

Nom, Prénom : Choqueuse, Vincent
Responsable pédagogique UTT :
Cédric Richard

Branche : GSID/RCE
Année : 2004
Semestre : Printemps

Utilisation d'outils statistiques pour l'individualisation des HRTF

Ce projet de fin d'étude s'est réalisé dans l'entreprise France Télécom R&D. Le sujet s'est déroulé en partenariat entre deux unités : l'unité « Traitements Statistiques pour les Télécommunications » et l'unité « Interaction de la Parole et du Son ». Le but du stage est de constater s'il est possible d'utiliser des outils statistiques pour individualiser les HRTF.

La méthodologie employée pour individualiser les HRTF est la suivante :

- Trouver le nombre et la localisation des HRTF suffisantes mais nécessaires à la reconstruction de l'ensemble des HRTF. Nous appelons ces HRTF : les HRTF représentatives.
- Créer un modèle à partir de réseaux de neurones permettant d'obtenir l'ensemble des HRTF d'un individu à partir des HRTF représentatives.

Le travail effectué s'est basé sur l'utilisation de réseaux de neurones (Cartes de Kohonen et MLP). Plusieurs programmes Matlab ont été réalisés permettant l'application de la méthodologie, l'interprétation et l'écoute des résultats

Entreprise : France Télécom R&D
Lieu : Lannion (22)
Responsable : Vincent Lemaire

Mots clés : Télécommunications
Recherche fondamentale
Traitement du signal
Statistiques



Remerciements

Je tiens tout d'abord à remercier Vincent Lemaire, mon maître de stage, pour m'avoir encadré lors de ce stage. Sa patience, sa disponibilité, ses explications et ses conseils au niveau rédactionnel m'ont permis d'évoluer que ce soit sur le sujet de stage ou bien sur mon rapport.

Je remercie Daniel Collobert pour m'avoir accueilli dans son service. Je remercie également l'ensemble de TNT pour leur accueil, leur bonne humeur et l'ambiance agréable qui règne dans le service. TNT un service où il fait bon travailler...

Je remercie Sylvain Busson, Fabrice Clérot, Rozenn Nicol et Françoise Fessant pour leurs conseils précieux et avisés. Merci à Sylvain pour m'avoir apporté les connaissances en audio et en spatialisation nécessaires à la réalisation du sujet et merci à Fabrice pour m'avoir prodigué des explications en statistiques et mathématiques et d'avoir su les adapter à mon niveau.

Merci à tous de m'avoir donné goût à la recherche.

Enfin un clin d'oeil aux stagiaires de France Télécom R&D et à l'équipe du Cosmos.

Table des matières

1	Présentation	7
1.1	Cadre du stage	7
1.1.1	France Télécom	7
1.1.2	La partie Recherche & Développement	7
1.1.3	Présentation des services	8
1.2	Problématique	9
1.3	Objet de l'étude	10
2	Données, Représentations, Evaluation	15
2.1	Introduction	15
2.2	Origine des données	16
2.2.1	Coordonnées spatiales des mesures	16
2.2.2	Spécification des HRTF	17
2.3	Modification des données	18
2.3.1	Modification de l'échelle des amplitudes	18
2.3.2	Modification de l'échelle des fréquences	19
2.3.3	Egalisation des données en champ "libre" et champ "diffus"	20
2.4	Evaluation des performances	21
2.4.1	Espace commun de comparaison	21
2.4.2	Visualisation synthétique des erreurs	22
2.4.3	Liste des différentes orientations possibles	23
3	Sélection des HRTF représentatives	25
3.1	Introduction	25
3.2	Méthodologie	26
3.2.1	Clustering par Carte de Kohonen	26
3.2.2	Clustering de la carte par CHA	28
3.2.3	Election des représentants	29
3.3	Un exemple : l'étude orientée "Signal-Champ Libre"	30
3.3.1	Clustering par Carte de Kohonen	30
3.3.2	Clustering par Carte de Kohonen avec Séparation des données Avant/Arrière	33
3.3.3	Clustering de la carte par CHA	36
3.3.4	Election des représentants	38
3.4	Comparaison des orientations	38

3.5	Intérêt de la démarche de sélection des représentants	40
3.6	Discussion	43
4	Modélisation d’HRTF par régression statistique	45
4.1	Introduction	45
4.2	Sélection des ensembles d’apprentissage, de validation et de test	46
4.3	Choix du vecteur d’entrée	46
4.3.1	Utilisation de tous les représentants	47
4.3.2	Utilisation d’un représentant	47
4.4	Type de normalisation utilisée	48
4.5	Estimation en régression linéaire	48
4.5.1	Utilisation d’un seul représentant	48
4.5.2	Utilisation des 26 représentants	49
4.6	Estimation en régression non linéaire	49
4.6.1	Choix du nombre de neurones de la couche cachée	50
4.6.2	Utilisation d’un seul représentant	50
4.6.3	Utilisation de tous les représentants	51
4.6.4	Comparaison des reconstructions suivant les vecteurs utilisés en entrée	51
4.7	Discussion	53
5	Individualisation des HRTF par régression statistique	55
5.1	Introduction	55
5.2	Sélection des ensembles d’apprentissage, de validation et de test	56
5.3	Choix du vecteur d’entrée	56
5.3.1	Utilisation d’un représentant orienté “Signal-Champ libre”	56
5.3.2	Utilisation d’un représentant orienté “Uniformement réparti à la surface de la sphère”	57
5.4	Type de normalisation utilisée	57
5.5	Modélisation en régression non linéaire : Représentants orientés “Signal- Champ Libre”	57
5.5.1	Modélisation utilisant 1 représentant parmi les 26 représentants	57
5.5.2	Influence du nombre de représentants	58
5.6	Modélisation en régression non linéaire : Représentants orientés “Unifor- mement répartis à la surface de la sphère”	58
5.6.1	Modélisation utilisant 1 représentant parmi les 26 représentants	58
5.6.2	Influence du nombre de représentants	59
5.7	Discussion	60
6	Conclusion	63
A	Sélection des bandes fréquentielles	65
B	Sélection des représentants uniformément répartis à la surface de la sphère	67
B.1	Transformation des coordonnées polaires interaurales des 1250 Points en coordonnées cartésiennes	67
B.2	Regroupement des points en K-cluster par l’algorithme des K-moyennes	69

B.3	Election d'un représentant par cluster	69
C	Introduction aux réseaux de neurones	71
C.1	Présentation	71
C.1.1	Qu'est-ce qu'un réseau de neurones ?	71
C.1.2	Que peut-on faire à l'aide d'un réseau de neurones ?	73
C.2	Apprentissage	73
C.2.1	Différentes règles d'apprentissage	73
D	Non individualisé : Modélisation en Regression non linéaire avec 1 représentant "Signal Champ-Libre"	77
D.1	Répartition fréquentielle des erreurs	77
D.2	Répartition spatiale des erreurs	78
E	Non individualisé : Modélisation en Regression non linéaire avec 26 représentants "Signal Champ-Libre"	81
E.1	Répartition fréquentielle des erreurs	81
E.2	Repartition spatiale des erreurs	82
F	Individualisé : Modélisation en Regression non linéaire avec 1 représentant "Signal Champ-Libre"	85
F.1	Répartition fréquentielle des erreurs	85
F.2	Repartition spatiale des erreurs	86
G	Individualisé : Modélisation en Regression non linéaire avec 1 représentant "Uniformément réparti à la surface de la sphère"	89
G.1	Répartition fréquentielle des erreurs	89
G.2	Repartition spatiale des erreurs	90
	Index	93
	Bibliographie	95

Chapitre 1

Présentation

1.1 Cadre du stage

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

1.1.1 France Télécom

Le groupe France Télécom est un opérateur de télécommunications. Avec plus de 112 millions de clients et implanté sur les cinq continents, France Télécom est aujourd'hui un des leaders mondiaux du domaine des télécommunications. Le groupe présente, en 2003, un chiffre d'affaire consolidé de 46,6 milliards d'euros. En s'appuyant sur des filiales internationales majeures comme Orange (acquisition en 2000), Wanadoo (acquisition en 2000), Equant ou GlobeCast, France Telecom propose essentiellement huit services afin de couvrir le vaste domaine d'activités des télécommunications :

- La téléphonie fixe (49,5 millions d'abonnés).
- La téléphonie mobile (49,9 millions d'abonnés).
- Les communications internationales.
- Les réseaux d'entreprises (3700 grandes entreprises multinationales).
- L'accès à Internet (10,1 millions d'abonnés).
- Les services de télédiffusion et de diffusion par câble.
- Les services en ligne.

Aujourd'hui, malgré de lourdes dettes dues à des investissements peu favorables, France Telecom apparaît comme extrêmement compétitif, avec un potentiel opérationnel intact. Le groupe a d'ailleurs lancé une stratégie de redressement pour reprendre en main son destin dès 2005.

1.1.2 La partie Recherche & Développement

Les laboratoires de France Télécom R&D sont à l'origine d'environ 70% des produits et services commercialisés par le Groupe. Avec près de 3700 chercheurs et ingénieurs,

France Télécom R&D est un atout majeur pour France Télécom. Ses missions consistent à anticiper les évolutions technologiques et les changements d'usages, à innover pour offrir aux clients le meilleur des technologies et explorer de nouvelles sources de croissance pour fournir au Groupe les assises d'un développement durable. Peu d'opérateurs dans le monde peuvent justifier d'un tel atout, pensé et organisé pour servir la stratégie globale de développement. Avec ses 8 sites en France, plus deux aux Etats-Unis (San Francisco et Boston), un au Japon (Tokyo) et un en Angleterre (Londres), il accompagne également le développement international de France Télécom en mettant son expertise au service de ses filiales étrangères, en s'impliquant auprès de grands groupes industriels et de la communauté scientifique mondiale et des instances internationales.

Pour répondre à ses ambitions, France Télécom R&D a accéléré sa réorganisation dès 1995 en se recentrant sur les domaines de recherche liés aux stratégies de marché de France Télécom et en donnant la priorité à la recherche appliquée. France Télécom R&D poursuit aujourd'hui son évolution et améliore en permanence sa capacité d'innovation à la fois technologique, marketing, managériale et organisationnelle. Ainsi, le centre de recherche contribue également à développer le capital humain du Groupe en jouant un rôle de vivier de compétences et en aidant les Ressources Humaines à anticiper les évolutions de métiers et les besoins en compétences.

Le stage s'est déroulé dans le centre de Recherche et Développement localisé à Lannion (22).

1.1.3 Présentation des services

France télécom est structuré en Directions. Ces directions comportent plusieurs laboratoires, eux-mêmes organisés en unité de Recherche et Développement. L'étude présentée dans ce rapport s'est effectuée en partenariat entre deux unités, l'unité Intégration de la Parole et du son(IPS), maîtrise d'ouvrage du projet, et l'unité Techniques Neuromimétiques pour les Télécommunications(TNT) la maîtrise d'oeuvre. Suite à la réorganisation récente (juillet 2004) de la structure de France Telecom R&D, nous présentons uniquement ici les objectifs et les activités de chaque service sans détailler leur direction.

Maitrise d'ouvrage du projet

La maitrise d'oeuvre du projet est la section Intégration de la Parole et du Son. Les activités d'IPS 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.

Les problèmes d'implémentation de la synthèse binaurale pour le grand public sont à l'origine du projet.

Maitrise d'oeuvre du projet

L'unité de Recherche et Développement TNT¹ (Techniques Neuromimétiques pour les Télécommunications) est chargée d'étudier, de proposer et de développer de nouvelles méthodes destinées aux services et aux réseaux de France Télécom, basées principalement sur des techniques d'apprentissage statistiques.

Ses domaines d'activité comportent en particulier le traitement de séquences d'images pour la communication de groupe, 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.

1.2 Problématique

L'étude développée dans ce rapport vient s'inscrire dans un objectif d'adaptation individuelle de la synthèse binaurale. La synthèse binaurale permet de recréer un espace sonore à 3 dimensions à partir d'un casque. Les applications possibles touchent des domaines tels que les télécommunications, la réalité virtuelle, le domaine militaire et le "design" sonore.

Pour reproduire et concevoir un système de spatialisation du son, il est important de bien comprendre les mécanismes mis en jeu lors du processus perceptif de localisation des sources sonores. Ces mécanismes sont complexes mais nous savons aujourd'hui sur quels indices ils se basent.

Les premiers indices de localisation découverts sont appelés indices binauraux car il n'existe 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 ipsilatérale) 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 (voir Mills [9]) : certaines positions spatiales possèdent le même couple ILD, ITD (voir figure 1.1).

Pour lever l'ambiguïté, il existe un autre indice appelé "indice monaural" (nécessitant un seul capteur sonore). C'est un indice spectral qui se caractérise par le filtrage du signal par notre morphologie. On définit ce filtrage par les fonctions de transfert de la tête : les HRTF (Head Related Transfer Function).

Le filtrage décrit 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.

¹Suite à la réorganisation, le service TNT est actuellement divisé en une section spécialisée en Traitement de l'image et une partie spécialisée en Traitement Statistique de l'information

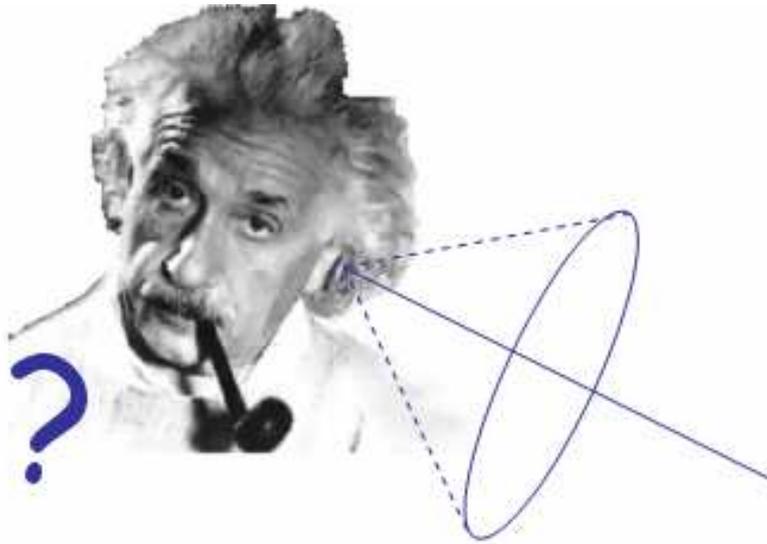


FIG. 1.1 – Cône de confusion

Elles sont obtenues en plaçant des micros à l'entrée du canal auditif et en faisant varier la position d'une source sonore large bande autour de l'auditeur. La connaissance du signal émis et du signal d'arrivée permet d'obtenir la fonction de transfert.

La limitation majeure actuelle de la synthèse binaurale réside dans le fait que les HRTF sont individuelles. 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. Cela se traduit par une perte d'externalisation des sources sonores et des confusions avant-arrière lorsqu'un individu utilise des HRTF qui ne sont pas les siennes.

Une solution serait de mesurer les HRTF de chaque individu, mais l'équipement et le temps nécessaires (plus de 2 heures de mesures) pour les obtenir rend cette solution inapplicable pour le grand public. Il faut donc trouver un moyen d'obtenir les HRTF d'un individu sans devoir pour autant les mesurer, c'est-à-dire trouver un moyen d'individualiser les HRTF.

1.3 Objet de l'étude

Les approches les plus couramment utilisées pour individualiser les HRTF sont présentées ci dessous (Voir Pernaux [10]) :

- *Sélection d'un jeu de HRTF au sein d'une base de données réduites par écoute interactive.* Cette approche est sans doute une des plus simples à mettre en oeuvre mais elle souffre cependant d'un champ d'application restreint et de performances perceptives réduites. Le jeu d'HRTF 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 (les positions testées) et éloigné pour d'autres.

- *Sélection de HRTF dans une base de données en fonction d'un critère de proximité morphologique.* Cette méthode proposée par Zoltkin and al [3] est également simple à mettre en place mais elle nécessite une base de données importante de HRTF, couplées à des mesures anthropométriques. De plus, les résultats de l'individualisation dépendent de manière importante de la proximité morphologique du sujet à individualiser avec un sujet de la base.
- *Modèle de signal des HRTF.* Cette méthode suivie par Middlebrooks [8] consiste à synthétiser le module d'un jeu de HRTF d'un sujet, à partir d'un jeu de HRTF, en effectuant des opérations de dilatation/compression (opération de warping) sur l'échelle des fréquences. A ce jour, les performances perceptives de cette approche d'individualisation n'ont pas été évaluées.
- *Modèle de HRTF basé sur une approche physique.* Cette approche est certainement la plus naturelle. Le but est de modéliser dans un environnement 3D la morphologie d'un individu puis de résoudre l'équation de propagation de l'onde entre la source et le tympan. Cette résolution se base le plus souvent sur la technique dite des éléments finis de frontière. Cependant, cette technique est très coûteuse en ressource informatique. Lorsque l'on augmente la précision du modèle (prise en compte de la forme du tympan, du torse etc), la complexité de résolution de l'équation augmente de façon exponentielle.
- *Modèles basés sur une décomposition linéaire de HRTF.* 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 de cette approche est d'obtenir les coefficients acoustiques qui permettront de reconstituer un jeu de 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 par Carlile, Jin et Leong [11].

Nous allons suivre une démarche proche de celle basée sur une décomposition linéaire des HRTF. Le but commun entre notre approche et celle de Jin et Leong est de répondre à la problématique suivante :

"Est-il possible de construire un modèle permettant d'obtenir, à partir de certains paramètres, certaines mesures etc..., les HRTF d'un individu ?"

L'approche menée par de Jin et Leong utilise les paramètres anthropomorphiques en entrée du modèle. Dans notre approche, nous allons évaluer s'il est possible d'obtenir l'intégralité des HRTF à partir d'un petit nombre d'HRTF. Les proximités spectrales entre des HRTF situées à des positions voisines peuvent être utilisées pour réduire le nombre de mesures à effectuer. Partant de cet a priori, nous pouvons reformuler notre problématique.

"Est-il possible de construire un modèle permettant d'obtenir, à partir d'un nombre restreint d'HRTF d'un individu, l'ensemble de ses HRTF ?(sachant que le modèle construit doit être utilisable quel que soit l'individu)"

Répondre à cette problématique revient à considérer les deux sous-problèmes suivants :

- Trouver le nombre et la localisation des HRTF suffisantes mais nécessaires à la reconstruction de l'ensemble des HRTF. Nous appellerons dans tout ce qui suit ces HRTF : les HRTF “représentatives”
- Créer un modèle permettant d'obtenir l'ensemble des HRTF d'un individu à partir des HRTF “représentatives”.

Pour répondre à ces deux problèmes, nous avons choisi les modes préférentiels suivants :

- Pour élire les HRTF “représentatives”, nous pouvons simplement prendre une nouvelle résolution spatiale de l'espace des mesures des HRTF. Par exemple, à la place d'utiliser des HRTF mesurées tous les 5° en élévation, nous pouvons garder une HRTF tous les 45° et ainsi diviser par 9 le nombre de mesures. Cette méthode ne tient pas compte du fait que la localisation sonore humaine n'est pas uniforme dans l'espace. Pour prendre en compte cette non-uniformité, nous allons diviser l'étape d'élection des HRTF “représentatives” en deux parties. Nous utiliserons une technique de clustering pour regrouper les HRTF proches dans l'espace des mesures. Parmi toutes les techniques existantes (k-moyennes, Partition Around Medoids, etc...), nous avons choisi les cartes de kohonen. L'avantage de ces cartes réside dans la visualisation sur un espace réduit (2D par exemple) de données de grandes dimensions. Ensuite nous allons désigner une HRTF “représentative” pour chaque cluster.
- Il existe plusieurs techniques qui permettent de construire un modèle permettant d'estimer l'ensemble des HRTF d'un individu à l'aide de ses HRTF “représentatives” : la régression linéaire, non linéaire, etc. Nous allons évaluer l'efficacité de ces méthodes et élire celle qui fonctionne le mieux.

Il n'existe pas, à l'heure actuelle, de modèles mathématiques pour synthétiser les HRTF. Notre approche consiste à essayer de construire une fonction de R^n vers R^n via des outils de régressions pour reconstruire les HRTF d'un individu.

En conséquence de ce qui précède, l'organisation de ce document est le suivant :

- Dans un premier temps, nous allons définir le terme HRTF “représentative”. Plusieurs types de représentants peuvent être choisis suivant les orientations de l'étude possibles. L'étude vise t-elle à reconstruire des HRTF en minimisant l'erreur entre les spectres reconstruits et ceux mesurés ? Devons-nous prendre en compte les propriétés de l'oreille humaine pour minimiser l'erreur perceptible ? Ces orientations vont permettre d'élire des HRTF “représentatives” dans des espaces différents (signal, perceptif, etc). Nous présentons également une manière de quantifier l'erreur de reconstruction des HRTF. L'erreur de reconstruction dépend d'un grand nombre de paramètres (fréquence, azimut, élévation, individu). Nous devons trouver une manière synthétique pour la visualiser.

- Une fois le terme “représentatif” défini, nous allons voir dans le chapitre 3, comment élire des représentants. Pour y arriver, nous avons décidé d’utiliser une méthode basée sur le clustering. Nous allons grouper les HRTF proches entre elles (au sens du domaine choisi dans le chapitre 2) via un outil statistique de clustering : les cartes de Kohonen. Une fois les groupements effectués, nous allons désigner au sein d’un même cluster qu’elle est l’HRTF la plus représentative. Cette opération va permettre d’avoir un représentant pour chaque cluster et donc des représentants pour l’ensemble des HRTF. Nous comparerons ces représentants avec des représentants choisis uniformément sur la sphère pour évaluer l’intérêt de notre méthode.
- Le chapitre 4 évalue les performances de plusieurs outils statistiques pour reconstituer les HRTF d’un individu. Il s’agira d’une étude préliminaire car les HRTF d’un seul individu seront reconstituées à partir de ses HRTF représentatives et une phase d’apprentissage réalisée sur une partie de ses HRTF. Plusieurs outils statistiques seront utilisés, puis comparés comme notamment la régression linéaire, non linéaire etc. Ce chapitre doit permettre de répondre à la question : “Est -il possible de modéliser des HRTF ?”
- Dans le chapitre 5, nous évaluerons les performances de différents modèles pour individualiser les HRTF d’un individu à partir de ses HRTF représentatives. Ce chapitre est le centre du rapport, il permettra de répondre à la question : “Est-il possible d’individualiser les HRTF ? et si oui, “comment perfectionner ces modèles pour reconstruire de façon la plus fidèle les fonctions de transfert ?”.

Chapitre 2

Données, Représentations, Evaluation

Ce chapitre présente l'origine des données que nous utilisons dans notre étude, les transformations apportées avant la sélection des HRTF représentatives et la méthodologie pour déterminer les erreurs d'estimation.

2.1 Introduction

Dans un premier temps, nous allons présenter dans ce chapitre l'origine des données utilisées dans notre étude. Nous décrivons la manière avec laquelle les mesures ont été effectuées ainsi que les traitements apportés pour obtenir les données sous forme d'HRTF.

Dans un second temps, nous devons définir le type de reconstruction d'HRTF désiré. Le but de notre étude est-il de reconstruire une HRTF en minimisant l'écart entre le spectre d'amplitude de l'HRTF reconstruite (obtenue par estimation) et de l'HRTF réelle (obtenue par mesures) ? Devons-nous mettre en valeur les petits écarts entre ces deux spectres d'amplitude et diminuer l'importance des grands écarts ? Devons-nous utiliser les caractéristiques de l'appareil auditif humain pour minimiser un écart "perceptif" entre ces deux spectres plutôt qu'un écart "signal" ? Plusieurs orientations sont possibles pour notre étude.

Enfin, nous présentons la méthodologie pour déterminer puis visualiser les erreurs d'estimation. D'une part, nous devons déterminer un critère de comparaison permettant de répondre à la question : Parmi deux spectres estimés, quel est le spectre fournissant une information de localisation la plus proche de l'HRTF réelle ? Il n'est pas trivial de répondre à cette question car à l'heure actuelle il n'existe pas de critère mathématique fiable à 100%. D'autre part, les erreurs d'estimation dépendent d'un nombre élevé de paramètres. Pour représenter les erreurs d'estimation, nous devons donc trouver une visualisation synthétique de leur répartition.

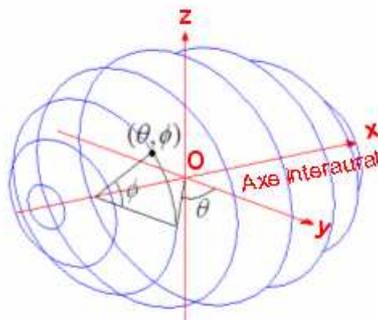


FIG. 2.1 – système polaire interaural

2.2 Origine des données

Plusieurs campagnes de mesures d’HRTF ont été menées notamment par l’IRCAM, le MIT et le CIPIC. Le processus de mesures étant très coûteux en temps et en équipement, nous avons exploité une base de donnée d’HRTF publique.

Nous nous sommes basés sur la campagne du CIPIC [14] dont les mesures décrivent les HRTF pour 1250 positions spatiales placées à un mètre de distance de la tête des sujets (l’éloignement ne joue pas un rôle essentiel sur l’allure de la fonction de transfert). La base de donnée des mesures présente les HRTF de 43 sujets (27 hommes et 16 femmes) aux morphologies différentes (les valeurs de plusieurs facteurs anthropomorphiques sont également disponibles).

2.2.1 Coordonnées spatiales des mesures

Pour chaque individu, les 1250 positions spatiales sont définies par leur azimut θ et par leur élévation ϕ . Cependant, Le repère du CIPIC est basé sur un système de coordonnées dit *polaire interaurale* (voir figure 2.1). En synthèse binaurale, les plans standard sont nommés :

- Le plan (xOy) ou *plan horizontal*
- Le plan (xOz) ou *plan frontal*
- L’axe (yOz) ou *plan médian*

Dans ce système de coordonnées, l’azimut est mesuré comme l’angle entre le plan médian et la source. On mesure l’élévation comme l’angle entre le plan horizontal et un plan passant par la source et l’axe des x , qui contient l’axe interaural. Les lignes d’azimut constant sont des arcs de cercles dont les centres sont répartis sur l’axe interaural.

Pour les mesures du CIPIC, l’élévation a été uniformément échantillonnée par pas de 5.625° ($360/64$) de -45° à $+230.625^\circ$. L’azimut a été échantillonné pour les valeurs $-80^\circ, -65^\circ, -55^\circ$ puis tous les 5° de -45° à 45° et enfin pour les valeurs $55^\circ, 65^\circ$ et 80° (figure 2.2).

