function derived from the criterion $e_i(\theta, \varphi)$ (CIPIC HRTF database, azimuth cone : 0°, 20° and 65°). The JND value is plotted as vertical dashed line (dark line : JND value of 1, light line : JND value of Protocol 2).

of interest to investigate whether the auditory system is able to discriminate the ITD variation of two individuals. Since ITD variation along a cone of confusion is useful, is it required to individualize the ITD variation or is it possible to use a mean variation of ITD ?

In order to answer this question, the mean ITD of the subjects of the CIPIC HRTF database is computed for each location :

$$ITD_{mean}(\theta,\varphi) = \frac{1}{N}\sum_{i=1}^{N} ITD_i(\theta,\varphi) \tag{E.4}$$

Then it is examined for each subject whether he (she) is able to discriminate this mean ITD from his (her) individualized ITD in terms of the JND of ITD. The criterion defined for the audibility of ITD variation is thus modified in order to account for the differences between the individualized ITD and the mean ITD :

$$e_{ind_i}(\theta,\varphi) = \frac{ITD_i(\theta,\varphi) - ITD_{mean}(\theta,\varphi)}{ITD_{i_{\max}}} \tag{E.5}$$

This criterion is computed for all the subjects and all the locations. In the same way as previously, these data are analysed in terms of cumulative distribution function for each cone of confusion. These functions are plotted in Fig. 11. The probability corresponding to the JND value is given in Table V. According to JND1, the mean variation of ITD is not discriminated from the individualized one, since the probability of non audibility is very close to 1. If JND2 is considered, the probability of non audibility decreases to 0.78 in the median plane and 0.64 for the 65° cone, which is still quite high. The requirement of individualizing the ITD variation is less strong than the requirement of ITD variation itself. However the individualization of ITD variation may be useful for lateral location in high quality context.

Tab. E.5 – Audibility of individual variation of ITD along cones of confusion

	Probablility that : $e_{ind} \leq$ JND (normalized)	
Protocol **Cone**	**1**	**2**
0°	0.99	0.78
20°	1	0.72
65°	0.99	0.64

– **Perceptual assessment of ITD modelling : SHM and EDF**

It has been shown that ITD variation and individualized variation are audible in some cases. This result highlights the interest of assessing various ITD modelling corresponding to increasing level of accuracy. Two models will be compared : the SHM with individualized head radius (Algazi, 2001 a) and the Ear Displacement Formula (EDF) (Busson *et al.*, 2004). The SHM provides no ITD variation along a cone of confusion, whereas the EDF gives individualized variation of ITD. Indeed this latter consists in shifting the ears with respect to the symmetry axis of the head. The optimal shift is computed from least square fit so that the modelled ITD best matches the ITD extracted from the measured HRTF for a given cone of confusion, which is chosen as the 65° cone in our study. An example of the ITD obtained by SHM and EDF is depicted for one individual in Fig. 12. In the same way as the perceptual difference between the mean variation of ITD and the

