Application de techniques d'apprentissages statistiques à la prédiction d'HRTF

Patrick Vovor

Mémoire pour le Master Sciences et Technologie
Université Pierre et Marie CURIE

Mention Informatique, spécialité SAR
Parcours ATIAM

Laboratoire d'accueil : France Telecom Recherche et Développement (Lannion)
Responsable du stage : Vincent Lemaire

01 Septembre 05

Table des matières

Introduction

La synthèse binaurale permet de reproduire des scènes sonores spatialisées en 3 dimensions à partir d'un casque d'écoute. Les applications possibles touchent des domaines tels que les télécommunications, la réalité virtuelle, le domaine militaire et le "design" sonore.

La connaissance des mécanismes mis en jeu dans la localisation auditive a permis l'élaboration des techniques de synthèse binaurale. Dans une situation d'écoute réelle : lorsqu'une onde sonore incidente se propage jusqu'au tympan, elle est diffractée par le corps de l'auditeur, en particulier par sa tête et son torse. Le son incident arrive alors aux tympans transformé, livrant ainsi au système auditif des indices caractéristiques de sa position. Pour chaque incidence, ces transformations peuvent être capturées, et implémentées sous forme de filtres audionumériques, qualifiés de Head-Related Transfer Functions (HRTF). Ces filtres constituent des empreintes spatiales qu'on appliquera au canal monophonique à spatialiser.

Ces techniques suscitent un grand intérêt du fait qu'elles se proposent de restituer aux tympans de l'auditeur le champ sonore qu'il aurait perçu dans une situation d'écoute réelle. Cependant elles présentent certaines difficultés pour une utilisation "grand public". Ces difficultés résident essentiellement dans la lourdeur des traitements audionumériques, et la variabilité des HRTF en fonction de l'individu. En effet, la perception d'une scène sonore spatialisée avec les HRTF d'un individu se dégrade considérablement pour un individu différent de celui-ci. Une conséquence direct de cette dépendance individuelle est que les systèmes de reproduction sonore 3D (utilisant les techniques de synthèse binaurale) à usage "grand public" nécessitent un grand nombre de mesures d'HRTF (au moins 1000) à effectuer par individu. L'étude développée dans ce rapport s'inscrit dans un objectif d'individualisation de la synthèse binaurale.Des solutions aux problèmes qui viennent d'être évoqués seront étudiées. Celles-ci s'appuyent sur l'analyse statistique d'une base de donnee HRTF.

Présentation des unités de Recherche et Développement TSI et SSTP

Ce stage s'est déroulé dans l'entreprise France Telecom Recherche et Développement au centre de Lannion.

L'étude présentée dans ce rapport s'est effectuée en partenariat entre deux unités de Recherche et Développement, l'unité Speech, Sound technologies and processing (SSTP), maîtrise d'ouvrage du projet, et l'unité Traitement Statistique de l'Information (TSI) la maitrise d'oeuvre. Nous présenterons ici les objectifs et les activités de chaque service.

Maitrise d'ouvrage

La maitrise d'ouvrage du projet est la section Speech and Sound Technologies and Processing.
Les activités de SSTP sont principalement la synthèse vocale, le codage de la voie, la reconnaissance vocale et l'acoustique.
La section acoustique est spécialisée dans la création d'environnements acoustiques virtuels utilisant des procédés de spatialisation tridimensionnelle.
Les différents modes de spatialisation 3D étudiés et implémentés sont l'holophonie, le rendu ambisonic et la synthèse binaurale.

Maitrise d'oeuvre du projet

L'unité de Recherche et Développement TSI (Traitement Statistique de l'Information) est chargée d'étudier, de proposer et de développer de nouvelles méthodes destinées aux services et aux réseaux de FranceTelecom, basées principalement sur des techniques d'apprentissage statistiques.
Ses domaines d'activité comportent en particulier la prévision de séries temporelles pour le trafic et les réseaux de communications, le traitement de données et le data mining.
Des connaissances à la pointe de la recherche dans le domaine des statistiques et de l'apprentissage statistique ont permis d'offrir les compétences nécessaires à la réalisation du projet.

Vincent Lemaire (TSI), Rozenn Nicol et Sylvain Busson (SSTP) ont assuré l'encadrement de ce stage.

Chapitre 1

L’évolution du système de prédiction d’une base de donnée d’HRTF

Ce premier chapitre a pour but d’apporter les notions nécessaires à la compréhension des travaux réalisés lors de ce stage. On y présentera le contexte, les enjeux et la problématique de celui-ci.

1.1 Introduction : vers un système de reproduction sonore 3D à usage "grand public"

La synthèse binaurale permet de recréer une scène sonore en 3 dimensions à partir d’un casque d’écoute. Les systèmes de spatialisation sonore doivent leurs existences à une meilleur compréhension des mécanismes mis en jeu lors du processus perceptif de localisation des sources sonores. L’étude de ces mécanismes, bien que complexe, a permis de mettre en évidence plusieurs indices jouant un rôle considérable dans le processus perceptif de localisation. Les premiers indices de localisation qui ont été découverts sont appelés indices binauraux car ils n’existent qu’en présence de deux capteurs sonores. Ces indices sont :

- **l’ILD (Interaural Level Difference) :** Indice basé sur la différence d’intensité entre un signal sonore arrivant à notre oreille la plus exposée de la source (l’oreille ipsélatérale) et l’oreille opposée (l’oreille contralatérale).
- **l’ITD (Interaural Time Difference) :** Indice basé sur la différence de temps d’arrivée d’un son entre nos deux oreilles.

Ces deux indices constituent la "Duplex Theorie" de Rayleigh. Malheureusement, la symétrie de notre tête pose plusieurs problèmes, notamment celui appelé cône de confusion : certaines positions spatiales possèdent le même couple ILD, ITD (voir la figure 1.1).

Pour lever l'ambiguité sur la localisation de sources sonores situées dans cette zone de

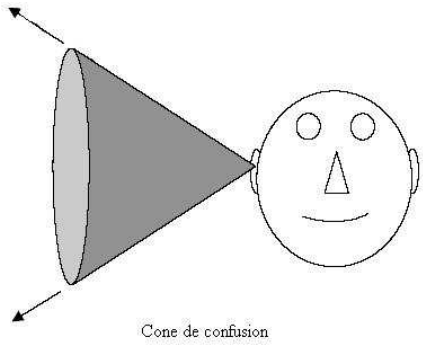

FIG. 1.1 – Le cône de confusion : tous les points pris sur un cercle appartenant à la surface de ce clône ont des ITD et ILD confondus.

confusion, il existe un troisième indice qualifié d'indice monaural (il ne nécessite qu'un seul capteur). Il s'agit d'un indice **spectral** qui caractérise le filtrage du signal sonore par notre morphologie. On définit ce filtrage par les fonctions de transferts de la tête : les **HRTF (Head Related Transfer Function).** Les HRTF décrivent les modifications du signal causées par l'oreille externe (le pavillon), le torse, les épaules et la tête entre son émission et son arrivée à l'entrée du canal auditif. On les mesure dans une chambre anéchoïque en plaçant des micros à l'entrée du canal auditif de l'auditeur (tête artificielle ou non) et en faisant varier la position de la source sonore (un signal impulsionnel large bande 20-20khz) autour de la tête de l'auditeur (voir figure 1.2). La connaissance du signal émis et du signal d'arrivée permet d'obtenir la fonction de transfert.

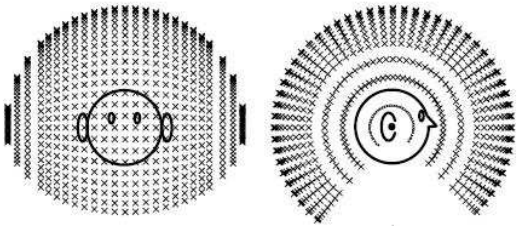

FIG. 1.2 – Position des points de mesures utilisés par le CIPIC (Center Image Processing and Integrated Computing) : Les croix représentent les différentes positions qu'occupent les Haut-Parleurs du dispositif de mesure d'HRTF.

1.1.1 La problématique du projet

Les HRTF caractérisent la position spatiale de la source sonore émettrice. Ainsi, pour donner l'impression à un auditeur qu'un signal sonore provient d'une position spatiale particulière (repèrée par une azimut θ et une élévation ϕ), on convolue ce signal monophonique par les HRIR (HRTF exprimée dans le domaine temporel) des oreilles droite et

gauche mesurées en cette position. En envoyant ce nouveau signal ainsi "spatialisé" dans un casque d'écoute, l'auditeur, recevra directement dans l'oreille interne une image artificielle du signal qu'elle aurait reçue si une véritable source sonore placée en (θ,ϕ) avait émit un son (voir figure 1.3). Son système cognitif interprétera alors ce signal artificiel comme provenant d'une source sonore située en (θ,ϕ).

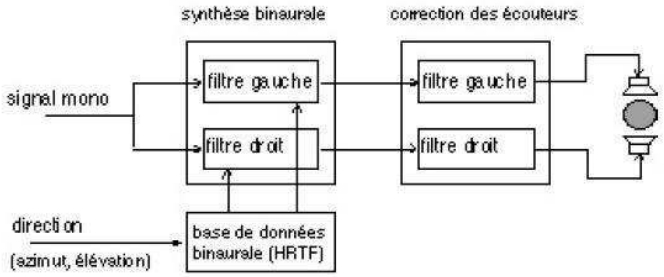

FIG. 1.3 – Principe de restitution binaurale à partir d'un enregistrement

Mais cette utilisation de l'indice spectral ne se fait pas sans difficultés. Celles-ci sont dûes à la nature mêmes des HRTF :

– Leur dépendance spatiale implique que pour reproduire le plus fidèlement possible une scène contenant des sons venant de n'importe où dans l'espace, il faut mesurer un maximum d'HRTF tout autour de la tête de l'auditeur. Ne pouvant mesurer les HRTF en tous les points d'une sphère (cas idéal) autour de la tête d'un sujet, on procède à un échantillonnage spatiale des mesures. Le nombre d'HRTF à mesurer reste cependant conséquent.
  Le Center Image Processing and Integrated Computing (CIPIC [10]) considère qu'avec 1250 mesures d'HRTF par individu, on a une très bonne résolution spatiale.

– L'autre inconvénient de l'utilisation des HRTF est qu'elles sont individuelles. En effet, que ce soit au niveau de notre pavillon, de nos épaules ou de notre torse, les morphologies changent assez pour modifier de manière considérable le filtrage des signaux sonores.

La somme de ces deux propriétés des HRTF constitue un véritable obstacle à la réalisation d'un système de spatialisation pour le "grand public". En effet, un tel système demanderait à l'utilisateur final de mesurer ses 1250 HRTF, or, l'équipement et le temps nécessaire (plus de deux heures de mesures) rend cette pratique irréalisable. Et le fait d'utiliser des HRTF génériques, autrement dit, d'utiliser des HRTF mesurées sur un individu différent de celui qui écoutera la scène sonore reconstituée via un casque, altère considérablement la qualité de la spatialisation. Cela se traduit par une perte d'externalisation des sources sonores ainsi que par une confusion avant/arrière à l'écoute au casque. **Une solution serait alors de trouver un moyen d'obtenir les HRTF d'un individu sans devoir pour autant les mesurer, ie, trouver un moyen d'individualiser les HRTF.** L'idée de modéliser des HRTF trouve ici tous son sens.

1.1.2 L'état de l'art des techniques d'individualisation des HRTF

On retrouve dans la littérature plusieurs approches pour individualiser les HRTF, les plus couramment utilisées figurent parmi celles présentées ci-dessous :

- *Sélection d'un jeu d'HRTF au sein d'une base de données réduite par écoute interactive* (voir [16]). Dans cette approche, on dispose d'une base de données de plusieurs jeux d'HRTF correspondant à différents groupes d'individus dont les facultés de localisation auditive sont similaires. Le jeu d'HRTF qui sera asscoié à l'auditeur est sélectionné à partir de tests de localisation sur plusieurs positions spatiales. Le risque est de choisir un jeu d'HRTF qui est proche de celui de l'auditeur pour certaines positions spatiales (celles testées) et éloigné pour d'autres. C'est une approche simple à mettre en oeuvre, mais elle offre des performances perceptives réduites.

- *Sélection dans une base de données en fonction d'un critère de proximité morphologique* (Zoltkin,Al [25]). Cette approche joue sur la proximité morphologique du sujet dont on doit individualiser les HRTF, avec un sujet de la base. Cette méthode est également simple à mettre en place, mais elle nécessite une base de données importante d'HRTF, couplées à des mesures anthropométriques.

- *Modèle de signal des HRTF* (Middlebrooks [15]). Cette méthode consiste à synthétiser le module d'un jeu d'HRTF d'un sujet, à partir d'un jeu d'HRTF connu, en effectuant des opérations de dilatation/compression (opération de warping) sur l'échelle des fréquences.

- *Modèle basé sur une décomposition linéaire d'HRTF* (Carlile, Jin, Leong [12]). Cette approche utilise une ACP (analyse en composantes principales) qui permet de constater que seul un petit nombre des coefficients obtenus permet de décrire les variations des grandeurs anthropométriques. L'objectif final est d'obtenir les coefficients acoustiques qui permettront de reconstituer un jeu d'HRTF "adapté" à un sujet donné, à partir des mesures d'un nombre réduit de paramètres anthropométriques. Cette technologie fait l'objet d'un brevet déposé.

- *Méthodes basées sur l'interpolation d'HRTF*. Cette méthode utilise un algorithme de Clustering qui permet de réduire le nombre de mesures d'HRTF à réaliser sur l'auditeur. Ensuite, une interpolation (linéaire ou non) est effectuée pour reconstruire ses HRTF manquantes. Fahn et Lo [6] utilisent l'algorithme LBG[1] pour la phase de clustering et effectue une interpolation linéaire pour reconstruire le jeu complet d'HRTF d'un individu. Cependant d'autres méthodes d'interpolations comme l'ACP[2] ou les réseaux de neurones (Nishino [18]) sont utilisées. Durant [8] utilise un algorithme génétique pour approximer les HRTF.

[1] algorithme de Linde, Buzo et Gray
[2] Analyse en Composantes Principales

L'équipe en charge du projet a quant à elle choisit une démarche proche de celle basée sur l'interpolation d'HRTF. Le but commun entre notre approche et celle de Fahn est de répondre à la problématique suivante : "Est-il possible de construire un modèle permettant d'obtenir, à partir de certains paramètres, certaines mesures,..., les HRTF d'un individu ?". Les recherches effectuées l'année dernière par les unités R&D TSI et SSTP ont montrés qu'en faisant apprendre à des outils d'apprentissage statistique, une base de donnée d'HRTF, on pouvait explorer cette voie. Mais avant d'en dire plus sur la technologie développée, nous allons justement décrire dans la partie suivante la base de donnée que nous avons utilisé.

1.2 Données et Notations

Nous présentons ici la base de donnée d'HRTF que nous avons utilisé pour réaliser nos études statistiques ainsi que les notations utilisées dans les parties et chapitres suivants.

1.2.1 Origine des Données

Plusieurs campagnes de mesures d'HRTF ont été menées notamment par l'IRCAM[3], le MIT[4] et le CIPIC[5]. Le processus de mesures étant très coûteux en temps et en équipement, nous avons exploité une base de donnée d'HRTF publique. On s'est basé sur la campagne du CIPIC dont les mesures décrivent les HRTF pour 1250 positions spatiales.

Coordonnées spatiales des mesures

Les 1250 positions spatiales sont définis avec l'azimuth θ et l'élevation ϕ du système de coordonnées dit polaire interaurale (voir la figure 1.4) qu'utilise la base CIPIC. Dans tout ce qui suit, on notera $HRTF_{\lambda,\theta,\phi}$, l'HRTF de l'individu λ mesurée à la position (θ, ϕ). On notera $HRTF^L_\lambda$ et $HRTF^R_\lambda$, les HRTF mesurées respectivement sur l'oreille gauche et droite pour un individu λ et une position quelconque. On manipule dans cette étude uniquement les HRTF de l'oreille gauche.

Cette base de donnée possède les 1250 HRTF de 45 individus (dont deux têtes artificiels), soit 56250 HRTF en tout. Les points de mesures du CIPIC ont été uniformément répartis par pas de 5.625 ^{o} ($= 360/64$) de -45 ^{o} à 230.625 ^{o} en élévation, et pour les valeurs -80 ^{o}, -65 ^{o}, -55 ^{o} puis tous les 5 ^{o} de -45 ^{o} à 45 ^{o} et enfin pour les valeurs 55 ^{o}, 65 ^{o} et 80 ^{o} en azimuth (voir [10]).

Spécification des HRTF

Il est important de séparer les informations spectrales et temporelles contenues dans les HRTF. Or, comme une HRTF peut être considérer comme un filtre causal et stable, on

[3] Institut de Recherche et Coordination Acoustique/Musique
[4] Massachusett Institute of Technology
[5] Center Image Processing and Integrated Computing

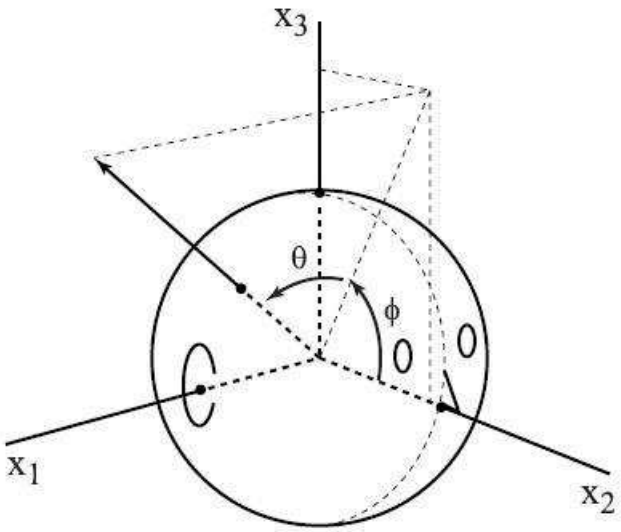

FIG. 1.4 – Système polaire interaural.

peut la décomposer en un produit de filtre à phase minimal H_{mph} (contenant les informations fréquentielles) et d'un filtre passe-tout H_{pt} (contenant les informations temporelles).

$$HRTF^L = H^L_{mph} * H^L_{pt} \tag{1.1}$$

$$HRTF^R = H^R_{mph} * H^R_{pt} \tag{1.2}$$

Whitman et Kistler [23] ont montré que l'on peut modéliser les composantes passe-tout par des retards purs, indépendants de la fréquence. On obtient ainsi :

$$ITD = H^L_{pt} - H^R_{pt} \tag{1.3}$$

L'amplitude $|H_{mph}|$ et la phase ϕ_{mph} d'un filtre à phase minimale sont reliés par la transformée de Hilbert :

$$\phi_{mph} = Im(Hilbert(-\log(|H_{mph}|))); \tag{1.4}$$

Une paire d'HRTF s'écrit donc :

$$HRTF^L(f) = \exp^{j\omega ITD} |H^L_{mph}(f)| * \exp^{j\phi^L_{mph}(f)} \tag{1.5}$$

$$HRTF^R(f) = |H^R_{mph}(f)| * \exp^{j\phi^R_{mph}(f)} \tag{1.6}$$

Pour notre étude, on s'intéresse seulement au module $|H_{mph}|$ de la composante à phase minimale. Par abus de langage, nous appellerons à partir de maintenant ce module HRTF. L'échelle des fréquences des HRTF est échantillonnée sur 100 fréquences linéairement espacées de 0Hz à 22.05kHz. On utilise l'indice i pour désigner le n-ième bin fréquentiel (i varie de 1 à 100). Finalement, on définit l'HRTF (ie, son module) d'une personne pour une position spatiale d'azimuth θ et d'élévation ϕ par le vecteur :

$$H_{\lambda,\theta,\phi} = \begin{pmatrix} H_{\lambda,\theta,\phi}(1) \\ \cdots\cdots\cdots \\ \cdots\cdots\cdots \\ H_{\lambda,\theta,\phi}(i) \\ \cdots\cdots\cdots \\ \cdots\cdots\cdots \\ H_{\lambda,\theta,\phi}(100) \end{pmatrix} \tag{1.7}$$

Dans les chapitres suivant, on désignera par $\hat{H}_{\theta,\phi,\lambda}$ l'HRTF estimée par nos modèles, celle-ci approxime l'HRTF réelle $H_{\theta,\phi,\lambda}$.

Prétraitements

Plusieurs prétraitements des HRTF peuvent être réalisés. Leur utilité est d'éliminer des HRTF, alors exprimées dans un format quelconque, la part d'information superflue (ie, ne jouant aucun rôle dans la localisation auditive). Afin d'étudier l'influence de ces prétraitements sur les performances de nos outils d'apprentissage statistique (voir la discussion concernant l'importance du choix de prétraitement dans la partie suivante), nous avons testé deux prétraitements :

– **Modification de l'échelle des amplitudes**.

Lorsque l'on travaille sur des données audio, on préfère utiliser une échelle des amplitudes logarithmique plus proche de notre perception auditive qu'une échelle linéaire. Pour éviter que la conversion en base logarithmique des amplitudes ait une borne inférieure qui tend vers $-\infty$ lorsque l'amplitude linéaire est égale à 0, on fixe la borne inférieure des amplitudes à $-60dB$ en échelle logarithmique (qui équivaut à 10^{-3} en échelle d'amplitude linéaire). Finalement, on transforme le vecteur d'entrée comme suit :

$$H^{l}_{\lambda,\theta,\phi} = \begin{pmatrix} 20\log_{10}(max(H_{\lambda,\theta,\phi}(1), 10^{-3})) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ 20\log_{10}(max(H_{\lambda,\theta,\phi}(i), 10^{-3})) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ 20\log_{10}(max(H_{\lambda,\theta,\phi}(100), 10^{-3})) \end{pmatrix} \tag{1.8}$$

– **L'égalisation des données "champs diffus" et le lissage spectrale**.