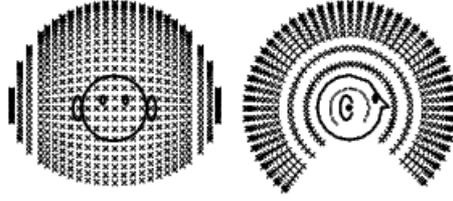


FIG. 2.2 – Coordonnées du CIPIC

2.2.2 Spécification des HRTF

Les informations du CIPIC sont données dans le domaine temporel sous forme de HRIR (Head Related Impulse Reponse) échantillonnées à 44.1kHz. Une FFT est appliquée sur les HRIR pour obtenir des HRTF.

Séparation amplitude/retard pur

Une HRTF est décrite par son module mag et sa phase ϕ (voir figure 2.3). Comme tout filtre causal et stable, chaque HRTF peut être divisée en une composante à

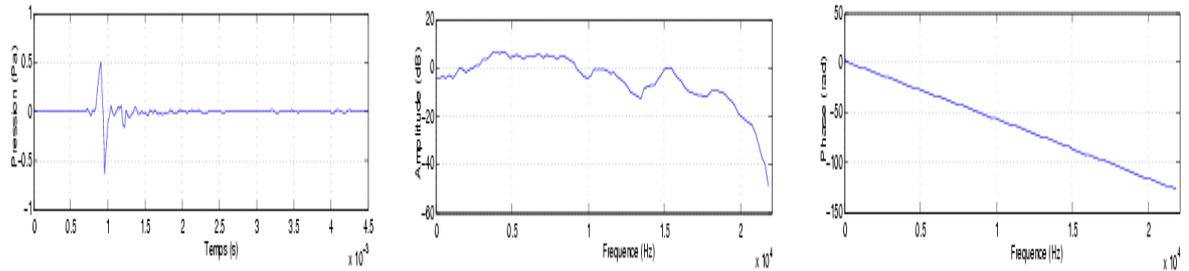


FIG. 2.3 – HRIR, module et phase de l’HRTF

phase minimal H_{min} et une composante à excès de phase H_{exc} :

$$HRTF = mag.exp^{j.\phi} = mag.exp^{j.mph}.exp^{j.eps} = H_{min}.H_{exc} \quad (2.1)$$

avec : $H_{min} = mag.exp^{j.mhp}$ et $H_{exc} = exp^{j.eph}$

La phase ϕ de chaque HRTF est ainsi divisée entre la phase de la composante minimale ou “phase minimale” notée mph et la partie à “excès de phase” noté eph .

La partie H_{exc} peut être décomposée par :

$$H_{exc} = H_{ap}.H_{lin} \quad (2.2)$$

avec H_{ap} la composante passe-tout et H_{lin} la composante à phase linéaire.

La phase de la composante à excès de phase étant quasiment linéaire, la composante passe-tout est négligée et la composante à phase linéaire est remplacée par un retard pur. La différence entre le retard à gauche et à droite donne l’ITD.

Pour notre étude, nous allons nous intéresser uniquement au module mag de la composante à phase minimale H_{min} . Par abus de langage, nous appellerons ce module $HRTF$.

Détail de l'échelle fréquentielle

L'échelle des fréquences des HRTF est échantillonnée sur 100 fréquences linéairement espacées de 0Hz à 22.05kHz. Nous utiliserons par la suite f_n pour désigner la n^e fréquence échantillonnée (n variant de 1 à 100). Nous noterons alors $HRTF(f_1)$ pour désigner l'amplitude de l'HRTF à la 1^{re} fréquence échantillonnée.

Finalement, nous définissons l'HRTF d'une personne λ pour une position spatiale définie par l'azimut θ et l'élévation ϕ par le vecteur :

$$\mathbf{V}_{\text{lin}}(\lambda, \theta, \phi) = \begin{pmatrix} HRTF_{\lambda, \theta, \phi}(f_1) \\ HRTF_{\lambda, \theta, \phi}(f_2) \\ \dots \\ HRTF_{\lambda, \theta, \phi}(f_{100}) \end{pmatrix} \quad (2.3)$$

La phase de clustering va permettre de regrouper les vecteurs proches entre eux. Sans rentrer dans les détails (voir chapitre 3), le calcul de proximité entre deux vecteurs $V_{(\lambda, \theta, \phi)}$ et $V_{(\lambda', \theta', \phi')}$ est basé sur la distance euclidienne¹ $d(V_{(\lambda, \theta, \phi)}, V_{(\lambda', \theta', \phi')})$.

$$\mathbf{d}_{\text{lin}}(\mathbf{V}_{\text{lin}}(\lambda, \theta, \phi), \mathbf{V}_{\text{lin}}(\lambda', \theta', \phi')) = \frac{1}{100} \sum_{n=1}^{100} (HRTF_{\lambda, \theta, \phi}(f_n) - HRTF_{\lambda', \theta', \phi'}(f_n))^2 \quad (2.4)$$

Cette distance est appelée fonction de coût. Elle est utilisée pour calculer une distance entre deux vecteurs. Il est important de souligner que toute modification de cette fonction ou des vecteurs modifie l'étape de clustering et donc l'orientation de l'étude. On préférera cependant modifier les données plutôt que la fonction de coût quand cela sera possible.

2.3 Modification des données

2.3.1 Modification de l'échelle des amplitudes

L'échelle des amplitudes des HRTF est initialement linéaire. J.O.Smith, dans une étude sur la synthèse des instruments du type violon [13], spécifie que l'écart entre deux spectres est corrélé avec l'écart en dB de leur amplitude. Ainsi, lorsque l'on travaille sur des données de nature audio, on préfère utiliser une échelle 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 -80dB en échelle logarithmique (qui équivaut à 10^{-4} en échelle d'amplitude linéaire). En terme de dynamique, nous considérons que le seuil de -80dB est largement suffisant, au-delà de cette limite l'oreille n'entend plus rien.

Finalement, nous allons transformer les vecteurs d'entrée comme suit :

¹La distance euclidienne d entre deux vecteurs $X = [x_1, x_2, x_3, \dots, x_n]^T$ et $Y = [y_1, y_2, y_3, \dots, y_n]^T$ est définie par la relation $d(X, Y) = \|X - Y\| = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2$

$$\mathbf{V}_{\log}(\lambda, \theta, \phi) = \begin{pmatrix} 20. \log_{10}(\max(HRTF_{\lambda, \theta, \phi}(f_1), 10^{-4})) \\ 20. \log_{10}(\max(HRTF_{\lambda, \theta, \phi}(f_2), 10^{-4})) \\ \dots \\ 20. \log_{10}(\max(HRTF_{\lambda, \theta, \phi}(f_{100}), 10^{-4})) \end{pmatrix} \quad (2.5)$$

Dans ce cas précis, la distance euclidienne d entre deux vecteurs $V_{log}(\lambda, \theta, \phi)$ et $V_{log}(\lambda', \theta', \phi')$ est définie par la relation :

$$\mathbf{d}_{\log}(\mathbf{V}_{\log}(\lambda, \theta, \phi), \mathbf{V}_{\log}(\lambda', \theta', \phi')) = \frac{1}{100} \sum_{n=1}^{100} \left(20. \log_{10} \left(\frac{\max(HRTF_{\lambda, \theta, \phi}(f_n), 10^{-4})}{\max(HRTF_{\lambda', \theta', \phi'}(f_n), 10^{-4})} \right) \right)^2 \quad (2.6)$$

2.3.2 Modification de l'échelle des fréquences

Les HRTF possèdent une échelle des fréquences linéaire. Or certaines disciplines, comme la psychoacoustique, spécifient que la résolution fréquentielle de l'oreille est plus complexe. Nous pouvons donc modifier nos données pour qu'elles soient perceptivement plus pertinentes.

Des études ont montré que les échelles fréquentielles telles que l'échelle de Barks (ou échelle des bandes critiques) ou l'échelle des ERB (Equivalent Rectangular Bandwidth en anglais) sont plus proches des propriétés de notre système auditif. De plus, il semble que l'échelle des ERB ait une meilleure validité théorique que l'échelle des Barks [2].

Pour ces échelles, la largeur de bande Δf est définie en fonction de la fréquence centrale f_n par :

$$\Delta f_{CB}(f_n) = 25 + 75 \left(1 + 1.4 \left(\frac{f_n}{1000} \right)^2 \right)^{0.69} \quad (2.7)$$

$$\Delta f_{CE}(f_n) = 0.108 \cdot f_n + 24.7 \quad (2.8)$$

De ces formules, nous pouvons déduire que plus la fréquence est élevée, plus la largeur de bande est élevée et moins l'oreille est sensible. La précision de l'oreille pour une fréquence donnée est alors inversement proportionnelle à la largeur de bande. Nous définissons alors une nouvelle erreur "perceptivement" pertinente entre deux spectres en appliquant un coefficient de pondération lors du calcul de la distance euclidienne. Ainsi, pour chaque fréquence f_n , il suffit de choisir un coefficient de pondération inversement proportionnel à la largeur de bande $\Delta f_{CE}(f_n)^2$.

Une distance "perceptive" d_{per} entre deux vecteurs $X = [x_1, x_2, x_3, \dots, x_n]^T$ et $Y = [y_1, y_2, y_3, \dots, y_n]^T$ peut être définie par la relation :

$$\mathbf{d}_{per}(\mathbf{X}, \mathbf{Y}) = \frac{1}{100} \sum_{n=1}^{100} \alpha_n (x_n - y_n)^2 \quad \text{ou} \quad \alpha_n = \frac{1}{\Delta f_{CE}(f_n)} \quad (2.9)$$

²Cette méthode s'inspire de Larcher[7] pour le calcul d'une erreur perceptive. On retrouve cette méthode dans la fonction "spectral distance" de la toolbox matlab de l'IRCAM

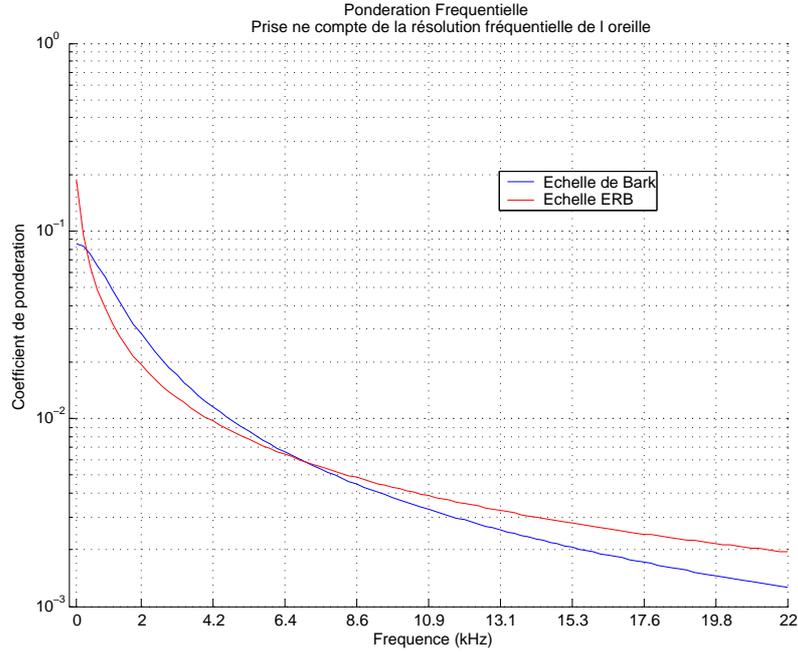


FIG. 2.4 – Coefficient de pondération

On effectue une normalisation pour que la somme de tous les poids soit égale à 1.

$$\alpha_n = \frac{1}{\Delta f_{CE}(n) \sum_{n=1}^{100} \frac{1}{\Delta f_{CE}(n)}}$$

La courbe qui suit (figure 2.4) présente les valeurs des coefficients suivant l'échelle de Barks et l'échelle ERB.

Nous pouvons directement inclure ces coefficients dans nos vecteurs d'entrée pour ne pas modifier la fonction de coût. L'étude utilisant ces nouveaux vecteurs sera appelée "étude orientée perceptif".

$$\mathbf{V}_{\text{per}}(\lambda, \theta, \phi) = \begin{pmatrix} 20 \cdot \sqrt{\alpha_1} \cdot \log_{10}(\max(HRTF_{az,el}(f_1), 10^{-4})) \\ 20 \cdot \sqrt{\alpha_2} \cdot \log_{10}(\max(HRTF_{az,el}(f_2), 10^{-4})) \\ \dots \\ 20 \cdot \sqrt{\alpha_{100}} \cdot \log_{10}(\max(HRTF_{az,el}(f_{100}), 10^{-4})) \end{pmatrix} \quad (2.10)$$

2.3.3 Egalisation des données en champ "libre" et champ "diffus"

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

- Le champ libre constitué d'une onde plane provenant d'une incidence donnée.
- Le champ diffus constitué d'ondes planes décorellées provenant d'incidences uniformément distribuées autour du récepteur.

Le but d'une égalisation champ 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 (Directional Transfer Function). L'égalisation champ diffus offre des avantages par rapport à l'égalisation champ libre car il permet, en plus, d'éliminer les artefacts de mesures indépendants de la direction, de réduire de façon significative les différences entre les sessions de mesures entre individus.

Pour égaliser nos données en champs diffus, nous avons utilisé la méthode de calcul proposée par Véronique Larcher [7] basée sur la moyenne énergétique des HRTF pour des mesures. La méthode consiste à estimer la HRTF en champ diffus par la moyenne en énergie des spectres d'amplitude des HRTF mesurées en champ libre. Une implémentation logicielle de cette méthode est disponible dans la toolbox binaurale de l'IRCAM.

2.4 Evaluation des performances

Pour évaluer les performances d'estimation des HRTF, nous définissons deux types d'erreur : 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 représentative (la localisation spatiale des HRTF représentatives est déterminée dans le chapitre suivant). Cette erreur permet de répondre à la question : Quelle est l'erreur commise lorsque l'on utilise un nombre de HRTF mesurées par individu inférieur à 1250 ? L'erreur de modélisation correspond à l'erreur commise lorsque l'on modélise une HRTF à partir des HRTF représentatives. La comparaison de la valeur de ces deux types d'erreur doit permettre d'évaluer l'efficacité de notre méthode de modélisation des HRTF.

Dans tous les cas, l'évaluation des performances est une opération complexe à deux niveaux :

- D'une part, les HRTF estimées peuvent provenir d'espaces différents³ et ne sont pas directement comparables ; comment comparer des données "Perceptives" avec des données "Signal" par exemple ? Pour comparer les performances d'estimation, nous devons déterminer un espace commun.
- D'autre part, les erreurs d'estimation dépendent d'un nombre important de paramètres (le numéro de l'individu λ , la position de la HRTF θ, ϕ et la fréquence f). Il est nécessaire de trouver une visualisation synthétique de la répartition des erreurs.

2.4.1 Espace commun de comparaison

Pour comparer les performances d'estimation, l'idéal serait de réaliser des tests perceptifs. Cette opération nécessite d'avoir à disposition les individus dont nous possédons les mesures d'HRTF. Ne disposant pas de ces individus, nous allons comparer les performances d'estimation mathématiquement. L'erreur mathématique que nous avons utilisée est basée sur la différence entre l'HRTF estimée et l'HRTF réelle dans l'espace "signal"

³certaines données n'ont pas le même sens physique, certaines sont des données "signals" d'autres des données "perceptives"

(voir 2.3.1).

Soit $V(\widehat{\lambda}, \widehat{\theta}, \phi)$ la HRTF qui approxime la HRTF réelle $V(\lambda, \theta, \phi)$.

Nous désignons l'erreur d'estimation de l'HRTF de position spatiale (λ, θ, ϕ) à la fréquence f par l'équation 2.11.

$$\mathbf{E}^{\text{approx}}(\lambda, \theta, \phi, \mathbf{f}) = V_{\log}(\widehat{\lambda}, \widehat{\theta}, \phi, f) - V_{\log}(\lambda, \theta, \phi, f) \quad (2.11)$$

ou $V_{\log}(\lambda, \theta, \phi, f)$ correspond à la valeur dans l'espace signal de la HRTF en θ, ϕ de l'individu λ à la fréquence $n^{\circ} f$.

Nous pouvons sommer l'erreur en valeur absolue sur toutes les fréquences (100 fréquences) pour obtenir une valeur indépendante de la fréquence.

$$\text{Erreur}(\lambda, \theta, \phi) = \frac{1}{100} \cdot \sum_{i=1}^{100} |V_{\log}(\widehat{\lambda}, \widehat{\theta}, \phi, i) - V_{\log}(\lambda, \theta, \phi, i)| \quad (2.12)$$

2.4.2 Visualisation synthétique des erreurs

L'erreur dépend de plusieurs paramètres qui sont l'individu λ , l'azimut θ , l'élévation ϕ et la fréquence f . Nous définissons les répartitions d'erreur suivantes.

- *Répartition spatiale de l'erreur* : Localisation des erreurs sur l'ensemble des mesures (c'est-à-dire une sphère)
- *Répartition fréquentielle de l'erreur* : Localisation des erreurs sur l'ensemble des fréquences (c'est-à-dire sur les 100 fréquences)
- *Répartition individuelle de l'erreur* : Localisation des erreurs sur l'ensemble des individus (44 individus du CIPIC)

Pour présenter les performances d'estimation des HRTF, il est important de trouver une visualisation synthétique de la répartition des erreurs. Nous ne représentons pas l'erreur individuelle qui sera synthétisée par la moyenne de l'erreur sur l'ensemble des individus.

Les deux visualisations que nous proposons sont les suivantes.

Répartition spatiale et fréquentielle

Le premier affichage représente l'erreur d'estimation spatiale pour plusieurs bandes fréquentielles. Nous avons effectué un regroupement des variables par outil statistique. Les différentes bandes utilisées (voir tableau 2.1) sont obtenues par carte de kohonen des variables (étude présentée en annexe A).

n° Bande	Largeur Bande	Nombre Fréquence
Bande 1	[0Hz-5.345kHz]	25
Bande 2	[5.568 kHz-10.468kHz]	23
Bande 3	[10.691kHz-17.150kHz]	30
Bande 4	[17.373kHz-19.155kHz]	9
Bande 5	[19.377 kHz-22.050kHz]	13

TAB. 2.1 – Bandes Spectrales

La figure 2.5 présente l’erreur d’une estimation dans une bande fréquentielle.

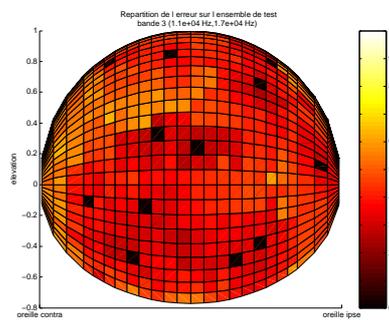


FIG. 2.5 – Répartition des erreurs lors d’une estimation dans une bande spectrale. Plus la surface de la sphère est noire, plus l’erreur d’estimation est faible.

Répartition fréquentielle

Le second affichage permet de visualiser plus précisément la répartition des erreurs sur l’ensemble des fréquences. Nous utilisons un “boxplot” (voir figure 2.6) de l’erreur calculée par l’équation 2.12. Cette outil de visualisation statistique présente la répartition des erreurs d’estimation pour une fréquence f donnée par :

- La valeur de la médiane des erreurs. Cette médiane est représentée par un trait rouge
- Les quartiles à 25% et 75%. Ces valeurs sont représentées par les extrémités du rectangle
- Les valeurs adjacentes correspondent à 1.5 fois la distance inter-quartile. Elles sont représentées par les “moustaches” du rectangle.
- Les erreurs supérieures et inférieures aux valeurs adjacentes. Elles sont affichées par une croix.

Nous indiquerons également sur les figures les extrémités de chaque bande par un trait noir (voir figure 2.7).

2.4.3 Liste des différentes orientations possibles

Nous pouvons suivre plusieurs orientations pour notre étude suivant les opérations apportées aux vecteurs d’entrée. L’étude n° 1 utilisant les données originales du CIPIC avec

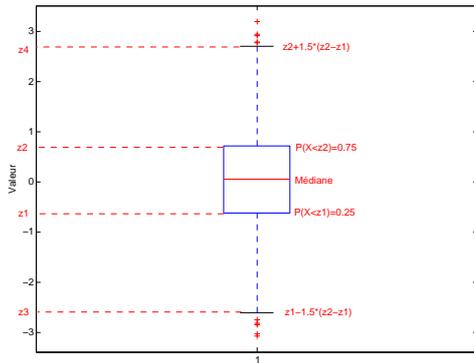


FIG. 2.6 – boxplot

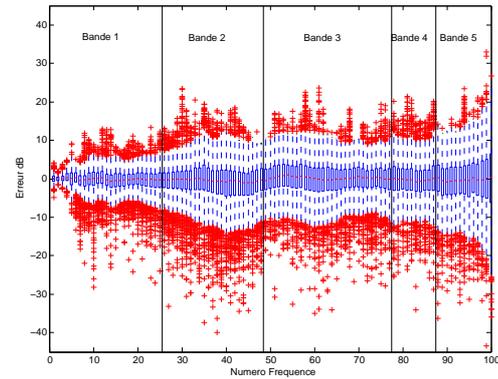


FIG. 2.7 – boxplot des erreurs

une échelle des amplitudes logarithmiques est appelée *étude orientée signal*. L'étude n° 2 utilisant une échelle logarithmique des amplitudes et une pondération ERB de l'échelle des fréquences est appelée *étude orientée perceptif*. Nous ajoutons le terme champ diffus au nom de l'orientation pour spécifier que les données ont été égalisées en champ diffus (étude n° 3 et n° 4).

Finalement, les 4 orientations possibles sont rappelées sur le tableau 2.2.

Numéro de l'étude	Echelle fréquence	Egalisation
Etude no 1	Linéaire	champ libre
Etude no 2	ERB	champ libre
Etude no 3	Linéaire	champ diffus
Etude no 4	ERB	champ diffus

TAB. 2.2 – Liste des différentes orientations

Chapitre 3

Sélection des HRTF représentatives

Le but de ce chapitre est de déterminer des HRTF “représentatives” qui seront par la suite utilisées pour estimer l’ensemble des HRTF.

3.1 Introduction

La démarche suivie pour déterminer les représentants, dite à deux étapes, est la suivante :

- Groupement des HRTF d’un individu les plus proches via une méthode de clustering.
- Sélection d’un représentant par cluster.

Le but des méthodes de clustering est de regrouper entre eux des individus “semblables” et de remplacer les groupes ainsi déterminés par un représentant (que ce représentant soit la moyenne, la médiane, le parangon du groupe n’est pas important). Un exemple de clustering est l’algorithme des k-moyennes. La grande difficulté de cet algorithme est sa non-reproductibilité et le choix difficile d’un nombre “optimal” de centres. Les techniques de classification hiérarchique ascendantes sont aussi des techniques de clustering, même s’il est plus difficile de trouver un représentant pour les classes (les classes ne sont pas nécessairement convexes et la moyenne de la classe peut être en dehors de la classe par exemple). Ces techniques sont difficiles à appliquer directement sur les données (trop d’individus). Souvent, ces techniques de classification hiérarchique, dont le but est plus de montrer la structure des données, sont appliquées après un passage de l’algorithme des k-moyennes avec k grand, ce qui permet d’éviter d’avoir à choisir au préalable le nombre de centres. Dans tous les cas, le défaut de ces techniques est que la description des données est encore faite dans un espace de grande dimension (même si c’est avec nettement moins d’individus) et qu’il reste à comparer ces individus entre eux pour prendre vraiment connaissance des données. La classification hiérarchique ascendante répond en

partie à cela en regroupant entre eux des groupes proches mais elle n’apprend rien par exemple sur les relations de plus ou moins grande proximité entre groupes classés dans des branches différentes de l’arbre. Pour répondre à ces problèmes, notre choix préférentiel s’est porté sur l’utilisation des *cartes de Kohonen* (voir [18]). Ces cartes permettent de visualiser un grand nombre de données dans un espace à faible dimension (souvent 2D) et de rendre compte des relations de proximité entre groupes. Suite au regroupement des HRTF proches par carte de Kohonen, nous allons désigner l’HRTF la plus “représentative” de chaque cluster.

Ce chapitre se décompose comme suit : Dans un premier temps la méthodologie pour élire les représentants est présentée, puis dans un second temps l’étude orientée “signal” est détaillée et permet au lecteur d’interpréter seul les résultats des autres orientations. Enfin, nous comparons la valeur de l’erreur d’approximation commise entre les HRTF et leur HRTF représentative (erreur que nous appelons *erreur de quantification*) pour l’ensemble des orientations. Cette comparaison permet d’élire l’orientation qui sera retenue pour la modélisation des HRTF par régression statistique.

3.2 Méthodologie

Cette première section expose la méthode à suivre pour élire les HRTF “représentatives”. Pour faciliter cette étude en terme de complexité de calcul et d’implémentation, nous avons établi deux hypothèses :

- La localisation des HRTF “représentatives” est la même pour tous les individus
- Seules les HRTF mesurées sur une oreille suffisent.

La première hypothèse permet de simplifier la démarche d’élection des HRTF en utilisant les données d’un *unique individu*. Il est beaucoup plus facile d’élire la position (θ', ϕ') des HRTF “représentatives” d’un individu dans un ensemble de couple (θ, ϕ) plutôt que d’élire les positions (θ', ϕ') des HRTF “représentatives” de tous les individus dans un ensemble de couples (λ, θ, ϕ) . Cette hypothèse pose néanmoins le problème de l’individualisation des HRTF “représentatives”

La seconde hypothèse nous permet de focaliser l’étude sur les HRTF mesurées sur une oreille uniquement (dans notre cas nous avons utilisé les HRTF de l’oreille droite). La morphologie humaine possède un plan de symétrie¹ suivant le plan médian, et par conséquent, l’ $HRTF(\lambda, \theta, \phi)$ mesurée sur l’oreille de droite est quasiment égale à l’ $HRTF(\lambda, -\theta, \phi)$ mesurée sur l’oreille gauche.

Compte tenu des hypothèses, l’objectif de cette section est de présenter la phase d’élection des HRTF “représentatives” d’un individu sur ses HRTF mesurées à l’entrée du canal auditif droit (1250 HRTF).

3.2.1 Clustering par Carte de Kohonen

Les cartes de Kohonen sont des réseaux de neurones qui appartiennent aux algorithmes de clustering. Nous allons les utiliser pour regrouper les 1250 HRTF d’un unique

¹Ce plan de symétrie est plus ou moins évident suivant les individus

individu. Une carte de Kohonen est composée d'un ensemble de k -points de 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. Il faut donc bien, dès à présent, comprendre que la topologie définit a 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 "à 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. Nous ne détaillons pas dans cette sous-section le processus d'apprentissage des cartes de Kohonen, le lecteur intéressé peut consulter l'ouvrage de référence sur les cartes de Kohonen [5].