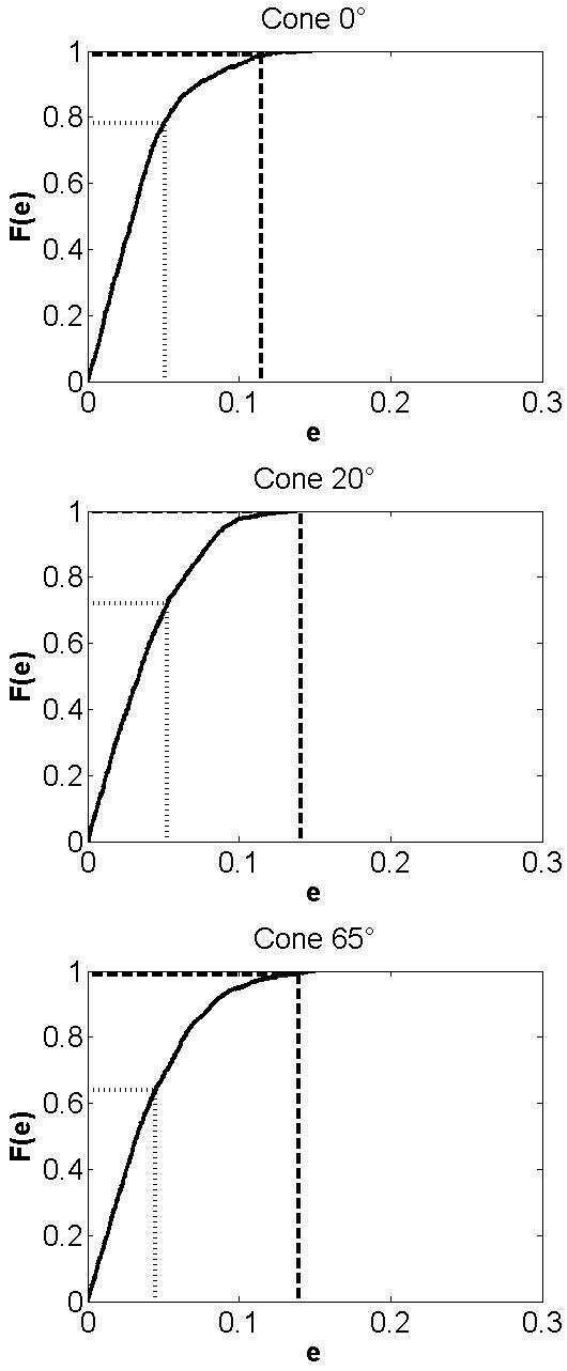


FIG. E.11 – Individualization of ITD variation : Cumulative distribution function derived from the criterion $e_{ind_i}(\theta, \varphi)$ (CIPIC HRTF database, azimuth cone : 0°, 20° and 65°). The JND value is plotted as vertical dashed line (dark line : JND value of Protocol 1, light line : JND value of Protocol 2).

individualized one was previously investigated, the following criteria are defined for the assessment of SHM and EDF :

$$e_{mod_i}(\theta, \varphi) = \frac{ITD_i(\theta, \varphi) - ITD_{mod}(\theta, 0)}{ITD_{i_{\max}}} \quad mod = SHM, EDF \tag{E.6}$$

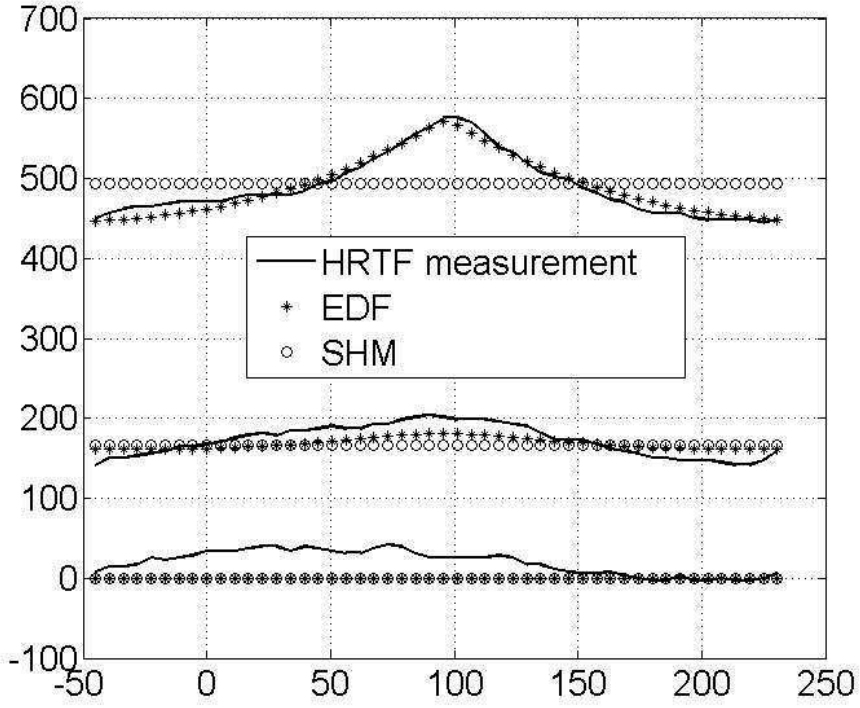


FIG. E.12 – ITD modelling : Comparison between SHM and EDF for one subject of the CIPIC HRTF database for the 0°, 20° and 65° cones (ITD estimate from HRTF measurement : ___ , ITD computed by SHM : ooo , ITD computed by EDF : ***).