On désigne par champ diffus le champ constitué d'ondes planes décorellées provenant d'incidences uniformément distribuées autour du récepteur. Le but d'une égalisation champs diffus est d'obtenir un champ moyen pour enlever des caractéristiques spectrales qui ne sont pas dues à la direction. On note ces filtres : DTF (Directionnal Transfer Function). L'égalisation champ diffus élimine les artefacts de mesures indépendants de la direction et réduit de façon significative les différences entre les sessions de mesures entre individus.

Pour égaliser nos données, nous avons utilisé la méthode de calcul proposée par Larcher [11]. Le lissage spectrale (voir figures 1.5 et 1.6) a été réalisé avec une méthode de la toolbox binaural de l'IRCAM.

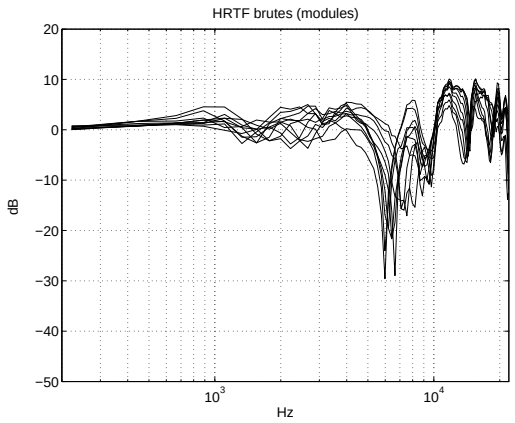

FIG. 1.5 – HRTF Brutes (modules) : pour une élévation nul et des azimuts qui varient entre -80 ^{o} et -15 ^{o}. L'axe des abscisses représente les fréquences en Hz, les ordonnées présente le gain en dB.

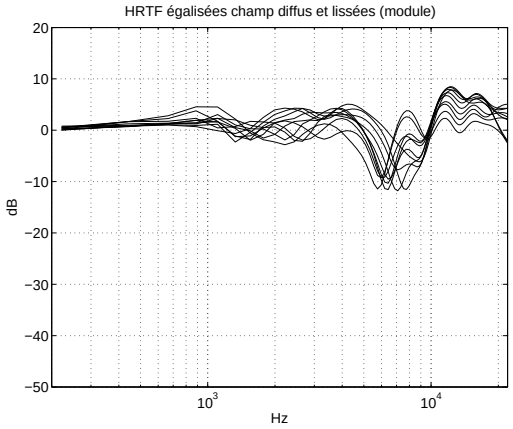

FIG. 1.6 – HRTF égalisées champ diffus et lissées : pour une élévation nul et des azimuts qui varient entre -80 ^{o} et -15 ^{o}. L'axe des abscisses représente les fréquences en Hz, les ordonnées présente le gain en dB

1.2.2 Découpage des données

Les outils d'apprentissage statistique que nous allons présenter dans les chapitres suivant (voir section 3.2 et chapitre 4) vont nous permettre d'extraire les informations pertinentes que contient la base de données CIPIC.
Dans leur processus d'apprentissage, ceux-ci construisent une représentation à partir d'un nombre fini d'exemples, l'ensemble d'apprentissage. Cependant, le plus important est qu'ils arrivent à généraliser cette représentation sur toutes les données, y compris celles n'appartenant pas à l'ensemble d'apprentissage. Une manière d'évaluer cette faculté de généralisation (encore appelée "erreur" de généralisation) consiste à évaluer les performances de ces outils sur des données représentatives du problème non apprises. La différence entre l'erreur d'apprentissage et l'erreur de généralisation représente une mesure de la qualité de l'apprentissage effectué.

Les méthodes qui évaluent l'erreur de généralisation sont presque toutes basées sur la partition de l'ensemble des données qu'on possède en plusieurs sous-ensembles. Par exemple un ensemble utilisé pour l'apprentissage et un autre pour la validation. L'ensemble de validation est utilisé pour contrôler et mesurer la généralisation des outils d'apprentissage statistique. Ces deux ensemble servent à déterminer l'architecture la plus appropriée pour le problème à traiter : pour différentes architectures (par exemple, le nombre de couches et neurones cachés d'un réseau de neurone) on contrôle l'erreur de validation et on choisit l'architecture pour laquelle elle est minimale. Une des techniques

suivant cette méthodologie consiste à réserver 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.
Notre choix s'est porté sur cette méthode du partage de l'ensemble des exemples en trois ensembles (apprentissage, validation, test) citée ci-dessus car nous préférons ne jamais "apprendre" les exemples de l'ensemble de validation. Ces trois ensembles ont les cardinalités suivantes :

- l'ensemble d'Apprentissage noté D_{App} : il contient les HRTF de 18 individus, soit 22500 HRTF.
- l'ensemble de Validation noté D_{Val} : il contient les HRTF de 8 individus, soit 10000 HRTF.
- l'ensemble de Test noté D_{Test} : il contient les HRTF de 8 individus, soit 10000 HRTF.

1.3 But et problématique de l'étude 2005

1.3.1 Etat des travaux des unités R&D SUSI et SSTP dans le domaine

L'approche proposée par l'unité R&D SUSI et SSTP conçoît qu'un système individualisant des HRTF consiste à construire un modèle permettant d'obtenir, à partir d'un nombre restreint d'HRTF d'un individu, l'ensemble de ses HRTF. Sachant qu'un tel modèle doit être utilisable quelque soit l'individu.

Les travaux de l'étude 2004 (voir [7]) ont montré qu'en utilisant des outils d'apprentissage statistique, il serait possible d'arriver à mettre au point un jour un tel système. L'étude effectuée résolvait les deux sous-problèmes suivants :

1. **Trouver le nombre et la localisation des HRTF suffisantes mais nécessaires à la reconstruction de l'ensemble des HRTF**.
   On appelera dans tout ce qui suit ces HRTF : les HRTF "représentantes". Cette tâche a été réalisée en utilisant une technique de clustering pour regrouper les HRTF par similarité fréquentielles.
   Pour élire les HRTF représentantes, l'algorithme de clustering utilisé regroupait les HRTF dans n clusters. Une HRTF "représentante" était ensuite élue pour chacun d'eux. Cette étape permettait donc d'obtenir n points de mesures nécessaires et suffisants aux lieu de 1250. Parmi toutes les techniques existantes (k-moyennes, Partition Around Medoids, etc...), ce sont les cartes de kohonen qui ont été préfèrées. L'avantage de ces cartes réside dans la visualisation sur un espace réduit (2D par exemple) de données de grandes dimensions (ici, les 100 composantes fréquentielles).
   En procédant ainsi, on exploite la relation existant entre les proximités spatiales et spectrales des HRTF pour réduire le nombre de mesures à effectuer.

   Le clustering a été effectué sur les 1250 HRTF d'un seul individu choisi arbitrairement parmi les 44 de la base. Les HRTF "représentantes" obtenues par cette mé-

thode ont ensuite été appliquées aux autres individus de la base de donnée. Cela a permis de vérifier si ces HRTF représentantes pouvaient être généralisées aux autres individus de la base.

2. **Créer un modèle permettant d'obtenir l'ensemble des HRTF d'un individu à partir de ses HRTF "représentantes"**.
   Cette phase a été réalisée en utilisant des outils de régression statistique. Les réseaux de neurones de type Multi Layer Perceptron (MLP) ont été choisis pour leurs propriétés d'approximation de fonction et de généralisation.
   Un modèle est construit à travers l'apprentissage d'une base de donnée de plusieurs milliers d'HRTF. Le modèle, lors de son utilisation, reçoit les HRTF représentative ainsi que la position de l'HRTF qu'il va prédire.

Les performances d'estimation des HRTF ont été mesurées avec deux types d'erreurs : l'**erreur de quantification** et l'**erreur de modélisation**. L'erreur de quantification correspond à l'erreur commise lorsque l'on approxime une HRTF par son HRTF représentante. Cette erreur permet de répondre à la question : Quelle est l'erreur commise lorsque l'on utilise un nombre d'HRTF mesurées par individu inférieur à 1250 ?
L'erreur de modélisation correspond à l'erreur commise lorsque l'on approxime une HRTF à l'aide d'un modèle construit sur une base de donnée d'HRTF. La comparaison de la valeur de ces deux types d'erreur doit permettre d'évaluer l'efficacité de notre méthode de modélisation des HRTF.

Afin de mesurer l'intérêt de cette démarche de sélection des représentants, il a été comparé l'erreur de quantification qu'obtiennent les n représentants obtenus par la méthode de clustering avec celle qu'obtiennent les n représentants uniformément répartis dans l'espace. Ces représentants uniforméments répartis sont choisis indépendemment de leur spectre. Ce mode d'élection d'HRTF est la manière la plus triviale de réduire le nombre d'HRTF. Si les erreurs de quantification et de modélisation calculées pour nos représentants sont supérieures à celles calculées pour les représentants uniforméments répartis, c'est que notre méthode de clustering est mauvaise. La méthodologie utilisée pour obtenir les n représentants uniformément répartis est présentée dans Chocqueuse [7].

1.3.2 Problématique de l'étude 2005

La démarche d'élection utilisée lors de l'étude 2004 a permis de réduire l'erreur de quantification de 10% pour l'individu 1 par rapport à des représentants uniforméments répartis à la surface de la sphère. Par contre, l'erreur calculée pour tous les individus de la base du CIPIC (44 individus dans cette étude), était plus faible avec les représentants uniformément répartis (amélioration de 8%). La figure 1.7 montrent les courbes d'erreur de quantification/modélisation moyenne en fonction du nombre de représentants pour l'ensemble des individus.

Les deux courbes d'erreur de modélisation montrent qu'en utilisant un réseau de neurone MLP pour l'individualisation des HRTF, il est possible de modéliser une HRTF correctement à partir des HRTF représentatives. Cependant, en ce qui concerne l'origine des HRTF représentatives, les meilleurs résultats semblent être obtenus avec les représentants

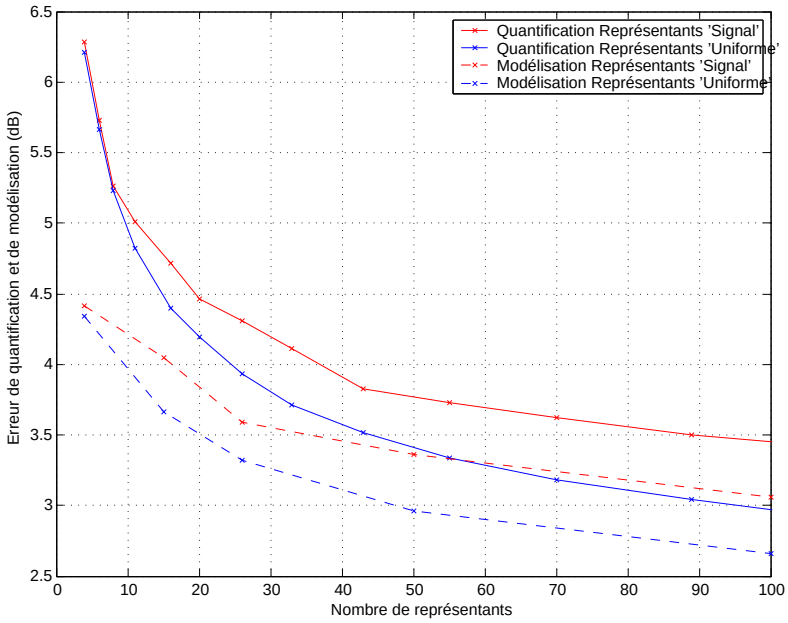

FIG. 1.7 – Resultats 2004 : Les erreurs de modélisation sont plus faibles que les erreurs de quantification. Les représentants uniformément répartis à la surface de la sphère offrent de meilleurs erreurs que les représentants calculés par notre méthode de clustering.

uniformément répartis à la surface de la sphère.

Les résultats de l'étude 2004 laissent entrevoir les trois champs d'amélioration possible que voici :

1. **Les données :**
   On se place ici en amont de notre système. Il s'agirait de trouver les meilleurs représentations et prétraitements possibles pour manipuler l'information de localisation contenue dans les HRTF. Les divers possibilités en la matière ont été évoqués dans la partie 1.2.

2. **La méthode d'élection des représentants et l'apprentissage de nos outils statistique :**
   On agit ici dans le coeur de notre système. L'efficacité de la méthode de sélection des HRTF représentatives utilisée en 2004 souffre d'un problème d'individualisation des positions des HRTF représentatives. Pour pallier à ce problème, plusieurs solutions s'offrent à nous :
   – On pourrait effectuer la même méthode de sélection sur un individu cette fois-ci mieux choisit. En effet, l'individu λ qui a été utilisé pour élire les HRTF représentatives n'est peut être pas le plus représentatif de la base. On pourrait mener une étude visant à déterminer l'existence d'un individu λ dont les représentants, déterminés via notre méthode d'élection, offriraient une meilleure erreur

de quantification que les représentants uniformément répartis.
– Il serait encore possible d'effectuer notre méthode de clustering sur **tous les individus de la base**. Cela nous permettra peut-être de trouver des positions d'HRTF représentatives valables pour tous les individus de la base CIPIC.

De plus, nos outils statistiques apprennent par des mesures de similarité au moyen d'une fonction de coût. On pourrait alors voir s'il est possible d'améliorer l'apprentissage de nos outils en utilisant différentes fonctions de coûts.

3. **Les critères d'évaluations**
On se place ici en aval de notre système. L'évaluation des performances de nos modèles se fait en calculant les erreurs de quantification et de modélisation. La recherche du critère d'évaluation le plus adapté à notre problème serait une voix à approfondir.

L'étude 2005 a permis d'explorer ces trois champs d'investigation.

1.3.3 Plan du rapport

Etude sur la représentation et les prétraitements des HRTF

Une HRTF est couramment décrite par son module et sa phase, cependant, il existe d'autres manières de représenter les HRTF. Il ya par exemple les hrir[6] (HRTF exprimées dans le domaine temporel) qui contiennent les informations de phase et de module des HRTF. Cette représentation revient à utiliser un filtre RIF[7] comme modèle paramétrique des HRTF. Cette représentation n'est pas adéquate parce qu'elle utilise 200 coefficients pour représenter une HRTF. Son utilisation par nos modèles serait trop coûteuse en mémoire RAM. Il ya aussi la représentation pôle-zeros qui consiste à déterminer (ie, localiser) les k pôles et zeros d'une HRTF alors considérée comme un filtre RII[8]. Cette dernière représentation est optimale pour la modélisation d'HRTF. Néanmoins, la modélisation paramétrique RII introduit une distorsion audible dont l'importance est inversement proportionnelle aux nombres de coefficients qu'elle utilise pour exprimer une HRTF. Des tests d'écoutes réalisés par l'unité R&D SSTP ont montré qu'il fallait finalement utiliser une centaine de coefficients pour obtenir un modèle paramétrique d'HRTF RII fidèle.
Pour toutes ces raisons, nous avons choisis pour cette étude de représenter les HRTF uniquement par leurs modules spectrales.

Par contre, afin d'étudier l'influence des prétraitements sur les performances de nos outils d'apprentissage statistique, nous avons effectuer les étapes d'élection de représentants et celle de modélisation sur deux types de données : les HRTF brute, puis sur les HRTF égalisées champs diffus lissé (voir section 1.2.1).

[6] Head Related Impulse Response
[7] filtre à Réponse Impulsionnelle Finie
[8] filtre à Réponse Impulsionnelle Infinie

Une nouvelle méthode d'élection de représentants

Nous avons choisis de tester une nouvelle méthode de sélection des HRTF représentatives. Cette nouvelle méthode consiste à réaliser un clustering sur **tous les individus de la base de donnée CIPIC**, afin d'élire indépendemment de ceux-ci, des positions d'HRTF représentantes offrant une meilleure erreur de quantification que les représentants uniformément répartis.

La figure 1.8 décrit l'étude réalisée cette année.

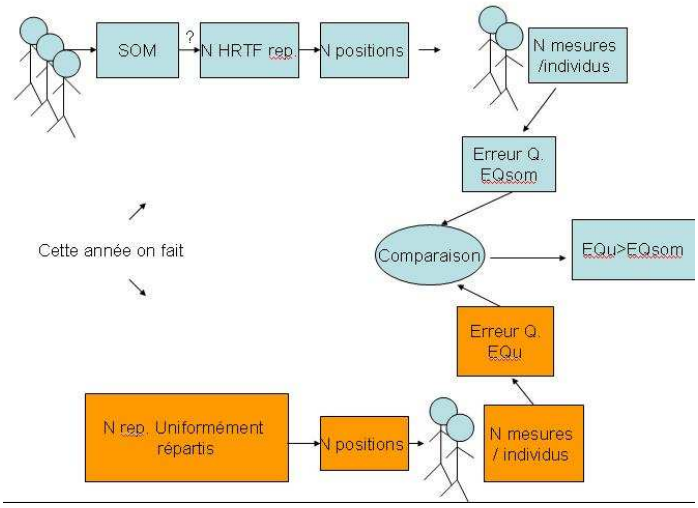

FIG. 1.8 – Schéma bloc décrivant l'étude réalisée en 2005. Equ est l'erreur de quantification calculée avec les représentants uniformes. Eqsom est l'erreur de quantification calculée avec les représentants issus de notre méthode de clustering (utilisant des cartes auto-organisatrices de Kohonen : SOM).

Choix du critère d'évaluation

Le laboratoire désire améliorer la qualité de ses modèles. Or, les techniques de modélisation d'HRTF cherchent à minimiser l'écart entre le spectre de départ et le spectre modélisé, au sens d'une certaine norme ou critère d'évaluation.

En introduisant un critère d'évaluation dans la fonction de coût d'un réseau de neurone qui réalise la modélisation, on espère améliorer les résultats. Cependant, cette opération impose des contraintes au niveau du choix du critère lui-même, et aussi au niveau du choix de la représentation des données. Il faut déja s'assurer que le critère adéquate soit minimisable. De plus, ce critère impose une représentation particulière des données qu'il compare. On voit là l'importance du choix du critère, sa détermination orientera les prochaines améliorations qui seront apportées au système de reproduction sonore 3D. Plusieurs critères d'évaluation ont été proposés dans la littérature, et étant donné la place centrale qu'occupe le choix du critère d'évaluation dans les futures travaux de l'équipe SSTP/SUSI, une étude était nécessaire pour déterminer ceux qui ressortent. La deuxième objectif de ce stage a donc été de mener une étude sur les critères d'évaluation afin de répondre aux questions suivantes :

1. Quel critère d'évaluation devons nous utiliser pour évaluer nos erreurs de modélisation ?
2. Quelle est l'influence de ce choix sur notre étude ?
3. l'introduction d'un critère d'évaluation dans les fonctions de coûts de nos outils d'apprentissage statistique permet-elle d'améliorer les résultats ?
4. Quelle représentation et quels prétraitements devrons nous utiliser pour diminuer nos erreurs de quantifications ?

Pour tenter de répondre à ces questions, nous allons comparer les résultats obtenus en évaluant, à l'aide de différents critères d'évaluation, les erreurs de quantifications des étapes d'élection pour diverses fonctions de coûts.

Chapitre 2

Critère d'évaluation et représentation des données

2.1 Introduction

Tout laboratoire travaillant sur la modélisation des HRTF cherche à mesurer les performances de ses modèles au moyen de critères d'évaluation. Pour rappel, ces critères évaluent l'erreur commise lorsqu'on approxime une HRTF quelconque par sa représentante ou son modèle. S'il on interprète ces critères en terme de mesure d'erreur de localisation, le critère idéal permettrait d'établir qu'une erreur de X dB calculée sur deux HRTF mesurées en des positions différentes correspond à une erreur de position de Y ^{o} entre les points de mesures des deux HRTF. On pourrait aussi les utiliser pour mesurer une erreur d'individualisation. Dans ce cas, une erreur de X dB calculée sur deux HRTF mesurées en la même position mais sur deux individus différents donnera le rapport de proportion existant entre les caractéristiques morphologiques de ces deux individus.
Hélas, la caractérisation fréquentielle des erreurs (de localisation ou d'individualisation) sur les HRTF n'étant pas encore maîtrisée, on est pas en mesure de proposer un tel critère. Cependant, plusieurs équipes scientifiques ont proposé leurs propres critères. Dans un premier temps, on a fait une étude sur ces différents critères, pour les définir, rappeler leurs propriétés mathématiques et établir leurs pertinence d'un point de vue perceptif. Puis, dans un second temps, nous avons proposé une procédure de tests (étalonnage) permettant de les évaluer. Le but ultime étant d'engager des discussions avec la communauté scientifique, au sujet du choix ou de l'élaboration d'un critère d'évaluation normalisé.

Nous présenterons dans une première partie les spécificités des différents critères d'évaluation[1] de l'état de l'art, pour après leur faire passer un banc de test. Les résultats de ces tests seront exposés dans la deuxième partie de ce chapitre.

[1] On pourra se reporter à la section 1.2.1 pour la définition des notations utilisées dans cette partie.