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.

Il n'existe pas de formule empirique pour définir la topologie d'une carte de Kohonen. Il faut tester différentes architectures puis choisir celle qui semble être la mieux adaptée aux données à regrouper. Nous utiliserons dans notre étude des cartes de Kohonen 2D à voisinage hexagonal le plus souvent de taille 12×12 . Un exemple de carte Kohonen 2D à voisinage hexagonal de taille 12×12 obtenue après apprentissage est présenté sur la figure 3.1.

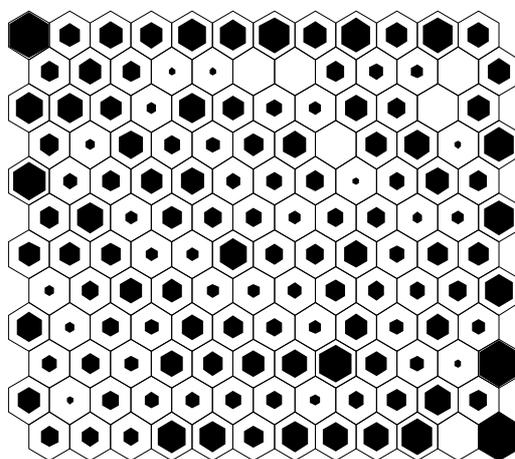


FIG. 3.1 – Carte de Kohonen à voisinage hexagonal de taille 12×12 neurones. La taille des losanges noirs indique le nombre d'HRTF compris dans chaque neurone.

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). 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". On indique le nombre d'HRTF "gagné" par chaque neurone via un losange noir

de taille variable.

Un des avantages des cartes de Kohonen est la possibilité de visualiser un grand nombre d'information sur la carte. Pour notre étude, on peut par exemple projeter l'information d'azimut et d'élévation de chaque HRTF sur la carte. Cette projection permet de répondre à la question : "Est ce que deux HRTF voisines sur la carte correspondent à deux HRTF mesurées à des positions spatiales proches ?".

Dans la sous-section suivante, nous présentons une façon de regrouper les neurones de la carte.

3.2.2 Clustering de la carte par CHA

Pour faciliter l'analyse quantitative de la carte [4], 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 les fusionne. Le critère de regroupement que nous avons choisi est la méthode de maximisation de la variance inter-classe et de minimisation de la variance intra-classe (Critère de Ward). Pour visualiser le comportement de la classification, nous construisons un arbre appelé *dendrogramme* (voir figure 3.2).

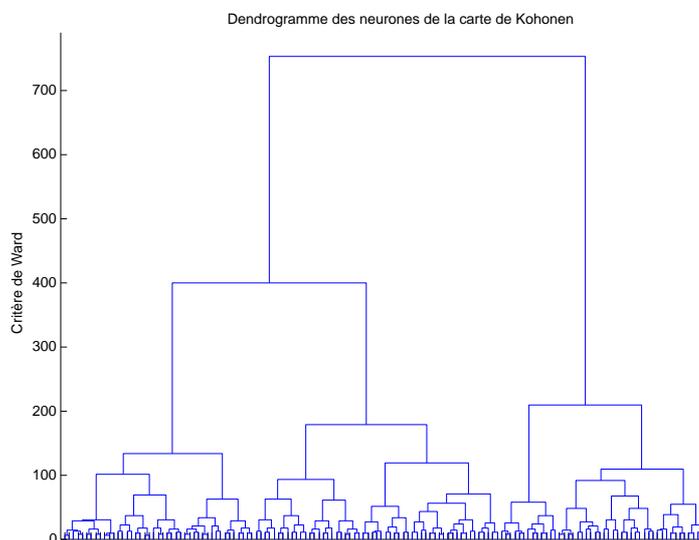


FIG. 3.2 – Exemple de dendrogramme. La base de l'arbre est composée de 144 clusters. A chaque étape de la construction de l'arbre, on regroupe les deux clusters les plus proches. Le sommet de l'arbre contient un seul cluster

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 ob-

jectif, d'obtenir un bon compromis entre le nombre de classes et la perte inévitable de la variance intra-classe.

La figure 3.3 présente l'évolution du nombre de classe en fonction du critère de ward.

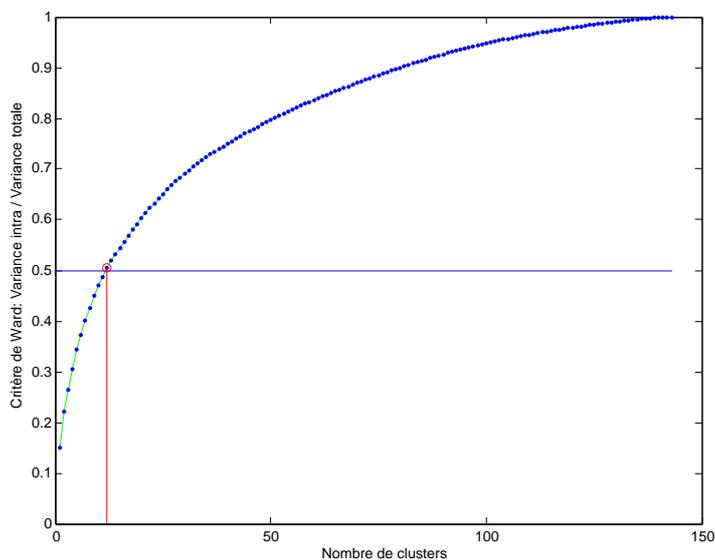


FIG. 3.3 – Evolution du critère de Ward en fonction du nombre de cluster. Pour un critère de ward égal à 0.5, nous obtenons 13 clusters

3.2.3 Election des représentants

Le but de cette sous-section est de répondre à la question : “Pour un cluster donné quel est la HRTF qui “représente” le mieux l’ensemble des HRTF appartenant à ce cluster ?” (voir figure 3.4)

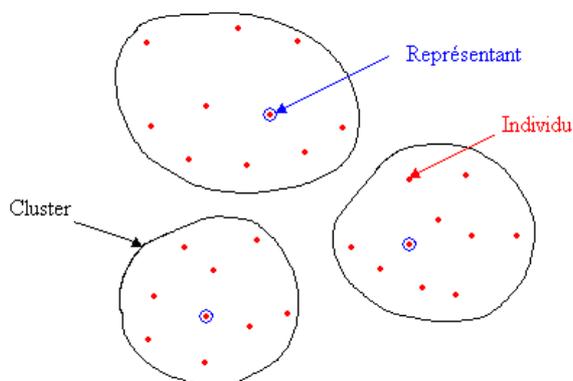


FIG. 3.4 – Des clusters et leur représentant

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.

$$\mathbf{r}_q = \underset{n \in C_q}{\operatorname{argmin}} \sum_{j=1}^{\operatorname{length}(C_q)} \sum_{i=1}^{100} (V_n(f_i) - V_j(f_i))^2 \quad (3.1)$$

Dans la section suivante, nous allons présenter une application de notre méthodologie pour sélectionner les HRTF “représentatives”.

3.3 Un exemple : l’étude orientée “Signal-Champ Libre”

La démarche à suivre pour sélectionner les HRTF “représentatives” de l’orientation signal est présentée dans cette section. Cette orientation utilise les vecteurs $V_{\log}(\theta, \phi)$ rappelés ci-dessous.²

$$\mathbf{V}_{\log}(\theta, \phi) = \begin{pmatrix} 20 \cdot \log_{10}(\max(HRTF_{\theta, \phi}(f_1), 10^{-4})) \\ 20 \cdot \log_{10}(\max(HRTF_{\theta, \phi}(f_2), 10^{-4})) \\ \dots \\ 20 \cdot \log_{10}(\max(HRTF_{\theta, \phi}(f_{100}), 10^{-4})) \end{pmatrix} \quad (3.2)$$

Nous allons regrouper ces vecteurs puis élire les positions (θ', ϕ') des HRTF représentatives.

3.3.1 Clustering par Carte de Kohonen

Après plusieurs tests expérimentaux, nous avons choisi d’utiliser une carte de Kohonen à voisinage hexagonal de topologie 12×12 dont la topologie est adaptée à nos données. La carte obtenue, après convergence de l’algorithme d’apprentissage, est présentée sur la figure 3.5.

²L’étude étant réalisée sur le premier individu, nous utilisons l’abus de notation $V_{\log}(\theta, \phi)$ pour désigner $V_{\log}(\lambda, \theta, \phi)$ avec $\lambda = 1$

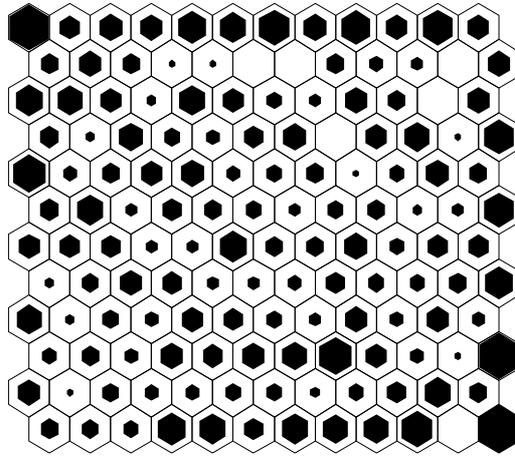


FIG. 3.5 – Carte de Kohonen regroupant les 1250 HRTF de l’espace signal champ libre en 12×12 neurones. La taille des losanges noirs indique le nombre d’HRTF compris dans chaque neurone. les hexagone blancs correspondent à des neurones vides

Cette première carte de Kohonen présente des problèmes d’homogénéité non souhaitables avec des regroupements de données spatialement éloignés dans l’espace des mesures et spatialement proches sur la carte de Kohonen. Ce problème d’homogénéité est présent à deux niveaux :

- Au niveau *Intra-neurone* : Nous caractérisons la non-homogénéité intra-neurone par un regroupement, au sein d’un même neurone, d’HRTF mesurées pour des positions spatialement éloignées.
- Au niveau *Inter-neurone* : Nous caractérisons la non-homogénéité inter-neurone par un regroupement d’HRTF spatialement éloignées dans des neurones voisins.

Pour visualiser les non-homogénéités inter-neurone sur la carte de Kohonen, nous pouvons projeté, via un code couleur, la moyenne des azimuts θ et des élévations ϕ de chaque neurone. Les non-homogénéités intra-neurone sont visualisées via une projection de l’écart type.

La figure 3.7 montre des non-homogénéités inter-neurone au Nord-Est³. Les élévations spatialement opposées sont réparties sur des neurones voisins (passage rapide du bleu au rouge).

La figure 3.9 montre des non-homogénéités intra-neurone visualisables par des neurones rouges (écart type de 120°). Les élévations spatialement opposées sont réparties sur un même neurone (présence de neurones rouges).

Pour éviter ces non-homogénéités, nous avons décidé de sectionner notre ensemble de vecteurs en deux sous-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

³S’agissant d’une carte, les expressions utilisées sont exactement les mêmes que celles utilisées en géographie “traditionnelle”

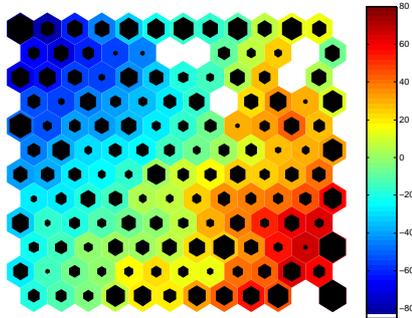


FIG. 3.6 – Moyenne de l'azimut : Les hexagones rouges correspondent à des neurones contenant en moyenne des HRTF situées à des azimuts élevés ($+80^\circ$) et les hexagones bleus, à des neurones contenant en moyenne des HRTF situées à des azimuts faibles (-80°).

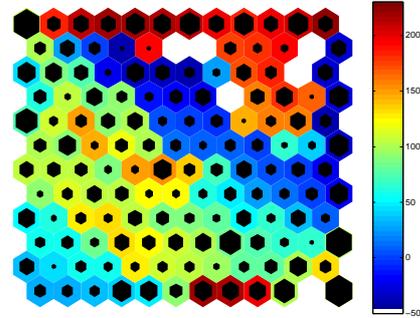


FIG. 3.7 – Moyenne de l'élévation : Les hexagones rouges correspondent à des neurones contenant en moyenne des HRTF situées à des élévations élevées (230°) et les hexagones bleus, à des neurones contenant en moyenne des HRTF situées à des élévations faibles (-50°).

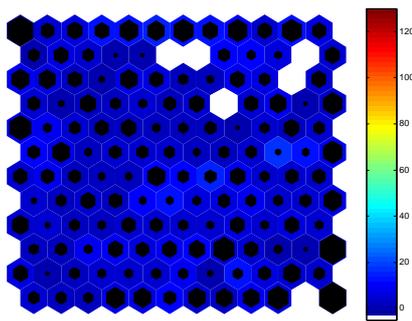


FIG. 3.8 – Ecart type de l'azimut : Les hexagones rouges correspondent à des neurones contenant des HRTF situées à des azimuts éloignées et les hexagones bleus, à des neurones contenant des HRTF situées à des azimuts proches.

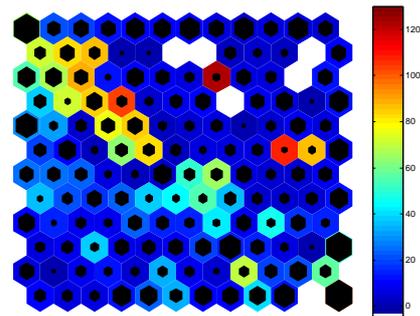


FIG. 3.9 – Ecart type élévation : Les hexagones rouges correspondent à des neurones contenant des HRTF situées à des élévations éloignées et les hexagones bleus, à des neurones contenant des HRTF situées à des élévations proches.

d'empêcher tout problème dans les regroupements en forçant les données de l'hémisphère avant et de l'hémisphère arrière à se séparer dès le départ.

$$\mathbf{V}_{\log}(\theta, \phi) = V_{\log}(\theta, -45 \leq \phi \leq 90) \cup V_{\log}(\theta, 90 \leq \phi \leq 230.625) \quad (3.3)$$

La sous-section qui suit présente l'étape de clustering pour ces deux ensembles d'HRTF.

3.3.2 Clustering par Carte de Kohonen avec Séparation des données Avant/Arrière

Nous allons utiliser deux cartes de Kohonen. Le choix de l'architecture et la taille de la carte sont déterminés en effectuant plusieurs essais. Une carte de taille 12×12 présente une bonne répartition des données avec peu de neurones vides et peu de non-homogénéités inter et intra-neurones. La répartition des HRTF sur deux cartes de Kohonen 12×12 à voisinage hexagonal est présentée sur les figures 3.10 et 3.11 respectivement pour les HRTF de devant et de derrière.

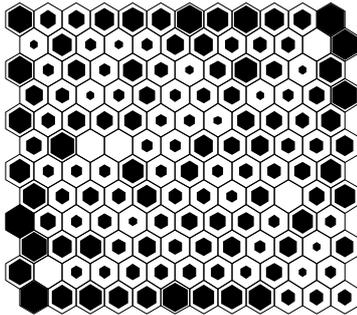


FIG. 3.10 – Répartition des HRTF de l'hémisphère avant sur sur une carte de Kohonen de taille 12×12 . La taille des losanges noirs indique le nombre d'HRTF compris dans chaque neurone.

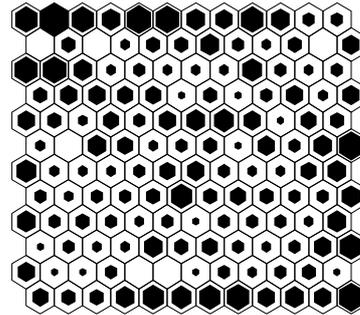


FIG. 3.11 – Répartition des HRTF de l'hémisphère arrière sur une carte de Kohonen de taille 12×12 . La taille des losanges noirs indique le nombre d'HRTF compris dans chaque neurone.

Nous pouvons projeter les azimuts et les élévations sur ces cartes pour détecter les non-homogénéités.

Répartition de l'azimut

Les cartes présentant la moyenne de l'azimut (Figures 3.12 et 3.13) ne sont pas non-homogènes dans le sens où l'on passe progressivement de l'oreille ipsilatérale à l'oreille contralatérale. Pour l'hémisphère avant, le passage de l'oreille ipsilatérale ($+80^\circ$) à l'oreille contralatérale (-80°) se situe sur la diagonale Sud-Ouest Nord-Est et pour l'hémisphère arrière, cette progression se fait sur la diagonale Nord-Ouest Sud-Est. La carte des écarts-types montre que les neurones contiennent des HRTF aux azimuts très proches (écart type maximum de 11.5° dans le cas de l'hémisphère avant et de 17.7° pour l'hémisphère arrière). Ces deux cartes ne présentent pas de non-homogénéités intra et inter-neurones.

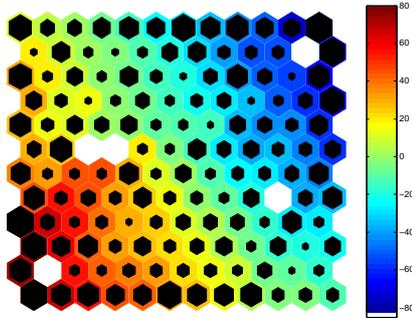


FIG. 3.12 – Moyenne de l'azimut sur la carte de l'hémisphère avant : Les hexagones rouges correspondent à des neurones contenant en moyenne des HRTF situées à des azimuts élevés ($+80^\circ$) et les hexagones bleus, à des neurones contenant en moyenne des HRTF situées à des azimuts faibles (-80°).

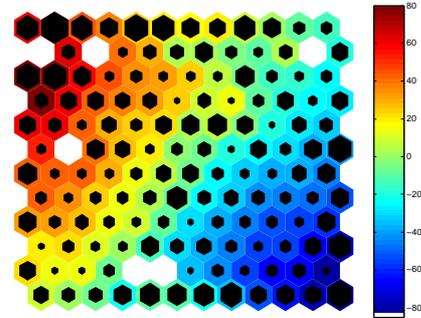


FIG. 3.13 – Moyenne de l'azimut sur la carte de l'hémisphère arrière : Les hexagones rouges correspondent à des neurones contenant en moyenne des HRTF situées à des azimuts élevés ($+80^\circ$) et les hexagones bleus, à des neurones contenant en moyenne des HRTF situées à des azimuts faibles (-80°).

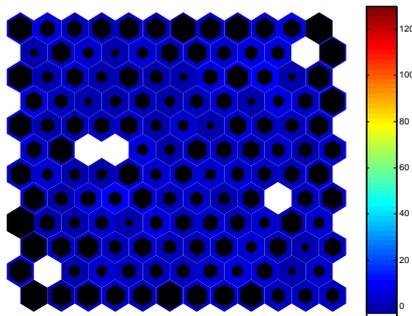


FIG. 3.14 – Ecart type de l'azimut sur la carte de l'hémisphère avant : Les hexagones rouges correspondent à des neurones contenant des HRTF situées à des azimuts éloignés et les hexagones bleus, à des neurones contenant des HRTF situées à des azimuts proches.

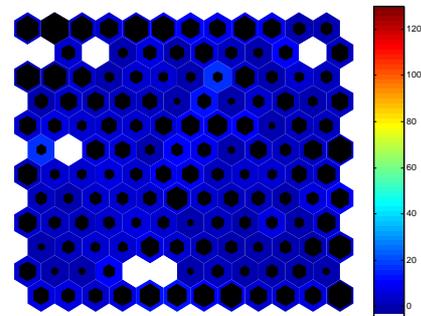


FIG. 3.15 – Ecart type de l'azimut sur la carte de l'hémisphère arrière : Les hexagones rouges correspondent à des neurones contenant des HRTF situées à des azimuts éloignés et les hexagones bleus, à des neurones contenant des HRTF situées à des azimuts proches.

Répartition de l'élévation

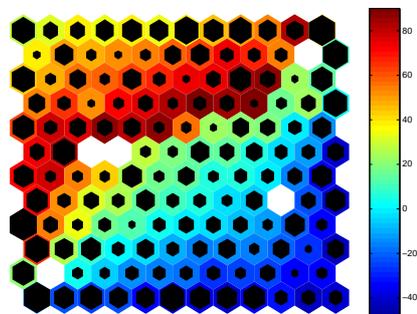


FIG. 3.16 – Moyenne de l'élévation sur la carte de l'hémisphère avant : Les hexagones rouges correspondent à des neurones contenant en moyenne des HRTF situées à des élévations élevés (90°) et les hexagones bleus, à des neurones contenant en moyenne des HRTF situées à des élévations faibles (-50°).

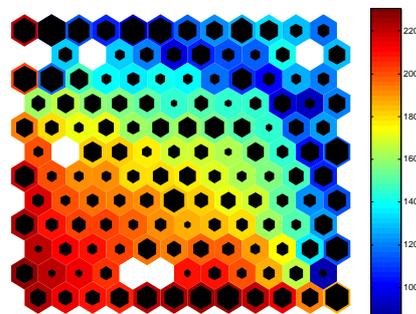


FIG. 3.17 – Moyenne de l'élévation sur la carte de l'hémisphère arrière : Les hexagones rouges correspondent à des neurones contenant en moyenne des HRTF situées à des élévations élevés (230°) et les hexagones bleus, à des neurones contenant en moyenne des HRTF situées à des élévations faibles (90°).

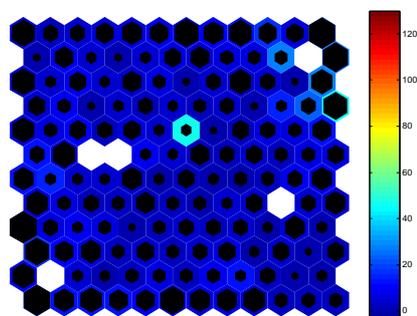


FIG. 3.18 – Ecart type de l'élévation sur la carte de l'hémisphère avant : Les hexagones rouges correspondent à des neurones contenant des HRTF situées à des élévations éloignées et les hexagones bleus, à des neurones contenant des HRTF situées à des élévations proches.

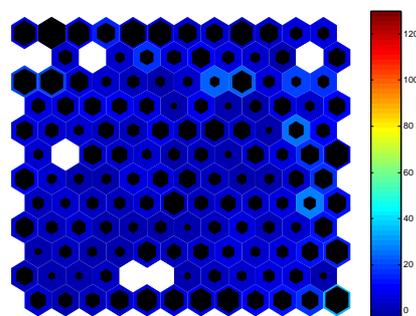


FIG. 3.19 – Ecart type de l'élévation sur la carte de l'hémisphère arrière : Les hexagones rouges correspondent à des neurones contenant des HRTF situées à des élévations éloignées et les hexagones bleus, à des neurones contenant des HRTF situées à des élévations proches.

Les cartes représentant l'élévation sont beaucoup plus perturbées (figures 3.16, 3.17, 3.18 et 3.19). Les différentes élévations sont rangées sur la diagonale Sud-Est Nord-Ouest pour l'hémisphère avant (3.12) et Nord-Est Sud-Ouest pour l'hémisphère arrière (3.13) mais leur évolution le long de ces diagonales n'est pas continue.

En ce qui concerne l'hémisphère avant, le centre de la carte comporte des neurones voi-

sins au sens de la topologie de la carte mais éloignés dans l'espace de mesures des HRTF (différence de 70° suivant l'élévation). La carte présentant les écarts type (3.14) montre également ce phénomène car certains neurones contiennent des HRTF éloignées en élévation (maximum de 47.7°). Cette non-homogénéité (intra et inter-neurone) suivant l'élévation est aussi présente pour l'hémisphère arrière. Les cartes de Kohonen présentent des proximités entre HRTF situées à des élévations éloignées. Cependant ces proximités étant moins nettes que pour la composante avant/arrière, nous estimons qu'il n'est pas nécessaire de sectionner l'ensemble des données par un plan horizontal situé à l'élévation 0.

Finalement, cette première étape de clustering par carte de Kohonen permet de regrouper les HRTF de l'hémisphère avant et de l'hémisphère arrière sur 2×144 neurones et de confronter la répartition des HRTF sur la carte avec celle de l'azimut et de l'élévation. Pour regrouper les neurones, nous utilisons un clustering par Classification hiérarchique ascendante.

3.3.3 Clustering de la carte par CHA

Nous allons utiliser une classification hiérarchique ascendante avec une agrégation réalisée à l'aide du critère de de ward pour regrouper les neurones de la carte de Kohonen. Le critère de ward que nous utilisons pour obtenir un nombre de classe doit être ni trop élevé (trop de classes) ni trop faible. Avec un critère faible, nous risquerions de regrouper des données situées à des élévations éloignées. Pour illustrer ce problème, la figure 3.20 présente la répartition des clusters obtenue avec un rapport entre l'inertie intra et inter-classe faible (rapport égal à 0.3) sur la carte de Kohonen. On remarque que chaque classe contient des neurones d'élévations éloignées.

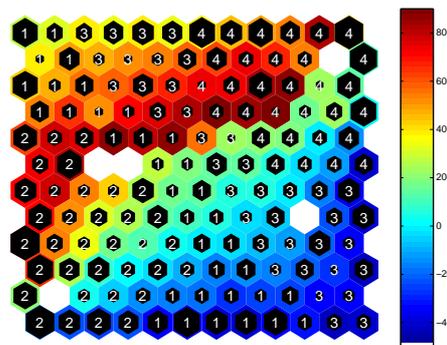


FIG. 3.20 – Répartition des clusters sur la carte 3.16 obtenue avec un critère de ward de 0.3. L'appartenance d'un neurone à un cluster est indiquée par un numéro qui correspond au numéro du cluster.

Pour éviter des variations intra-classe élevées au niveau de la répartition de l'élévation, nous avons utilisé un rapport égal à 0.5 pour regrouper les neurones de la carte de l'hémisphère avant et de l'hémisphère arrière. Ce rapport nous a semblé suffisant car il permet de regrouper dans les deux cas les 144 neurones en 13 classes⁴. Le clustering des

⁴Le fait que nous obtenons 13 classes pour l'hémisphère avant et pour l'hémisphère arrière est une coïncidence. Les données sont indépendantes, nous aurions pu avoir un nombre de classes différent entre ces deux

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 HRTF, en 2×144 neurones puis en 2×13 classes. Les clusters finaux sont présentés sur les figures 3.21 et 3.22 et les HRTF de chaque cluster sont visualisables en annexe.