The cumulative distribution functions estimated from the data collected for the two models are plotted in Fig. 13 and 14. The probability corresponding to the JND value is given in Table VI. According to JND1, both SHM and EDF are perceptually equivalent to ITD estimate from HRTF measurement. However, EDF stands out when the results are examined in terms of JND2. Indeed, whereas the probability of non audibility decreases to 0.49 for SHM for the 65° cone, EDF reaches a probability of 0.96. For the 65° cone, the ITD computed by EDF is not discriminated from the ITD even according to JND2. This result is not surprising since the parameters of EDF are optimized for this cone, but it is striking that EDF succeeds noticeably well in individualizing the ITD variation for lateral locations, where SHM gives poor modelling. modelling. Moreover, for the 0 and 20° cones, the probability greater than 0.7. Therefore ITD modelling by EDF perceptually matches well the measured ITD.

In addition, the ITD modelling by applying EDF with a set of parameters optimized for the mean ITD[6] is assessed. The associate cumulative distribution function and the probability values are given in Fig. 15 and Table VI. The results are very are very close to SHM except for the 65° cone where EDF with parameters matches better the measured ITD than SHM, but worse than EDF with individualized parameters. **VI.**

[6]The parameters are the following : the head radius is 91 mm, the ears are shifted of 3° in azimuth toward back hemisphere and 19° in elevation toward up hemisphere with respect to the interaural axis.

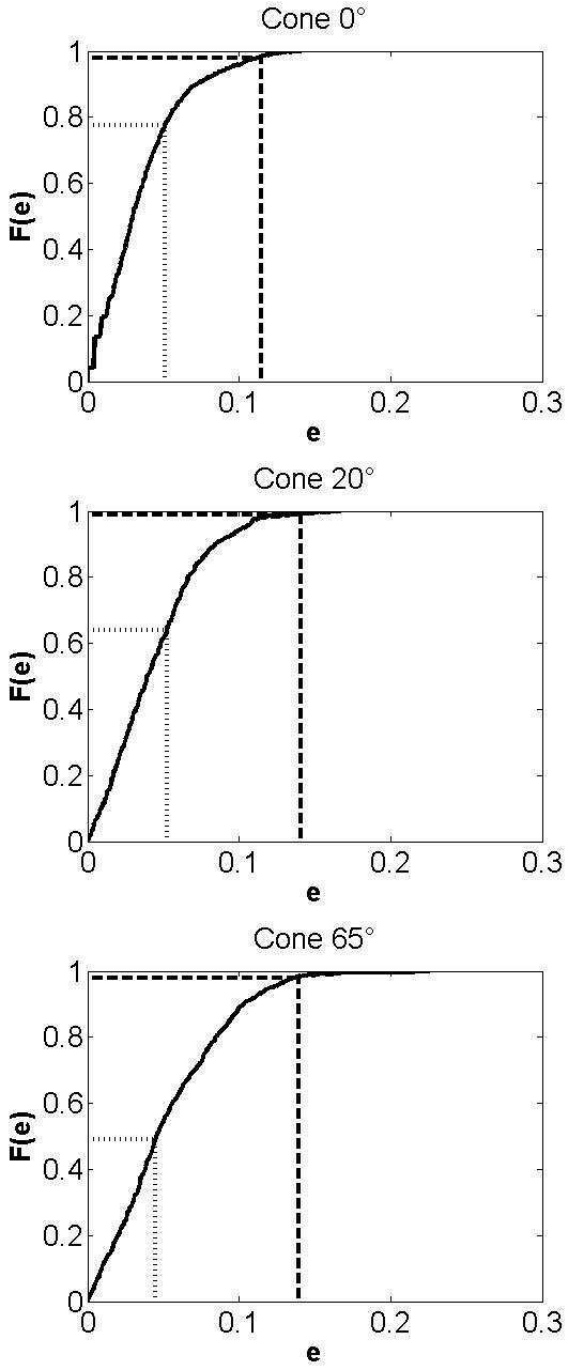


FIG. E.13 – Perceptual assessment of ITD modelling by SHM : Cumulative distribution function derived from the criterion $e_{SHM_i}(\theta, \varphi)$ (CIPIC HRTF database, azimuth cone : 0°, 20° and 65°). The JND value is plotted as vertical dashed line (dark line : JND value of Protocol 1, light line : JND value of Protocol 2).