2.2 Critère d'évaluation : contexte et problèmes

Les techniques de modélisation de filtres cherchent à minimiser l'écart entre le spectre de départ, H_1, et le spectre modélisé, H_2, au sens d'une certaine norme ou critère d'évaluation.
Ce critère d'évaluation n'est peut-être pas adapté à nos données. L'erreur entre deux HRTF H_1 et H_2 de la base CIPIC, à cause de la nature même des HRTF, peut être dûe aux trois raisons suivantes :

1. soit elles ont été mesurées sur le même individu mais à des positions spatiales différentes : On parle alors d'erreur de localisation. Les études de Langendijk [1] montrent que les zones de creux et pics spectraux situés dans la bande de fréquence 4-16khz contiennent les information de localisation. Les indices spectraux situés entre 5.7-11.3 khz servent à la localisation Haut-bas tandis que la bande de fréquence 8-16khz sert pour la localisation avant/arrière.
2. soit elles ont été mesurées sur des individus différents mais pour une même position spatiale : On parle alors d'erreur d'individualisation. Middlebrooks [15] caractérise cette erreur par une translation des creux et ventres spectraux des HRTF sur l'échelle fréquentielle.
3. soit elles ont été mesurées sur des individus différents et pour des positions différentes. Il est dans ce cas difficile de distinguer l'origine de ce type d'erreur. Larcher [11] montre que l'erreur d'individualisation étant d'amplitude plus élevée que l'erreur de localisation, l'une masque l'autre.

Il serait par exemple intéressant de trouver un critère d'évaluation capable de distinguer les erreurs de localisation et les erreurs d'individualisation. Un tel critère pourrait nous servir tant pour l'évaluation de l'erreur de modélisation que pour l'apprentissage par nos modèles de la base de donnée CIPIC.
En effet, à mesure qu'ils parcourent les exemples d'apprentissage, nos modèles identifient la caractérisation fréquentielle des variations spatiales et inter-individus des HRTF. En utilisant pour cette phase d'apprentissage un critère d'evaluation tenant compte de toutes les considérations perceptives citées plus haut, on s'assurerait que nos modèles apprennent la seule part d'information utile à la perception auditive humaine de scènes sonores spatialisées.

2.3 Analyse des critères d'évaluation

Dans tout ce qui suit, $\hat{H}_{\theta,\phi,\lambda}$ désignera l'HRTF qui approxime (par un modèle quelconque) l'HRTF $H_{\theta,\phi,\lambda}$.

2.3.1 Le critère MSE

Définition et propriétés mathématiques

La Mean Square Error (MSE) est un critère très prisé à cause de la propriété que lui confère la relation de Parseval, son calcul est indépendant du choix de la représentation (fréquentielle ou temporelle) des deux vecteurs à comparer. L'erreur d'approximation (moyennée sur les fréquences) d'un module d'HRTF $H_{\theta,\phi,\lambda}$ s'exprime au moins du critère MSE par la formule 2.1.

$$E_{MSE}(\theta,\phi,\lambda) = \frac{1}{100}\sum_{i=1}^{100}(H^l_{\theta,\phi,\lambda}(i) - \hat{H}^l_{\theta,\phi,\lambda}(i))^2 \tag{2.1}$$

De plus, le terme d'erreur de ce critère étant quadratique, on s'assure que ce critère possède un minimum global, et est donc minimisable par la descente des gradient si on l'utilise comme fonction de coût dans un réseau de neurone.

Sensibilité et adéquation psycho-acoustique du critère

S'il on choisit la représentation fréquentielle pour les données que comparera ce critère, on peut remarquer que à cause de son terme d'erreur quadratique, ce critère ne tient compte que des fortes erreurs. En effet, la valeur de l'erreur obtenue en appliquant ce critère aux deux HRTF à comparer, va accumuler (par la somme des carrés) les contributions des fortes erreurs et très vite, écrasera la contribution des faibles erreurs.

2.3.2 Le critère de Chocqueuse

Définition et propriétés mathématiques

Ce critère calcule la différence entre l'HRTF estimée et l'HRTF réelle. L'erreur d'approximation (moyennée sur les fréquences) d'une HRTF $H_{\theta,\phi,\lambda}$ s'exprime au moins du critère de Chocqueuse par la formule 2.2.

$$E_{chocqueuse}(\theta,\phi,\lambda) = \frac{1}{100}\sum_{i=1}^{100}|H^l_{\theta,\phi,\lambda}(i) - \hat{H}^l_{\theta,\phi,\lambda}(i)| \tag{2.2}$$

2.3.3 Le critère BARK

Définition et propriétés mathématiques

Le critère de BARK consiste à introduire dans le critère Chocqueuse des magnitudes exprimées dans l'échelle fréquentielle de BARK. Cela revient à pondérer par l'inverse des coefficients de bark α_i la somme des erreurs par points fréquentiels. Ce critère ne possédant pas de terme quadratique, on s'est assuré qu'il est minimisable en effectuant le calcul de la descente de gradient. L'erreur d'approximation (moyennée sur les fréquences) d'une HRTF $H_{\theta,\phi,\lambda}$ s'exprime au moins du critère de Bark par la formule 2.3.

$$E_{bark}(\theta,\phi,\lambda) = \frac{1}{100}\sum_{i=1}^{100}\alpha_i|H^l_{\theta,\phi,\lambda}(i) - \hat{H}^l_{\theta,\phi,\lambda}(i)| \tag{2.3}$$

Sensibilité et adéquation psycho-acoustique du critère

Ce critère utilise l'échelle fréquentielle Bark qui est plus proche des propriétés de notre système auditif que ne l'est l'échelle de fréquences linéaires. Ce critère est donc à même de mesurer une erreur perceptuellement pertinente entre deux spectres. A cause de ses coefficients de pondération, ce critère tend à ne pas considérer les erreurs entre deux HRTF situées sur les points fréquentiels élevés (hautes fréquences).

2.3.4 Le critère de Durand

Définition et propriétés mathématiques

Le critère de Durand [8] calcule pour une position (θ, ϕ) donnée, la variance de la distribution de l'erreur sur les 100 points fréquentiels. $\bar{d}$ est l'erreur moyennée sur les 100 points fréquentiels de deux HRTF mesurées sur une position (θ, ϕ). Le calcul de ce critère revient à mesurer la variance de la distribution fréquentielle de l'erreur pour une position (θ, ϕ). Ce critère possède un terme quadratique, il est donc minimisable par la méthode des gradients. L'erreur d'approximation (moyennée sur les fréquences) d'une HRTF $H_{\theta,\phi,\lambda}$ s'exprime au moins du critère de Durand par la formule 2.4

$$E_{durand}(\theta, \phi, \lambda) = \frac{1}{100} \sum_{i=1}^{100} w_i \{20 \log_{10} \{\frac{H_{\theta,\phi,\lambda}(i)}{\hat{H}_{\theta,\phi,\lambda}(i)}\} - \bar{d}\}^2 \tag{2.4}$$

avec

$$\bar{d} = \frac{1}{100} \sum_{i=1}^{100} H^l_{\theta,\phi,\lambda}(i) - \hat{H}^l_{\theta,\phi,\lambda}(i) = \frac{1}{100} \sum_{i=1}^{100} 20 \log_{10} \{\frac{H_{\theta,\phi,\lambda}(i)}{\hat{H}_{\theta,\phi,\lambda}(i)}\} \tag{2.5}$$

Sensibilité et adéquation psycho-acoustique du critère

Pour donner une dimension perceptive a son critère, Durant utilise un coefficient de pondération w_n pour donner plus d'importances aux erreurs commises sur des points fréquentielles appartenant à la bande 900hz-10khz. Durant se base sur les travaux de Blommer et Wakefield pour considèrer que les points fréquentiels situés dans cette bande sont les plus important pour localiser des sources sonores. La fenêtre w_n est alors choisie de tel manière qu'une erreur de 10 dB située en un point fréquentiel hors-bande et une erreur de 1 dB dans la bande aient un même poids. Si i est l'indice d'un point fréquentiel de la bande 900hz-10khz et j l'indice d'un point fréquentiel situé en dehors de cette bande, alors $\frac{w_i}{w_j} = 10$.

2.3.5 Le critère Algazi

Définition et propriétés mathématiques

Le critère d'ALGAZI [5] calcule en la position (θ, ϕ) la "mean-squared error[2]" normalisée entre le module du spectre désiré, et le module du spectre modélisé.l'application de la fonction $\log_{10}$ à cette erreur augmentée d'un terme unité permet d'obtenir un résultat

[2]Le terme en anglais est laissé car communément utilisé

de 0-dB pour une modélisation parfaite. On s'est assuré qu'il est minimisable en effectuant le calcul de la descente de gradient. L'erreur d'approximation d'une HRTF $H_{\theta,\phi,\lambda}$ s'exprime au moins du critère d' Algazi par la formule 2.6.

$$E_{Algazi}(\theta,\phi,\lambda) = 10\log_{10}\left[1+\left[\frac{\sum_{i=1}^{M}(H_{\theta,\phi,\lambda}(i)-\hat{H}_{\theta,\phi,\lambda}(i))^2}{\sum_{i=1}^{M}H_{\theta,\phi,\lambda}(i)^2}\right]\right] \tag{2.6}$$

Sensibilité et adéquation psycho-acoustique du critère

A cause de son terme normalisateur, le critère Algazi est en théorie beaucoup plus sensible aux faibles erreurs (sur les points fréquentiels) que ne l'est la MSE. Le critère BARK+ALGAZI est la variante perceptive de ce critère. On l'obtient en pondérant les modules d'HRTF par les coefficients de Bark dans le calcul du critère d'ALGAZI.

2.3.6 Le critère Fahn (linéaire)

Définition et propriétés mathématiques

Le critère Fahn lin. [6] calcule en la position (θ,ϕ) la "mean-squared error" normalisée entre le module du spectre désiré, et le module du spectre modélisé.L'utilisation de ce critère en échelle d'amplitude logarithmique se nomme "Fahn log". Le critère Nishino [18] est une variante du critère Fahn log., ils sont de signes opposés. On s'est assuré que ces critères sont minimisable en effectuant sur chacun d'eux le calcul de la descente de gradient. L'erreur d'approximation d'une HRTF $H_{\theta,\phi,\lambda}$ s'exprime au moins du critère d' Algazi par la formule 2.6.

$$E_{Fahnlinaire}(\theta,\phi,\lambda) = \frac{\sum_{i=1}^{M}(H_{\theta,\phi,\lambda}(i)-\hat{H}_{\theta,\phi,\lambda}(i))^2}{\sum_{i=1}^{M}H_{\theta,\phi,\lambda}(i)^2} \tag{2.7}$$

$$E_{Fahnlog.}(\theta,\phi,\lambda) = 10\log_{10}\left[\frac{\sum_{i=1}^{M}(H_{\theta,\phi,\lambda}(i)-\hat{H}_{\theta,\phi,\lambda}(i))^2}{\sum_{i=1}^{M}H_{\theta,\phi,\lambda}(i)^2}\right] \tag{2.8}$$

Sensibilité et adéquation psycho-acoustique du critère

A cause de son terme normalisateur, le critère Fahn est en théorie beaucoup plus sensible aux faibles erreurs (sur les points fréquentiels) que ne l'est la MSE.

2.4 Tests de différents critères d'évaluation

Pour tester les critères d'évaluation présentés plus haut, deux types d'indicateurs ont été étudié :

- Les indicateurs servant à mesurer la sensibilité du critère à la caractérisation spectrale d'une erreur de localisation
- Les indicateurs servant à mesurer la sensibilité du critère à la caractérisation spectrale d'une erreur d'individualisation

Pour pouvoir comparer les critères entre eux, il est nécessaire de connaitre les ordres de grandeurs de ces critères : savoir pour chaque critère, à partir de quelle valeur l'erreur calculée devient audible. Des tests d'écoutes auraient été adéquates, mais on ne disposa pas du temps ni des sujets nécessaires à leurs réalisations. On a alors choisit de prendre comme valeur de référence, l'erreur calculée entre une HRTF brute et une HRTF lissée[3]. Des tests perceptifs réalisés par Nicol et Busson ont montré qu'il n'ya pas de différence perceptive entre les HRTF brute et lissées. Ainsi, en mesurant cette erreur, on **approche** la valeur minimal d'un critère, à partir de laquelle, l'erreur sera perçu par un auditeur. Cette valeur approchée sera notée E_{min}, notons toutefois qu'elle nous garantit seulement qu'une erreur de modélisation inférieure à celle-ci ne sera pas audible. Par contre, on n'est pas sur qu'une erreur supérieur à E_{min} sera forcément audible.

2.4.1 Choix d'un critère pour mesurer l'erreur de localisation

Les travaux de Langendijk et Bronkhorst [1] montrent que les creux et ventres spectraux situés entre 4khz et 16khz aident à la localisation avant/arrière et haut/bas. Une HRTF $\hat{H}$ modélisant une HRTF H lui sera perceptivement identique si elle conserve la position (sur l'échelle fréquentielle) et l'amplitude des creux et ventres spectraux de H. Un critère servant à mesurer des erreurs de localisation sera d'autant plus efficace perceptivement s'il est sensible aux erreurs sur les creux et ventres. En se basant sur cette observation, on a choisit de mesurer la sensibilité d'un critère aux creux et ventres spectraux.

La procédure de test consiste à mesurer pour chaque critère, l'erreur E_1 commise entre deux HRTF particulière H et $\hat{H}_1$ puis l'erreur E_2 commise entre H et $\hat{H}_2$. H est tirée de la base CIPIC, $\hat{H}_1$ est obtenu à partir de H en introduisant une erreur de 1 db sur tous ses points fréquentiels. En procédant ainsi, on s'assure que l'erreur introduite dans $\hat{H}_1$ conserve l'amplitude et la position des creux et ventres spectraux de H. $\hat{H}_1$ représente ainsi, au sens de Langendijk, une "bonne" modélisation de H. $\hat{H}_2$ est obtenu à partir de H en introduisant une erreur de 1 db sur tout les points fréquentiels sauf sur le 25^{ime} bins fréquentiel où l'erreur est augmentée de 10 db. On introduit ainsi un ventre spectrale dans $\hat{H}_2$ qui n'apparaissait pas sur le 25^{ime} bins fréqentiel de H (voir figure 2.1). Ce 25^{ime} bins fréquentiel a été choisi arbitrairement, toutefois, on a évité de choisir un bins fréquentiel des hautes fréquences à cause des coeffients de pondération des échelles perceptives. En effet, ces échelles affectent un coefficient quasi nul pour les hautes fréquences, les critères utilisant ces échelles ne seraient donc pas capable de prendre en compte l'erreur introduite sur le N^{ime} bins (avec N proche de 100) de $\hat{H}_2$. La modélisation $\hat{H}_2$ peut donc être jugée mauvaise au sens de Langendijk.

La comparaison des erreurs E_1 et E_2 va alors nous permettre de savoir si un critère fait la distinction entre les deux types d'erreurs E_1 et E_2. On saura ainsi lequel de ces critères est sensible à la conservation des creux ou ventres de l'HRTF modélisée. Si l'écart entre E_1 et E_2 se trouve très grand devant E_{min}, alors le critère qui a servit à les mesurer a été sensible à l'erreur introduite dans $\hat{H}_2$. Cela implique que ce critère tient compte de l'importance de la conservation des pics spectraux dans la modélisation. Si

[3] cf section 1.2.1

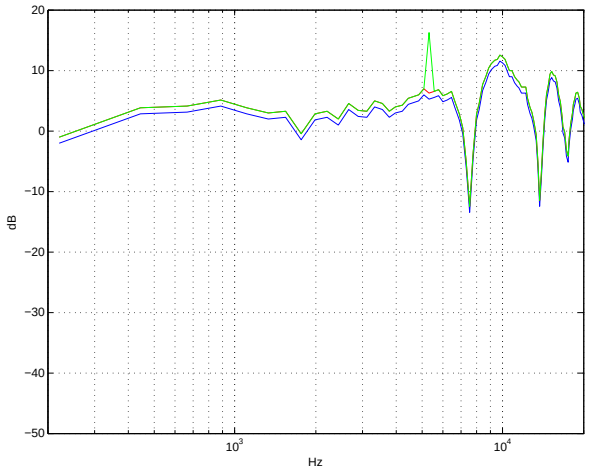

FIG. 2.1 – Représentation des HRTF de test. H est la courbe en bleu, les courbes en rouges sont celles de $\hat{H}_1$ et $\hat{H}_2$. On repère bien l'erreur introduite pour la fréquence 5.5khz sur $\hat{H}_2$

à l'inverse l'écart entre E_1 et E_2 est du même ordre que E_{min}, alors c'est que le critère utilisé pour les mesurer n'a pas été sensible à l'erreur introduite dans $\hat{H}_2$. Dans ce cas, le critère d'évaluation ne tient pas compte des erreurs sur les pics et ventres spectraux.

Une autre manière de comparer les critères consiste à les évaluer en considérant, au lieu de l'HRTF modélisée, une HRTF mesurée dans une autre direction. On a donc utilisé les HRTF d'un individu obtenues dans le plan horizontal. On les a regroupées par paires caractérisées par l'écart angulaire entre ces deux HRTF. On a évaluer les critères d'erreurs pour des paires d'écarts angulaires croissants, celles-ci correspondent ainsi à des paires d'HRTF de plus en plus différentes.

Résultats

Le tableau 2.1 présente les résultats de la procédure de test visant à mesurer la sensibilité des critères à l'erreur de localisation.
On voit que pour tous les critères sauf Algazi et Bark+Algazi, $|E_1 - E_2| < E_{min}$. Le critère d'Algazi obtient une erreur relative $|E_{1,algazi} - E_{2,algazi}| = 0.28dB$ légèrement supérieur à $E_{min,algazi}$. Ces critères n'ont donc été sensible à l'introduction de l'erreur dans H_2.

	Erreurs tests			
Critères d'évaluation	E_1	E_2	$\|E_1 - E_2\|$	E_{min}
$Chocqueuse$	1	1.1	0.1	2.0
$Bark$	0.010	0.011	0.001	0.007
$Algazi$	0.06	0.35	0.28	0.21
$Bark + Algazi$	0.08	0.12	0.04	0.02
Mse	1	2.2	1.2	11.5
$Durant$	0	0.99	0.99	10.8
$Durant + Bark$	0	0.009	0.009	0.04
$Fahnlinaire$	0.01	0.08	0.07	0.05
$Fahnlogarithmique$	-18.3	-10.8	-7.5	-13.1
$Nishino$	18.3	10.8	7.5	13.1

TAB. 2.1 – tableau des erreurs E_1 et E_2 évaluées par l'ensemble des critères d'évaluation

Une représentation par boxplot se révèle utile pour visualiser les distributions d'erreurs par paires d'écarts angulaires fixes.
Les figures 2.2 et 2.3 tracent pour deux critères d'évaluation, les boxplot des erreurs mesurées sur des paires d'HRTF d'écarts angulaires croissants. L'axe des abscisses représente l'écart angulaire en degrés entre les HRTF de la paire obtenue.
La représentation des Boxplot permet de visualiser et comparer les distributions d'erreurs pour des écarts angulaires constants en affichant :

- La valeur de la médiane des énergies des HRTF est représentée par un trait horizontale dans la "boîte à moustache" : la moitié de la distribution a des valeurs inférieurs à celle de la médiane.
- Les quartiles à 25% et 75%. Ces valeurs sont représentées par les bords de la "boîte à moustache".
- Les valeurs adjacentes correspondent à 1.5 fois la distance inter-quartile. Elles sont représentées par les "moustaches" de la "boîte à moustache".
- Les outliers : énergies supérieures et inférieures aux valeurs adjacentes. Elles sont affichées par une croix.

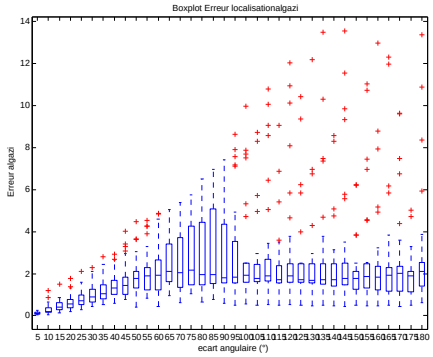

FIG. 2.2 – Boxplot des erreurs de localisation mesurées pour le critère Algazi.

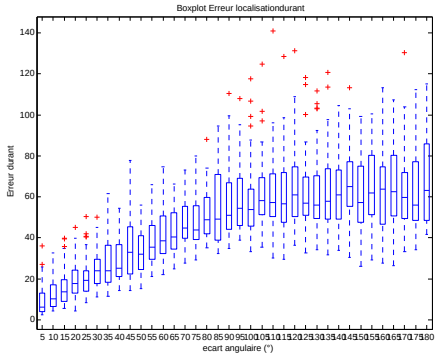

FIG. 2.3 – Boxplot des erreurs de localisation mesurées pour le critère Durant.

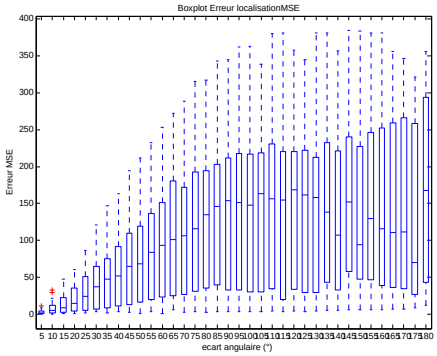

FIG. 2.4 – Boxplot des erreurs de localisation mesurées pour le critère MSE.

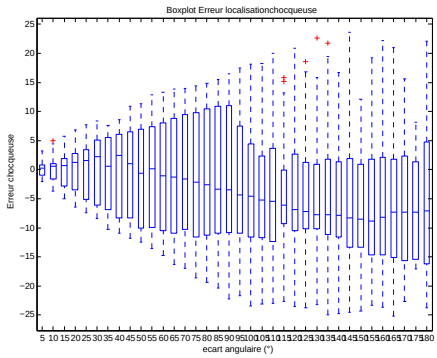

FIG. 2.5 – Boxplot des erreurs de localisation mesurées pour le critère Chocqueuse.

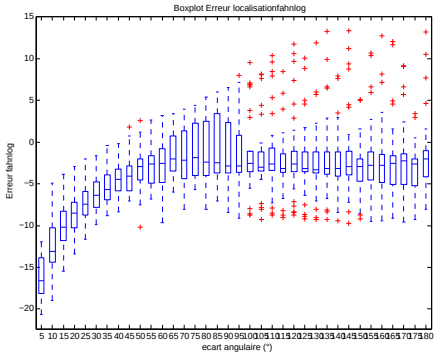

FIG. 2.6 – Boxplot des erreurs de localisation mesurées pour le critère Fahn logarithmique.