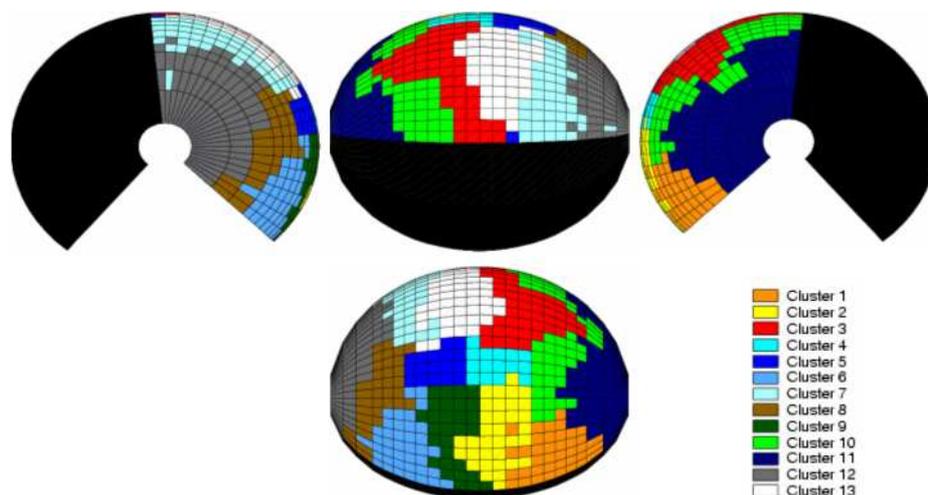


FIG. 3.21 – Cluster de l'hémisphère avant : Les zones de même couleur appartiennent au même cluster.

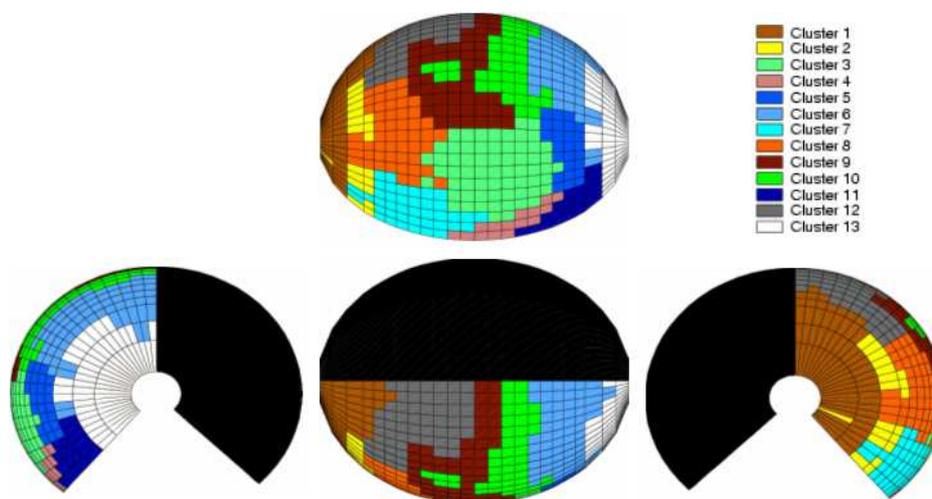


FIG. 3.22 – Cluster de l'hémisphère arrière : Les zones de même couleur appartiennent au même cluster.

Pour chacun des clusters, nous allons élire une HRTF représentative.

3.3.4 Election des représentants

Les positions géographiques des HRTF “représentatives” sont déterminées à partir de l’équation 3.1. Elles sont présentées sur la figure 3.23.

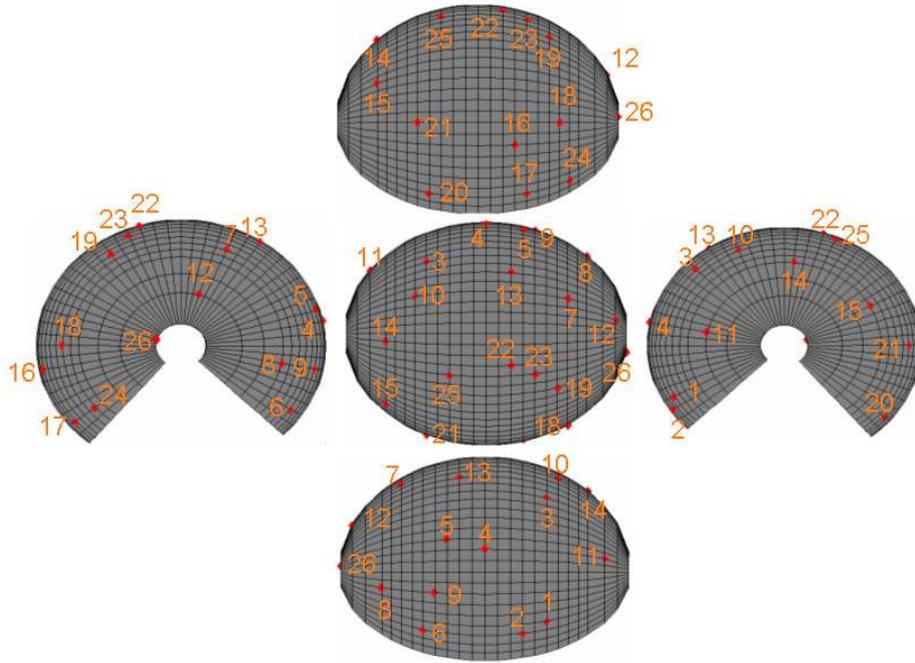


FIG. 3.23 – Localisation des représentants : Les numéros correspondent au numéro du cluster représenté.

Cette phase d’élection des représentants permet d’obtenir une liste de positions spatiales correspondant aux points de mesures des HRTF représentatives. Nous noterons par la suite (θ', ϕ') les coordonnées de l’HRTF “représentative” de l’HRTF mesurée aux coordonnées (θ, ϕ) .

La section suivante définit un critère qui permet de comparer l’efficacité des différentes orientations en fonction des positions (θ', ϕ') obtenues.

3.4 Comparaison des orientations

Les différentes orientations possibles sont rappelées dans le tableau 3.1 :

Numéro de l’étude	Echelle fréquence	Egalisation
Etude no 1	linéaire	champ libre
Etude no 2	ERB	champ libre
Etude no 3	linéaire	champ diffus
Etude no 4	ERB	champ diffus

TAB. 3.1 – Liste des orientations

Nous avons effectué la phase de sélection des HRTF “représentatives” pour toutes ces orientations. Les paramètres utilisés pour l’opération de clustering et d’élection sont les

suivants :

- Séparation des Données Avant/Arrière
- Clustering utilisant 2 cartes de Kohonen de topologie 12×12 à voisinage hexagonal
- Clustering par CHA utilisant une agrégation par critère de ward et une troncature pour un critère égal à 0.5
- Election des représentants avec le critère n° 1

Nous obtenons, pour chaque orientation, un nombre de couples (θ', ϕ') décrivant les positions spatiales des HRTF “représentatives”. La question qui se pose est la suivante : Quelle orientation devons-nous privilégier ? et Quels sont les meilleurs représentants ?

Les figures 3.24 3.25 3.26 et 3.26 présentent la répartition de l’erreur fréquentielle de quantification entre les HRTF et leur représentant dans l’espace “signal”. Chaque figure correspond à une orientation.

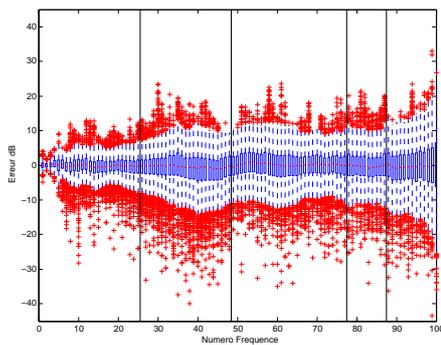


FIG. 3.24 – Erreur de quantification sur le premier individu. Représentants “Signal-Champ Libre”

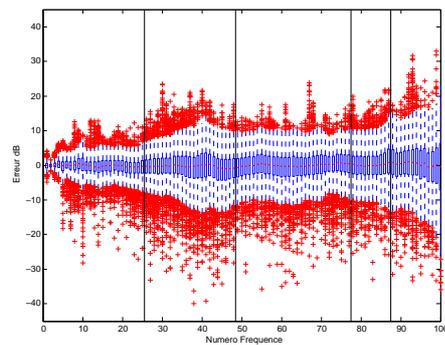


FIG. 3.25 – Erreur de quantification sur le premier individu. Représentants “Signal-Champ Diffus”

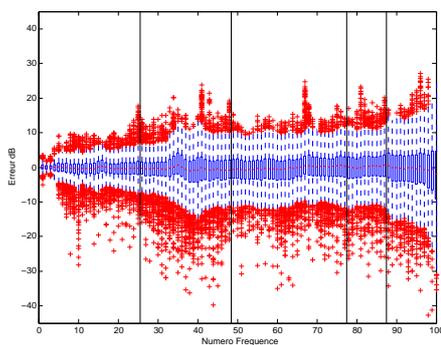


FIG. 3.26 – Erreur de quantification sur le premier individu. Représentants “Perceptif-Champ Libre”

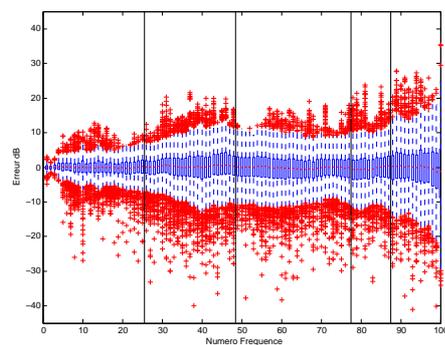


FIG. 3.27 – Erreur de quantification sur le premier individu. Représentants “Perceptif-Champ Diffus”

On remarque que l’erreur semble être symétrique avec toutefois une erreur légèrement plus élevée sur le bas. Le tableau 3.2 présente la valeur moyenne de l’erreur définie par

la relation 2.12 pour chaque orientation. L'erreur est calculée pour toutes les HRTF de l'individu n° 1.

n° Etude	Nombre Cluster	Erreur Moy	Erreur Min(5%)	Erreur Méd(50%)	Erreur Max(95%)
1	26	3.40 dB	1.68 dB	3.43 dB	5.17 dB
2	26	3.36 dB	1.63 dB	3.38 dB	5.15 dB
3	27	3.41 dB	1.63 dB	3.42 dB	5.28 dB
4	28	3.36 dB	1.69 dB	3.35 dB	5.23 dB

TAB. 3.2 – Erreur d'approximation

Les études n° 3 et n° 4 utilisent respectivement 27 et 28 représentants contre 26 représentants pour les études n° 1 et n° 2. Les résultats en terme d'erreur entre ces deux catégories d'études sont très proches. Nous allons donc privilégier les études n° 1 et n° 2 orientées "signal" qui offre le meilleur rapport nombre de représentant sur l'erreur moyenne. Les études n° 1 et n° 2 ont des erreurs proches même si l'orientation champ diffus présente des résultats légèrement meilleurs (moyenne de 3.368 dB contre 3.405 dB soit 1% de différence).

Pour la suite de l'étude, nous utiliserons les HRTF "représentatives" de l'orientation "Champ-Libre". Les études n° 1 et n° 2 étant proches en terme d'erreur, nous préférons utiliser l'orientation nécessitant le moins de transformation des données de départ.

3.5 Intérêt de la démarche de sélection des représentants

L'intérêt de notre démarche d'élection des représentants peut être évalué en comparant les 26 représentants que nous avons obtenus avec 26 représentants uniformément répartis dans l'espace. Ces représentants uniformément répartis sont choisis indépendamment de leur spectre. La méthodologie pour les obtenir, présentée en annexe, utilise l'agorithme des k-moyennes.

L'erreur calculée par la relation 2.12 est présentée via des boxplots pour les HRTF de l'individu n° 1 (figure 3.28, 3.29) et pour l'ensemble des HRTF de tous les individus (figure 3.30 et 3.31). Ces erreurs sont calculées pour les représentants de l'orientation "signal-champ libre" et pour des représentants uniformément répartis à la surface de la sphère.

L'erreur définie par l'équation 2.12 est présentée sur le tableau 3.3.

Individu	Type Orientation	Erreur Moy	Erreur Min(5%)	Erreur Méd(50%)	Erreur Max(95%)
n° 1	Signal-Champ Libre	3.40 dB	1.68 dB	3.43 dB	5.17 dB
	Uniforme	3.79 dB	1.66 dB	3.72 dB	6.95 dB
Tous	Signal-Champ Libre	4.31 dB	1.95 dB	4.21 dB	6.98 dB
	Uniforme	3.93 dB	1.81 dB	3.86 dB	7.06 dB

TAB. 3.3 – Erreur d'approximation

Notre démarche d'élection réduit l'erreur de 10% pour l'individu n° 1 par rapport à

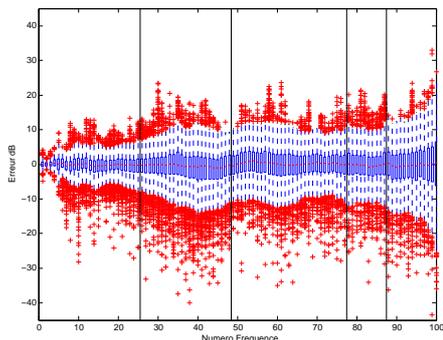


FIG. 3.28 – Erreur de quantification sur le premier individu. Représentants “Signal-Champ Libre”.

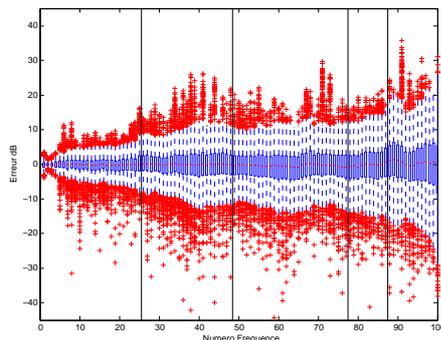


FIG. 3.29 – Erreur de quantification sur le premier individu. Représentants “Uniformément répartis”.

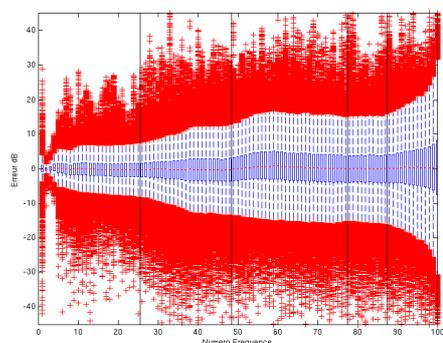


FIG. 3.30 – Erreur de quantification sur tous les individus. Représentants “Signal-Champ Libre”.

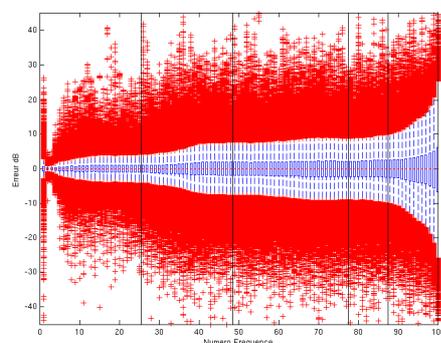


FIG. 3.31 – Erreur de quantification sur le tous les individus. Représentants “Uniformément répartis”.

des représentants uniformément répartis à la surface de la sphère. Par contre, lorsque l’on calcule l’erreur pour tous les individus de la base du CIPIC (44 individus), l’erreur commise est plus faible avec les représentants uniformément répartis (amélioration de 8%). Ces différences de performances dépendent du nombre de représentants. Nous avons donc dressé la courbe de l’erreur de quantification moyenne en fonction du nombre de représentants pour l’individu n° 1 (figure 3.32) et pour l’ensemble des individus (figure 3.33). Le nombre de représentants obtenu par carte de Kohonen évolue directement en fonction du critère de ward. Nous avons fait évoluer la valeur de ce critère pour obtenir différents nombres de clusters puis appliquer notre démarche d’élection des représentants sur chaque cluster obtenu. En toute rigueur, nous aurions dû construire une nouvelle carte de Kohonen à chaque fois.

Les courbes présentant l’erreur de quantification en fonction du nombre de représentants semblent préconiser l’utilisation des HRTF uniformément réparties pour la modélisation des HRTF. Cependant, l’individu λ utilisé pour élire les HRTF représentatives n’est peut être pas le plus représentatif de la base. Cela nous amène à considérer la question suivante : “ Existe-t’il un individu λ dont les représentants, déterminés via notre méthode d’élection, offriraient une meilleure erreur de quantification que les représentants unifor-

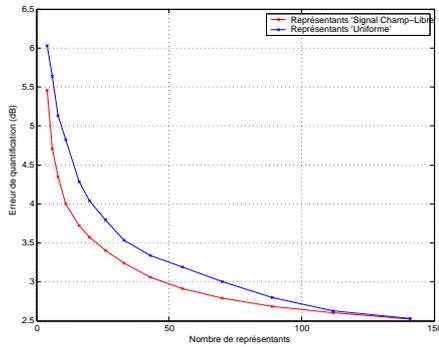


FIG. 3.32 – Erreur de quantification sur l'individu n° 1 : Les deux courbes sont décroissantes ce qui signifie que l'erreur décroît en fonction du nombre de représentants. L'erreur commise avec les représentants des cartes de Kohonen est plus faible qu'avec les représentants uniformément répartis.

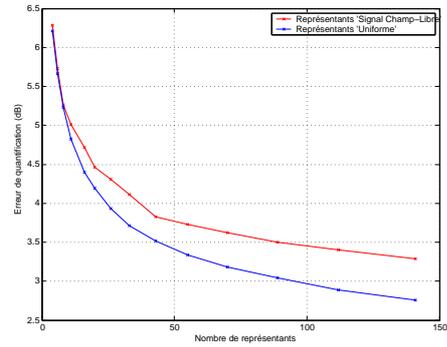


FIG. 3.33 – Erreur de quantification sur l'ensemble des individus : Les deux courbes sont décroissantes ce qui signifie que l'erreur décroît en fonction du nombre de représentants. L'erreur commise avec les représentants uniformément répartis est plus faible qu'avec les représentants déterminés par carte de Kohonen.

mement répartis ?”

Pour répondre à cette problématique, il faudrait déterminer par carte de Kohonen les représentants de chaque individu de la base du CIPIC. Cette étude coûteuse en ressource informatique⁵ n'a pas été menée lors du stage.

⁵Une carte de Kohonen des HRTF provenant d'un hémisphère d'un unique individu est construite en 6H

3.6 Discussion

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.

La phase clustering des HRTF permet également de visualiser des rapprochements entre les spectres des HRTF et leur position spatiale. Wightman [19] indique que les confusions avant-arrière sont essentiellement causées par une symétrie par rapport au plan frontal de l'ILD et de l'ITD. Notre étude permet d'enrichir le débat sur les confusions avant-arrière en précisant que les HRTF, tout comme l'ILD et l'ITD, présentent une symétrie par rapport au plan frontal. Cette symétrie semble provenir du fait que notre tête, notre corps, et dans une moindre mesure, notre pavillon de l'oreille externe, présente une symétrie suivant le plan frontal. De même, le phénomène de confusion haut-bas semble être lié à une symétrie des HRTF par rapport au plan horizontal. Ces confusions sont moins fréquentes que les confusions avant-arrière car il est probable que la symétrie par rapport au plan frontal est plus nette que la symétrie par rapport au plan horizontal. En effet les signaux sonores captés par l'oreille ipsilatérale contiennent, dans le cas où la source est située en bas de la tête ($\phi < 0^\circ \cup \phi > 180^\circ$), des réflexions du signal sonore provoquées par le corps. A contrario, ces réflexions ne sont pas captées lorsque la source sonore est située en haut de la tête ($0^\circ < \phi < 180^\circ$).

En ce qui concerne l'efficacité de notre méthode de sélection, elle souffre d'un problème d'individualisation des positions des HRTF "représentatives". Les représentants uniformément répartis à la surface de la sphère offrent de meilleurs résultats que nos représentants. Cela vient du fait que les HRTF sont intraséquentement individuelles.

Nous pouvons envisager une étude supplémentaire pour déterminer si la position des HRTF représentatives d'un individu λ de la base offrent une erreur de quantification meilleure que celle obtenue avec les représentants uniformes. Nous pensons que les têtes artificielles peuvent donner de bons résultats car elles possèdent des facteurs anthropomorphiques moyens par rapport à l'ensemble de la population.

Chapitre 4

Modélisation d'HRTF par régression statistique

Dans ce chapitre applicatif, nous allons mettre en oeuvre des outils statistiques pour reconstruire les HRTF d'un individu. Le but ici est de constater s'il est possible d'obtenir les HRTF d'un individu à partir de ses HRTF représentatives et d'un apprentissage réalisé sur plusieurs de ses HRTF. Cette étude est une étude préliminaire car nous allons considérer les HRTF d'un *unique* individu, le modèle n'effectue aucune individualisation.

Des connaissances en réseaux de neurones de type Multi Layer Perceptron sont nécessaires pour aborder la lecture de ce chapitre (Une introduction sur les réseaux de neurones est disponible en annexe).

4.1 Introduction

Avant de commencer les prédictions pour individualiser les HRTF d'un individu à partir des HRTF représentatives, nous allons tester différents modèles statistiques pour prédire les HRTF d'un individu à partir de ses HRTF représentatives et d'une phase d'apprentissage réalisée sur ses HRTF. Ces premiers tests doivent permettre de répondre à la question : Est ce qu'un modèle statistique permet de reconstruire des HRTF qu'il n'a pas apprises ? Nous décomposons ce chapitre comme suit :

Dans un premier temps, nous allons sectionner nos données en plusieurs ensembles. Pour évaluer objectivement les performances d'une régression obtenue avec un outil statistique, la méthode la plus simple est de couper les données à estimer en deux ensembles indépendants : le premier, noté \mathcal{D}_{Appr} , est utilisé pour l'apprentissage et le second, noté \mathcal{D}_{Vali} sert à mesurer les performances. Ce second ensemble est appelé ensemble de validation. Lorsque l'on teste différentes architectures du réseau, on pourrait croire qu'il suffit de mesurer pour chaque méthode la performance empirique faisant appel à l'ensemble de validation. Cela serait pourtant commettre la faute de faire dépendre cette mesure du réglage de la méthode alors qu'elle doit en être indépendante. En procédant de la sorte, on s'arrange pour minimiser le risque mesuré sur l'échantillon de test et l'on règle la mé-

thode en fonction de cet échantillon. A force de vouloir minimiser ce risque, on risque d’adapter étroitement la méthode d’apprentissage à cet échantillon. Cela est dangereux car il se peut qu’à tant poursuivre ce but, on s’éloigne d’une diminution du risque réel [1]. C’est pourquoi, on prévoit, en plus de l’ensemble d’apprentissage et de validation, un ensemble de test \mathcal{D}_{Test} sur lequel on évalue la performance réelle de la méthode.

Dans un second temps, nous définissons les vecteurs d’entrée utilisés pour la prédiction. Les données utilisées en entrée des outils de régression sont les HRTF “représentatives” du chapitre précédant. Mais comment devons-nous présenter ces données aux outils statistiques de régression ? Devons-nous préciser la localisation des HRTF “représentatives” ? Est-il judicieux de donner toutes les HRTF “représentatives” pour estimer une HRTF ?

Enfin, dans un troisième temps, nous présentons l’estimation des HRTF obtenues avec 2 outils de régressions statistiques, la régression linéaire et la régression non linéaire obtenues avec un réseau de neurones type perceptron multi-couche.

4.2 Sélection des ensembles d’apprentissage, de validation et de test

Nous avons divisé l’ensemble des HRTF, noté \mathcal{D}_{HRTF} en 4 ensembles disjoints :

- Les HRTF représentatives \mathcal{D}_{Rep}
- L’ensemble d’apprentissage \mathcal{D}_{Appr}
- L’ensemble de validation \mathcal{D}_{Vali}
- L’ensemble de test \mathcal{D}_{Test}

Les ensembles \mathcal{D}_{Appr} , \mathcal{D}_{Vali} et \mathcal{D}_{Test} sont pris des proportions respectives de 50%, 25% et 25% dans l’ensemble $\mathcal{D}_{HRTF} - \mathcal{D}_{Rep}$. La répartition de ces ensembles est présentée sur la figure 4.1. L’ensemble \mathcal{D}_{Rep} est représenté par des losanges rouges, l’ensemble \mathcal{D}_{Appr} par des ronds bleus, l’ensemble \mathcal{D}_{Vali} par des ronds jaunes et l’ensemble \mathcal{D}_{Test} par des ronds noirs.

4.3 Choix du vecteur d’entrée

Les représentants que nous allons utiliser sont ceux de l’espace “Signal-Champ Libre”. Nous rappelons que dans l’espace signal, l’HRTF(θ, ϕ) est désignée par le vecteur $V_{log}(\theta, \phi)$

Nous allons émettre l’hypothèse que l’importance d’une HRTF “représentative”, lors de l’estimation d’une HRTF, est inversement proportionnelle à la distance spatiale entre ces deux HRTF. Nous allons donc pondérer l’importance des HRTF “représentatives” en fonction de la distance spatiale qui la sépare de l’HRTF à estimer. Pour ce faire, nous ajoutons comme information aux entrées les distances angulaires ($\theta' - \theta$) et ($\phi' - \phi$)

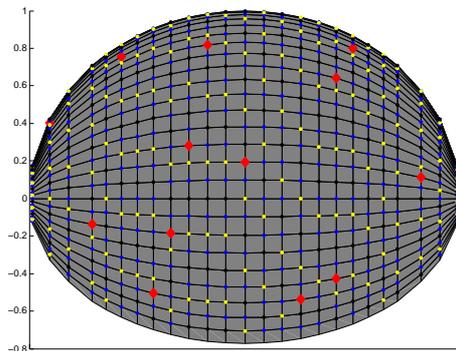


FIG. 4.1 – Répartition des ensembles d’apprentissage, de validation et de test. Les points rouges de la sphère indiquent la localisation spatiale des HRTF représentatives. La localisation spatiale des HRTF appartenant à l’ensemble d’apprentissage sont représentées en bleu. Celles appartenant à l’ensemble de validation sont représentées en jaune et celles appartenant à l’ensemble de test, en noir.

Nous proposons 2 jeux de vecteurs d’entrée qui diffèrent par le nombre de représentants que nous utilisons.

4.3.1 Utilisation de tous les représentants

Pour estimer une HRTF, nous allons donner comme entrée au réseau de neurones l’ensemble des HRTF “représentatives” de l’individu. Ces HRTF représentatives correspondent aux HRTF de l’individu mesurées aux positions θ' et ϕ' déterminées dans la sous-section 3.3.4.