CONCLUSION

This paper describes psychoacoustic experiments investigating the JND value of ITD along the cones of confusion for binaural synthesis. A first experiment using an adaptive 3-AFC procedure gives JND values within the range [74 – 91 μs] (JND1), which is

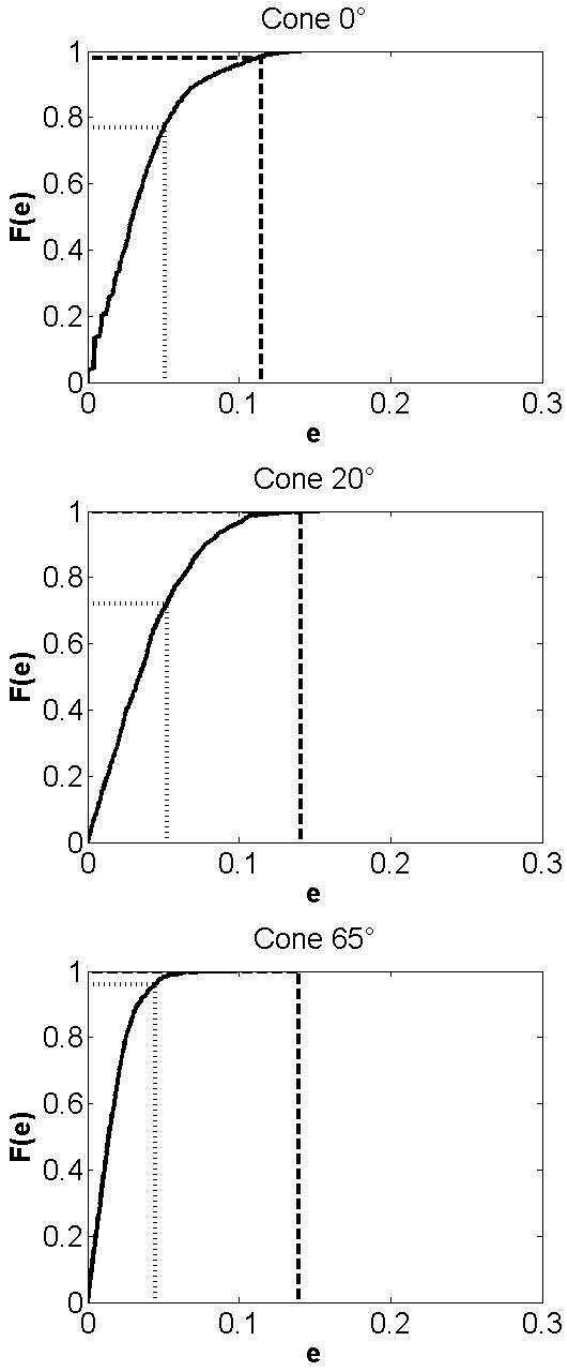


FIG. E.14 – Perceptual assessment of ITD modelling by EDF : Cumulative distribution function derived from the criterion $e_{EDF_i}(\theta, \varphi)$ (CIPIC HRTF database, azimuth cone : 0°, 20° and 65°). The JND value is plotted as vertical dashed line (dark line : JND value of Protocol 1, light line : JND value of Protocol 2).