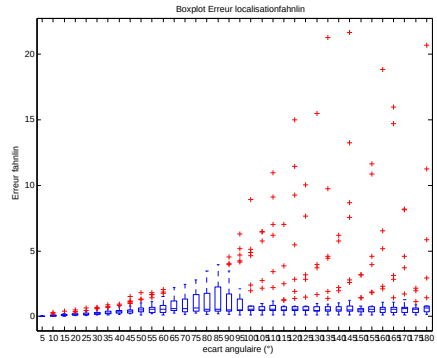

FIG. 2.7 – Boxplot des erreurs de localisation mesurées pour le critère Fahn linéaire.

Elles montrent qu'il n'est pas évident de trancher entre les critères. On observe que les critères Durant et Chocqueuse ont une variation monotone en fonction de l'écart angulaire alors que les autres stagnent. Les critères se distinguent aussi par l'étendue et la linéarité des variations. Pour le critère d'Algazi les valeurs restent très faibles, la moyenne varie entre 0 et 4, alors que pour le critère de durand, les valeurs en moyenne montent jusqu'à 80. Durant possède la meilleur linéarité, on voit bien une droite se dessiner à mesure qu'on augmente les écarts angulaires (voir figure 2.3). Considèrons maintenant le fait que les critères doivent donner une information cohérente pour des paires d'HRTF présentant le même écart angulaire Cela revient à vérifier sur chaque critère que les erreurs mesurées sur des paires d'écarts angulaires constant ont des valeurs proches.

Les distributions d'erreurs calculées avec la MSE sont illustrées sur la figure 2.4. On remarque qu'elles sont très dispersées. De plus sa croissance est quasi linéaire puis stagne pour des écarts angulaires supérieurs à 90 ^{o}.

Les figures 2.5, 2.3 montrent les distributions d'erreurs des critères Chocqueuse et Durant. Parmi les différents critères, ce sont eux qui ont le moins d'outliers. Cependant, le critère Durant est moins dispersé que le critère Chocqueuse. En effet, le critère Chocqueuse a des distances inter-quartiles beucoup plus élevées que celles du critère durand. Or, plus ces distances interquartiles sont élevées, moins le critère est robuste. C'est donc le critère Durand qui apportent l'information la plus cohérente sur l'erreur de localisation. Ainsi, on minimise l'erreur commise lorsqu'on étalonne ce critère. Par exemple, la figure indique que pour le critère durant, la valeur de 22.5 correspond à une erreur de localisation de 20 ^{o}. On pourrait effectuer le même raisonnement avec le critère chocqueuse, mais en commettant un erreur d'estimation plus grande.

D'autre part, l'étude des bornes des critères peut aussi nous aider à les comparer. Le critère Fahn Log tend vers $-\infty$ pour une bonne approximation d'HRTF. Un tel critère ne peut être utilisé comme fonction de coût d'un réseau de neurone. La méthode des gradients qu'utilise notre modèle impose en effet que la fonction de coût utilisée soit minimisable, ie, qu'elle tende vers zéro. Les critères Fahn lin., Algazi, Durant, Choqueuse et Bark sastisfont cette condition. Cette étude permet donc de retirer de la compétition les critères Fahn log. et Nishino.

2.4.2 Choix d'un critère pour mesurer l'erreur d'individualisation

Les critères d'évaluation sont ici interprétés en termes d'erreur d'individualisation. L'idée est de les évaluer en considérant, au lieu de la HRTF modélisée, l'HRTF mesurée à la même position, mais sur un individu différent. Ainsi, on a considéré les HRTF des 30 individus[4] mesurées pour un azimuth et une élévation. Ces HRTF sont regroupées par paires, on évalue les critères d'erreur pour des individus dont la différence morphologique croit. Middlebrooks montre que les dimensions de la conque jouent un rôle important dans les variations spectrales inter-individuelles des HRTF. Partant de ce constat, les donnees morphologiques prises en compte dans cette étude sont la largeur, la longueur et la profondeur de la conque. On désigne respectivement par $d_1(\lambda), d_2(\lambda), d_3(\lambda)$ les di-

[4]notre base contient 34 individus, mais 4 d'entres eux n'ont pas leurs données morphologiques répertoriés dans la base CIPIC

mensions de la conque citées précedemment pour l'individu λ. Ainsi les vecteurs d_1, d_2 et d_3 rassemblent les dimensions de la conque pour tous les individus. Pour chacunes de ces données morphologiques que nous fournit la base CIPIC, on classe les individus par ordre croissant selon la donnée morphologique considérée. Pour la largeur de la conque par exemple, l'individu λ^* ayant la plus petite valeur $d_1^{\lambda=\lambda^*}$ sera pris pour référence. On détermine λ^* en calculant $\lambda^* = argmin_\lambda(d_1(\lambda))$. On forme ensuite 29 paires de la forme $(H_{\lambda^*,\theta,\phi}, H_{\lambda,\theta,\phi})$ avec $\lambda \neq \lambda^*$. Ainsi, à partir de ces paires d'HRTF, on est capable d'évaluer les critères d'erreurs pour des écarts de dimensions morphologiques croissants, ce qui correspond à des HRTF de plus en plus "différentes".

Résultats

Les figures 2.8 et 2.9 (voir aussi en annexe B) présentent les erreurs d'individualisation mesurées avec le critère Fahn linéaire et (respectivement) Durant. L'axe des abscisses est indiciel, il représente l'écart croissant de dimensions morphologiques entre les individus sur lesquels ont été mesurées les HRTF de la paire considérée. La dimension morphologique considérée est la largeur de la conque. Pour un même écart morphologique, les résultats sont moyennés sur toutes les paires évaluées (courbe en bleu), l'ensemble des valeurs obtenues est également reproduit par des points verts.
Ces figures montrent que les critères, dans leur ensemble, ne sont pas sensibles à l'erreur d'individualisation causée par la variation d'une unique dimension morphologique[5].

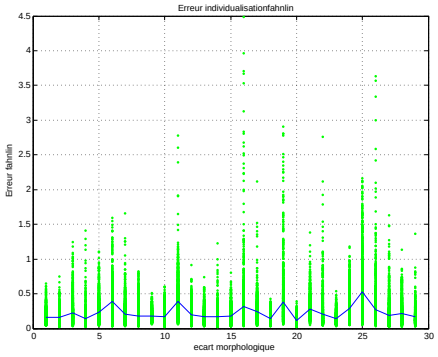

FIG. 2.8 – Erreur d'individualisation mesurées pour le critère Fahn linéaire

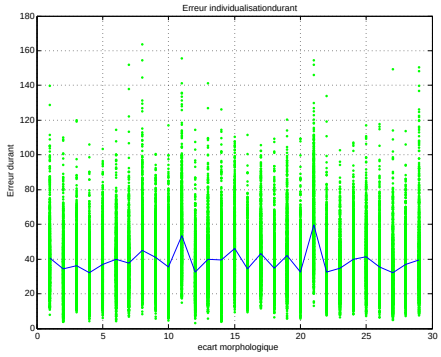

FIG. 2.9 – Erreur d'individualisation mesurées pour le critère Durant

En effet, on remarque que tous ces critères ont une évolution en moyenne constante, c'est à dire qu'ils mesurent la même erreur pour des paires d'HRTF mesurées sur des individus dont la différence morphologique croit. Cela est sûrement dû au fait que l'erreur d'individualisation est la conséquence d'une variation de non pas une, mais **plusieurs données anthropomorphiques**. On ne peut donc pas dire ici que pour une erreur de valeur x mesurée sur une paire d'HRTF $(H_{\lambda*,\theta,\phi}, H_{\lambda,\theta,\phi})$, on a un facteur y reliant $d_1(\lambda*)$ à $d_1(\lambda)$ par la relation : $d_1(\lambda*) = y(x).d_1(\lambda)$.

[5]la dimension considérée pour cette étude est la longueur de la conque

2.4.3 Discussion

Bien que les critères Algazi et Durant semblent se démarquer des autres critères, cette étude n'a pas permit d'élire un critère en particulier. C'est surtout l'élimination des critères Fahn log et Nishino qui transparait. Ces études nous permettent par contre d'envisager plusieurs pistes pour des travaux futurs.

Il serait intéressant pour les prochaines études de classer les individus par ordre croissant de différenciation morphologique en considérant cette fois toutes les données antropomorphiques fournies par la base CIPIC. En désignant $D_\lambda(i)$ le vecteur contenant la i-ème donnée morphologique de l'individu λ, on pourrait calculer quel est l'individu λ^*qui a la morphologie la plus "commune".
Un tel calcul se réaliserait à l'aide de la formule :

$$r_q = argmin_{\lambda^*} \sum_{\lambda=1}^{30} \sum_{i=1}^{15} ((D_{\lambda^*}(i) - D_\lambda(i))^2)$$

Une fois cette individu déterminé, on évaluerait les écart morphologiques entre tous les autres individus et celui-ci par un calcul de distances.
On pourra ainsi établir un classement des individus en fonction de leur ressemblance morphologique par rapport à celle de l'individu "commun". En procédant ainsi, on est sur qu'un tel classement s'est effectué en considérant l'ensemble des données morphologiques de la base CIPIC. On pourra alors calculer les erreurs d'individualisation en suivant la méthodologie décrite dans la section 2.4.2.

D'autre part, ce travail sur l'erreur d'individualisation a toutefois permis de constater que les observations de Middlebrooks [15] sont fondées. Les figures 2.10 montrent en effet que les HRTF d'une même position mais appartenant à des individus différents ont la même forme et se diffèrencient par translation sur l'échelle des fréquences.

Il serait alors intéressant de réfléchir sur la manière dont on pourrait utiliser une technique de warping [14] dans notre système. Celle-ci pourrait être introduite dans notre critère d'évaluation, ou même dans la fonction de coût du modèle. On remarque toutefois que comme pour les erreurs tracées sur la figure 2.10, le sens de translation des HRTF oscille entre gauche et droite quand on fait croitre la dissimilarité morphologique des individus sur lesquels elles ont été mesurées.

Dans cette optique où l'on se retrouve encore avec plusieurs critères sous la main, les phases d'élection d'HRTF représentantes et de modélisation serviront ,entre autres, à tester l'influence de ceux-ci sur la qualité de l'apprentissage effectué par nos outils statistique. On pourra alors confirmer, ou, infirmer la piste "Durant"/"Algazi". Cette étude sera détaillée dans la section 3.2.

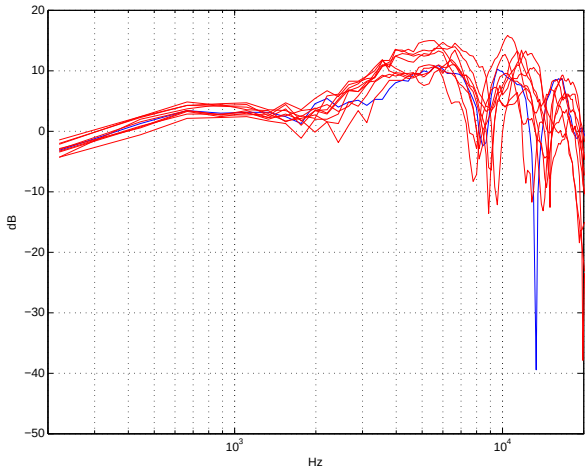

FIG. 2.10 – Caractérisation de l'erreur d'individualisation : on constate que les HRTF mesurées sur un même position mais pour des individus différents se translatent sur l'axe fréquentiel

Chapitre 3

Détermination d'HRTF représentatives

Le but de l'étude présentée dans ce chapitre est d'observer les variations spatiales et spectrales ainsi que les dépendances individuelles des HRTF.

3.1 Introduction

L'objectif de cette étude, est de réduire de façon significative le nombre de mesures d'HRTF nécessaires à la restitution fidèle d'une scène auditive. Cette phase passe par exemple par un travail de clustering sur la base de donnees CIPIC[1] visant à regrouper entre elles les HRTF "semblables spectralement" et de remplacer les N groupes déterminés par N HRTF représentantes. Ce travail, a l'année dernière été réalisé sur un seul individu, en conjecturant que les positions des HRTF représentantes de cet individu étaient valables pour les 45 individus de la base. Autrement dit, cette hypothèse revenait implicitement à supposer que si on effectuait la même étude sur les HRTF de tous les individus de la base, on verrait des regroupement d'HRTF effectués par rapport à leurs seuls positions et communs à tous les individus. Mais on pourrait obtenir toute autre chose, par exemple un regroupement des HRTF par individus, ie, trouver des groupes d'individus aux morphologies semblables qui possédent les mêmes positions d'HRTF représentantes, on pourrait aussi tout simplement ne retrouver aucun regroupement...
Notre méthode de clustering nous permettra de répondre à la question suivante :

– Peut-on obtenir des regroupements géographiques d'HRTF communes à tous les individus ?

Le deuxième objectif de cette étude est de mesurer l'influence de certaines représentations d'HRTF et certains critères d'évaluation sur le processus d'apprentissage des cartes de Kohonen. En effet, les cartes de Kohonen mesurent la similarité fréquentielle des HRTF pour les regroupées. Cette mesure de similarité à été effectuée au moyen de cinq

[1] La base de donnée CIPIC est décrite dans la section 1.2

critères : BARK, MSE, ALGAZI, BARK+ALGAZI et Durant[2]. Il a donc été construit cinq cartes de Kohonen, chacune utilisant un des critères d'évaluation comme mesures de similarités pour son processus d'apprentissage. La qualité des regroupements, ie, le choix + ou - judicieux des positions de représentants obtenues pour ces différentes cartes sera évaluée par un calcul d'erreurs de quantifications. Cette évaluation sera aussi réalisée avec différents critères (les mêmes que précédemment).
Aussi, pour pouvoir évaluer l'influence des prétraitements sur notre étude, nous avons utilisé deux jeux de cartes de kohonen. Le premier jeu contient cinq cartes de kohonen utilisant chacunes un critère d'évaluation (BARK, MSE, ALGAZI, BARK+ALGAZI ou Durant[3]. Ces cartes utiliseront les données brutes de la base CIPIC, les études de ce jeu seront alors qualifiées de "Clustering (+*nom du critère utilisé*) sur données brutes". Le deuxième jeu de cartes contient le même nombre de cartes utilisant les mêmes critères pour apprendre, par contre, elles sont destinées à effectuer une analyse exploratoire sur des données égalisées champs diffus et lissées[4]. On qualifiera alors les études de ce jeu par "Clustering (+*nom du critère utilisé*) sur données égalisées champs diffus et lissées (ECDL)" Ce chapitre se décompose comme suit : dans un premier temps, nous présenterons la méthodologie utilisée pour obtenir les regroupements, à travers l'étude "Clustering BARK+ALGAZI sur les données CIPIC égalisées champs diffus et lissées", cela permettra au lecteur de comprendre notre démarche expérimentale. Enfin, nous comparerons la qualité des regroupements par leurs erreurs de quantifications.

3.2 Méthodologie du Clustering

Cette partie développe à travers un exemple l'étude statistique qui a été effectuée sur tous les individus de la base CIPIC Chacunes des quatres études que nous avons menées s'organisent autour de trois étapes :

- Le regroupement des HRTF en fonction de leurs composantes spectrales : clustering par carte de kohonen et par CHA
- La projection et la visualisation des variables cibles (autres que fréquentielles) sur les cartes de kohonen
- Le calcul de l'erreur de quantification pour évaluer notre méthode de réduction du nombre de mesures

3.2.1 Clustering par Carte de Kohonen

Cette étude, a porté sur les 1250 HRTF des 34 individus de la base CIPIC (une fois nettoyée, voir l'annexe C), soit 42500 HRTF. L'utilisation de cartes de kohonen va nous permettre d'évaluer aisément les variations fréquentielles et spatiales inter-individus des HRTF.

Les cartes de Kohonen sont des réseaux de neurones qui appartiennent aux algorithmes de clustering. Une carte de Kohonen se compose d'un ensemble de k-points de

[2] voir section 2.3
[3] voir section 2.3
[4] ce prétraitement est décrit dans la section 1.2.1 p. 11

l'espace liés entre eux par des relations de voisinage. L'ensemble de ces relations de voisinage constitue la topologie de la carte. Ces k points de l'espace sont appelés neurones et sont des vecteurs avec autant de composantes que les vecteurs d'entrée à classer.
En ce qui concerne la topologie de la carte , il en existe plusieurs en 2D ou 3D, en tore ou avec des voisinages hexagonaux et avec des tailles plus ou moins grandes. Chocqueuse [7] démontre que la topologie de la carte de Kohonen la plus adaptée aux données que nous avons à regrouper, est une carte 2D à voisinage hexagonal de topologie 12x12.
Il s'avère qu'une carte 12x12 présente une bonne répartition des données avec peu de neurones vides. Une fois la topologie choisit, la carte de Kohonen est prête pour effectuer un apprentissage non supervisé.

L'apprentissage de la carte de Kohonen met en correspondance l'espace des entrées, espace dans lequel ont été les données d'entrée, et la carte. Cette phase consiste à adapter les composantes des neurones de la carte de telle manière que des exemples proches dans l'espace d'entrée soit associés au même neurone ou à des neurones proches dans la carte. Il faut donc bien, dés à présent, comprendre que la topologie définit à priori des relations de voisinage entre points de la carte et que ces relations de voisinage n'ont a priori aucun caractère métrique : a priori, des "voisins" au sens de la topologie peuvent être très éloignés dans l'espace et de même, a priori, des points très proches de la carte peuvent ne pas être voisins. On insiste sur les mots "a priori" : en effet, tout le processus d'apprentissage de la carte de Kohonen consiste d'une certaine façon à faire en sorte que, quand il est possible, le voisinage topologique de la carte corresponde à une proximité métrique.
Un exemple de carte de Kohonen 2D à voisinage hexagonal de taille 12x12 obtenue après apprentissage est présenté sur la figure 3.1.

Sur la figure 3.1, chaque hexagone est un neurone. Un neurone est représenté par un vecteur, appelé vecteur poids, qui possède autant de composantes que les vecteurs d'entrée (dans notre cas, 100 composantes spectrales). Après convergence de l'algorithme d'apprentissage, on évalue pour chaque HRTF quel est le vecteur poids le plus proche (au sens d'une distance). Le neurone possédant le vecteur poids le plus proche est appelé "neurone gagnant" ou BMU (Best Matching Unit). On indique le nombre d'HRTF "gagné", (on parle alors de Hits) par chaque neurone via un losange noir de taille variable.

L'algorithme d'apprentissage se déroule ainsi :

1. Initialisation des vecteurs poids (neurones) de la carte de Kohonen
2. On présente un vecteur d'entrée à la carte
3. On calcule la distance entre ce vecteur d'entrée et tous les vecteurs poids de la carte : c'est la phase de compétition entre tous les neurones de la carte
4. On détermine le neurone gagnant
5. On modifie le vecteur poids du neurone gagnant ainsi que ceux des neurones de son voisinage de tel sorte que ceux-ci se rapproche encore plus du vecteur d'entrée
6. On retourne à l'étape 2

Afin d'éviter le "sur-apprentissage", phénomène qui conduit le réseau à apprendre par coeur les données de l'ensemble d'apprentissage et ainsi perdre toutes ses facultés de généralisation, nous avons forcé l'algorithme d'apprentissage a s'arrêter au début de la phase de stabilisation de la fonction d'erreur. Pour ce faire, nous avons dû vérifier toutes les 10 000 itérations, la diminution de l'erreur de quantification, si celle-ci devenait inférieur à

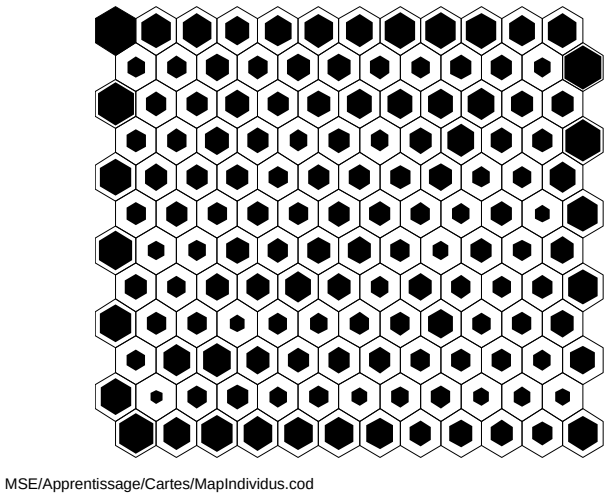

FIG. 3.1 – Carte de Kohonen à voisinage hexagonal de taille 12×12 neurones. La taille des losanges noirs indique le nombre d'HRTF compris dans chaque neurone.

une valeur seuil $(10^{-}4)$ on arrêtait l'apprentissage. La figure 3.2 montre par exemple que la phase d'apprentissage sur l'ensemble d'apprentissage a été stoppé au bout de 650000 itérations.

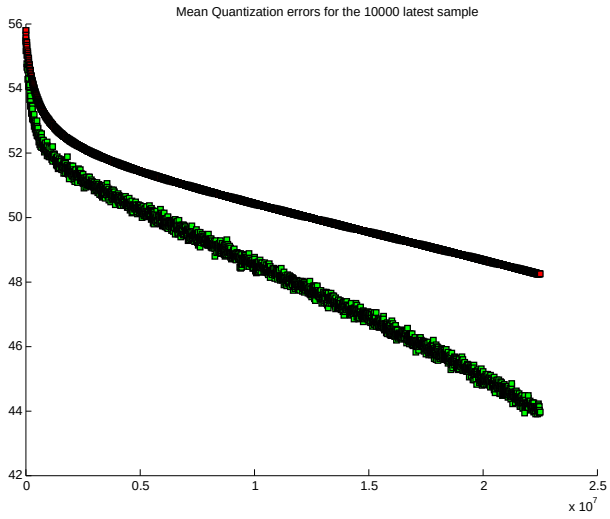

FIG. 3.2 – Convergence d'une carte de Kohonen : tracé de l'erreur en fonction du nombre d'itérations.