Soit n , le n^{e} représentant de l’espace signal dont les coordonnées spatiales sont (θ'_n, ϕ'_n) . Le vecteur $Vect_1$ envoyé à l’outil de régression statistique est défini par :

$$\mathbf{Vect}_1 = [\theta, \phi, V_{log}(\theta'_1, \phi'_1), (\theta'_1 - \theta), (\phi'_1 - \phi), V_{log}(\theta'_2, \phi'_2), \dots, (\theta'_n - \theta), (\phi'_n - \phi)] \quad (4.1)$$

Ce vecteur présente un nombre d’informations élevé dont certaines jouent un rôle très faible voire nul pour l’estimation¹. Pour éviter de charger trop d’informations qui n’ont qu’un impact mineur, nous allons considérer d’autres vecteurs d’entrée.

4.3.2 Utilisation d’un représentant

Pour modéliser une HRTF en (θ, ϕ) , nous déterminons dans un premier temps le cluster qui contient la position spatiale (θ, ϕ) (voir sous-section 3.3.3). Dans un second temps, nous déterminons la position spatiale (θ', ϕ') du représentant du cluster (voir sous-section 3.3.4). Le vecteur d’entrée $Vect_2$ utilise uniquement la valeur de l’HRTF en (θ', ϕ') pour estimer une HRTF en (θ, ϕ) .

¹Par exemple, l’impact des représentants situés vers l’oreille contralatérale pour l’estimation d’une HRTF située vers l’oreille ipsilatérale est très faible voire nul

Soit (θ', ϕ') la position spatiale du représentant de l'HRTF (θ, ϕ) . le vecteur $Vect_2$ est défini par :

$$\mathbf{Vect}_2 = [\theta, \phi, V_{log}(\theta', \phi'), (\theta' - \theta), (\phi' - \phi)] \quad (4.2)$$

Pour simplifier le travail du réseau de neurones, nous allons reconstruire dans ce cas précis la différence entre $V(\theta, \phi)$ et $V(\theta', \phi')$. La tâche de prédiction sert à estimer les modifications à apporter à la HRTF représentative pour obtenir l'HRTF en (θ, ϕ) .

Ces données sont couramment normalisées avant d'effectuer une régression.

4.4 Type de normalisation utilisée

Avant d'effectuer une régression, nous allons normaliser nos données. Pour toutes les expériences réalisées dans cette partie, les données d'apprentissage et de test sont centrées par rapport à la moyenne de toute la série, puis normalisées entre $[-1,1]$. Ainsi, toutes les valeurs sont ramenées dans les bornes de la fonction sigmoïdale utilisée lors de la régression non linéaire. Pour éviter que certaines valeurs soient biaisées par des bruit de mesure lors des mesures des HRTF, nous proposons la méthode de normalisation suivante :

- Soit X une variable
- On recherche z_1 telle que $P(X < z_1) = 0.01$.
- On recherche z_2 telle que $P(X > z_2) = 0.99$.
- On normalise X par $X_{norm} = -1 + \frac{2 \cdot (X - z_1)}{z_2 - z_1}$

4.5 Estimation en régression linéaire

La régression linéaire est la forme la plus simple de régression. Nous allons tester ses performances avec les vecteurs d'entrée $Vect_1$ et $Vect_2$ pour reconstruire les HRTF de l'individu n° 1. La régression non linéaire est obtenue avec un réseau de neurones MLP utilisant des fonctions d'activation linéaires et un seul neurone dans la couche cachée.

4.5.1 Utilisation d'un seul représentant

Les erreurs de reconstruction par bande sont présentées dans le tableau 4.1

Numero Bande	Erreur Moyenne	Erreur Minimum (5%)	Erreur Médiane	Erreur Maximum (95%)
n° 1	1.95 dB	0.90 dB	1.75 dB	3.58 dB
n° 2	4.71 dB	2.44 dB	4.71 dB	7.17 dB
n° 3	3.57 dB	1.38 dB	3.37 dB	6.60 dB
n° 4	3.40 dB	0.96 dB	3.06 dB	6.95 dB
n° 5	4.92 dB	1.89 dB	4.60 dB	8.65 dB
Toutes	3.59 dB	2.29 dB	3.60 dB	5.04 dB

TAB. 4.1 – Erreur de reconstruction

L’erreur de modélisation est plus élevée que l’erreur de quantification (voir tableau 3.3). La modélisation par régression linéaire utilisant comme vecteur d’entrée le vecteur $Vect_2$ ne fonctionne pas.

4.5.2 Utilisation des 26 représentants

Les erreurs de reconstruction par bande sont présentées dans le tableau 4.2

Numero Bande	Erreur Moyenne	Erreur Minimum (5%)	Erreur Médiane	Erreur Maximum (95%)
n° 1	2.07 dB	0.99 dB	1.96 dB	3.69 dB
n° 2	4.79 dB	2.42 dB	4.74 dB	7.13 dB
n° 3	4.10 dB	1.72 dB	3.74 dB	7.92 dB
n° 4	3.86 dB	1.17 dB	3.39 dB	7.69 dB
n° 5	5.07 dB	2.20 dB	4.72 dB	9.67 dB
Toutes	3.85 dB	2.36 dB	3.85 dB	5.24 dB

TAB. 4.2 – Erreur de reconstruction

Nous obtenons également une erreur de reconstruction élevée. Il semble que les modèles en régression linéaire offrent de mauvais résultats². Nous allons tester dans la section qui suit des modèles utilisant des régressions non linéaires.

4.6 Estimation en régression non linéaire

Les réseaux de neurones MLP sont des “approximateurs universels”. Ils peuvent approximer avec une précision arbitraire n’importe quelle transformation continue d’un espace à dimension finie, s’il possède suffisamment de neurones formels cachés (voir [12]). Nous présentons ici les résultats de la prédiction en régression non linéaire. Dans un premier temps, nous déterminons le nombre de neurones de la couche cachée offrant les meilleurs résultats. Dans un second temps, nous effectuons une prédiction par MLP utilisant le nombre de neurones de la couche cachée optimal. Les résultats obtenus sont analysés via la répartition de l’erreur de reconstruction.

²Ce phénomène peut venir du fait que les informations en entrée se trouvent compressées car nous utilisons un seul neurone dans la couche cachée. Pour vérifier cette hypothèse, une étude utilisant une régression non linéaire pour chaque fréquence peut être menée.

4.6.1 Choix du nombre de neurones de la couche cachée

Le nombre de neurones cachés est déterminé expérimentalement en dressant l'évolution du coût³ sur l'ensemble de test en fonction du nombre de neurones de la couche cachée. Le vecteur $Vect_2$ est utilisée en entrée du réseau de neurones

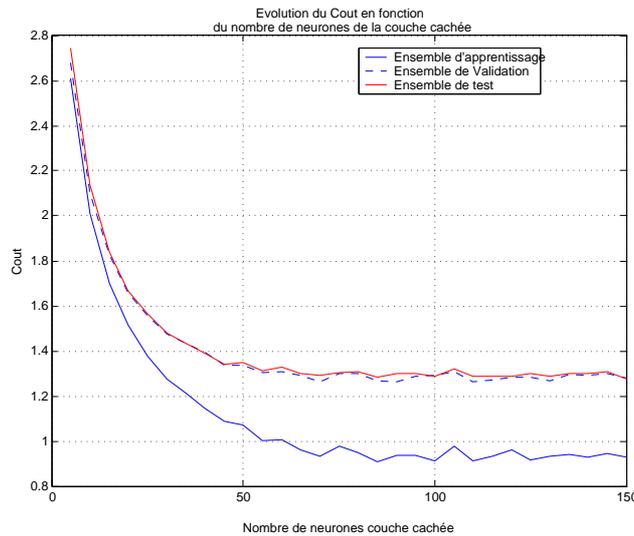


FIG. 4.2 – Evolution du coût en fonction du nombre de neurones de la couche cachée. Le coût est déterminé sur la base d'apprentissage, de validation et de test (ensemble défini dans la section 4.2)

Plus le nombre de neurones cachés augmente, plus l'erreur sur l'ensemble d'apprentissage, de validation et de test semble diminuer. Cela est vrai jusqu'à un certain niveau. Lorsque le nombre de neurones de la couche cachée est trop important, les performances sur la base de test se dégradent tandis que l'erreur continue à décroître sur la base d'apprentissage. Le réseau apprend "par coeur" les données, c'est ce que l'on appelle le sur-apprentissage.

Nous avons décidé d'utiliser par la suite 100 neurones cachés. A partir de cette valeur on peut considérer que l'intérêt d'augmenter le nombre de neurones de la couche cachée est faible. Nous pouvons faire un rapprochement entre les 100 neurones dans la couche cachée et la dimension des vecteurs de sortie (une HRTF est définie sur 100 échantillons fréquentiels). Nous allons faire l'hypothèse par la suite que cette valeur convient indépendamment du vecteur d'entrée $Vect_1$ ou $Vect_2$ choisi.

Dans les sous-sections suivantes, nous allons évaluer les performances de reconstruction avec les entrées $Vect_1$ et $Vect_2$

4.6.2 Utilisation d'un seul représentant

Les erreurs de reconstruction par bande sont présentées dans le tableau 4.3.

³Pour une entrée donnée, le coût est la différence entre la sortie désirée et la sortie réelle obtenue par le réseau de neurone

Numero Bande	Erreur Moyenne	Erreur Minimum (5%)	Erreur Médiane	Erreur Maximum (95%)
n° 1	0.79 dB	0.36 dB	0.64 dB	1.69 dB
n° 2	1.72 dB	0.72 dB	1.43 dB	3.81 dB
n° 3	1.93 dB	0.67 dB	1.58 dB	4.36 dB
n° 4	2.46 dB	0.71 dB	2.11 dB	5.86 dB
n° 5	3.66 dB	1.27 dB	3.46 dB	7.03 dB
Toutes	1.88 dB	0.92 dB	1.71 dB	3.24 dB

TAB. 4.3 – Erreur de reconstruction

Les erreurs obtenues sont relativement faibles⁴. En comparant avec le tableau 3.3, nous réduisons l'erreur entre une HRTF et son représentant de 3.40 dB à 1.88 dB en moyenne (diminution de 40% de l'erreur). L'annexe D présente de façon plus approfondie la répartition des erreurs au niveau spatial et fréquentiel. Au niveau fréquentiel, on remarque que, plus la fréquence est élevée, plus l'erreur de reconstruction augmente. Les erreurs sont élevées à partir de la bande fréquentielle n° 2. Au niveau spatial, les erreurs élevées sont présentes, principalement du côté contralatéral et dans l'hémisphère avant.

4.6.3 Utilisation de tous les représentants

Les erreurs de reconstruction par bande sont présentées dans le tableau 4.4.

Numero Bande	Erreur Moyenne	Erreur Minimum (5%)	Erreur Médiane	Erreur Maximum (95%)
n° 1	0.83 dB	0.41 dB	0.70 dB	1.77 dB
n° 2	1.72 dB	0.72 dB	1.51 dB	3.54 dB
n° 3	2.08 dB	0.87 dB	1.88 dB	4.10 dB
n° 4	2.68 dB	0.80 dB	2.31 dB	5.69 dB
n° 5	3.64 dB	1.37 dB	3.37 dB	6.74 dB
Toutes	1.95 dB	1.17 dB	1.82 dB	3.02 dB

TAB. 4.4 – Erreur de reconstruction

Les erreurs sont légèrement plus élevées que dans le test précédant (erreur de 1.95 dB contre 1.88 dB). L'annexe E présente de façon plus approfondie la répartition des erreurs au niveau spatial et fréquentiel.

4.6.4 Comparaison des reconstructions suivant les vecteurs utilisés en entrée

Lorsque l'on compare les performances de la régression linéaire réalisée avec les vecteurs $Vect_1$ et $Vect_2$, on remarque que l'erreur moyenne est assez proche (respectivement 1.88 et 1.95). Nous pouvons émettre l'hypothèse que l'erreur de modélisation n'est pas liée au nombre de représentants utilisés en entrée. Lorsque l'on mesure l'importance des

⁴On peut considérer qu'à 1 ou 2 dB, nous sommes au niveau du bruit de mesure

entrées via une technique exposée dans Lemaire [16], nous constatons que seules deux entrées sont utiles pour le réseau : l'entrée $n^{\circ} 1$ qui correspond à la valeur de θ et l'entrée $n^{\circ} 2$ qui correspond à la valeur de ϕ . Nous pouvons supposer que le réseau de neurones construit une fonction f telle que $f(\theta, \phi) = \widehat{HRTF}(\theta, \phi)$. Pour vérifier cette hypothèse, nous avons réalisé un apprentissage avec, comme vecteurs d'entrée, les couples (θ, ϕ) des HRTF à reconstruire. Nous obtenons une erreur moyenne de $1.81dB$. L'erreur étant légèrement plus faible que celles obtenues avec les vecteurs $Vect_1$ et $Vect_2$, notre hypothèse est vérifiée : le réseau modélise une HRTF à partir de θ et ϕ uniquement.

4.7 Discussion

Ce chapitre préliminaire sur l'estimation des HRTF pour un individu permet de constater que la régression linéaire offre de moins bonnes performances que la régression non linéaire. Nous utiliserons par la suite uniquement des régression non linéaires pour reconstruire les HRTF.

La régression non linéaire offre de bonnes performances de reconstruction avec des erreurs proches du bruit de mesure (écoutes formelles actuellement en cours). Nous avons montré que cet outil statistique modélise une HRTF à partir de θ et ϕ uniquement.

Comme ces modèles utilisent uniquement les entrées θ et ϕ , nous pouvons en extraire un nouveau mode d'individualisation des HRTF. Nous pouvons envisager de mesurer un petit nombre d'HRTF d'un individu λ et considérer que ces HRTF correspondent à l'ensemble d'apprentissage d'un modèle individuel. Le but de la modélisation serait de construire une fonction f_λ telle que $f_\lambda(\theta, \phi) = \widehat{HRTF}(\lambda, \theta, \phi)$. Il serait intéressant d'évaluer les performances de ces modèles individuels. Combien d'HRTF appartenant à l'ensemble d'apprentissage sont nécessaires pour obtenir une erreur de modélisation faible ? Ces modèles fonctionnent-ils pour tous les individus ?

Chapitre 5

Individualisation des HRTF par régression statistique

Dans ce chapitre applicatif, nous allons mettre en oeuvre des outils statistiques pour individualiser les HRTF. Le but est de constater qu'il est possible d'individualiser les HRTF d'un individu à partir de ses HRTF représentatives et d'un apprentissage réalisé sur d'autres individus.

5.1 Introduction

Nous allons évaluer l'intérêt d'utiliser des réseaux de neurones pour réduire l'erreur de quantification. Nous avons vu dans le chapitre précédant que ces outils permettent de modéliser les HRTF d'un individu λ à partir des coordonnées θ et ϕ de l'HRTF à reconstruire et d'une phase d'apprentissage réalisée sur plusieurs de ses HRTF. Nous avons également remarqué que la régression non linéaire offrait de meilleurs résultats que la régression linéaire. Dans le cas présent, nous n'allons pas utiliser des HRTF de l'individu λ à individualiser dans l'ensemble d'apprentissage car ces HRTF sont supposées inconnues. Pour permettre aux réseaux de neurones d'individualiser les HRTF de l'individu λ , nous allons lui transmettre un petit nombre d'HRTF (les HRTF "représentatives") mesurées sur cet individu en entrée.

Pour effectuer ce travail, nous décomposons ce chapitre comme suit :

Tout d'abord, nous allons sectionner notre ensemble de données en plusieurs ensembles : l'ensemble d'apprentissage de validation et de test. Ensuite nous définissons le type de normalisation utilisée au niveau du réseau de neurones puis les vecteurs d'entrée utilisés pour la modélisation. Enfin, nous présentons les résultats de la modélisation des HRTF obtenue par régression non linéaire.

5.2 Sélection des ensembles d'apprentissage, de validation et de test

Nous avons divisé l'ensemble des HRTF, noté $\mathcal{D}_{\mathcal{HRTF}}$ en 4 ensembles disjoints :

- Les HRTF représentatives \mathcal{D}_{Rep} correspondant aux HRTF de tous les individus mesurées aux positions spatiales définies dans la sous section 3.3.4.
- L'ensemble d'apprentissage \mathcal{D}_{Appr} correspondant aux HRTF non représentatives de l'individu 1 à l'individu 22 de la base.
- L'ensemble de validation \mathcal{D}_{Vali} correspondant aux HRTF non représentatives de l'individu 23 à l'individu 22 de la base.
- L'ensemble de test \mathcal{D}_{Test} correspondant aux HRTF non représentatives de l'individu 34 à l'individu 44 de la base.

Finalement 50% des HRTF appartiennent à l'ensemble \mathcal{D}_{Appr} , 25% à l'ensemble \mathcal{D}_{Vali} et 25% à l'ensemble \mathcal{D}_{Test} . Nous allons définir dans la section suivante les vecteurs qui seront utilisés en entrée de la régression.

5.3 Choix du vecteur d'entrée

Nous allons pondérer l'importance des HRTF “représentatives” en fonction de la distance spatiale qui la sépare de l'HRTF à estimer. Cette pondération est effectuée en ajoutant comme information les distances angulaires $(\theta' - \theta)$ et $(\phi' - \phi)$

Nous proposons 2 jeux de vecteurs d'entrée qui diffèrent par l'origine des représentants : orientés “Signal-Champ Libre” et “Uniformément répartis à la surface de la sphère”. Dans les deux cas, nous allons utiliser un *unique représentant* pour modéliser une HRTF¹. Pour une orientation donnée, les HRTF représentatives sont mesurées aux mêmes positions spatiales quel que soit l'individu λ à individualiser.

5.3.1 Utilisation d'un représentant orienté “Signal-Champ libre”

Chaque HRTF est située dans un cluster qui est représenté par une HRTF “représentative” (voir sous-section 4.3.2). Le vecteur d'entrée $Vect_1$ utilise uniquement ce représentant pour estimer une HRTF

Soit (θ', ϕ') la position spatiale du représentant “Signal-Champ libre” de l' $HRTF(\theta, \phi)$. le vecteur $Vect_1$ est défini par :

$$\mathbf{Vect}_1 = [\theta, \phi, V_{log}(\theta', \phi'), (\theta' - \theta), (\phi' - \phi)] \quad (5.1)$$

Pour simplifier le travail du réseau de neurones, nous allons modéliser la différence entre $V(\theta, \phi)$ et $V(\theta', \phi')$. Nous modélisons les modifications à apporter à la HRTF représentative pour obtenir l'HRTF en (θ, ϕ) .

¹Nous utilisons uniquement une HRTF représentative pour modéliser une HRTF. L'utilisation de plus d'une HRTF est possible mais est très coûteuse en mémoire RAM

5.3.2 Utilisation d'un représentant orienté "Uniformement réparti à la surface de la sphère"

Nous avons montré dans le chapitre précédant que l'erreur de quantification était plus faible avec les représentants uniformement répartis à la surface de la sphère. Nous pouvons penser qu'ils offriront également une meilleure erreur de modélisation.

Soit (θ'', ϕ'') la position spatiale du représentant "Uniformement réparti à la surface de la sphère" de l' $HRTF(\theta, \phi)$. le vecteur $Vect_2$ est défini par :

$$\mathbf{Vect}_2 = [\theta, \phi, V_{log}(\theta'', \phi''), (\theta'' - \theta), (\phi'' - \phi)] \quad (5.2)$$

Nous modélisons en sortie la différence entre $V(\theta, \phi)$ et $V(\theta'', \phi'')$. Les normalisations appliquées à ces vecteurs sont présentées dans la section suivante.

5.4 Type de normalisation utilisée

Avant d'effectuer une régression nous allons normaliser nos données. Pour toutes les expériences réalisées dans cette partie, les données d'apprentissage et de test sont centrées par rapport à la moyenne de toute la série, puis normalisées entre [-1,1]. Nous utilisons le même type de normalisation que dans le chapitre précédant rappelé ci dessous :

- Soit X une variable
- On recherche z_1 telle que $P(X < z_1) = 0.01$.
- On recherche z_2 telle que $P(X > z_2) = 0.99$.
- On normalise X par $X_{norm} = -1 + \frac{2 \cdot (X - z_1)}{z_2 - z_1}$

5.5 Modélisation en régression non linéaire : Représentants orientés "Signal-Champ Libre"

Nous allons utiliser les représentants de l'orientation signal champ libre pour individualiser les HRTF d'un individu. L'élection de ces représentants est détaillée dans le chapitre 3.3. Nous utiliserons, pour la régression non linéaire, 100 neurones dans la couche cachée.

5.5.1 Modélisation utilisant 1 représentant parmi les 26 représentants

Nous allons évaluer l'erreur de modélisation lorsque nous utilisons le vecteur d'entrée $Vect_1$. L'ensemble des représentants \mathcal{D}_{Rep} correspond aux HRTF des individus 1 à 22 mesurées sur les positions spatiales obtenues avec l'orientation signal champ libre. Les erreurs de modélisation par bande sont présentées dans le tableau 5.1

La répartition fréquentielle et spatiale des erreurs de modélisation est présentée en annexe F.

Numero Bande	Erreur Moyenne	Erreur Minimum (5%)	Erreur Médiane	Erreur Maximum (95%)
n° 1	1.63 dB	0.79 dB	1.42 dB	3.12 dB
n° 2	3.36 dB	1.29 dB	3.09 dB	6.35 dB
n° 3	4.41 dB	2.04 dB	4.15 dB	7.76 dB
n° 4	4.50 dB	1.30 dB	3.95 dB	9.75 dB
n° 5	5.27 dB	1.91 dB	4.89 dB	9.95 dB
Toutes	3.59 dB	2.09 dB	5.48 dB	3.47 dB

TAB. 5.1 – Erreur de reconstruction

5.5.2 Influence du nombre de représentants

Le nombre de représentants appartenant à l'ensemble \mathcal{D}_{Rep} a un impact important sur l'erreur de modélisation. Plus le nombre de représentants est élevé, plus le réseau de neurones possède d'information pour individualiser les HRTF. Le tableau 5.2 dresse l'évolution de l'erreur de modélisation et de l'erreur de quantification en fonction du nombre de représentants.

Nombre de Représentants	Erreur de Quantification EQ	Erreur de Modelisation EM	Amélioration EM-EQ /EQ
4	6.28 dB	4.41 dB	42%
15	4.81 dB	4.04 dB	16%
26	4.31 dB	3.59 dB	16%
50	3.77 dB	3.36 dB	10%
100	3.46 dB	3.06 dB	11%

TAB. 5.2 – Erreur de modélisation

5.6 Modélisation en régression non linéaire : Représentants orientés “Uniformement répartis à la surface de la sphère”

Les représentants uniformement répartis à la surface de la sphère offriraient une erreur de quantification plus faible que les représentants orientés “Signal-Champ libre”. Nous pouvons nous attendre à ce que ces représentants offrent également une erreur de modélisation plus faible.

5.6.1 Modélisation utilisant 1 représentant parmi les 26 représentants

Le tableau 5.3 présente la répartition de l'erreur de modélisation dans chaque bande. La répartition spatiale et fréquentielle de l'erreur est présentée en annexe G.

Numero Bande	Erreur Moyenne	Erreur Minimum (5%)	Erreur Médiane	Erreur Maximum (95%)
n° 1	1.52 dB	0.71 dB	1.33 dB	3.00 dB
n° 2	3.02 dB	1.03 dB	2.75 dB	5.96 dB
n° 3	4.02 dB	1.79 dB	3.70 dB	7.44 dB
n° 4	4.19 dB	1.18 dB	3.57 dB	9.22 dB
n° 5	5.08 dB	1.83 dB	4.68 dB	9.78 dB
Toutes	3.32 dB	1.82 dB	3.16 dB	5.34 dB

TAB. 5.3 – Erreur de reconstruction

En comparant l'erreur de modélisation obtenue avec les deux types de représentants, nous constatons que les représentants uniformément répartis à la surface de la sphère offrent de meilleurs résultats (3.32dB contre 3.59dB).

5.6.2 Influence du nombre de représentants

Le tableau 5.4 dresse l'évolution de l'erreur de quantification et de l'erreur de modélisation en fonction du nombre de représentants appartenant à l'ensemble \mathcal{D}_{Rep} .

Nombre de Représentants	Erreur de Quantification EQ	Erreur de Modelisation EM	Amélioration EM-EQ /EQ
4	6.13 dB	4.34 dB	29%
15	4.50 dB	3.66 dB	18%
26	3.93 dB	3.32 dB	15%
50	3.47 dB	2.96 dB	14%
100	2.99 dB	2.66 dB	11%

TAB. 5.4 – Erreur de modélisation

5.7 Discussion

Ce chapitre sur l'utilisation des réseaux de neurones MLP pour l'individualisation des HRTF montre qu'il est possible de modéliser une HRTF correctement à partir des HRTF représentatives. La modélisation est effectuée à partir de réseaux de neurones MLP. L'intérêt de la modélisation par réseau de neurones est exposé dans la figure 5.1. Les erreurs de modélisation sont plus faibles que les erreurs de quantification.

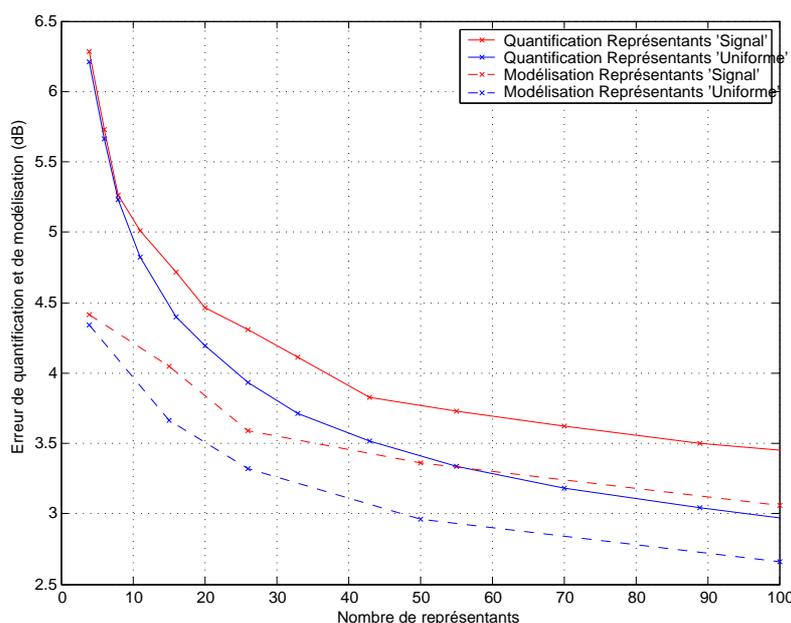


FIG. 5.1 – Les erreurs sont décroissantes en fonction du nombre de représentants. Les erreurs de modélisation sont plus faibles que les erreurs de quantification. Les représentants uniformément répartis à la surface de la sphère offrent de meilleures erreurs que les représentants “Signal-Champ Libre”.