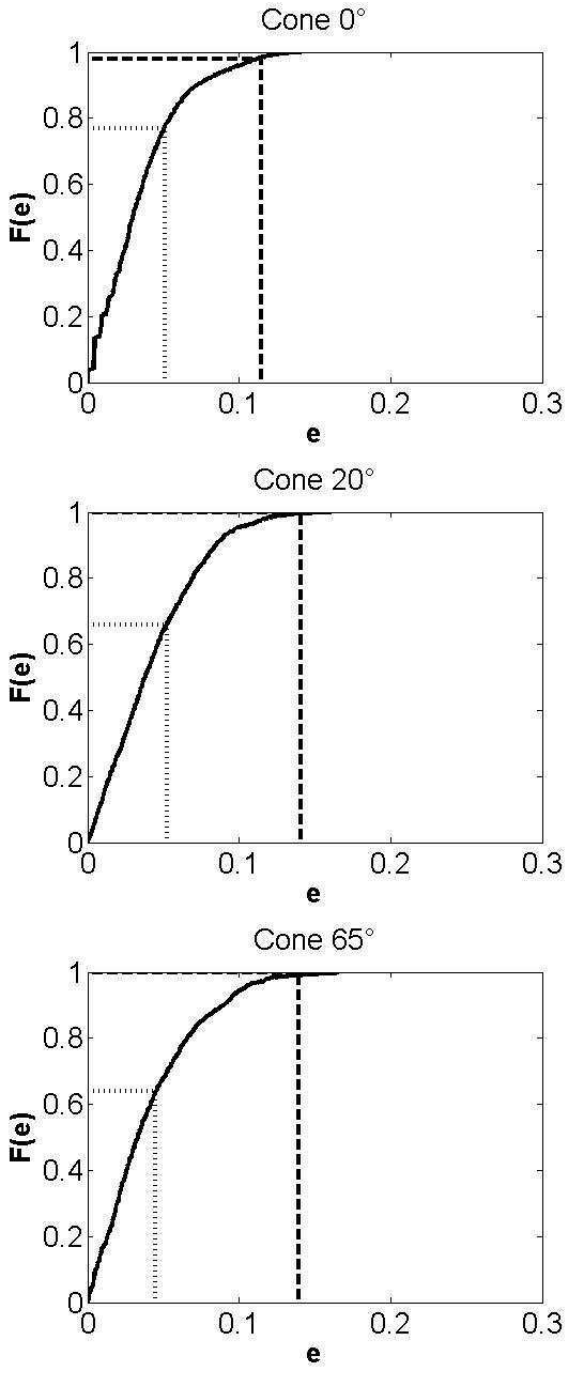


FIG. E.15 – Perceptual assessment of ITD modelling by EDF applied with the parameters optimized for the mean ITD : Cumulative distribution function derived from the criterion $e_{EDFopt_i}(\theta, \varphi)$ (CIPIC HRTF database, azimuth cone : 0°, 20° and 65°). The JND value is plotted as vertical dashed line (dark line : JND value of Protocol 1, light line : JND value of Protocol 2).

Tab. E.6 – Perceptual validity of ITD modelling (SHM, EDF)

SHM	**Probablility that :** $e_{SHM} \leq$ **JND (normalized)**	
Protocol	**1**	**2**
Cone		
0°	0.98	0.775
20°	0.99	0.64
65°	0.98	0.49
EDF	**Probablility that :** $e_{EDF} \leq$ **JND (normalized)**	
Protocol	**1**	**2**
Cone		
0°	0.98	0.77
20°	1	0.72
65°	1	0.96
EDF mean	**Probablility that :** $e_{EDFmean} \leq$ **JND (normalized)**	
Protocol	**1**	**2**
Cone		
0°	0.98	0.77
20°	1	0.66
65°	1	0.64

rather high in comparison with the literature. A second experiment based on a non-adaptive 2-AFC procedure reports JND values two times lower [20 – 40 μs] (JND2). Further investigation and new experiments would be required to fully understand this discrepancy between the two experiment protocols, but this issue is beyond the scope of the present paper.

The objective of the study is to infer from the JND value of ITD the level of accuracy required for the modelling and the individualization of the ITD. According to JND2 ([20 – 40 μs]), the ITD variation along the cones of confusion are audible for lateral locations (65° cone), which motivates ITD modelling which accounts for ITD variation. The EDF, which provides ITD variation by shifting the ears with respect to the symmetry axis of the head, is one solution. Two ITD modelling, EDF and SHM, have been assessed in the light of the JND value of ITD. It is shown that EDF succeeds very well in rendering individualized variation of ITD.