On ne détaillera pas plus l'algorithme, le lecteur intéressé pourra se reporter à l'ouvrage de référence dans ce domaine. Cependant, nous allons nous arrêtez sur la distance qu'utilisent nos cartes de Kohonen pour déterminer le neurone gagnant.

La carte de kohonen standard effectue son apprentissage non supervisé avec une distance euclidienne classique :

$$Erreur(H, \hat{H}) = \sum_{i=1}^{100}((H(i) - \hat{H}(i))^2) \tag{3.1}$$

Pour pouvoir comparer l'influence des critères d'évaluation sur les résultats du clustering, nous avons choisis d'utiliser ces critères pour évaluer la proximité entre les vecteurs d'entrées et les neurones de la carte de Kohonen. Nous avons donc eu à réaliser cinq simulations, la première apprenant avec la MSE, la deuxième avec la distance du critère de BARK, la troisième avec la distance du critère d'Algazi, puis la quatrième avec une distance combinant le critère de BARK et d'ALGAZI et enfin la cinquième apprenant avec le critère Durant.

Pour faire apprendre la carte avec les quatres critères choisis, on avait le choix entre :
- remplacer la distance euclidienne par la distance du critère.
- modifier les donnees d'entrée de manière à retrouver la distance du critère d'évaluation lorsqu'on leurs applique la distance euclidienne.

Par exemple, pour passer d'une distance euclidienne $\sum_{i=1}^{100}((H(i) - \hat{H}(i))^2)$ à la distance du critère perceptif de Bark $\sum_{i=1}^{100} \alpha_i((H(i) - \hat{H}(i))^2)$, on peut :

- soit remplacer directement les deux distance et ne pas toucher aux données d'entrées
- soit multiplier la matrice de données par un coefficient $\sqrt{\alpha}$

En effet, en posant $H' = H.\sqrt{\alpha}$ et $\hat{H}' = \hat{H}.\sqrt{\alpha}$, on voit facilement que :

$$\sum_{i=1}^{100} \alpha_i((H(i) - \hat{H}(i))^2) = \sum_{i=1}^{100}((\sqrt{\alpha_i}H(i) - \sqrt{\alpha_i}\hat{H}(i))^2) = \sum_{i=1}^{100}((H'(i) - \hat{H}'(i))^2) \tag{3.2}$$

Pour pouvoir comparer les cartes de Kohonen (obtenues après apprentissage) de ces cinq études, nous avons choisis de ne pas toucher aux données d'entrées, pour apprendre avec les distances des critères d'évaluation. En effet, s'il on avait choisit de faire rentrer les critères dans les données, on se retrouverait avec quatres cartes de Kohonen ayant appris avec la même distance mais sur des données d'entrées différentes. Cela aurait pour conséquence de modifier la répartition des observations sur la carte de Kohonen pendant la phase de compétition.

Enfin, afin de déterminer le format le plus utile des HRTF pour être appliqué en entrée du modèle (voir le chapitre 4), on a effectué l'apprentissage de cinq cartes (une par distance) sur une base de donnée d'HRTF brutes et cinq autres (utilisant les mêmes distances) sur une base d'HRTF égalisées champs diffus et lissées.

3.2.2 Projection et visualisation des variables cibles

L'étape d'apprentissage aboutit au regroupemment des 22500 HRTF de l'ensemble d'apprentissage sur les 144 neurones des cartes de Kohonen (voir figure 3.1). L'étape de

projection consiste à proposer en entrée de la carte des vecteurs de l'ensemble de validation ou de test, puis par comparaison de ces exemples avec tous les vecteurs poids de la carte, à ranger chacunes des HRTF dans le neurone ayant des composantes qui s'en rapproche le plus au sens de la distance utilisée (ici, MSE, BARK, ALGAZI et BARK+ALGAZI, DURANT). On obtient ainsi pour les cinq études, deux nouvelles cartes de Kohonen 12x12 regroupant respectivement les HRTF des ensembles de validation et de test en 144 neurones.

Ces regroupements ont été réalisés en ne considérant que les caractéristiques spectrales des HRTF. Or les HRTF dépendent de l'individu sur lesquelles on les mesures, et de la position spatiale où elles ont été mesurées (exprimée en terme d'azimuth et d'élévation) et l'on cherche à savoir si des HRTF mesurées dans des zones géographiques proches se regroupent dans les mêmes neurones ou dans des neurones voisins sur la carte de Kohonen. Afin de répondre à cette question, pour chacun des neurones de la carte, on va récupérer l'azimuth et l'élévation de toutes les HRTF gagnées, puis on va calculer les moyennes puis les écarts-types pour ces deux variables, et cela, pour tous les neurones de la carte.

Sur les figures 3.3 et 3.4, on observe la visualisation des moyennes et respectivement, la visualisation des écarts types des azimuth calculées, pour chaque neurone de la carte, sur toutes les HRTF qu'il a gagné. Les hexagones clairs correspondent à des neurones contenant en moyenne des HRTF situées à des azimuts élevés (+80 ^{o}) et les hexagones foncés, à des neurones contenant en moyenne des HRTF situées à de azimuts faibles (-80 ^{o}).

Ces graphes de visualisation vont aussi et surtout nous permettre observer les problèmes d'homogénéité non souhaitables avec des regroupements de données spatialement éloignées dans l'espace de mesures et spatialement proches sur la carte de Kohonen. Ce problème d'homogénéité s'observe sur la carte de deux manières :

- on constate une non-homogénéité au niveau Intra-neurones : on la caractérise par un regroupement au sein d'un même neurone, d'HRTF mesurées pour des positions spatialement éloignées.
- on constate une non-homogénéité au niveau Inter-neurones : on la caractérise par un regroupement d'HRTF spatialement éloignées dans des neurones voisins.

L'interprétation de ces graphes consistera alors à comparer la visualisation de la moyenne (révélant l'existence de non-homogénéité Inter-neurones) avec la visualisation de l'écart-type (révélant l'existence de non-homogénéité Intra-neurones), et cela pour chaque variable, afin de constater selon quelles variables s'est effectué le regroupement des HRTF. Les variables que l'on visualisera sur la cartes de Kohonen sont :

- les variables de positions : l'azimut θ et l'élévation ϕ, pour savoir si les regroupements des HRTF se sont faits par rapport à leurs proximité spatiales.
- les variables morphologiques : la largeur de la tête notée $x1$ et la profondeur de la tête notée $x3$ [5], pour savoir si les regroupements d'HRTF se sont fait par type de

[5] Ces données anthropomorphiques sont fournies par la base de donnée CIPIC, cf [10]

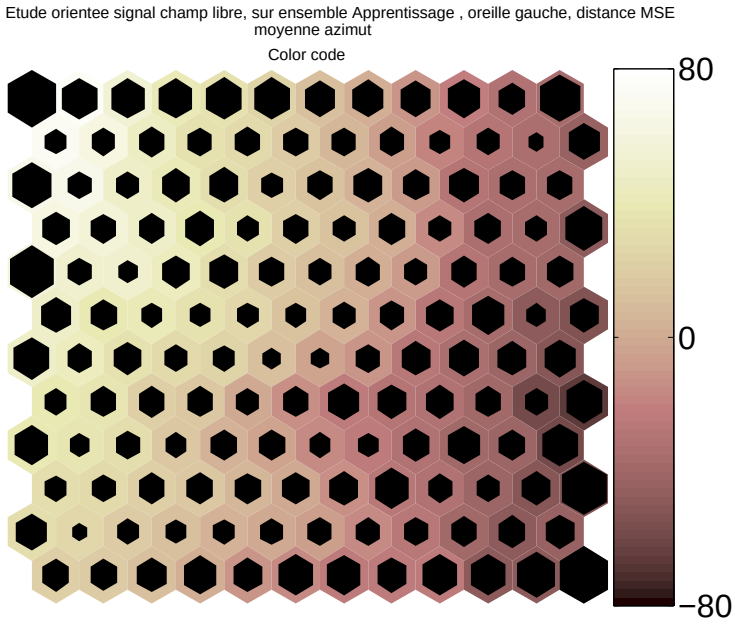

FIG. 3.3 – Visualisation des moyennes des azimuts sur une carte de Kohonen de taille 12×12. Les hexagones clairs correspondent à des neurones contenant en moyenne des HRTF situées à des azimuts élevés ($+80^o$) et les hexagones sombres, à des neurones contenant en moyenne des HRTF situées à des azimuts faibles (-80^o).

morphologies. concernant ce choix de données antropomorphiques, on se base sur les travaux de Middlebrooks (voir [15]) qui montrent qu'elles jouent un rôle considérable dans la localisation auditive.

3.2.3 Clustering de la carte par CHA

Pour faciliter l'analyse quantitative de la carte, nous allons diminuer le nombre de cluster à étudier en regroupant les neurones proches. Ce regroupement est réalisé en utilisant une classification hiérarchique ascendante (CHA) sur les neurones de la carte.
La CHA se présente comme un arbre avec comme base plusieurs classes. Dans notre cas, les classes de départ correspondent aux neurones obtenus par cartes de Kohonen. A chaque étape de la construction de l'arbre, on recherche les deux classes les plus proches au sens d'une distance et on fusionne. Le critère de regroupement que nous avons choisi est la méthode de maximisation de la variance intra-classe (Critère de Ward).
Pour visualiser le comportement de la classification, nous construisons un arbre appelé dendogramme 3.5.

Nous effectuons ensuite une troncature de l'arbre qui correspond à une valeur du critère de ward. Il n'existe pas de méthode systématique pour déterminer la valeur optimale de ce critère, il faut donc essayer plusieurs valeurs expérimentalement avec, comme objectif, d'obtenir un bon compromis entre le nombre de classes et la perte inévitable de la

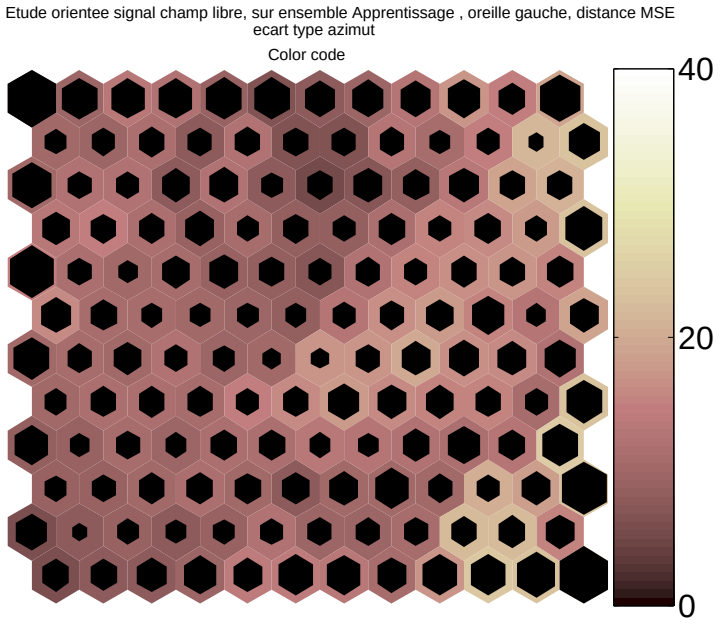

FIG. 3.4 – Visualisation des écarts-types des azimuts sur une carte de Kohonen de taille 12×12. Les hexagones clairs correspondent à des neurones contenant des HRTF situées à des azimuts éloignées et les hexagones sombres, à des neurones contenant des HRTF situées à des azimuts proches.

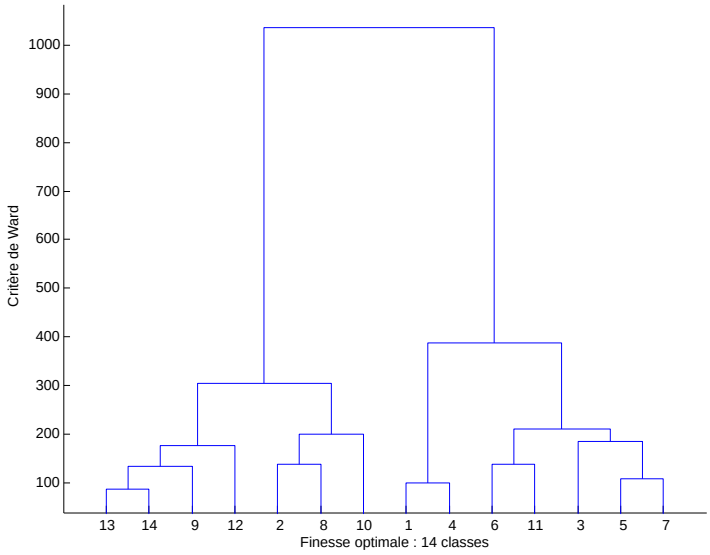

FIG. 3.5 – Exemple de dendrogramme. La base de l'arbre est composée de 144 clusters. A chaque étape de la construction de l'arbre, on regroupe les deux clusters les plus proches. Le sommet de l'arbre contient un seul cluster

variance intra-classe. Nous avons choisis la valeur 0.5 pour ce critère, on passe de 144 cluster à un nombre entre 11 et 15 cluster selon la carte de kohonen utilisée et les dimensions du tableau d'entrée[6] de celle-ci.

3.2.4 Election des positions représentantes

Le but de cette sous-section est de répondre à la question : "pour un cluster donné quel est l'HRTF qui "représente" le mieux l'ensemble des HRTF appartenant à ce cluster ?" (voir figure 3.6)

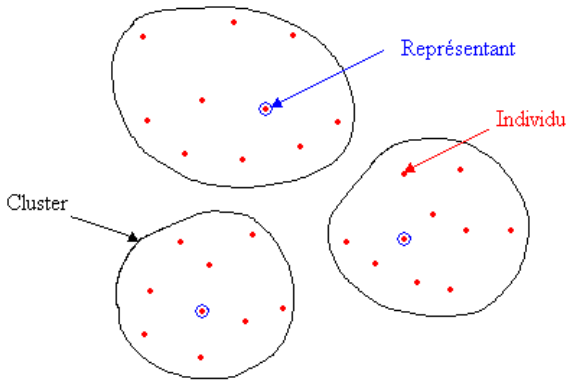

FIG. 3.6 – Des clusters et leur représentant

Nous proposons le critère d'élection suivant : Le représentant r_q du Cluster C_q est l'HRTF dont le vecteur est le plus proche de tous les vecteurs du cluster. Le calcul s'effectue au moyen d'une distance d $\backslash : (X, Y) \in \mathcal{R}^{100} \times \mathcal{R}^{100} \longmapsto d(X, Y) \in R\mathcal{R}$. Cette distance sert à mesurer les écarts entre les HRTF d'un même cluster. Pour une carte de Kohonen qui a effectué son apprentissage avec le critère "X" (ie, MSE,BARK,ALGAZI ou DURANT) comme mesure de similarité, on utilisera comme distance le critère "X".

$$r_q = argmin_{n \in C_q} \sum_{j=1}^{length(C_q)} (d(H_{\theta_n,\phi_n,\lambda_n}, H_{\theta_j,\phi_j,\lambda_j})) \tag{3.3}$$

Cette phase d'élection des représentants permet d'obtenir une liste de n positions spatiales correspondant aux points de mesures des HRTF représentantes.

Nous noterons par la suite (θ^*, ϕ^*) les coordonnées de l'HRTF "représentante" de l'HRTF mesurée à la position (θ, ϕ).

La figure 3.7 offre quant à elle, une vue plane de la position des représentants élus, ainsi que des positions moyennes des 144 HRTF (les vecteurs poids en fait) de la carte de Kohonen. On pourra ainsi se faire une idée des zones de l'émisphère "importantes", ie, dans lesquelles, quelques mesures d'HRTF sur l'utilisateur final seraient nécessaires et suffisantes pour la restitution d'une scène sonore fidèle.

[6] On procédera à un scindage du tableau d'entrée, voir section 3.4

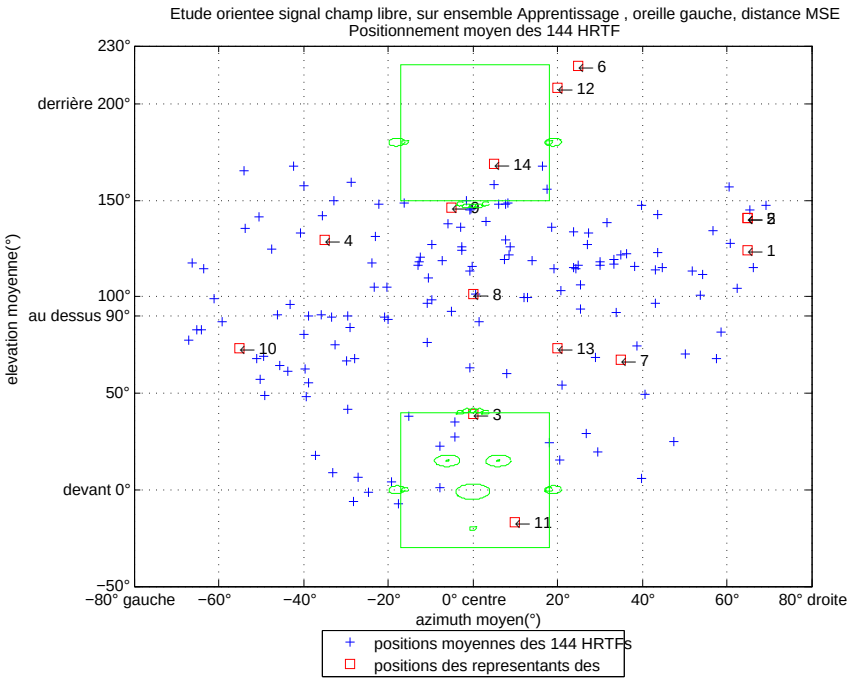

FIG. 3.7 – Position spatiale des représentants. les croix indiquent la position spatiale d'un représentant issue de l'apprentissage de la carte de kohonen (un vecteur poids). Les carrés indiquent la position d'un représentant issue de la CHA

L'erreur de quantification[7] va nous permettre d'évaluer les performances de notre méthode d'élection des HRTF représentantes. On calcule cette erreur à l'aide des quatres distances utilisées pour l'apprentissage des cartes[8].

la démarche utilisée pour déterminer les n représentants des n clusters issues de la CHA peut être utilisée pour élire les représentants des 144 clusters (neurones) de la carte de Kohonen. Une fois que le clustering+CHA sur les quatres distances a été réalisé, et que les représentants ont été élus pour les 144 neurones et les n clusters de la CHA, on procède aux calcul des erreurs de quantification. L'évaluation des HRTF représentantes obtenues par les différentes cartes de Kohonen consistera à calculer les erreurs de quantification.

On pourra alors répondre aux questions suivantes :

– L'utilisation d'un des critères pour l'apprentissage des cartes de kohonen permet-elle d'améliorer considérablement l'erreur de quantification ?
– Le clustering sur tous les individus de la base CIPIC permet-il de trouver des positions de représentants offrant une erreur de quantification moyenne meilleure que celle obtenue avec les positions uniformément réparties sur la sphère de mesure.

Dans la partie suivante, nous allons interpréter les résultats obtenus après avoir appliquer la même méthodologie sur la carte de kohonen réalisant le clustering BARK+ALGAZI sur données égalisées champs diffus/lissées.

[7] définie dans la section 1.3.1

[8] les formules des distances utilisées figurent dans la section 2.3 p. 20

3.3 Analyse des resultats de l'étude "Clustering sur la carte BARK+ALGAZI ECDL"

3.3.1 Répartition des HRTF sur la Carte de Kohonen 12x12

La carte de Kohonen construite (avec la distance BARK+ALGAZI) sur les données d'apprentissage et observable sur la figure 3.8 ne présente aucun neurone vide, on constate au contraire une bonne répartition des HRTF sur tous les neurones de la carte. La figure 3.9 présente la projection des données de test sur la carte de Kohonen, on constate qu'elles sont similaire, c'est à dire que pour chaque neurones, on retrouve la même proportion d'HRTF gagnées, on s'assure donc que les facultés de généralisation de la carte ont été conservées (il n'y a pas eu de sur-apprentissage).

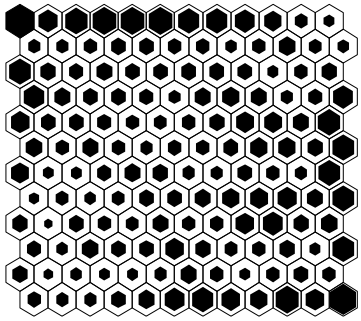

FIG. 3.8 – Répartition des HRTF de l'ensemble d'apprentissage sur une carte de Kohonen de taille 12×12. La taille des losanges noirs indique le nombre d'HRTF compris dans chaque neurone.

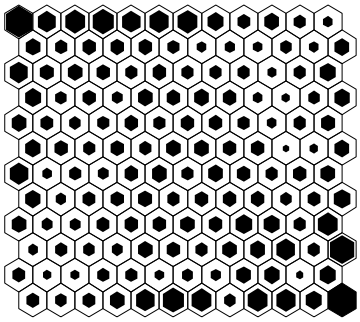

FIG. 3.9 – Répartition des HRTF de l'ensemble de test sur une carte de Kohonen de taille 12×12. La taille des losanges noirs indique le nombre d'HRTF compris dans chaque neurone.