En ce qui concerne l'origine des HRTF représentatives, les meilleurs résultats semblent être obtenus avec les représentants uniformément répartis à la surface de la sphère. Cette supposition dépend de l'individu utilisé pour l'élection des HRTF représentatives “Signal-Champ Libre”. Comme nous l'avons exposé dans la discussion du chapitre précédent, il existe peut être un individu λ pour qui la position des HRTF représentatives offrirait une erreur de quantification plus faible sur l'ensemble des individus que les HRTF uniformément réparties à la surface de la sphère. Dans ce cas, il est probable que les représentants de l'individu λ offrirait également une meilleure erreur de modélisation sur l'ensemble des individus.

Au niveau du nombre de données utilisées en entrée des réseaux de neurones, nous utilisons uniquement une HRTF représentative par HRTF à modéliser. L'utilisation de plusieurs représentants peut aider le réseau de neurones à modéliser de façon plus fine l'HRTF désirée. Une étude supplémentaire sur l'influence du nombre de représentants utilisés pour modéliser une HRTF devrait être menée.

Enfin, l'architecture du réseau fonctionne mais n'est pas optimale. Nous pouvons faire évoluer les différents paramètres constitutifs du réseau et voir leur impact sur l'erreur de modélisation. Quel est l'impact du nombre de neurones dans la couche cachée ? Est ce que d'autres types de normalisation offrirait de meilleurs résultats ? Au vu des propriétés de l'appareil auditif humain, il serait également approprié d'utiliser une fonction de coût plus proche des propriétés psychoacoustiques. Les erreurs commises à des amplitudes faibles ne sont pas aussi importantes que les erreurs commises à des amplitudes élevées au niveau perceptif.

Chapitre 6

Conclusion

Ce projet de fin d'étude s'est orienté sur deux disciplines : le traitement statistique de l'information et l'audio numérique. Nous avons pu évaluer l'intérêt d'utiliser des outils statistiques pour résoudre des problèmes liés à la synthèse binaurale. Nous avons montré qu'il était possible de modéliser des HRTF par réseau de neurones et que les réseaux étaient également capables d'individualiser les HRTF. Notre méthodologie suivie pour individualiser les HRTF est composée des étapes suivantes :

- Mesure de quelques HRTF d'un individu λ . Ces HRTF peuvent être mesurées à des positions spatiales uniformément réparties sur une sphère ou bien sur des positions déterminées par outils statistiques (Carte de Kohonen et CHA). Les positions uniformément réparties offrent de meilleurs résultats.
- Création d'un modèle permettant de reconstruire l'intégralité des HRTF de l'individu λ à partir de ses HRTF mesurées. Le modèle est créé à partir d'un réseau de neurones MLP.

Cette méthodologie a donné lieu à un brevet et nous évaluons en ce moment les performances des modèles au niveau perceptif via des écoutes formelles.

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

- *lissage perceptif des HRTF* : Les HRTF lissées sont utiles pour la modélisation. Il semble important de les lisser avant la phase de sélection des représentants pour "gommer" les variations du signal en haute fréquence. Ces variations en hautes fréquences sont peu perceptibles par l'oreille qui possède une résolution faible sur la partie haute du spectre. En effectuant des essais de lissage, nous avons remarqué une amélioration de l'erreur de modélisation de 25%. Ce pré-traitement est donc nécessaire.
- *Election de représentants plus performants que ceux uniformément répartis à la surface de la sphère* : Les représentants uniformément répartis à la surface de la sphère offrent de meilleurs résultats que les représentants déterminés par carte de Kohonen et CHA. Cependant ces derniers représentants ont été sélectionnés à partir d'une étude statistique réalisée sur les HRTF d'un seul individu. Il existe peut être

un individu “moyen” pour qui la démarche d’élection offrirait de meilleurs résultats que ceux obtenus avec les représentants uniformément répartis. Pour élire cet individu, nous pouvons nous aider des facteurs anthropomorphiques.

- *Optimisation de la structure du réseau de neurones* : Les phases de modélisation d’HRTF par réseau de neurones utilisent une HRTF représentative en entrée pour modéliser une HRTF. En utilisant un plus grand nombre de représentants lors de l’estimation d’une HRTF, nous pouvons penser que l’erreur de modélisation va diminuer. En ce qui concerne la fonction de coût, elle peut être améliorée pour tenir compte des propriétés psychoacoustiques de l’appareil auditif humain. Cette erreur est importante au sens psychoacoustique sur les amplitudes élevées mais moins importante sur les amplitudes faibles.
- *Utilisation de 2 modélisations, l’une pour l’hémisphère avant et l’autre pour l’hémisphère arrière* : Les HRTF situées à l’avant et l’arrière sont proches au sens spectral mais éloignées au sens géographique. Pour aider le réseau à distinguer ces deux hémisphères, nous pensons qu’il serait judicieux de construire deux modèles un pour l’avant et l’autre pour l’arrière.

D’autres axes sont également possibles mais s’éloignent de la méthodologie de départ. Nous les mentionnons car ils peuvent offrir de bonnes performances au vu de certains résultats présentés dans ce rapport.

- *Construction de modèles individuels* : Nous avons construit dans le chapitre 5 un modèle générique pour tous les individus utilisant les HRTF représentatives comme information individuelle. Au vu de certains résultats exposés dans le chapitre 4, il est possible de construire des modèles individuels où les HRTF mesurées seraient des HRTF appartenant à l’ensemble d’apprentissage. Le but de ces modèles serait de construire des fonctions individuelles f_λ telles que $f_\lambda(\theta, \phi) = \widehat{HRTF}(\lambda, \theta, \phi)$.
- *Modélisation dans l’espace pôles et zéros* : En utilisant une transformation pour modifier les HRTF (filtre FIR) en IIR, nous obtenons une série de coefficients décrivant l’allure d’une HRTF (localisation des pics et des bosses de la fonction de transfert). Ces informations apportent une information importante au cerveau pour localiser l’origine spatiale de la source sonore. Nous pensons que la conversion des HRTF en filtre IIR permettrait d’augmenter les performances de modélisation du réseau de neurones.

Le problème de l’individualisation des HRTF pour la synthèse binaurale est un problème complexe. Nous pensons que l’approche statistique est une solution intéressante. Elle doit être améliorée puis comparée aux autres techniques d’individualisation exposées dans l’état de l’art.

Annexe A

Sélection des bandes fréquentielles

L'objectif est de regrouper des intervalles fréquentiels aux comportements semblables sur l'ensemble des HRTF de la base. Nous ferons l'hypothèse que ces intervalles traduisent des différences inter-individuelles a priori indépendantes.

Le clustering des variables est réalisé dans l'espace signal et est construit à partir des HRTF de tous les individus de la base du CIPIC (soit 144×1250 HRTF).

L'opération de regroupement des variables est très proche de l'opération du regroupement des HRTF présentée dans le chapitre 3. Le lecteur désirant en connaître plus sur la méthodologie pour regrouper les variables peut se référer à des articles comme [17].

Nous avons utilisé une carte de Kohonen à voisinage hexagonal de taille 5×5 . Cette taille est déterminée expérimentalement et semble être adaptée à nos données. La carte de Kohonen des variables est présentée sur la figure A.1. La classification hiérarchique ascendante réalisée sur les neurones de la carte utilise une agrégation par critère de Ward. L'évolution du critère de Ward est présentée sur la figure A.2.

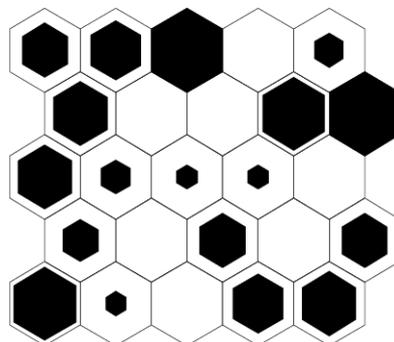


FIG. A.1 – carte de Kohonen des variables

Nous fixons la troncature du dendrogramme à 0.65. Cette troncature offre un bon compromis entre le nombre de classes (5 classes) et la variance intra-classes. La figure A.3 présente la localisation des classes sur les neurones de la carte de Kohonen des variables.

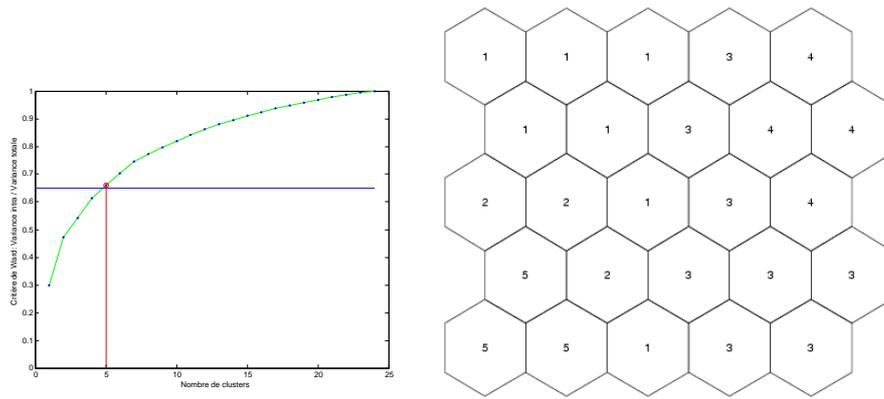


FIG. A.2 – Evolution du critère de Ward
 FIG. A.3 – Clusters obtenus par CHA (Ward=0.65)

Le tableau A.1 présente le regroupement des 100 fréquences.

n° Cluster	Fréquences incluses
Cluster 1	$0Hz \cup [10.691kHz - 17.150kHz]$
Cluster 2	$[17.373kHz - 19.155kHz]$
Cluster 3	$223Hz \cup [5.568kHz - 10.468kHz]$
Cluster 4	$[445Hz - 5.345kHz]$
Cluster 5	$[19.377kHz - 22.050kHz]$

TAB. A.1 – Clustering des Variables

Nous allons extraire à partir de ces regroupements de variables des bandes fréquentielles

Sur le tableau A.1, on remarque que certains clusters contiennent des fréquences qui ne se suivent pas (Cluster1 et Cluster3). L'article du CIPIC qui détaille le protocole expérimental de mesures des HRTF [6] spécifie que les fréquences entre [0Hz-400Hz] sont bruitées et qu'elles n'ont pas de significations physiques. Nous pouvons donc inclure les fréquences 0 Hz et 445 Hz dans le Cluster n° 4 pour faciliter la sélection de bandes fréquentielles. Finalement les bandes obtenues sont présentées dans le tableau A.2.

n° Bande	Largeur Bande	Nombre Fréquence
Bande 1	[0Hz-5.345kHz]	25
Bande 2	[5.568 kHz-10.468kHz]	23
Bande 3	[10.691kHz-17.150kHz]	30
Bande 4	[17.373kHz-19.155kHz]	9
Bande 5	[19.377 kHz-22.050kHz]	13

TAB. A.2 – Bandes Spectrales

Annexe B

Sélection des représentants uniformément répartis à la surface de la sphère

Lors de notre étude, nous utilisons des représentants dont les positions spatiales sont uniformément réparties à la surface de la sphère. Ces représentants uniformément répartis sont choisis à partir de l'ensemble des points de mesures du CIPIC (1250 points). Nous avons décidé de les sélectionner en minimisant la distance géodésique (distance réelle à la surface de la sphère) entre les représentants.

La méthodologie pour les élire comporte plusieurs phases :

- Conversion des coordonnées polaires interaurales (θ, ϕ) des points en coordonnées cartésiennes (x, y, z) .
- Regroupement des points en K-clusters par l'algorithme des K-moyennes.
- Election d'un représentant par cluster.

Nous présentons, dans les sections suivantes, comment ses 3 phases permettent d'obtenir des points uniformément répartis à la surface de la sphère.

B.1 Transformation des coordonnées polaires interaurales des 1250 Points en coordonnées cartésiennes

La position spatiale d'un point est définie par ses coordonnées polaire interaurale θ et ϕ . Il est possible d'obtenir les coordonnées cartésiennes x, y, z du point en effectuant les transformations suivantes (le rayon de la sphère est 1) :

$$x = \sin(\theta), y = \cos(\theta) \cdot \cos(\phi), z = \cos(\theta) \cdot \sin(\phi) \quad (\text{B.1})$$

Nous obtenons donc une liste de 1250 points chacun décrit par leur coordonnée x, y, z . Nous allons regrouper ces points en cluster de façon à ce que la distance géodésique entre

points de la même classe soit minimale. Nous savons que dans le cas particulier d'une sphère, minimiser la distance géodésique entre deux points revient à minimiser la distance euclidienne.

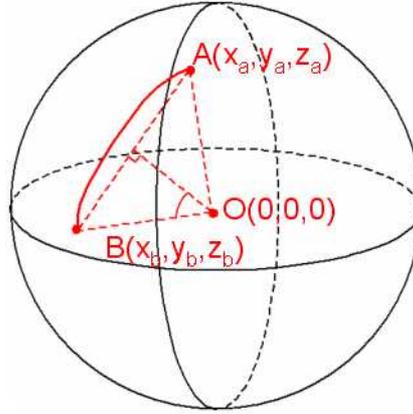


FIG. B.1 – Distance géodésique sur une sphère

Preuve : La distance euclidienne d^{eucl} entre deux points A et B (voir figure B.1) de coordonnées respectives (x_A, y_A, z_A) et (x_B, y_B, z_B) est défini par :

$$d_{AB}^{eucl} = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2}$$

La distance géodésique est obtenue à partir de l'équation (r étant le rayon de la sphère) :

$$d_{AB}^{go} = \widehat{AOB} \cdot r \quad (B.2)$$

L'angle \widehat{AOB} est calculé à partir de la distance euclidienne entre A et B.

$$\sin\left(\frac{\widehat{AOB}}{2}\right) = \frac{d_{AB}^{eucl}}{2 \cdot r} \Rightarrow \widehat{AOB} = 2 \cdot \arcsin\left(\frac{d_{AB}^{eucl}}{2 \cdot r}\right)$$

Finalement, on obtient dans B.2 $d_{AB}^{go} = 2 \cdot r \cdot \arcsin\left(\frac{d_{AB}^{eucl}}{2 \cdot r}\right)$.

la fonction \arcsin étant croissante, nous obtenons bien l'implication : Si $d_{AB}^{eucl} < d_{AC}^{eucl}$ alors $d_{AB}^{go} < d_{AC}^{go}$.

Au vu de cette implication, nous allons regrouper les points de façon à minimiser la distance euclidienne entre les points d'un même cluster. Il existe plusieurs algorithmes statistiques permettant de réaliser ce regroupement comme notamment les cartes de kohonen présentées dans ce rapport. Nous avons décidé d'utiliser l'algorithme des k-moyennes qui est un outil simple et rapide.

B.2 Regroupement des points en K-cluster par l’algorithme des K-moyennes

L’algorithme des K-moyenne est détaillé ci dessous :

- 1 : Définir K centroids (vecteurs (x, y, z)) initialisés au hasard.
- 2 : “Attribuer” chaque point au centroïde le plus proche au sens de la distance euclidienne.
- 3 : Calculer les coordonnées de chaque centroïde comme la moyenne des points qui lui sont attribués
- 4 : Tant que les centroïdes évoluent, aller en 2

Nous obtenons après application de cet algorithme, un ensemble de k cluster regroupant des points proches au sens euclidien et donc au sens géodésique. La dernière phase consiste à élire un représentant par cluster.

B.3 Election d’un représentant par cluster

Pour élire le point représentant du cluster C , nous utilisons le critère mathématique B. Ce critère permet d’élire la position spatiale située au centre du cluster.

$$\mathbf{r}_n = \underset{C}{\operatorname{argmin}} \sum_{i=1}^{\operatorname{length}(C)} (x_i - x_n)^2 + (y_i - y_n)^2 + (z_i - z_n)^2 \quad (\text{B.3})$$

Finalement nous obtenons un ensemble de K cluster représenté chacun par un représentant. Ces représentants ne sont pas uniques. En effectuant une rotation de centre O des représentants, nous obtenons également des représentants uniformément répartis à la surface de la sphère.

Annexe C

Introduction aux réseaux de neurones

Cette annexe est un extrait de la thèse [15] de Vincent Lemaire. Elle est reproduite avec son aimable autorisation.

Les développements récents de la théorie des réseaux neuronaux et de leurs applications pratiques en font une approche naturelle pour résoudre de nombreux problèmes non linéaires dans les télécommunications, tel par exemple le problème du contrôle d'admission des connexions (CAC) évoqué au chapitre précédent. Nous présentons en détails dans la première section ce qu'est un réseau de neurones artificiels.

Pour l'apprentissage supervisé d'un perceptron multicouche par correction d'erreur, l'algorithme le plus utilisé est l'algorithme de descente de gradient. Le calcul du gradient se fait en utilisant l'algorithme de la rétro-propagation de l'erreur. L'algorithme d'apprentissage utilisant ce procédé reste encore aujourd'hui la méthode d'apprentissage la plus largement utilisée et nous la détaillons au cours de la deuxième section.

C.1 Présentation

C.1.1 Qu'est-ce qu'un réseau de neurones ?

En tout premier lieu, lorsque nous parlons de réseaux de neurones, nous devrions plutôt dire “réseaux de neurones artificiels”. En effet, les réseaux de neurones biologiques sont de loin plus complexes que les modèles mathématiques utilisés pour les réseaux de neurones artificiels que nous utilisons. Cependant, il est usuel d'oublier le mot *artificiels* et de dire “réseaux de neurones”. Il n'y a pas de définition universelle pour caractériser un réseau de neurones. Mais la plupart de leurs utilisateurs ou concepteurs s'accordent sur le fait qu'un réseau de neurones est un réseau d'unités élémentaires interconnectées à fonctions d'activation linéaires ou non-linéaires. Ces unités se décomposent (pour les réseaux multicouches) en au moins deux sous-ensembles de neurones : un sous-ensemble de neurones d'entrée, un autre de neurones de sortie et éventuellement un ensemble de neurones cachés.

Il existe de nombreux modèles de réseaux de neurones : les réseaux de Hopfield, les réseaux de Kohonen, les réseaux à fonctions de bases radiales, les perceptrons multicouches (MLP pour Multi Layer Perceptron) ayant des architectures allant de la plus simple à la plus complexe (les différentes unités sont interconnectées aux autres, soit complètement, soit partiellement). Dans cette section nous définirons brièvement le perceptron multicouche.

L'ensemble des liens, convergeant vers une unité, constitue les connexions entrantes de l'unité. Ceux qui divergent vers d'autres unités constituent les connexions sortantes. A chaque *connexion*, on associe un *poids* ou efficacité *synaptique*. Le poids d'une connexion représente la force de l'influence d'une unité sur celle à laquelle est reliée sa sortie. En fait, la "*connaissance*" incluse dans le réseau de neurones est "mémorisée" dans les poids et l'architecture du réseau.

Le perceptron multicouche est structuré en couches. Chaque couche est entièrement connectée à la suivante et son graphe de connectivité ne possède pas de cycle.

Nous parlerons toujours de réseau de neurones qui fonctionne en temps discret et où chaque pas de temps est appelé cycle. Un cycle comprend tout le processus qu'effectue le réseau pour traiter une donnée (un motif d'entrée). Le fonctionnement du réseau est donc cadencé par la présentation des motifs en entrée. Un vecteur de scalaires présenté au temps t à toutes les unités d'entrées sera appelé exemple. A cet exemple sont aussi associées les valeurs (le vecteur de sortie) que l'on désire apprendre. Les poids des connexions sont éventuellement modifiés d'un cycle à l'autre par un mécanisme d'apprentissage.

Fixons à présent les notations que nous utiliserons dans la suite de ce chapitre. On appellera, au cycle t , w_{ij}^t le poids de la connexion reliant la sortie s_j d'un neurone j à l'entrée du neurone i . Le calcul de la sortie des unités à partir de leurs entrées s'appelle la règle de transition. Les règles de transition sont fondées principalement sur le modèle de Mc Culloch et Pitts. Cette transition peut être décomposée en deux étapes : calcul de l'activation présente à l'entrée de l'unité puis calcul de la sortie de l'unité suivant la fonction d'activation qu'elle possède.

D'une façon plus générale on peut définir un neurone par les quatre éléments suivants :

1. La nature de ses entrées : elles peuvent être binaires (0,1), ou prendre une valeur réelle.
2. La fonction d'entrée totale $a(\cdot)$, qui définit le pré-traitement effectué sur les entrées. Elle peut être affine, polynomiale, booléenne ou linéaire. Dans ce dernier cas elle s'écrit généralement comme la somme pondérée des entrées :

$$a_i = \sum_{p=1}^n w_{ip}^t x_{ip} \quad (\text{C.1})$$

où a_i est l'activité du neurone i à l'instant t , x_{ip} l'entrée p du neurone i , w_{ip}^t le poids de la connexion de l'entrée x_{ip} , n le nombre d'entrées du neurone. L'entrée totale ne dépend donc que des entrées extérieures, des poids et des sorties des neurones.

3. La fonction d'activation $f(\cdot)$ du neurone ou fonction de transfert qui définit son état de sortie en fonction de son entrée totale a , est telle que $s_i = f(a_i)$. Cette fonction peut être une fonction binaire à seuil (fonction de Heaviside ou fonction signe), une fonction sigmoïde, une fonction stochastique, ou toute autre fonction

généralement choisie croissante et impaire. L'utilisation de telles fonctions permet d'introduire une non-linéarité dans le réseau de neurones. Toutes les fonctions non-linéaires, par définition, atteignent ce but et les plus utilisées sont les fonctions sigmoïdes et gaussiennes. Pour les neurones de sortie, il faut choisir une fonction d'activation adaptée à la distribution des valeurs cibles. Mais si les valeurs n'ont pas de bornes connues il est préférable d'utiliser une fonction non bornée, comme par exemple la fonction identité (si les valeurs peuvent être positives ou négatives) ou l'exponentielle (si elles sont exclusivement du même signe).

Par la suite la sortie s_i va être redirigée soit vers l'environnement extérieur, soit vers d'autres unités où elle contribuera au calcul de leurs états d'activation.

4. Enfin, la nature de ses sorties : comme les entrées, elles peuvent être binaires ou réelles.

Pour terminer la définition du perceptron multicouche, il nous reste à ajouter deux éléments qui lui permettent d'apprendre des données : une fonction de coût ou de perte C et un algorithme d'apprentissage.

Dans ce mémoire, nous utiliserons des entrées et des sorties à valeurs réelles, les activations seront calculées à l'aide d'une somme pondérée linéaire et les fonctions d'activation seront sigmoïdales.

C.1.2 Que peut-on faire à l'aide d'un réseau de neurones ?

En principe, les réseaux de neurones sont des approximateurs universels, c'est-à-dire qu'ils peuvent approximer n'importe quelle fonction. Cependant, bien qu'ils puissent en théorie approximer n'importe quelle transformation continue d'un espace à dimension finie vers un autre espace à dimension finie (s'ils possèdent suffisamment de neurones cachés) aussi finement qu'on le souhaite, si on ne possède pas assez d'exemples, voire de puissance de calcul, le problème peut ne pas avoir de solution acceptable (au regard d'autres méthodes).

Suivant les architectures et les fonctions d'activation utilisées, le perceptron multicouche peut instancier beaucoup de modèles statistiques bien connus, comme la régression linéaire, la régression linéaire multi-variée, l'analyse discriminante, la régression polynomiale, l'analyse en composante principale. De plus, le perceptron multicouche, utilisant des fonctions d'activation non-linéaires et au moins une couche cachée, permet d'instancier des fonctions non-linéaires. Ces deux propriétés en font une *machine à apprendre* performante.

C.2 Apprentissage

C.2.1 Différentes règles d'apprentissage

L'apprentissage est le processus d'adaptation des paramètres d'un système, ici un réseau de neurones, pour donner une réponse désirée à une entrée. Le choix du processus d'adaptation des paramètres, c'est-à-dire la procédure d'apprentissage, est conditionné très fortement par la tâche que le réseau de neurones doit réaliser. On peut citer :

1. **approximation** : supposons que nous ayons une application non-linéaire qui lie à une entrée une sortie de la forme :

$$d = g(x) \tag{C.2}$$

où x est le vecteur d'entrée et d le scalaire de sortie. La fonction $g(\cdot)$ est inconnue. On veut construire un réseau qui approxime la fonction $g(\cdot)$ non-linéaire grâce à un ensemble de couples d'exemples entrée-sortie $(x_1, d_1), (x_2, d_2) \dots (x_P, d_P)$.

2. **classification** : Le problème consiste à ranger des entrées données dans un certain nombre de classes. La discrimination est une méthode de classification pour laquelle le but est la construction d'un classifieur qui à chaque observation x associe une classe parmi un ensemble fixé de classes. Les réseaux de neurones sont capables de construire de façon non-paramétrique des frontières de décision séparant ces classes : ils offrent donc la possibilité de résoudre des problèmes de classification extrêmement complexes. Par exemple, on peut construire un classifieur en associant une classe à chaque sortie d'un perceptron multicouche à q sorties (q représentant le nombre de classes).

3. **prédiction**

On possède, au temps t , des échantillons passés d'un signal :

$$x(t-1), x(t-2), \dots, x(t-M) \tag{C.3}$$

et on veut prévoir $x(t)$. On peut le faire par exemple par correction d'erreur de manière supervisée en imposant $x(t)$ comme sortie désirée.

Toutes ces tâches consistent fondamentalement à construire une application à partir d'un certain nombre d'exemples de cette application. Il existe de nombreuses méthodes d'apprentissage qui peuvent être utilisées pour la réalisation de ces tâches, parmi lesquelles on citera notamment :

1. **L'apprentissage par correction d'erreur** a pour but final de minimiser une fonction de coût C construite à partir de la base d'exemples qu'on possède. Une fois que la fonction de coût a été choisie l'apprentissage devient un problème d'optimisation pour lequel il existe quantité de techniques. Une méthode connue de minimisation est la descente de gradient.
2. **L'apprentissage de Hebb** qui est en fait un postulat tiré de la biologie et est la plus connue et la plus vieille des règles d'apprentissage.
3. **L'apprentissage compétitif** ou les neurones de sortie du réseau "décident" lequel d'entre eux sera actif. Il n'existe donc qu'un seul neurone de sortie actif à chaque cycle.
4. **L'apprentissage de Boltzmann** qui repose sur un algorithme stochastique issu de la thermodynamique et de la théorie de l'information (car on va chercher un maximum de vraisemblance).