The conclusions differ if JND1 ([74 – 91 μs]) is considered. Indeed, according to JND1, the ITD variation along the cones of confusion is not audible. Assuming that JND2 denotes discrimination performance of audio experts, whereas JND1 denotes non-expert performance, this result suggests that the requirement of modelling and individualization of ITD variation along the cones of confusion is weaker for standard applications which do not require a high level of accuracy. In this context, ITD modelling by SHM is perceptually sufficient,.

In addition it should be noticed that this study is based on a particular estimate of ITD (Algazi et al., 2001 a). An issue of interest for future research would be to investigate to what extent the previous conclusions may depend on the choice of the method of ITD

estimate (Busson et al., 2005).

ACKNOWLEDGMENTS

The authors are very grateful to D.Pressnitzer, E. Rio, G. Vandernoot and T. Caulkins for their precious help.

REFERENCES

Algazi, V. R., Avendano, C., and Duda, R. O. (**2001 a**). "Estimation of a spherical-head model from anthropometry," J. Audio. Eng. Soc. **49**(6), 472-478.

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001 b**). "The CIPIC HRTF Database," *Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*, Mohonk Mountain House, New Paltz, NY, 99-102.

Algazi, V. R., Avendano, C., and Thompson, D. (**1999**). "Dependence of subject and measurement position in binaural signal acquisition," J. Audio Eng Soc. **47**(11), 937-947.

Best, V., van Schaik, A., and Carlile, S. (**2004**). "Separation of concurrent broadband soun sources by human listeners", J. Acoust. Soc. Am. **115**(1), 324-336.

Blauert, J. (**1983**). *Spatial Hearing*, edited by MIT press, Cambridge MA.

Busson, S., Nicol, R., and Warusfel, O. (**2004**). "Influence of the ears location on spherical head model for the individualized interaural time difference," *Proc. of the Joint Meeting of CFA and DAGA 2004*, Strasbourg, France.

Busson, S., Nicol, R., and Katz, F. G. (**2005**). "Subjective investigations of the interaural time difference in the horizontal plane," presented at the *118th Audio Engineering Society Convention*, convention preprint 6324.

Domnitz, R. H. (**1973**). "The interaural time jnd as a simultaneous function of interaural time and interaural amplitude," J. Acoust. Soc. Am. **53**(6), 1549-1552.

Domnitz, R. H., and Colburn, H. S. (**1977**). "Lateral position and interaural discrimination," J. Acoust. Soc. Am. **61**(5), 1586-1598.

Duda, R. O., Avendano, C., and Algazi V. R. (**1999**). "An adaptable ellipsoidal head model for the interaural time difference," ICASSP'99 *Proc. IEEE International Conference on Acoustics Speech and Signal Processing*, **II**, 965-968.

Ericson, M.A., and McKinley, R. L. (**1989**). "Auditory localization cue synthesis and human performance," NAECON 89 ; *Proc. of the IEEE National Aerospace and Electronics Conference*, Dayton, OH, 718-725.

Grantham D. W., Hornsby B. W. Y., and Erpenbeck E. A. (**2003**). "Auditory spatial resolution in horizontal, vertical, and diagonal planes," J. Acoust. Soc. Am. **114**(2), 3030–3038.

Hafter, E. R., and Maio, J. D. (**1975**)."Difference thresholds for interaural delay," J. Acoust. Soc. Am., **57**(1), 181-187.

Henning, G. B. (**1974**). "Detectability of interaural delay in high-frequency complex waveforms," J. Acoust. Soc. Am. **55**(1), 84-90.

Hershkowitz, R. M., and Durlach, N. I. (**1969**). "Interaural time and amplitude jnds for a 500-Hz tone," J. Acoust. Soc. Am. **46**(6), 1464-1468.

Kistler D. J., and Wightman F. L. (**1992**), "A model of head related transfer function based on principal components analysis and minimum-phase reconstruction", J. Acoust. Soc. Am. **91**, 1637-1647.

Klump, R. G., and Eady, H. R. (**1956**). "Some measurements of interaural time difference thresholds", J. Acoust. Soc. Am. **53**(6), 859-860.

Laakso, T. I., Välimäki, V., Karjailanen, M., and Laine, K. U. (**1996**). "Splitting the

unit delay : tools for fractional delay filter design", IEEE Signal Processing Magazine, vol. 13, no. 1, 30–60.