3.3.2 Interprétation des graphes de visualisation

L'observation de les figures 3.10 et 3.11 sur laquelle figure les moyennes et écart-types de la variable des azimuths, permet d'affirmer que le regroupement des HRTF c'est au moins effectué par rapport à leur proximité selon les azimuth. En effet, ces deux cartes ne présentent aucunes non-homogénéités. La carte de la figure 3.11 présente des écart-types très faible (entre 5 et 10 pour des valeurs d'azimuth allant de -80 ^{o} à 80 ^{o}), tandis que sur la carte de la figure 3.10, on observe une très bonne répartition progressive entre les HRTF mesurées en des azimuths élevés (situés sur la région Nord Ouest de la carte), et les HRTF mesurées en des azimuths faibles (situés sur la région Sud Est de la carte). Ainsi, la faible valeur des écart-types prouve que la moyenne des azimuth, calculée sur les HRTF de chaque neurones, donne une idée assez précise de la distribution des azimuth. On s'assure ainsi du fait que, dans les différents clusters formé après la CHA, les HRTF

de l'ensemble d'apprentissage seront regroupées par proximité selon les azimuth.

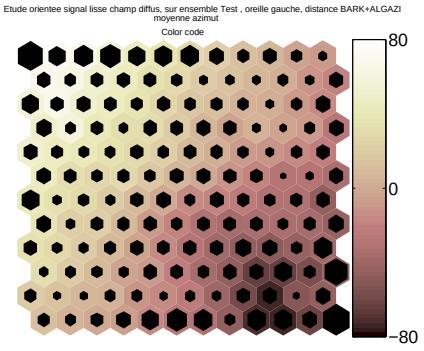

FIG. 3.10 – Moyenne de l'azimut : Les hexagones clairs correspondent à des neurones contenant en moyenne des HRTF situées à des azimuts élevés ($+80^o$) et les hexagones sombres, à des neurones contenant en moyenne des HRTF situées à des azimuts faibles (-80^o).

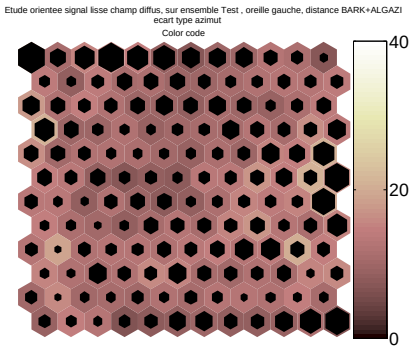

FIG. 3.11 – Ecart type de l'azimut : Les hexagones clairs correspondent à des neurones contenant des HRTF situées à des azimuts éloignées et les hexagones sombres, à des neurones contenant des HRTF situées à des azimuts proches.

Une autre façon d'obtenir des informations sur ces visualisations serait de les comparer à la carte des U-matrice de la figure 13, on constate alors que les HRTF aux azimuth élevées se place dans la région Nord-Ouest de la carte et que les HRTF aux azimut proches du plan médian (0 ^{o}) se place sur la diagonale Sud-Ouest Nord-Est, région où, les vecteurs poids sont très différent d'un voisin à un autre. Cela semble vouloir dire que la variation fréquentielle des HRTF augmente quand les azimut des positions de mesures se rapprochent du plan médian.

En effectuant la même analyse sur les figures 3.13 et 3.14, on observe que l'hypothèse affirmant que la répartition des HRTF s'effectue aussi selon l'élévation, est beaucoup moins évidente. En effet, on peut constater sur les visualisations de la figure 14 que sur la diagonale Nord-Ouest Sud-Est de la carte de Kohonen, on observe des non-homogénéités aussi bien au niveau inter-neurones qu'intra neurone.

D'un point de vue physique, ces résultats nous laissent déduire que des HRTF mesurées en des positions extrêmes en élévation (vers -50 ^{o} et 230 ^{o}) sont fréquentiellement proches. Cela veut aussi dire que les ondes sonores réfléchies sur le sol avant d'arriver aux oreilles de l'auditeur sont très peu utilisées par celui-ci pour localiser la source sonore. Cependant, au regard de la U-matrice de la figure 3.12, les neurones ayant gagnés des HRTF mesurées en des élévations élevées ont des vecteurs poids différents de ceux qui ont gagné des HRTF mesurées en des élévations faibles. Cette dernière observation nous empêche de conclure qu'en considérant les valeurs extrêmes d'élévation -50 ^{o} et 230 ^{o} comme équivalente, on obtiendrait une répartition homogène et continue sur toute la carte.

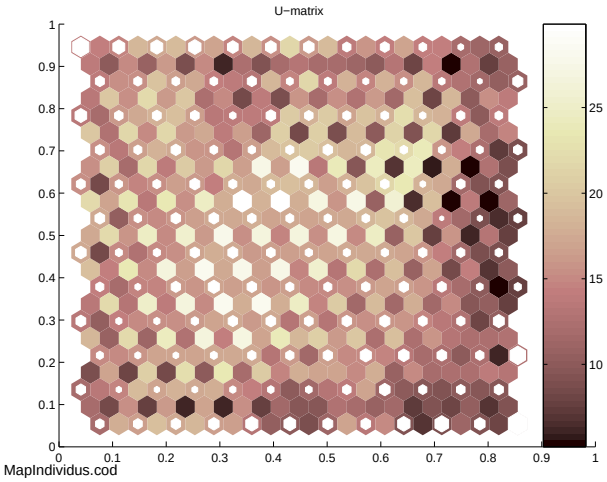

FIG. 3.12 – Carte des Umatrice : étude "Clustering sur la carte BARK+ALGAZI ECDL"

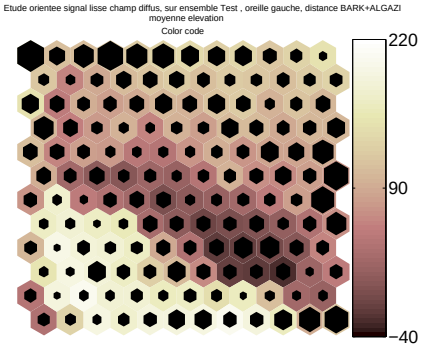

FIG. 3.13 – Moyenne de l'élevation : Les hexagones clairs correspondent à des neurones contenant en moyenne des HRTF situées à des élévations élevés (230^o) et les hexagones sombres, à des neurones contenant en moyenne des HRTF situées à des élévations faibles (-50^o).

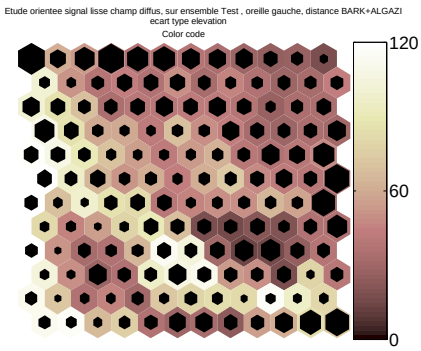

FIG. 3.14 – Ecart type élévation : Les hexagones clairs correspondent à des neurones contenant des HRTF situées à des élévations éloignées et les hexagones sombres, à des neurones contenant des HRTF situées à des élévations proches.

Pour éviter ces non-homogénéités, nous avons opté pour la solution proposée par Choqueuse [7], à savoir, sectionner notre ensemble de vecteurs en deux ensembles : un, contenant les HRTF situées à l'avant et l'autre, contenant les HRTF situées à l'arrière. Notre découpage, suivant le plan frontal, permet d'empêcher tout problème dans les regroupements en forçant les données de l'émisphère avant et de l'émisphère arrière à se séparer dès le départ.

3.4 Analyse des resultats du Clustering par Carte de Kohonen BARK+ALGAZI avec Séparation des données Avant/Arrière

3.4.1 Interprétation des graphes de visualisation

Le scindage de la base de données, tout en conservant les résultats obtenus sur l'analyse de la carte des Hits, la U-matrice, la carte des moyennes et écart-types des n ^{o} d'individus et de l'azimut, améliore considérablement la répartition en élévation des HRTF sur la carte. On obtient en effet des cartes représentant l'élévation beaucoup moins perturbées (voir figures 3.15, 3.16, 3.17, 3.18). Les différentes élévations sont rangées sur la diagonale Sud-Ouest Nord-Est pour les émisphères avant et arrière, et leur évolution le long de cette diagonale est cette fois-ci continue. La carte des écart types présentant pour les deux émisphères une teinte bleu dominante (indiquant de faibles valeurs d'écart-type), on obtient alors un regroupement homogène des HRTF selon les élévations, et cela aussi bien au niveau intra-neurone qu'inter-neurones.

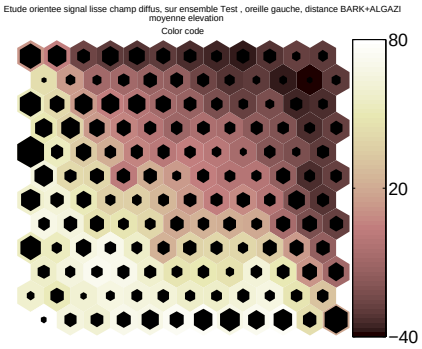

FIG. 3.15 – Moyenne de l'élévation sur la carte de l'hémisphère avant.

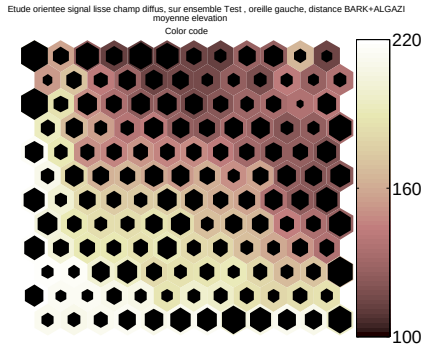

FIG. 3.16 – Moyenne de l'élévation sur la carte de l'hémisphère arrière.

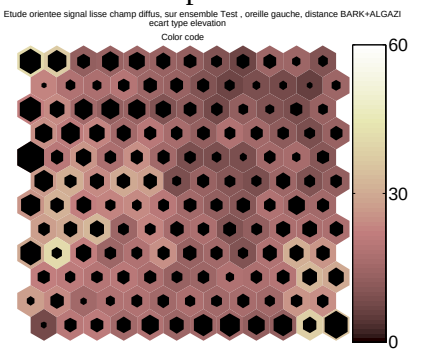

FIG. 3.17 – Ecart type de l'élévation sur la carte de l'hémisphère avant.

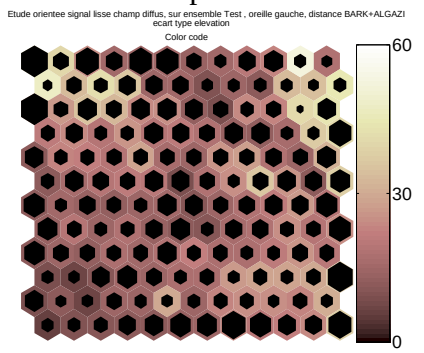

FIG. 3.18 – Ecart type de l'élévation sur la carte de l'hémisphère arrière.

3.4.2 Résultats du Clustering par CHA et Positionnement des HRTF représentantes

Les figures 3.19 et 3.20 présente la répartition fréquentielle des erreurs de quantification obtenues avec cette carte.

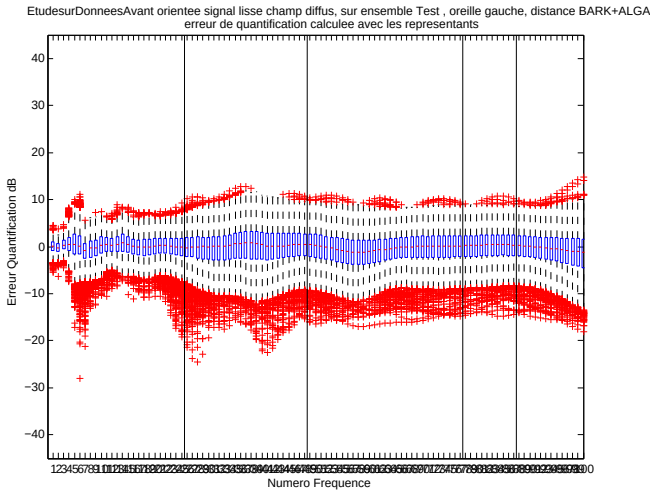

FIG. 3.19 – Erreur de quantification sur tous les individus de l'ensemble de Test (carte de l'hémisphère avant)

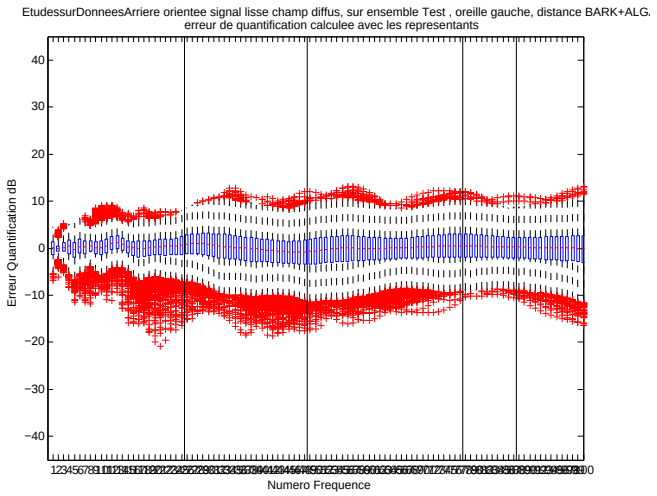

FIG. 3.20 – Erreur de quantification sur tous les individus de l'ensemble de Test (carte de l'hémisphère arrière)

L'utilisation d'un critère de Ward égal à 0.5 (pour éviter les variations intra-classe élevées au niveau de la répartition de l'élévation, cf CHOQUEUSE) ainsi que l'interprétation des graphes de visualisation présentée ci-dessus nous assure que ces super-classes regroupent en leurs seins des HRTF spatialement proches. Ce rapport semble suffisant car il permet de regrouper les 144 neurones de la carte de l'émisphère avant en 10 classes, et

ceux de la carte de l'émisphère arrière en 11 classes. Le clustering des HRTF obtenu par carte de kohonen et par classification hiérarchique ascendante permet de réduire consécutivement l'ensemble des HRTF composé de deux ensembles de 625*8 HRTF en 2x144 neurones puis 10 et 11 classes. La carte des clusters de l'émisphère avant est observable sur la figure 3.21.

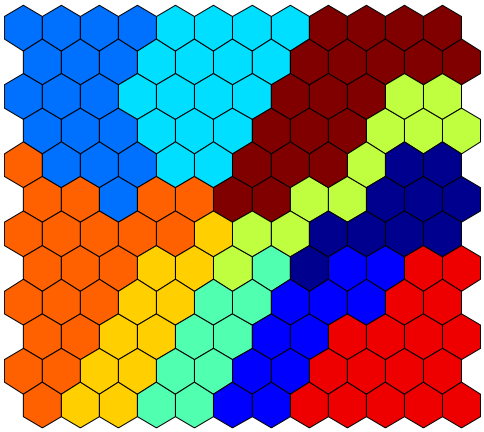

FIG. 3.21 – Résultat de la Classification Hiérarchique Regroupement des 144 neurones en Clusters

On retrouve sur les graphes 3.22 et 3.23 des positions d'HRTF représentatives formant une sorte d'ellipse autour de la tête.

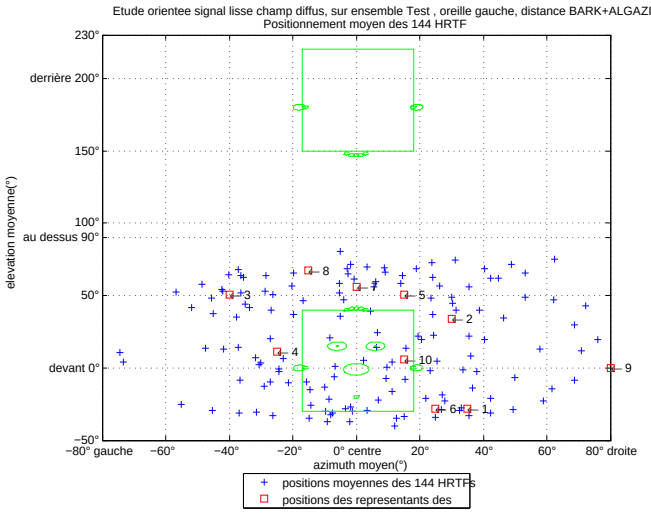

FIG. 3.22 – Positions des HRTF représentantes obtenues avec la carte de kohonen "émisphère avant"

Cela voudrait dire que l'information spectrale de localisation auditive se concentre dans les HRTF mesurées sur un "lasseau entourant la tête de l'auditeur".

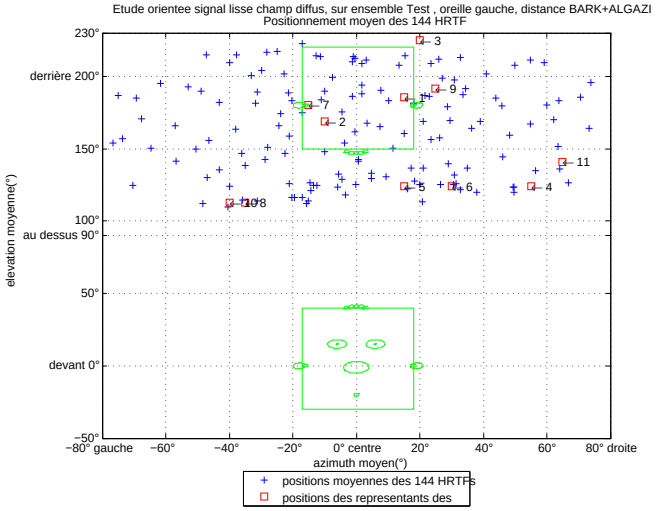

FIG. 3.23 – Positions des HRTF représentantes obtenues avec la carte de kohonen "émi-sphère arrière"

Les sept autres études comprenant l'apprentissage des cartes MSE, BARK et ALGAZI sur les données brutes et ECDL du CIPIC ne seront pas développées.
En effet, on retrouve pour ces études quasiment les mêmes graphes de visualisation que pour le "clustering BARK+ALGAZI sur les données ECDL". On ne détaillera donc pas plus l'interprétation de leurs résultats puisqu'on peut faire les mêmes observations que pour l'étude précédente sur le regroupement des HRTF.

3.5 Résultats des quatres études

3.5.1 Interprétation des graphes de visualisation

On obtient les mêmes graphes de visualisations pour toutes les cartes de kohonen construites. On constate donc que ces cinq cartes de kohonen apprisent sur des données scindées, présentent des proximités entre HRTF proches spatialement.En démontrant ainsi que des HRTF mesurées sur des individus différents, mais étant spectralement proches, ont été mesurées en des positions géographiques proches, cette étude approuve l'hypothèse de CHOCQUEUSE, selon laquelle les positions des représentant seraient dans des régions de l'émisphère (clusters) communes à tous les individus. Par contre, cette étude ne nous aura pas permis d'évaluer l'influence des différents critères d'évaluation sur la phase d'apprentissage des cartes.

3.5.2 Analyse de l'influence du critère d'evaluation sur l'Erreur de quantification

Dans le tableau 3.1, les lignes correspondent aux critères d'évaluations qui ont été utilisés pour calculer les erreurs de quantifications, tandis que les colonnes correspondent aux représentants déterminés par différentes carte de Kohonen. Ainsi, une case a_{ij} du tableau correspond à l'évaluation par le critère i des représentants obtenus par la carte j. Si l'on se restreint à l'étude d'un critère, prenons la Ligne E.Q. BARK par exemple, on constate que les erreurs sont sensiblement les mêmes quelque soit la carte utilisée. Encore une fois, l'étude montre que pour l'étape de clustering, l'utilisation de la distance MSE suffit.

	Cartes de Kohonen					
Critères d'évaluation	MSE	$BARK$	$ALGAZI$	$B.+A.$	$DURANT$	**RU**
$Chocqueuse$	2.52	2.64	2.58	2.65	2.72	2.55
$Bark$ (x10^{-2})	2	2.02	2.02	2.05	2.04	1.99
$Algazi$	0.36	0.38	0.38	0.40	0.45	0.32
$B.+A.$	0.21	0.23	0.23	0.22	0.21	0.19
Mse	10.80	11.53	11.14	11.80	12.8	10.60
$Durant$	8.82	8.9	9.24	9.24	8.52	8.62

TAB. 3.1 – Evaluation des positions de représentants issus des CHA sur différentes cartes de Kohonen "émisphère avant" : erreurs de quantification évaluées par l'ensemble des critères d'évaluation (RU : Représentants Uniformes, B+A : Bark + ALGAZI)

	Cartes de Kohonen					
Critères d'évaluation	MSE	$BARK$	$ALGAZI$	$B.+A.$	$DURANT$	**RU**
$Chocqueuse$	1.94	1.95	1.92	1.95	1.99	1.93
$Bark$ (x10^{-2})	1.76	1.72	1.74	1.73	1.79	1.63
$Algazi$	0.25	0.25	0.25	0.25	0.28	0.27
$B.+A.$	0.20	0.20	0.20	0.20	0.21	0.17
Mse	6.47	6.44	6.35	6.49	6.73	6.23
$Durant$	6.02	5.95	5.88	6.03	5.73	5.66

TAB. 3.2 – Evaluation des 144 positions de représentants issus des différentes cartes de Kohonen "émisphère avant" : erreurs de quantification évaluées par l'ensemble des critères d'évaluation (RU : Représentants Uniformes, B+A : Bark + ALGAZI)

	Cartes de Kohonen					
Critères d'évaluation	MSE	$BARK$	$ALGAZI$	$B.+A.$	$DURANT$	**RU**
$Chocqueuse$	2.66	2.66	2.62	2.64	2.71	2.60
$Bark$ (x10^{-2})	2.01	2.02	1.88	1.92	2.04	2.03
$Algazi$	0.40	0.41	0.40	0.39	0.44	0.35
$B.+A.$	0.2	0.22	0.17	0.18	0.20	0.21
Mse	11.73	11.69	11.69	11.76	12.56	11.72
$Durant$	9.92	9.31	9.4	9.66	9.16	9.18

TAB. 3.3 – Evaluation des représentants issus des CHA sur différentes cartes de Kohonen "émisphère arrière" : erreurs de quantification évaluées par l'ensemble des critères d'évaluation

	Cartes de Kohonen					
Critères d'évaluation	MSE	$BARK$	$ALGAZI$	$B.+A.$	$DURANT$	**RU**
$Chocqueuse$	1.93	1.94	1.94	1.95	2.02	1.96
$Bark$ (x10^{-2})	1.63	1.55	1.66	2	1.69	1.62
$Algazi$	0.25	0.24	| 0.26	0.25	0.27	0.24
$B.+A.$	0.15	0.13	0.15	0.14	0.16	0.15
Mse	6.49	6.55	6.51	6.58	6.99	6.50
$Durant$	6.04	6.11	6.07	6.13	5.94	5.99

TAB. 3.4 – Evaluation des 144 positions de représentants obtenus avec différentes cartes de Kohonen "émisphère arrière" : erreurs de quantification évaluées par l'ensemble des critères d'évaluation

En observant les erreurs de quantifications sur une ligne, on ne note pas de différence notable pour les différentes cartes utilisées. Pourtant, notre analyse des critères (section 2.3) nous laissais présager que l'introduction des critères dans les fonctions de coûts des cartes allait modifier considérablement leur apprentissage.