On peut aussi distinguer deux classes d'apprentissage :

1. **L'apprentissage non-supervisé** où il n'y a pas de maître pour contrôler de l'extérieur le déroulement de l'apprentissage ; c'est pourquoi on l'appelle quelquefois auto-apprentissage. Le réseau va essayer de s'adapter aux régularités statistiques des données d'entrée. Il va donc automatiquement coder des classes dans sa représentation interne. L'apprentissage non-supervisé est réalisé, par exemple, par la règle de l'apprentissage compétitif vue plus haut. Il s'agit alors d'une stratégie appelée **winner-take-all**.
2. **L'apprentissage supervisé** : La méthode classique pour l'apprentissage supervisé consiste à se procurer un ensemble d'exemples, c'est-à-dire un ensemble fini de couple de vecteurs (x_i, y_i) . Dans un tel couple, x_i désigne l'entrée du réseau et y_i la sortie désirée pour cette entrée. On écrit alors la fonction calculée par le réseau sous une forme paramétrique : $f(x, w)$ désigne la sortie du réseau quand on lui présente en entrée le vecteur x et qu'il utilise les poids synaptiques contenus dans la matrice w . Enfin on se donne une distance sur l'espace vectoriel de sortie, c'est-à-dire un moyen de mesurer l'erreur commise en un point par le réseau. Si cette distance est notée d , on cherche alors à trouver la valeur de w qui minimise l'erreur totale commise par le réseau, c'est-à-dire la somme des distances entre les sorties obtenues et les sorties désirées, c'est-à-dire la somme des $d(f(x_i, w), y_i)$. Cette erreur est une fonction de w et on peut utiliser les techniques classiques d'optimisation de fonctions pour trouver son minimum.

Annexe D

Non individualisé : Modélisation en Regression non linéaire avec 1 représentant “Signal Champ-Libre”

Les figures qui suivent présentent l'évaluation des performances pour une prédiction utilisant une régression non linéaire et 1 représentant.

D.1 Répartition fréquentielle des erreurs

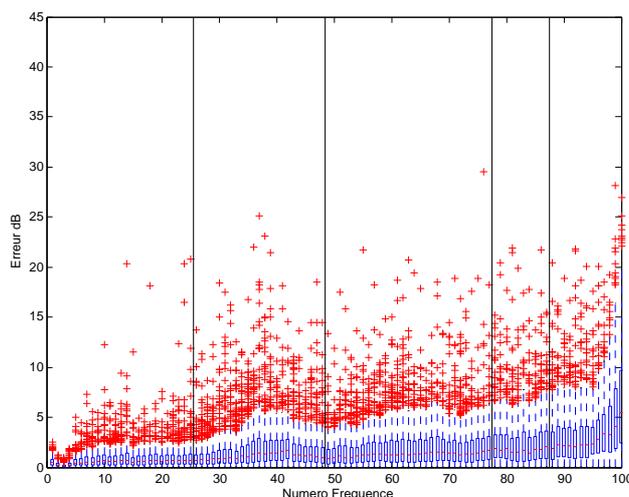


FIG. D.1 – Répartition fréquentielle des erreurs

D.2 Répartition spatiale des erreurs

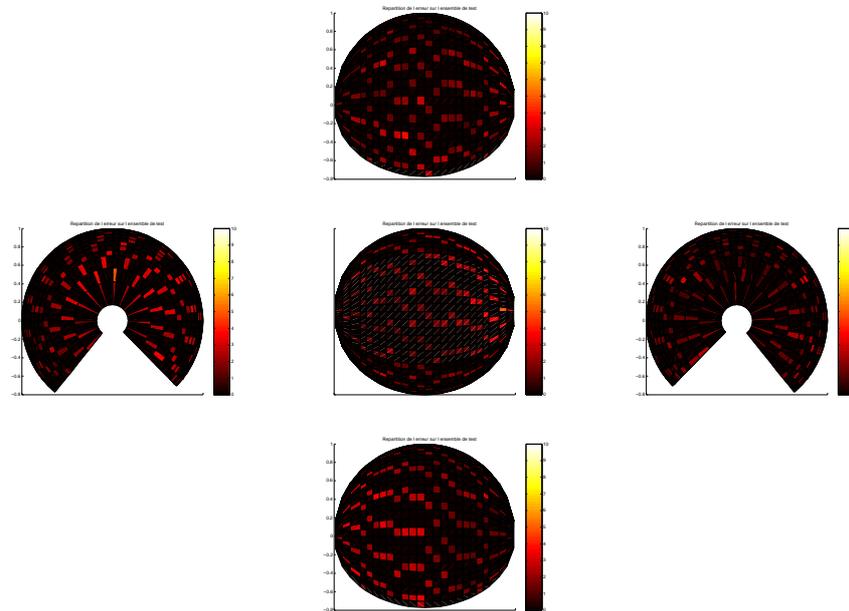


FIG. D.2 – Répartition des erreurs sur l'ensemble de test

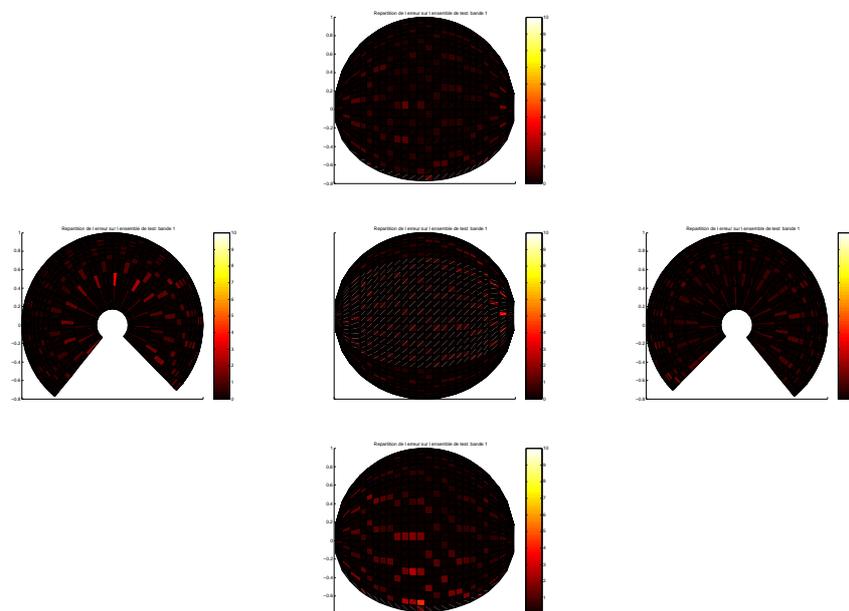


FIG. D.3 – Répartition des erreurs sur l'ensemble de test dans la bande n° 1

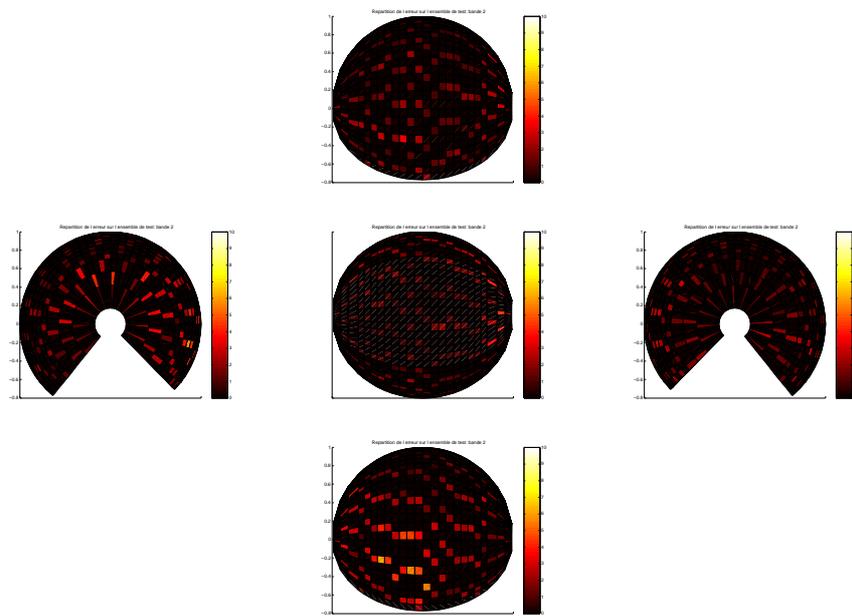


FIG. D.4 – Repartition des erreurs sur l'ensemble de test dans la bande n° 2

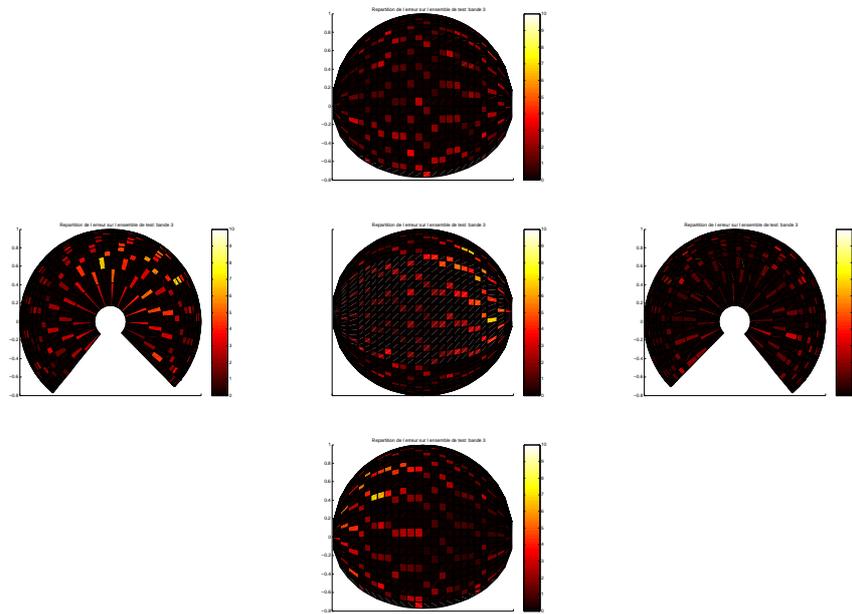


FIG. D.5 – Repartition des erreurs sur l'ensemble de test dans la bande n° 3

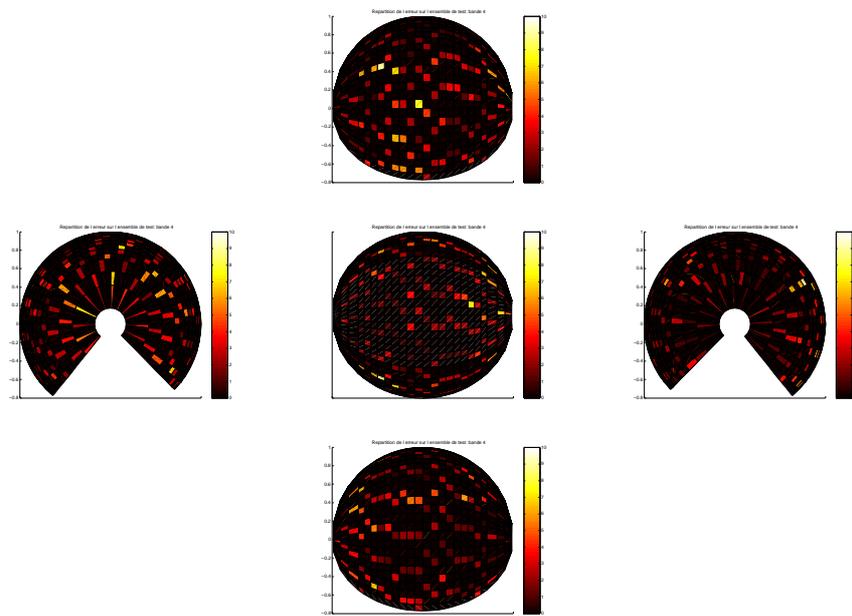


FIG. D.6 – Repartition des erreurs sur l'ensemble de test dans la bande n° 4

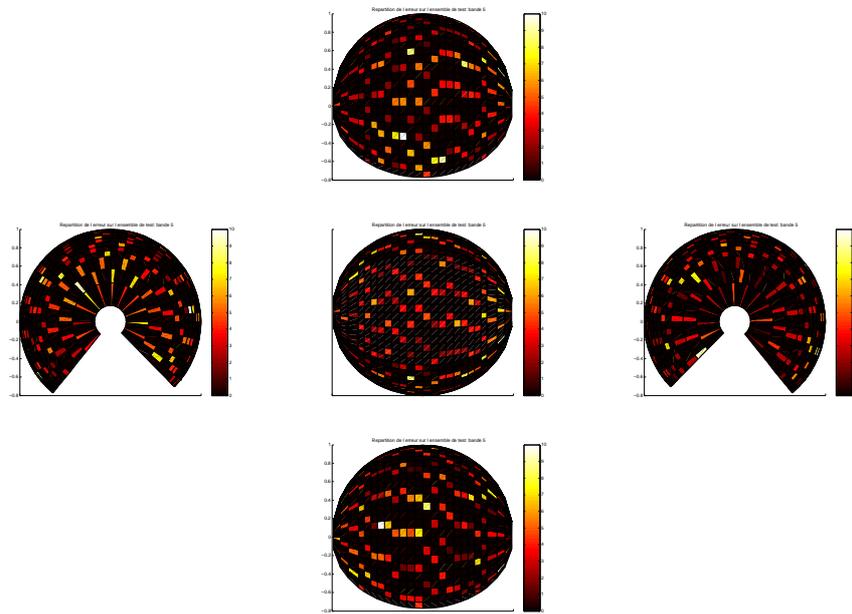


FIG. D.7 – Repartition des erreurs sur l'ensemble de test dans la bande n° 5

Annexe E

Non individualisé : Modélisation en Regression non linéaire avec 26 représentants “Signal Champ-Libre”

Les figures qui suivent présentent l'évaluation des performances pour une prédiction utilisant une régression non linéaire et 26 représentants.

E.1 Répartition fréquentielle des erreurs

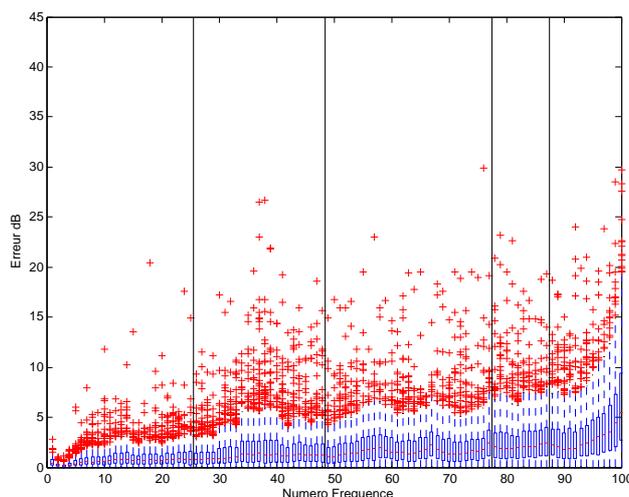


FIG. E.1 – Répartition fréquentielle des erreurs

E.2 Repartition spatiale des erreurs

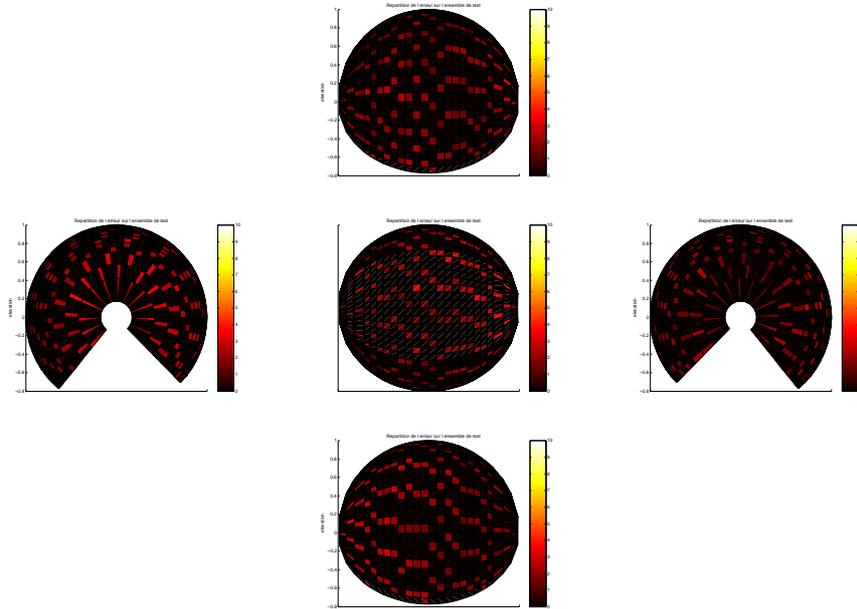


FIG. E.2 – Repartition des erreurs sur l'ensemble de test

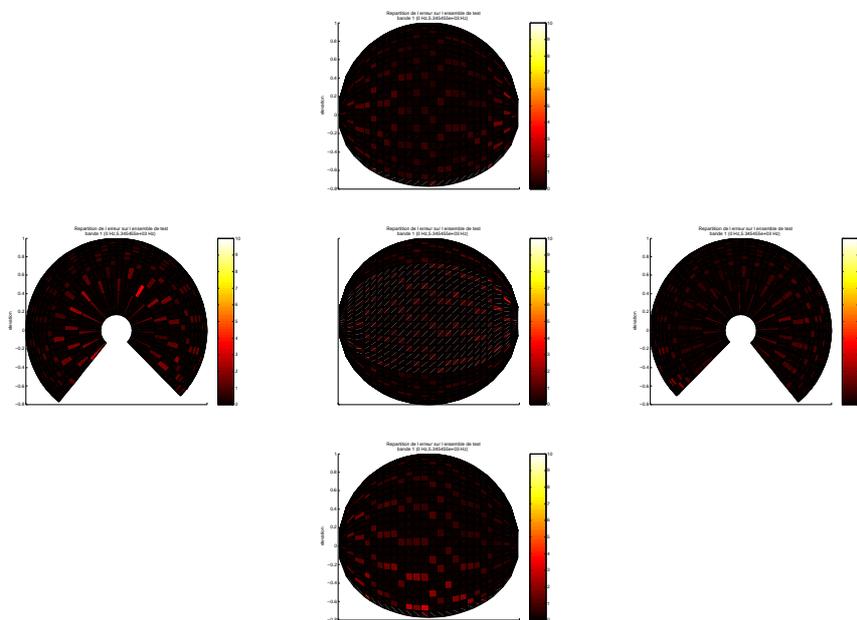


FIG. E.3 – Repartition des erreurs sur l'ensemble de test dans la bande n° 1

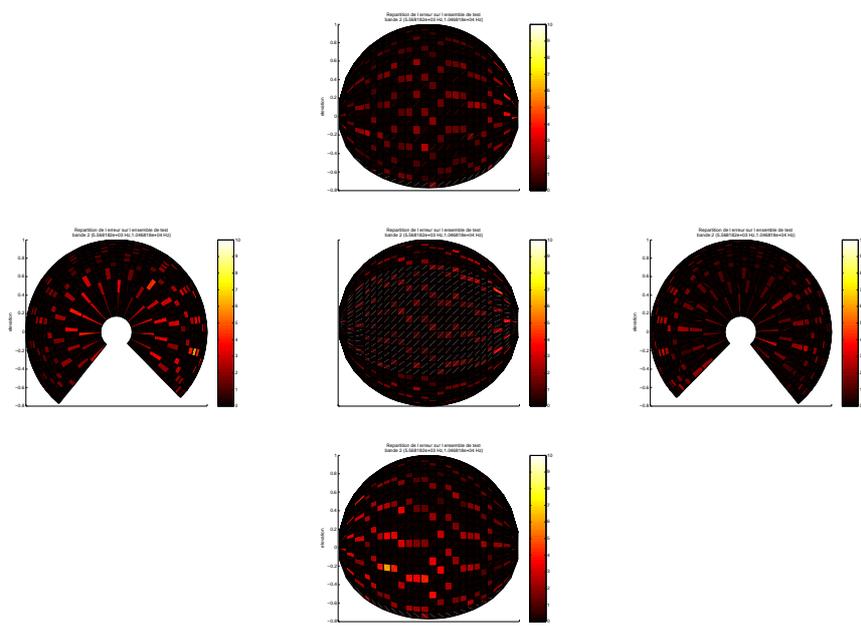


FIG. E.4 – Repartition des erreurs sur l'ensemble de test dans la bande n° 2

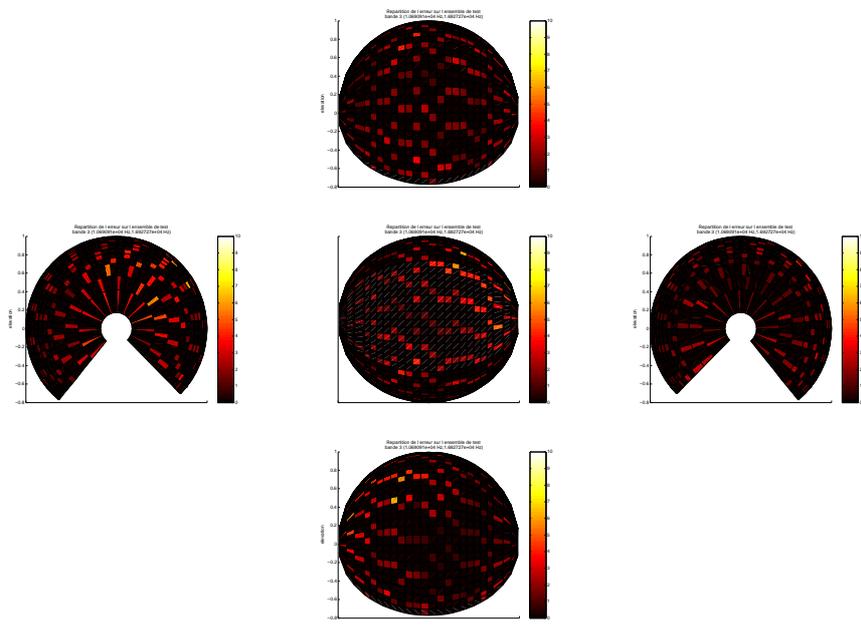


FIG. E.5 – Repartition des erreurs sur l'ensemble de test dans la bande n° 3

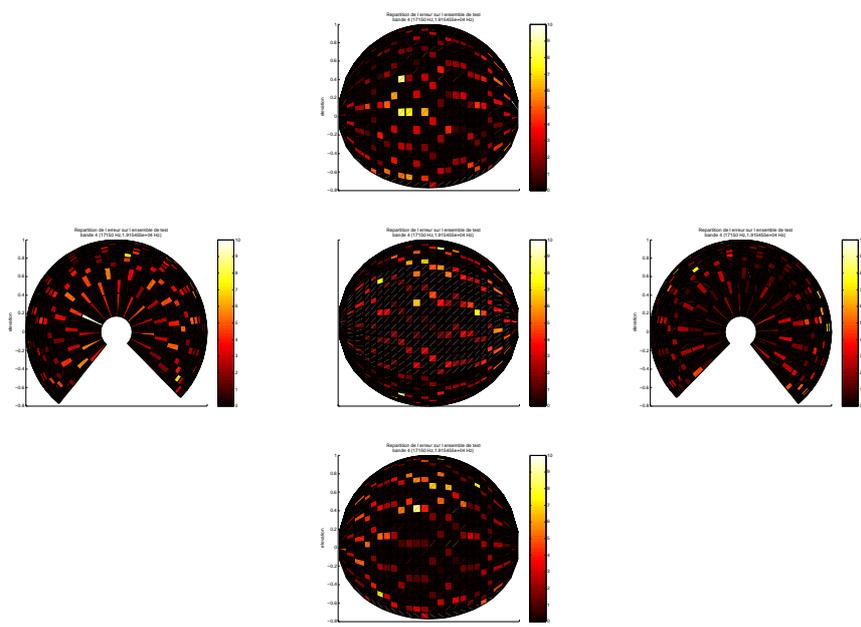


FIG. E.6 – Repartition des erreurs sur l'ensemble de test dans la bande n° 4

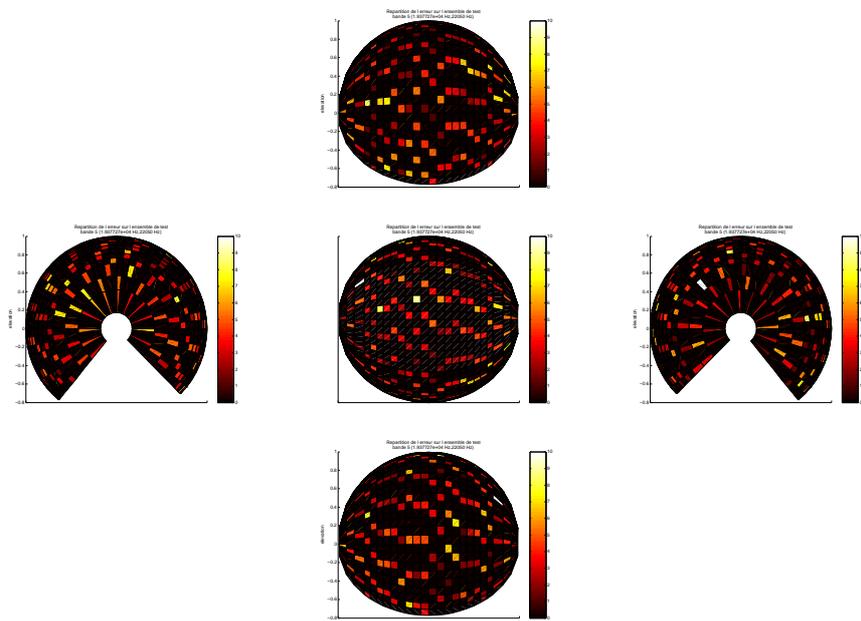


FIG. E.7 – Repartition des erreurs sur l'ensemble de test dans la bande n° 5

Annexe F

Individualisé : Modélisation en Regression non linéaire avec 1 représentant “Signal Champ-Libre”

Les figures qui suivent présentent l'évaluation des performances pour une prédiction utilisant une régression non linéaire et 26 représentants uniformément répartis à la surface de la sphère.