Larcher, V., and Jot, J. M. (**1999**). "Techniques d'interpolation de filtres audionumériques : Application à la reproduction spatiale des sons sur écouteurs", *Congrès français d'acoustique*, Marseille, France.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**(2), 467-477.

Mills, A. W. (**1958**). "On the minimum audible angle," J. Acoust. Soc. Am **30**(4), 237-246.

Minnaar, P., Plogsties, J., Olesen, S. K. , Christensen, F., and Moller H. (**2000**) . "The interaural time difference in binaural synthesis", presented at the *108th Audio Engineering Society Convention*, convention preprint 5133.

Oldfield, S. R., and Parker, S. P. A. (**1984**). "Acuity of sound localisation : a topography of auditory space. I. Normal Hearing," Perception, **13**, 581-600.

Pernaux, J-M. (**2003**). *Spatialisation du son par les techniques binaurales : application aux services de télécommunications*, Ph-D thesis, Institut National de Physique de Grenoble, Grenoble, France.

Savioja, L. Huopaniemi, J. Lokki, T., and Väänänen, R. (**1999**). "Creating interactive virtual acoustic environments," J. Audio Eng. Soc. **47**(9), 675-705.

Shinn-Cunningham, B G., Santarelli S. and Kopco, N. (**2000**). "Tori of confusion : Binaural localization cues for sources within reach of a listener**,"** J. Acoust. Soc. Am. **107**(3), 1627-1636.

Tollin, D. J., and Henning, G. B. (**1998**). "Some aspects of the lateralization of echoed sound in man. I. The classical interaural-delay based precedence effect," J. Acoust. Soc. Am. **104**, 3030–3038.

Trahiotis, C., Bernstein, L. R., Buell, T. N., and Spektor, Z. (**1990**). "On the use of adaptative procedures in binaural experiments", J. Acoust. Soc. Am. **87**(3), 1359.

Woodworth, R. S., and Schloesberg, H. (**1962**). *Experimental Psychology*, edited by Holt, Rinehard and Winston, NY, 348-361.

Zurek, P. M. (**1985**). "Spectral Dominance in sensitivity to interaural delay for broadband stimuli," J. Acoust. Soc. Am. Suppl. 1.**78,** S18.

abstract

Binaural synthesis is a sound spatialization technology, which is the closest to natural hearing. It allows the spatialization of a monophonic sound source with only two filters for a given position. The filters are defined by the HRTFs (Head Related Transfer Function) corresponding to the left and right ear of the listener. The major drawback of binaural synthesis is that the HRTF, which are related to the listener's morphology, are strongly individual. Listening with non-individual HRTF could lead to audible artifacts. Therefore binaural rendering of high quality requires individualized HRTF. This thesis tackles the problem of the individualization of binaural synthesis in the framework of its implementation as a pure delay, the interaural time difference (ITD), and a minimal phase filter determined by the magnitude of the HRTF. The work conducted on the ITD validates the implementation chosen even for the positions where the HRTF are poorly minimum phase filters. In addition the ITD calculation methods which are close to perception are pointed out. An experimental study is also undertaken to investigate the resolution of the ITD with the elevation angle along the cones of confusion. Perceptual results indicate that the ITD variation with the elevation angle needs to be reproduced. In order to account for this variation, a new formula is proposed on the basis of the spherical head model. Optimization of the parameters of this formula for a whole ITD database provides an average formulation which is appropriate for a large number of subjects and for many applications. Concerning the modeling of the spectral cues (HRTF magnitude), the Boundary Element Method (BEM) has been examined. It is concluded that BEM methods are useful in combination with measurement for the modeling of the low frequency part. A new approach, which involves statistical learning technique, is proposed for the HRTF prediction. A neural network is built to compute HRTF in any direction from a limited set of measured HRTF. Preliminary assessment of this modeling shows that the neural network succeeds well in individualizing spectral cues. This result suggests a simplified protocol of HRTF measurement : HRTF are measured for only a few directions and the HRTF for the other locations are obtained by the neural network.

Key words : Binaural synthesis, 3D sound, HRTF, minimum phase filter, psychoacoustics, ITD, JND, BEM, neural network.