Prenons par exemple les critères Bark et Durant :

- Le critère BARK minimise l'erreur commise sur les hautes fréquences[9]. Lorsqu'on utilise ce critère pour l'apprentissage d'une carte de Kohonen, celui-ci va regrouper des HRTF ayant un spectre similaire uniquement sur la bande de fréquence $0 - 15$Khz. Autrement dit, cette mesure de similarité regrouperait des HRTF très différentes sur les bandes spectrales hautes fréquences. On devrait donc avoir des distributions d'erreurs par points fréquentiels d'écarts interquartiles croissant (ie, des erreurs élevées pour des fréquences élevées) en fonction de la fréquence. L'observation des figures 3.24 et 3.25 montre qu'en pratique, la pondération par l'inverse des coefficients de BARK n'a pas joué son rôle. C'est que les hautes fréquences ne servent pas pour l'apprentissage des cartes.

- Le critère Durant calcule la variance de la distribution de l'erreur sur les points fréquentiels[10]. En théorie, l'utilisation d'un tel critère pour l'apprentissage d'une carte de Kohonen devrait regrouper des HRTF ayant un spectre similaire sur toute la bande fréquentielle. On devrait alors améliorer la qualité des clusters en regroupant des HRTF ayant les mêmes indices spectraux (en particulier la position et l'amplitude des creux et ventres spectraux).En effet, contrairement au critère de BARK, il tend à minimiser l'erreur sur tous les points fréquentiels. On devrait donc avoir des distributions d'erreurs par points fréquentiels plates et peu dispersées par rapport à celles obtenues avec un apprentissage utilisant la MSE. L'observation des figures 3.24 et 3.26 montre qu'en pratique, les propriétés de ce critère ont jouées leur rôle seulement sur les hautes fréquences.

Puisque les différentes cartes apprises effectuent les mêmes regroupements quelque soit la distance d'apprentissage utilisée, on peut considérer que l'utilisation de la MSE suffit pour l'étape de clustering. Or, cette distance ne tient compte que des grosses erreurs (à cause de son terme au carré), cela veut donc dire que la variation des positions de mesures des HRTF se traduit uniquement par de fortes variations spectrales.

[9] voir section 2.3
[10] voir section 2.3

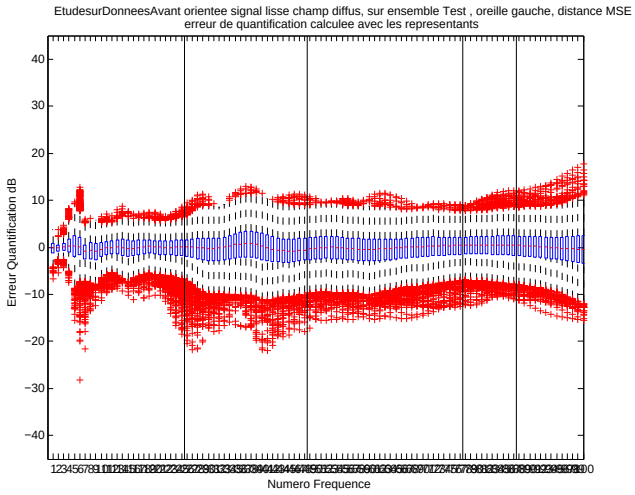

FIG. 3.24 – Evaluation des représentants obtenus par la carte MSE (sur emisphère avant) :Erreur de quantification moyennée sur tous les individus de l'ensemble de Test.

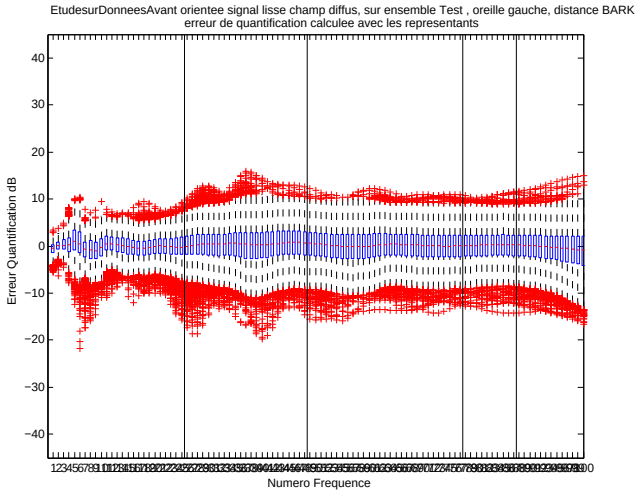

FIG. 3.25 – Evaluation des représentants obtenus par la carte BARK (sur emisphère avant) :Erreur de quantification moyennée sur tous les individus de l'ensemble de Test.

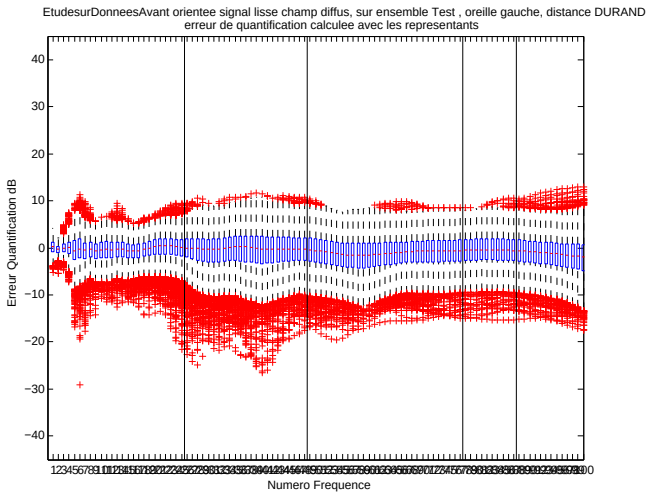

FIG. 3.26 – Evaluation des représentants obtenus par la carte DURANT (sur emisphère avant) : L'erreur est moyennée sur tous les individu de l'ensemble de Test.

3.6 Discussion et perspectives

Notre étude décomposée en une phase de clustering utilisant respectivement les cartes de Kohonen, des CHA et une phase d'élection nous a permis d'extraire les HRTF "représentatives" qui seront utilisées pour l'estimation de l'ensemble des HRTF. Cette étude a aussi permis de visualiser les rapprochements entre les spectres des HRTF et leurs positions spatiales, en mettant en évidence la symétrie des HRTF par rapport au plan frontal. Le problème de confusion avant/arrière provient de cette symétrie, sur ce point, on ne constate pas d'amélioration par rapport aux travaux de CHOQUEUSE. Les tableaux d'erreurs de quantifications nous apprennent que l'influence des critères d'évaluation sur l'apprentissage des cartes est négligeable. L'utilisation de la MSE comme fonction de coût suffit. Par contre, l'evaluation de nos méthodes de réduction du nombre de mesures au moyen du critère Durant semble prometteuse. Notons aussi que le jeu de cartes ayant utilisé des HRTF églisées champs diffus améliore l'erreur de quantification de 34% par rapport au jeu ayant utilisé des HRTF brutes.
Les représentants uniformes obtiennent des résultats équivalents à ceux déterminés par la méthode de clustering. On pourra se pencher dans les travaux futures vers la recherche d'un individu dont les HRTF représentantes pourraient dépasser les résultats des représentants uniformément répartis. Il serait aussi intéressant d'identifier de quels individus proviennent les HRTF représentantes issues du clustering. On pourra alors savoir si cette étude réalisée sur tous les individus de la base CIPIC a retrouver l'individu ayant des HRTF représentantes communes à tous les autres. La partie suivante, présente la phase de modélisation.

Chapitre 4

Modélisation des 1250 HRTF nécessaires et suffisantes pour un individu

4.1 Introduction

La phase de clustering explicitée dans le chapitre précédent permet d'effectuer n mesures d'HRTF au lieu de 1250 mesures sur chaque individu. Elle nous fournit les positions de ces n points de mesures. La phase de modélisation consiste à construire un modèle pouvant individualiser les $1250 - n$ HRTF manquantes d'un individu à partir de la seule connaissance de ses n HRTF représentantes. Les outils statistiques que nous utilisons pour modéliser des HRTF sont les réseaux de neurones de type perceptron multicouche (MLP). Pour permettre à ces réseaux de neurones d'individualiser les HRTF de l'individu λ, nous allons leurs transmettre en entrée les n HRTF ("représentantes") mesurées sur celui-ci. Dans son processus d'apprentissage, un réseau de neurone MLP va approximer une fonction $f(vecteur_{entree}(\theta, \phi, HRTF_{\theta^*,\phi^*,\lambda})) = \hat{HRTF}_{\theta,\phi,\lambda}$ sur les 22500 HRTF des 18 individus de l'ensemble d'apprentissage[1]. La fonction $vecteur_{entree}$ construit le vecteur d'entrée en associant à la position (θ, ϕ) l'HRTF représentante (de position (θ^*, ϕ^*)) qui lui correspond le mieux.
On présentera dans ce chapitre uniquement la méthodologie pour construire le vecteur d'entrée du réseau de neurone.

4.2 Construction du vecteur d'entrée

Nous avons utilisé deux jeux de vecteurs d'entrée qui diffèrent par l'origine des positions représentantes : les positions représentantes issus de la méthode de clustering et celles "Uniformément répartis à la surface de la sphère". Dans les deux cas, nous allons utiliser un *unique représentant* pour modéliser une HRTF. En effet, l'utilisation de plus

[1] voir section 1.2.2

d'une HRTF représentante est possible mais très coûteuse en mémoire RAM.

4.2.1 Utilisation d'un représentant issus du clustering sur les données ECDL

Le clustering effectué sur tous les individus de la base CIPIC a regrouper des HRTF provenant d'individus différents en n clusters.
Ensuite n positions représentantes $(\theta_1^*, \phi_1^*), \ldots, (\theta_n^*, \phi_n^*)$ ont été élues pour chacun de ces cluster.Puisqu'on utilise qu'un unique représentant, on doit choisir quelle position de représentant (θ_i^*, ϕ_i^*) (avec $i \in [[1 : n]]$ associer à la position (θ, ϕ) qu'on veut modéliser. On aurait pu tout simplement regarder dans quel cluster tombe l'HRTF mesurée à la position (θ, ϕ) et lui associer le représentant de celui-ci. Mais cela n'est pas possible puisque nos clusters se chevauche spatialement (voir figure 4.1). En effet, deux clusters différents peuvent posséder des positions d'HRTF communes mais mesurées sur des individus différents. De même, des HRTF mesurées à la même position mais appartenant à des individus différents peuvent se retrouver dans le même cluster. Ainsi, puisqu'une position (θ, ϕ) peut se retrouvée dans plusieurs clusters à la fois, on détermine le cluster possédant le plus d'HRTF mesurées à cet position. On associe alors la position représentante de ce cluster à la position (θ, ϕ).
On dresse ainsi un tableau associant à chaque position (θ, ϕ) une position (θ_i^*, ϕ_i^*) $(oi \in [[1 : n]])$ parmi les n possible. En procédant ainsi, on s'assure que le choix de la position (θ_i^*, ϕ_i^*) tient compte des regroupements par similarité spectrales des HRTF.

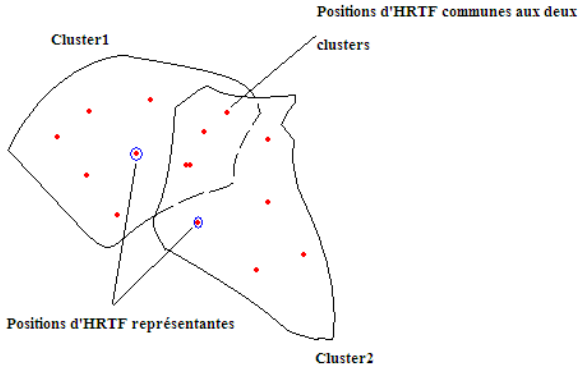

FIG. 4.1 – Chevauchement des clusters

Soit (θ^*, ϕ^*) la position spatiale du représentant de l'$HRTF_{\theta,\phi}$, le vecteur d'entrée du réseau de neurone est défini par :

$$Vect_{Entree} = [\theta, \phi, H^l_{\theta^*,\phi^*}]; \tag{4.1}$$

4.2.2 Utilisation d'un représentant parmi les représentants uniformément réparti à la surface de la sphère

Pour associer l'une des n positions d'HRTF représentantes à une position (θ, ϕ), le choix se porte ici sur la position d'HRTF représentantes la plus proche, au sens géodésique, de la position d'HRTF (θ, ϕ). Soit (θ^*, ϕ^*) la position spatiale du représentant de l'$HRTF_{\theta,\phi}$, le vecteur d'entrée du réseau de neurone est défini par :

$$Vect_{Entree} = [\theta, \phi, H^l_{\theta^*,\phi^*}]; \tag{4.2}$$

4.3 Discussion et perspectives

Nous avons présenté une manière de construire le vecteur d'entrée du réseau de neurone à l'aide des positions représentantes obtenues par Clustering. On pourrait aussi choisir d'associer à une position quelconque, la position représentante la plus proche spatialement. On pourra, lors de travaux futurs, etudier l'influence de cette étape sur les résultats de la modélisation.

On pourra aussi réaliser plusieurs modélisation afin de tester deux choses :

1. Les positions de représentants obtenues par notre méthode de clustering : Ces positions diminuent-elles l'erreur de modélisation ?
2. L'influence des critères d'évaluation, alors utilisés comme fonction de coût sur la qualité du modèle construit.

Chacunes de ces modélisations utilisera les représentants obtenus par une des 10 cartes de Kohonen construites précédemment. Afin de tester l'influence des critères d'évaluation sur la modélisation, on construira, pour une fonction de coût donnée, plusieurs réseaux de neurones utilisant les différents "types" de représentants.

Chapitre 5

Conclusion et Perspectives

Ce stage de recherche s'est orienté sur deux disciplines : le traitement statistique de l'information et l'audionumérique. On a montré à travers l'étude de clustering que l'utilisation de données égalisées champs diffus permet de diminuer l'erreur de quantification de 34%. On montre aussi que la distance MSE suffit à regrouper les HRTF de la base CIPIC par proximité spatiale et spectrale. L'étude sur les critères a priori a quant à elle permit de dégager les spécificités des différents critères d'évaluation. On retiendra que les critères Durant, Algazi et Chocqueuse offrent la meilleur réactivité aux erreurs de localisation.

Nous proposons les axes de recherche suivants pour améliorer les performances globales de notre méthode :

- **Election de représentants plus performants que ceux uniformément répartis à la surface de la sphère de mesures :** on pourrait rechercher un individu aux positions communes. On pourra aussi construire une carte de kohonen utilisant la norme de chebychev comme mesure de similarité. En effet cette norme minimise l'erreur maximum que l'on retrouve souvent sur les creux et ventres spectraux des HRTF. Ces creux et ventres spectraux servant à la localisation auditive, on peut s'attendre à améliorer les résultats avec une telle norme.

- **Recherche de la fonction de coût idéale pour la modélisation :** on pourra utiliser les critères Durant et Algazi comme fonction de coût de nos réseaux de neurones. On pourrait aussi introduire un coefficient de pondération fonction de la position spatiale afin de tenir compte du caractère non uniforme de la localisation auditive. En procédant ainsi, le réseau de neurone, lors de son processus d'apprentissage, tiendra compte du fait qu'il existe certaines régions de l'espace dont les HRTF jouent un rôle minimal dans le processus de localisation auditive. Les erreurs d'approximation mesurées pour des HRTF appartenant à ces régions de l'espace pourront alors être négligées, ie, pondérées par un coefficient proche de zéro.
  Pour lutter contre l'erreur d'individualisation, les techniques de warping fréquentiels sont à envisager.

Le problème de l'individualisation des HRTF pour la synthèse binaurale est un pro-

blème complexe. Nous pensons que l’approche statistique reste une voie prometteuse.

Remerciements

Je tiens tout d’abord à remercier Vincent Lemaire, mon maître de stage, pour m’avoir encadré lors de ce stage. Sa disponibilité, ses explications et ses conseils m’ont permis d’améliorer mes connaissances des techniques d’apprentissages statistiques.

Je remercie Sylvain Busson, Rozenn Nicol, Fabrice Clérot et Françoise Fessant pour leurs conseils précieux et avisés. Merci à Sylvain et Rozenn pour m’avoir apporté les connaissances en spatialisation sonore nécessaires à la réalisation du sujet.

Bibliographie

[1] A. Bronkhorst A. Langendijk. Contribution of spectral cues to human sound localization. 2002.

[2] J.F. Gyss B. S. Shinn-Cunningham, T. Streeter. Neural network : a comprehensive foundation, prentice hall, 1999.

[3] J.F. Gyss B. S. Shinn-Cunningham, T. Streeter. Perceptual plasticity in auditory displays. *Proceedings of the 2001 International Conference on Auditory Display*, 2001.

[4] M. Blommer. Pole-zero approximations for head-related transfer functions using a logarithmic error criterion. *IEEE Transaction on Speech and Audio Processing, Vol. 5*, 1997.

[5] R. O. Duda C. Avendano and V. R. Algazi. Modeling the contralateral hrtf. *AES 16th International Conference on Spatial Sound Reproduction*, 2001.

[6] Y.-C. Lo C. S. Fahn. On the clustering of head-related transfer functions used for 3-d sound localization. *journal of Information Science and Engineering 19*, 2003.

[7] Chocqueuse. Utilisation d'outils statistiques pour l'individualisation des hrtf. 2004.

[8] H. Wakefield E. Durant. Efficient model fitting using a genetic algorithm : Pole-zero approximatios of hrtfs. *IEEE Transaction on Spech and Audio Processing, Vol. 10*, 2002.

[9] Smith Huopaniemi. Spectral and time-domain preprocessing and the choice of modeling error criteria for binaural digital filters. *AES 16th International Conference on Spatial Sound Reproduction*, 2001.

[10] CIPIC Interface Laboratory. Documentation for the ucd hrir files. *Technical report, University of California at Davis*, 1998.

[11] V. Larcher. Techniques de spatialisation des sons pour la réalité virtuelle. 2001.

[12] S.Carlile P.H.W Leong. Methods for spherical data analysis and visualisation. *j.Neurosci Methods*.

[13] E. Macpherson J. Middlebrooks. Vertical-plane sound localization probed with ripple-spectrum noise. 2003.

[14] J. Middlebrooks. Individual differences in external-ear transfer functions reduced by scaling in frequency. 1999.

[15] J. Middlebrooks. Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. 1999.

[16] J.M Pernaux. Spatialisation du son par les techniques binaurales : Application aux services de télécommunications. 2003.

[17] 2001 Self Organizing Maps, Springer.

[18] S. Mase T. Nishino. Interpolating hrtf for auditory virtual reality.

[19] S. Busson R. Nicol V. Choqueuse V. Lemaire, F. Clerot. Individualized hrtfs from few measurements : a statistical learning approach. 2005.

[20] C. Avendano V. R. Algazi and R. O. Duda. Elevation localization and head-related transfer function analysis at low frequencies. *J. Acoust. Soc. Am., Vol. 109, No. 3, pp. 1110-1122*, 2001.

[21] R. Duraiswami N. A. Gumerov et Z. Tang V. R. Algazi, R. O. Duda. Approximating the head-related transfer function using simple geometric models of the head and torso. *J. Acoust. Soc. Am., Vol.112, pp. 2053-2064*, 2002.

[22] R.J. Dalton D.M. Thompson V.R. Algazi, R.O. Duda. Motional-tracked binaural sound for personal music players. *ISM2005 (IEEE International Symposium on Multimedia)*, 2005.

[23] F.L Wightman. Resolution of front-back ambiguity in spatial hearing by listener and source movement. *J.Acousti.Soc.Am*, 1999.

[24] Y. Kaneda N. Kitawasaki Y. Haneda, S. Makino. Common-acoustical-pole and zero moeling of head-related transfer functions. *IEEE Transactions on Speech and Audio Processing, Vol. 7, NO. 2*, 1999.

[25] D. Zoltkin. Customizable auditory display.

Annexe A

Boxplot des erreurs de localisation

Les figures qui suivent, présentent pour différents critères d'évaluation, l'erreur mesurée sur des paires d'HRTF d'écarts angulaires croissants.