F.1 Répartition fréquentielle des erreurs

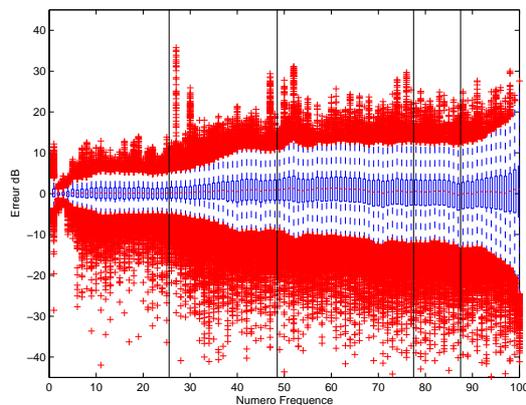


FIG. F.1 – Répartition fréquentielle des erreurs

F.2 Repartition spatiale des erreurs

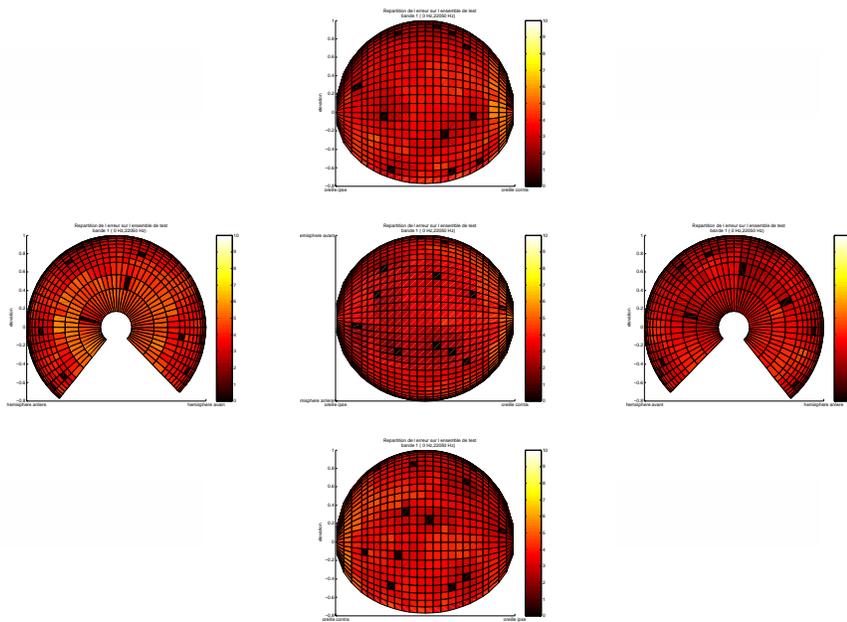


FIG. F.2: Repartition des erreurs sur l'ensemble de test

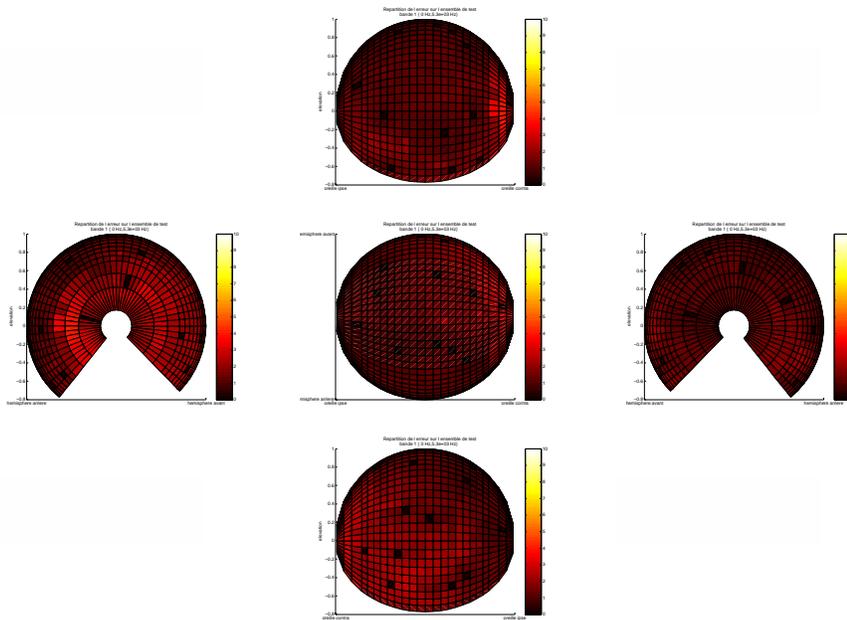


FIG. F.3: Repartition des erreurs sur l'ensemble de test dans la bande n° 1

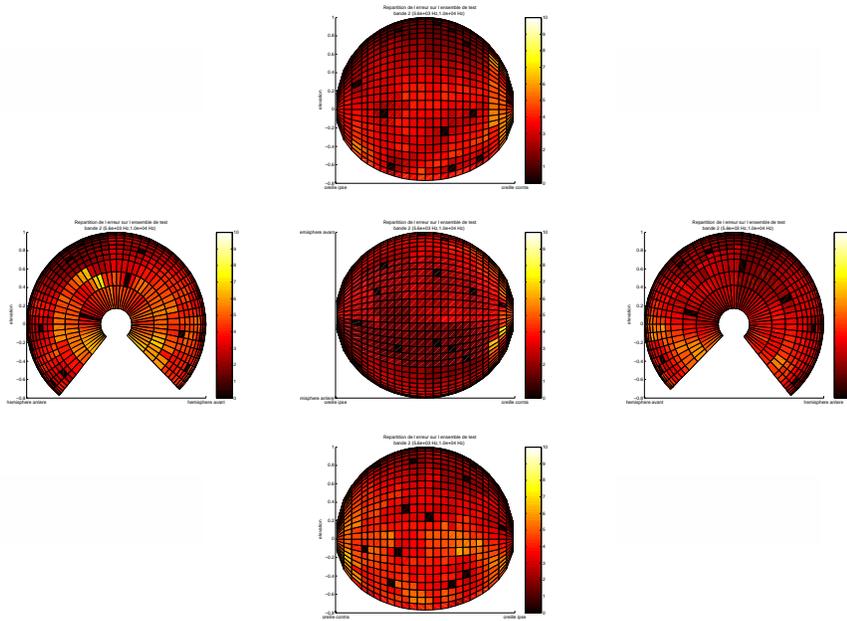


FIG. F.4: Repartition des erreurs sur l'ensemble de test dans la bande n° 2

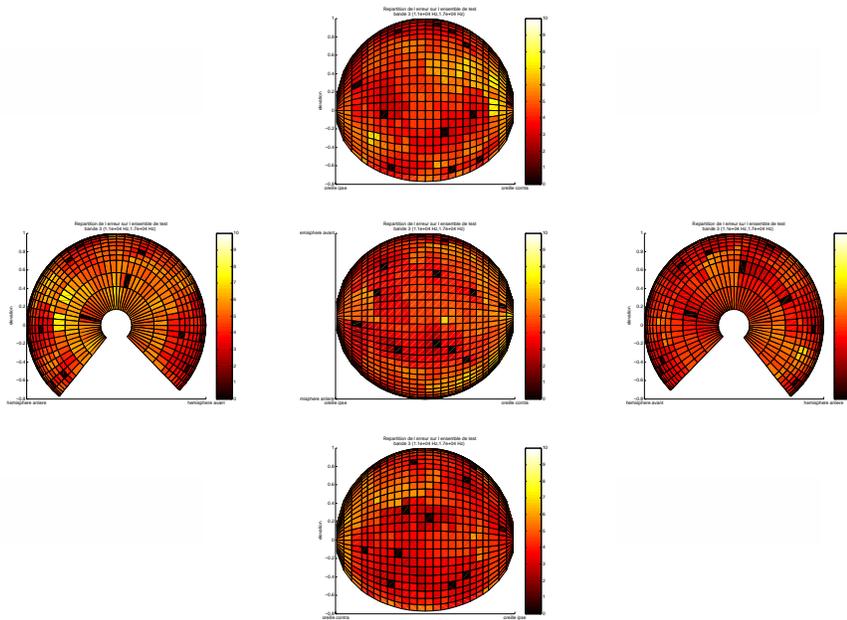


FIG. F.5: Repartition des erreurs sur l'ensemble de test dans la bande n° 3

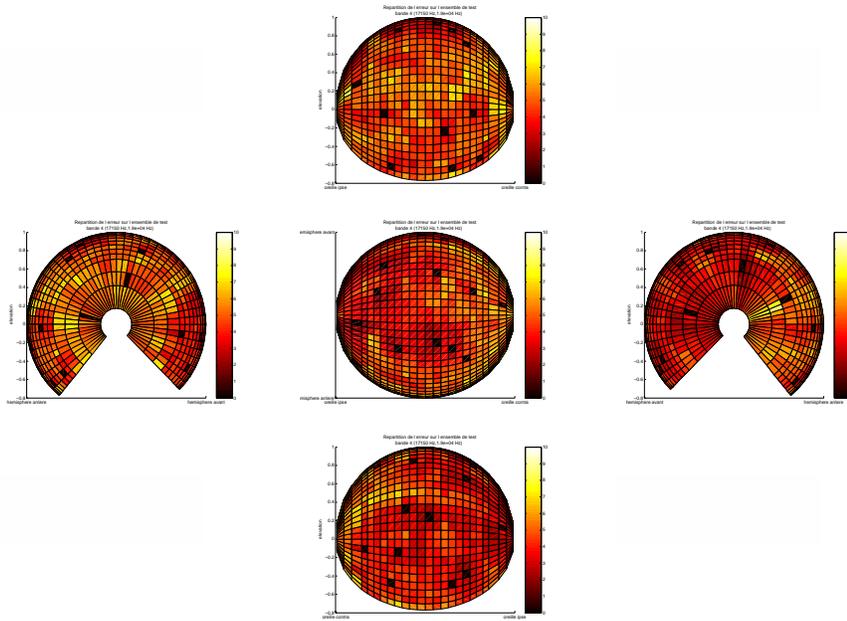


FIG. F.6: Repartition des erreurs sur l'ensemble de test dans la bande n° 4

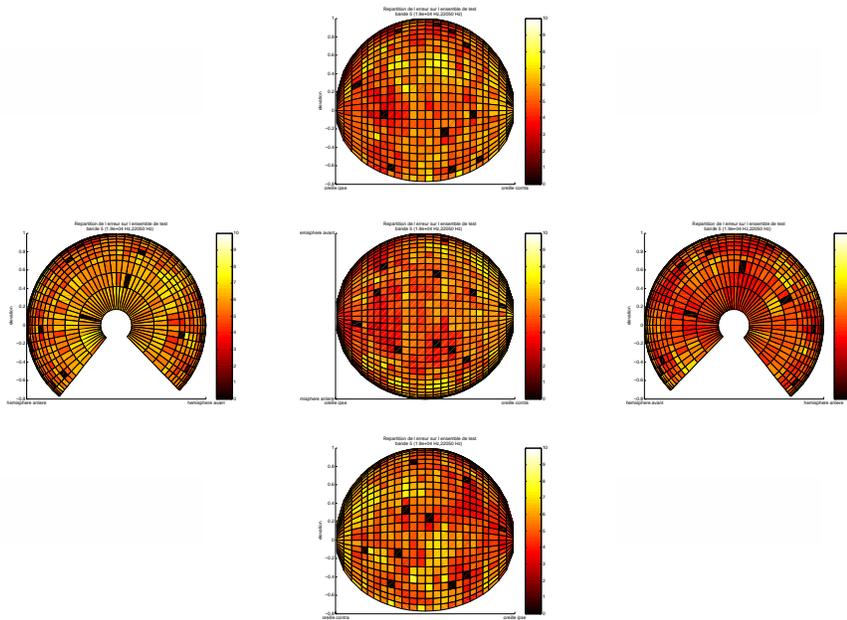


FIG. F.7: Repartition des erreurs sur l'ensemble de test dans la bande n° 5

Annexe G

Individualisé : Modélisation en Regression non linéaire avec 1 représentant “Uniformément réparti à la surface de la sphère”

Les figures qui suivent présentent l'évaluation des performances pour une prédiction utilisant une régression non linéaire et 26 représentants "Signal-Champ Libre".

G.1 Répartition fréquentielle des erreurs

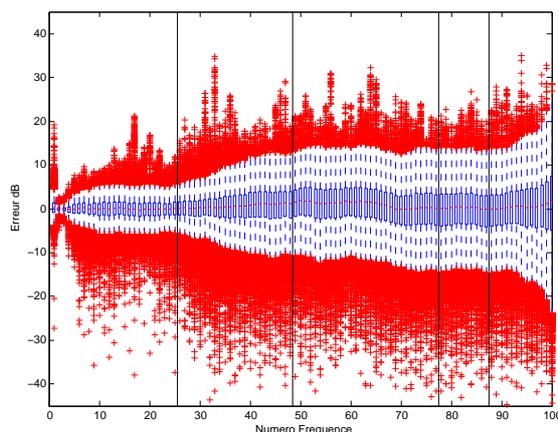


FIG. G.1 – Répartition fréquentielle des erreurs

G.2 Repartition spatiale des erreurs

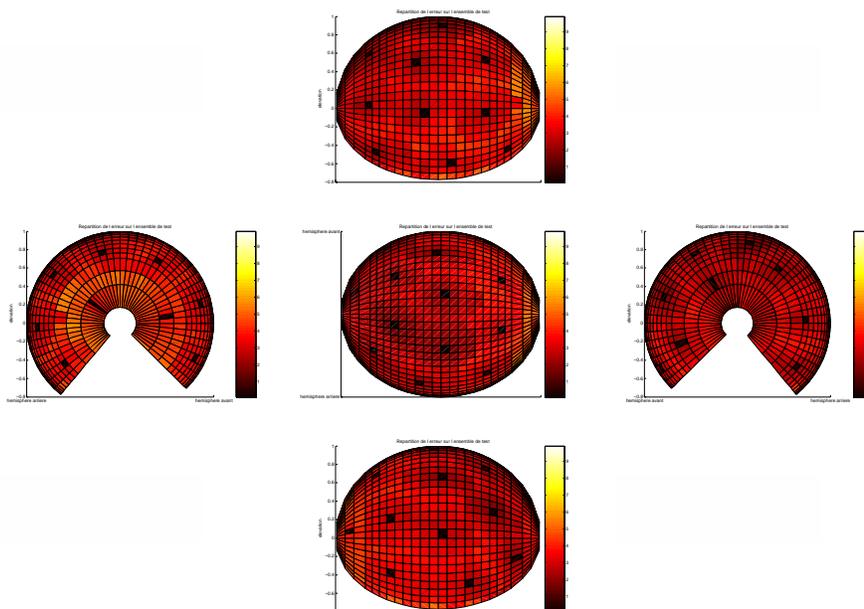


FIG. G.2: Repartition des erreurs sur l'ensemble de test

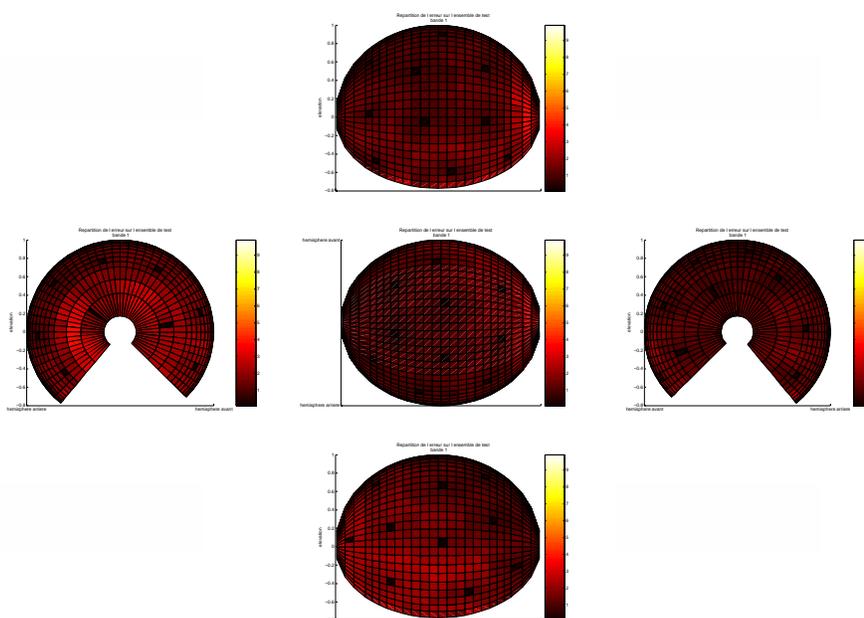


FIG. G.3: Repartition des erreurs sur l'ensemble de test dans la bande n° 1

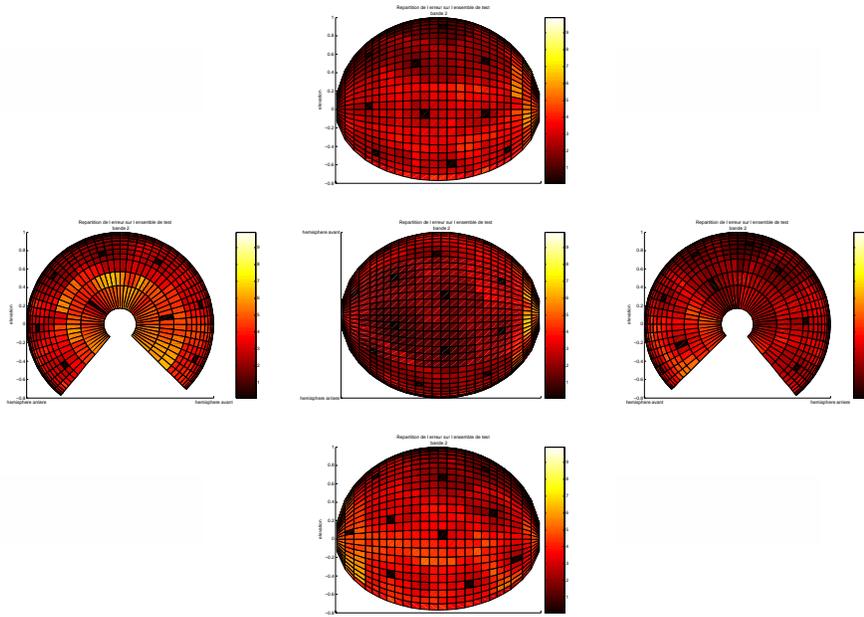


FIG. G.4: Repartition des erreurs sur l'ensemble de test dans la bande n° 2

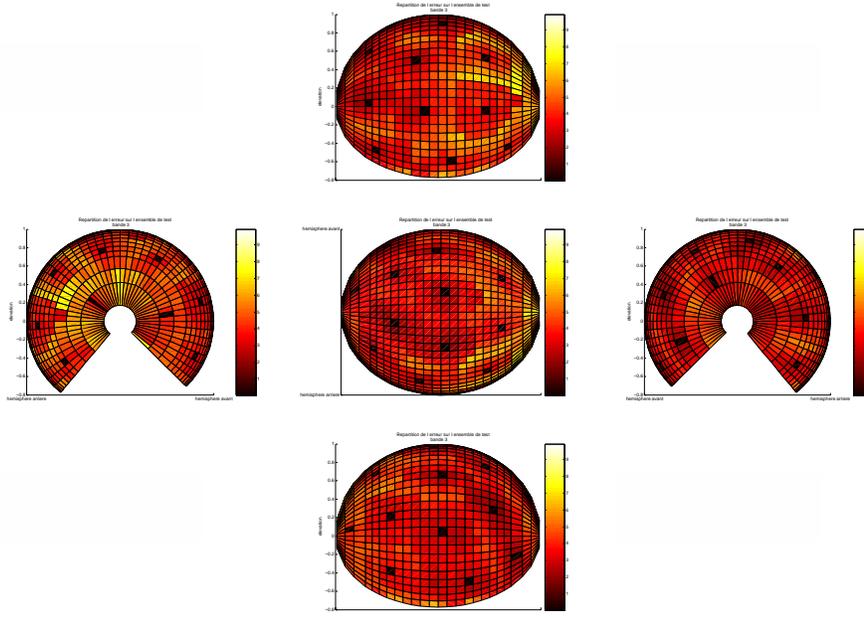


FIG. G.5: Repartition des erreurs sur l'ensemble de test dans la bande n° 3

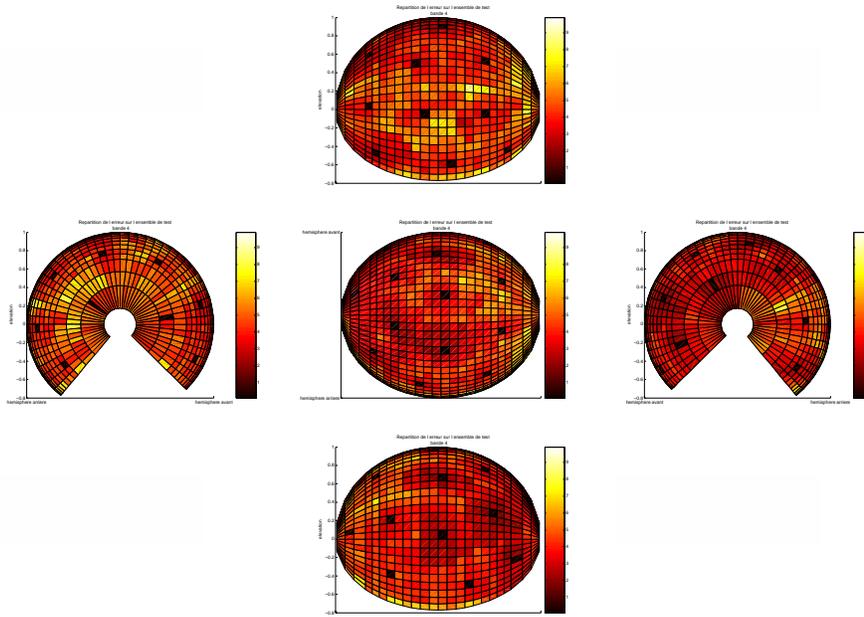


FIG. G.6: Repartition des erreurs sur l'ensemble de test dans la bande n° 4

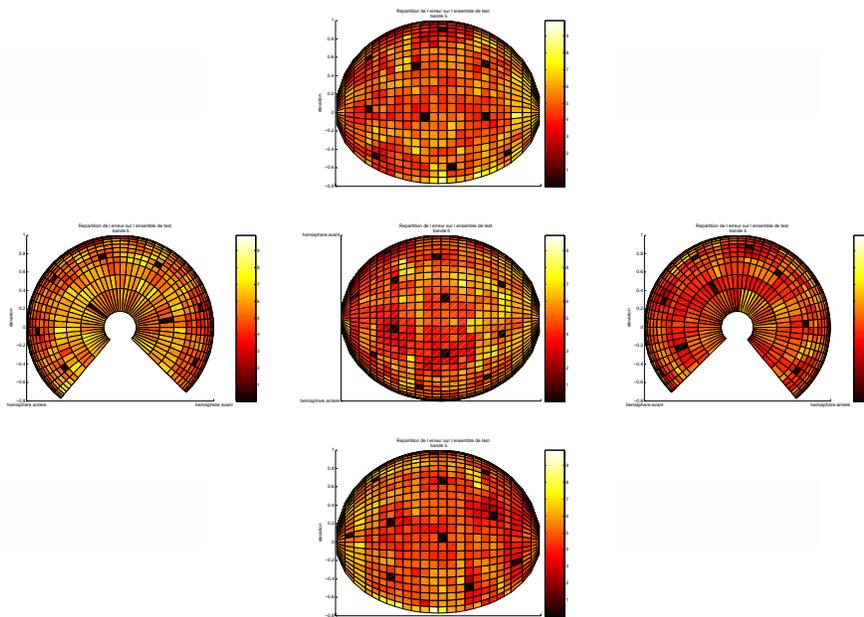


FIG. G.7: Repartition des erreurs sur l'ensemble de test dans la bande n° 5

Index

- ACP : Analyse Composantes Principales, 11
- ANN : Artificial Neural Network, 71
- CHA : Classification Hiérarchique Ascendante, 28
- CIPIC : Center for Image Processing and Integrated Computing, 16
- DTF : Directionnal Transfer Function, 21
- ERB : Equivalent Rectangular Bandwidth, 19
- FFT : Transformée de Fourier Rapide, 17
- FIR : Finite Impulse Response, 64
- HRIR : Head Related Impulse Response, 17
- HRTF : Head Related Transfer Function, 9
- ILD : Interaural Level Difference, 9
- Infinite Impulse Response, 64
- IPS : Intégration de la Parole et du Son, 8
- Ipsélatérale : Oreille la plus proche de la source sonore, 9
- Ipsélatérale : Oreille la plus éloignée de la source sonore, 9
- IRCAM : Institut de Recherche et Coordination Acoustique/Musique, 16
- ITD : Interaural Time Difference, 9
- MIT : Massachusetts Institute of Technology, 16
- PAM : Partition Around Medoids, 12
- Plan frontal : plan (xOz) dans un repère sphérique, 16
- Plan Horizontal : plan (xOy) dans un repère sphérique, 16
- Plan médian : plan (yOz) dans un repère sphérique, 16
- RN : Réseaux de neurones, 71
- TNT : Techniques Neuromimétiques pour les Télécommunications, 8

Bibliographie

- [1] L.Miclet A.Cornuéjols. *Apprentissage artificiel : Concepts et algorithmes*. Eyrolles, 2002.
- [2] R.W.Peters B.C Moore. A revision of swiker's loudness model. *Acta Acoustica*, 1996.
- [3] D.Zoltkin. Customizable auditory display. In *International Conference on Auditory Display*.
- [4] J.Alhoniemi E.Vesanto. Clustering of the selg organizing map. In *IEEE Transactions on Neural Networks*.
- [5] T Kohonen. *Self Organizing Maps*. Springer, 2001.
- [6] CIPIC Interface Laboratory. Documentation for the ucd hrir files. Technical report, University of California at Davis, 1998.
- [7] V Larcher. *Techniques de spatialisation des sons pour la réalité virtuelle*. PhD thesis, Université de Paris VI, 2001.
- [8] J.C Middlebrooks. Spectral shape cues for localization. binaural and spatial hearing in real and virtual environments. *R.H.Gilkey, Anderson, T.R.Mahwah, Lawrence Erlbaum Associates*, 1997.
- [9] A.W Mills. Auditory localisation. foundation of modern auditory theory. In *Academic*.
- [10] J.M Pernaux. *Spatialisation du son par les techniques binaurales : Application aux services de télécommunications*. PhD thesis, INPG, 2003.
- [11] S.Carlile P.H.W Leong. Methods for spherical data analysis and visualisatio. In *j.Neurosci Methods*.
- [12] S.Haykin. *Neural Network : A comprehensive foundation*. Prentice Hall, 1999.
- [13] J.O Smith. *Techniques for digital filtering design and system identification with the violin*. PhD thesis, CCRMA, Stanford, 1983.
- [14] D. M. Thompson V. R. Algazi, R. O. Duda and C. Avendano. The cipic hrtf database. In *IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*.
- [15] V.Lemaire. *Une nouvelle fonction de coût régularisante dans les réseaux de neurones artificielles : Application à l'estimation des temps de blocage dans un noeud ATM*. PhD thesis, France Telecom R&D, Paris IV, 1999.

- [16] F.Clérot V.Lemaire. An input variable importance definition based on empirical data probability and its use in variable selection.
- [17] F.Clérot V.Lemaire. Som-based clustering for on-line fraud behavior classification : a case study.
- [18] F.Clérot V.Lemaire. Detections des fraudes sur la carte france télécom. Technical report, France Telecom, 2002.
- [19] F.L Wightman. Resolution of front-back ambiguity in spatial hearing by listener and source movement. *J.Acousti.Soc.Am*, 1999.