A.1 Boxplot des erreurs de localisation calculées par différents critères d'évaluation

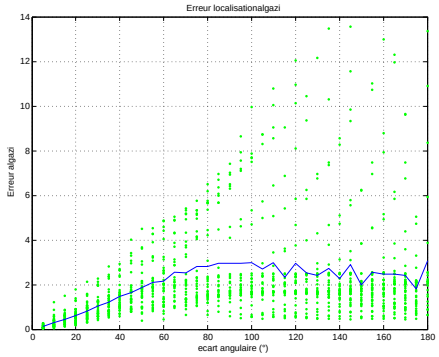

FIG. A.1 – Erreur de localisation mesurées pour le critère Algazi. : l'axe des abscisses représente l'écart angulaire en degrés entre les HRTF de la paire obtenue. Pour un même écart angulaire, les résultats sont moyennés sur toutes les paires évaluées (courbe en trait continu), l'ensemble des valeurs obtenues est également reproduit par des pointillés.

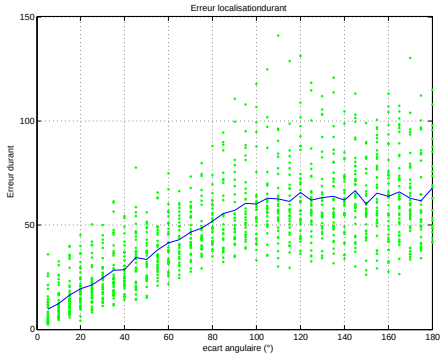

FIG. A.2 – Erreur de localisation mesurées pour le critère Durant. : l'axe des abscisses représente l'écart angulaire en degrés entre les HRTF de la paire obtenue. Pour un même écart angulaire, les résultats sont moyennés sur toutes les paires évaluées (courbe en trait continu), l'ensemble des valeurs obtenues est également reproduit par des pointillés.

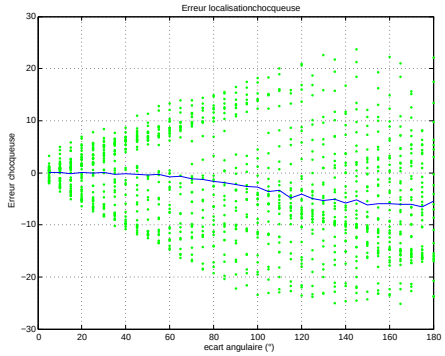

FIG. A.3 – Erreur de localisation mesurées pour le critère Chocqueuse : l'axe des abscisses représente l'écart angulaire en degrés entre les HRTF de la paire obtenue. Pour un même écart angulaire, les résultats sont moyennés sur toutes les paires évaluées (courbe en trait continu), l'ensemble des valeurs obtenues est également reproduit par des pointillés.

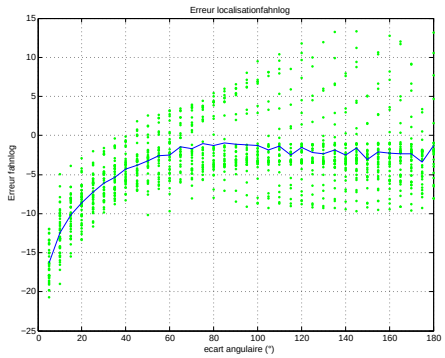

FIG. A.4 – Erreur de localisation mesurées pour le critère Fahn logarithmique : l'axe des abscisses représente l'écart angulaire en degrés entre les HRTF de la paire obtenue. Pour un même écart angulaire, les résultats sont moyennés sur toutes les paires évaluées (courbe en trait continu), l'ensemble des valeurs obtenues est également reproduit par des pointillés.

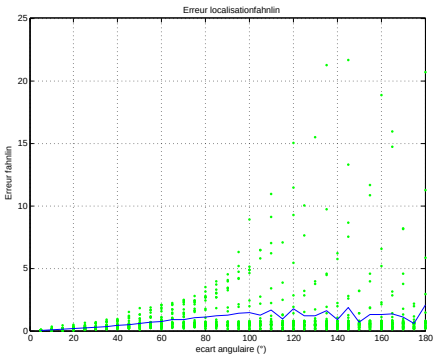

FIG. A.5 – Erreur de localisation mesurées pour le critère Fahn linéaire : l'axe des abscisses représente l'écart angulaire en degrés entre les HRTF de la paire obtenue. Pour un même écart angulaire, les résultats sont moyennés sur toutes les paires évaluées (courbe en trait continu), l'ensemble des valeurs obtenues est également reproduit par des pointillés.

Annexe B

Boxplot des erreurs d'individualisation

Les figures qui suivent, présentent pour différents critères d'évaluation, l'erreur mesurée sur des paires d'HRTF provenant d'individus de différence morphologique croissante.

B.1 Boxplot des erreurs d'individualisation calculées par différents critères d'évaluation

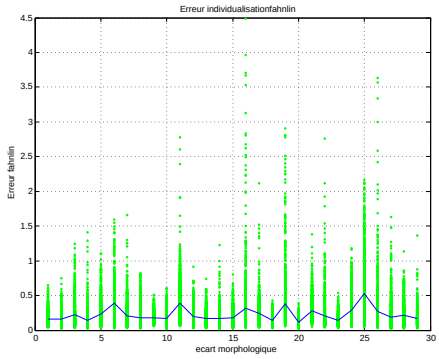

FIG. B.1 – Erreur d'individualisation mesurées pour le critère Fahn lin. : L'axe des abscisses est indiciel, il représente l'écart croissant de dimensions morphologiques entre les individus sur lesquels ont été mesurées les HRTF de la paire considérée. La dimension morphologique considérée est la largeur de la conque. Pour un même écart morphologique, les résultats sont moyennés sur toutes les paires évaluées (courbe en bleu), l'ensemble des valeurs obtenues est également reproduit par des points verts.

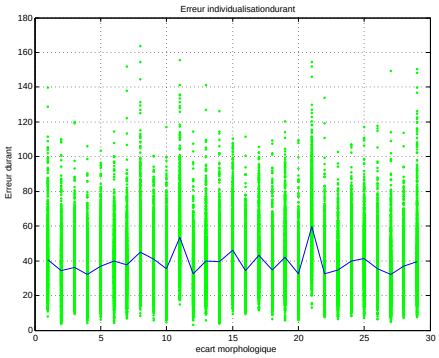

FIG. B.2 – Erreur d'individualisation mesurées pour le critère Durant : L'axe des abscisses est indiciel, il représente l'écart croissant de dimensions morphologiques entre les individus sur lesquels ont été mesurées les HRTF de la paire considérée. La dimension morphologique considérée est la largeur de la conque. Pour un même écart morphologique, les résultats sont moyennés sur toutes les paires évaluées (courbe en bleu), l'ensemble des valeurs obtenues est également reproduit par des points verts.

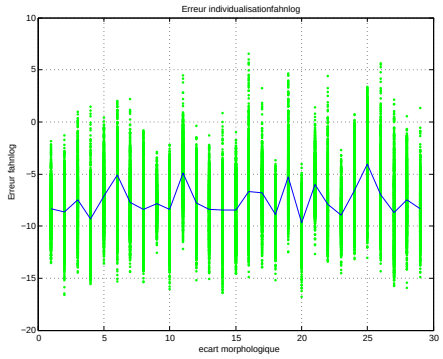

FIG. B.3 – Erreur d'individualisation mesurées pour le critère Fahn log. : L'axe des abscisses est indiciel, il représente l'écart croissant de dimensions morphologiques entre les individus sur lesquels ont été mesurées les HRTF de la paire considérée. La dimension morphologique considérée est la largeur de la conque. Pour un même écart morphologique, les résultats sont moyennés sur toutes les paires évaluées (courbe en bleu), l'ensemble des valeurs obtenues est également reproduit par des points verts.

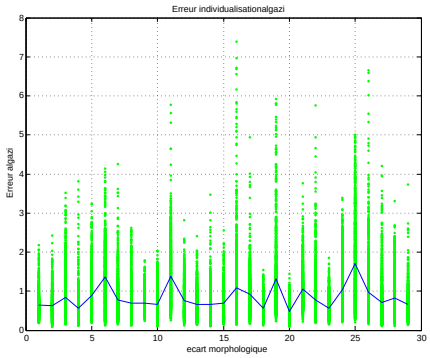

FIG. B.4 – Erreur d'individualisation mesurées pour le critère Algazi : L'axe des abscisses est indiciel, il représente l'écart croissant de dimensions morphologiques entre les individus sur lesquels ont été mesurées les HRTF de la paire considérée. La dimension morphologique considérée est la largeur de la conque. Pour un même écart morphologique, les résultats sont moyennés sur toutes les paires évaluées (courbe en bleu), l'ensemble des valeurs obtenues est également reproduit par des points verts.

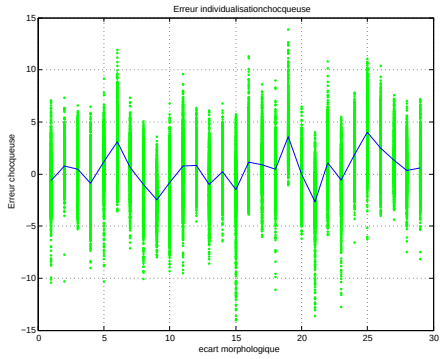

FIG. B.5 – Erreur d'individualisation mesurées pour le critère Chocqueuse : L'axe des abscisses est indiciel, il représente l'écart croissant de dimensions morphologiques entre les individus sur lesquels ont été mesurées les HRTF de la paire considérée. La dimension morphologique considérée est la largeur de la conque. Pour un même écart morphologique, les résultats sont moyennés sur toutes les paires évaluées (courbe en bleu), l'ensemble des valeurs obtenues est également reproduit par des points verts.

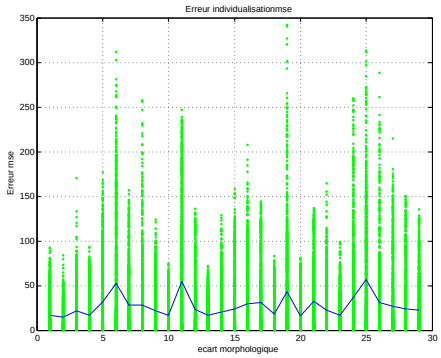

FIG. B.6 – Erreur d'individualisation mesurées pour le critère MSE : L'axe des abscisses est indiciel, il représente l'écart croissant de dimensions morphologiques entre les individus sur lesquels ont été mesurées les HRTF de la paire considérée. La dimension morphologique considérée est la largeur de la conque. Pour un même écart morphologique, les résultats sont moyennés sur toutes les paires évaluées (courbe en bleu), l'ensemble des valeurs obtenues est également reproduit par des points verts.

Annexe C

Nettoyage de la Base de Donnee CIPIC

C.0.1 Le but

Il a été observé des erreurs de mesures dans la base de données CIPIC. Celles-ci peuvent se traduire, soit par des différences de gain notables entre points de mesures, soit entre individus. D'autres erreurs peuvent apparaître, comme des mouvements des sujets pendant les mesures, mais celles-ci sont plus difficiles à repérer. Afin de constituer une base de données ne comportant pas de points aberrants (ie, génant lors d'un apprentissage statistique sur cette base de donnée), nous avons essayé de retirer ces points de mesures et donc les individus auxquels ils appartiennent.

C.0.2 Méthodologie

Nous avons tracer des graphes afin d'observer les différences de gain notables entre HRTF. L'observation des HRIR du plan de la base CIPIC, pour chaque individu est illustrée sur les figures C.1 et C.2.

N'ayant pu trouver les HRIR suspectes, nous sommes passés à l'analyse des HRTF. Cette analyse s'est effectuée à travers l'observation des énergies spectrales des HRTF du plan horizontale pour tous les individus (voir les figures C.3 et C.4). Cette énergie a été calculée en utilisant la formule C.1 :

$$Energie_{\lambda,\theta,\phi} = \sqrt{\sum_{i=1}^{100} H^2_{\lambda,\theta,\phi,i}} \tag{C.1}$$

L'analyse des graphes d'énergie a permis d'observer des différences de gain notable entre les HRTF. Pour confirmer ces observations, nous avons alors penser à utiliser la représentation des boxplots pour repérer les HRTF d'énergie extrémale. Cette représentation permet de visualiser et comparer la distribution des énergies des HRTF pour un individu donné (voir figure C.5) en affichant :

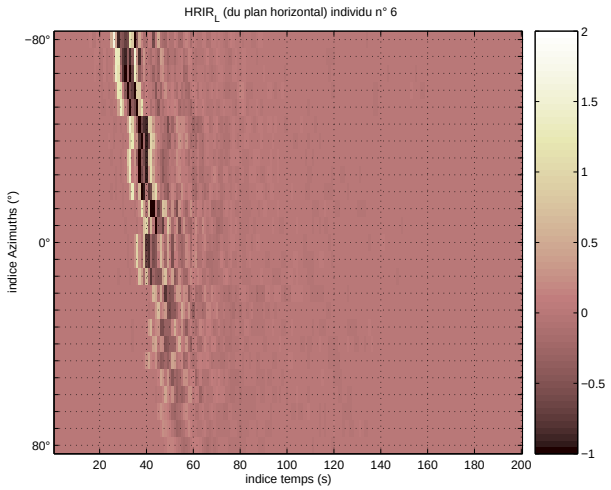

FIG. C.1 – Projection des 25 HRIR (celles du plan horizontale) mesurées sur l'oreille gauche de l'individu 6. L'axe des abscisses représente le temps, l'axe des ordonnées représente l'azimuth à laquelle l'HRIR a été mesurée tandis que le niveau de gris renseigne sur l'amplitude de la HRIR.

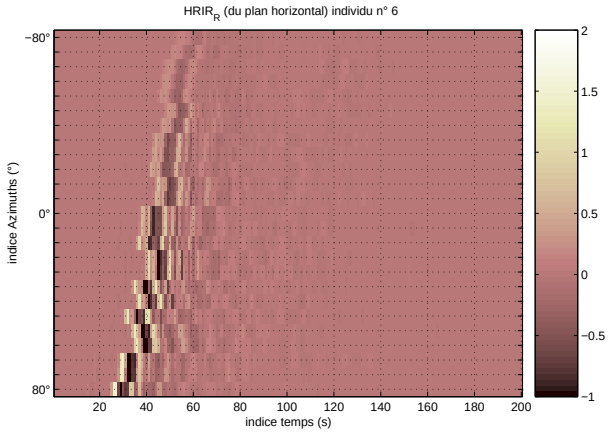

FIG. C.2 – Projection des 25 HRIR (celles du plan horizontale) mesurées sur l'oreille droite de l'individu 6. L'axe des abscisses représente le temps, l'axe des ordonnées représente l'azimuth à laquelle l'HRIR a été mesurée tandis que le niveau de gris renseigne sur l'amplitude de la HRIR.

- La valeur de la médiane des énergies des HRTF est représentée par un trait horizontale dans la "boîte à moustache" : la moitié de la distribution a des valeurs inférieurs à celle de la médiane.
- Les quartiles à 25% et 75%. Ces valeurs sont représentées par les bords de la "boîte à moustache".
- Les valeurs adjacentes correspondent à 1.5 fois la distance inter-quartile. Elles sont représentées par les "moustaches" de la "boîte à moustache".

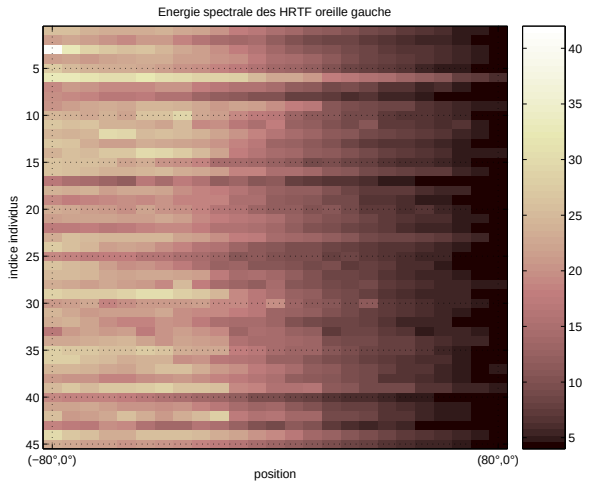

FIG. C.3 – Energie spectrale des 25 HRTF (celles mesurées à partir de l'oreille gauche sur le plan horizontale) de tous les individus de la base : l'axe des abscisses représente l'azimuth en lequel les HRTF ont été mesurés sur le plan horizontale(l'élévation est constante et nulle), l'ordonnée représente le numéro d'individus sur lequel ont été mesurées les HRTF, et le niveau de gris est associée à la valeur de l'énergie de l'HRTF mesurée.

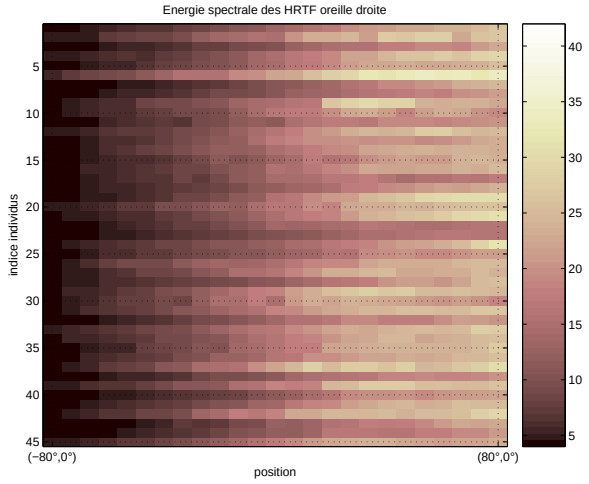

FIG. C.4 – Energie spectrale des 25 HRTF (celles mesurées à partir de l'oreille droite sur le plan horizontale) de tous les individus de la base : l'axe des abscisses représente l'azimuth en lequel les HRTF ont été mesurés sur le plan horizontale(l'élévation est constante et nulle), l'ordonnée représente le numéro d'individus sur lequel ont été mesurées les HRTF, et le niveau de gris est associée à la valeur de l'énergie de l'HRTF mesurée.

– Les outliers : énergies supérieures et inférieures aux valeurs adjacentes. Elles sont affichées par une croix.

Sur la figure C.5 on peut observer que la distribution des énergies de l'individu 6 se démarque par le fait que son premier quartile (25%) a une valeur supérieur à 6. Les distributions des autres individus ont leur médiane autour de six et s'écartent donc de

celle-ci. La suppression de l'individu 6 de la base se trouve alors justifié. On supprime de la même manière les individus ayant des distributions atypiques(ie, possédant des outliers, ou ayant une distance interquartile trop forte...).

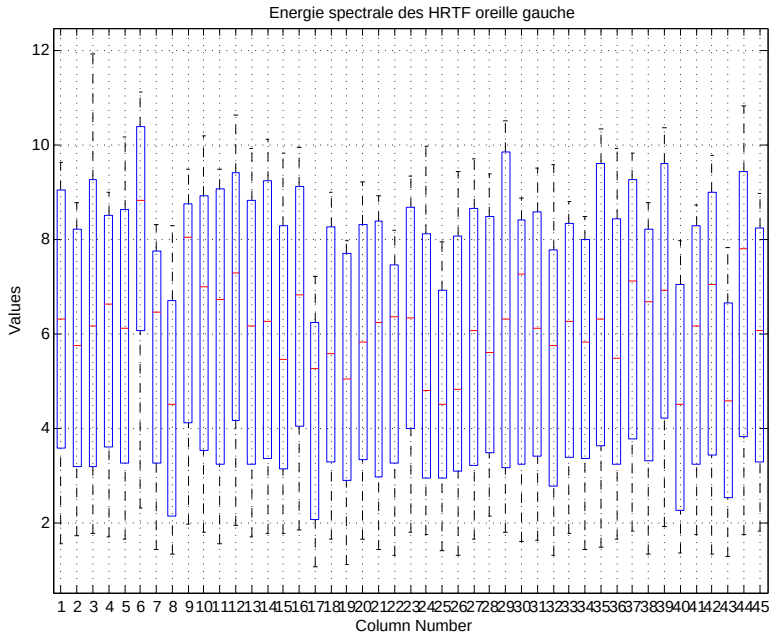

FIG. C.5 – Boxplot sur les distributions d'Energies spectrales des 25 HRTF (celles mesurées à partir de l'oreille gauche sur le plan horizontale) de tous les individus de la base.

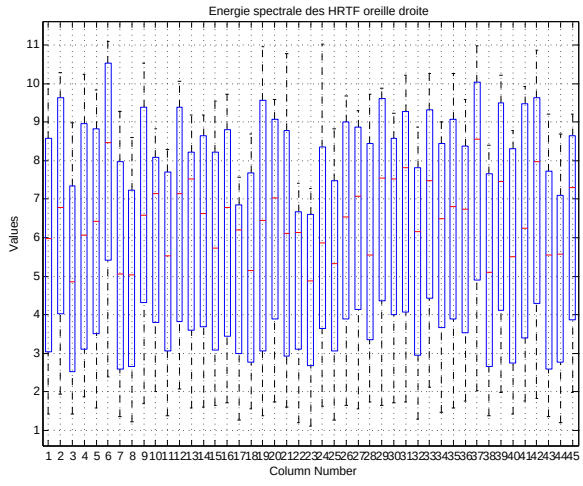

FIG. C.6 – Boxplot sur les Energie spectrale des 25 HRTF (celles mesurées à partir de l'oreille droite sur le plan horizontale) de tous les individus de la base.

C.0.3 Résultats

La représentation des Boxplot nous a ainsi permis d'identifier précisément les individus ayant des HRTF dont l'énergie spectrale est extrême. On a relevé 11 individus possédant des HRTF ayant une énergie spectrale anormalement élevée.
On obtient donc au final une base d'HRTF nettoyée et maintenant composée des 42500 HRTF des 34 individus restant.