Personnalisation de l'écoute binaurale par modèle déformable d'oreille

Slim Ghorbal

DOCTORAT BRETAGNE LOIRE / MATHSTIC

CentraleSupélec

THÈSE DE DOCTORAT DE

CENTRALESUPÉLEC

ÉCOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *Signal, Image, Vision*

Par
Slim GHORBAL

Personnalisation de l'écoute binaurale par modèle déformable d'oreille

Thèse présentée et soutenue à Rennes, le 2 avril 2021
Unité de recherche : IETR
Thèse N° : 2021CSUP0002

Rapporteurs avant soutenance :

Rozenn NICOL	Ingénieur de Recherche, Orange Labs
Olivier WARUSFEL	Chargé de recherche, IRCAM

Composition du Jury :

Président :	Kidiyo KPALMA	Professeur des Universités, IETR
Examinateurs :	Rozenn NICOL	Ingénieur de Recherche, Orange Labs
	Olivier WARUSFEL	Chargé de recherche, IRCAM
Dir. de thèse :	Renaud SÉGUIER	Professeur, CentraleSupélec

Invités :

Marc EMERIT	Ingénieur de Recherche, Orange Labs
Xavier BONJOUR	Responsable Produit, Microleed

À ma fille, Livia

Remerciements

Ne croire à ses talents que pour en remercier Dieu, c'est sanctifier l'amour-propre.

- Jean-Antoine Petit - *Bluettes et boutades*

Je tiens bien évidemment à remercier mon directeur de thèse, Renaud Séguier, Professeur à CentraleSupélec et responsable de l'équipe FAST, pour le suivi de ce travail et son accompagnement durant toutes ces années. Ses conseils, ses encouragements et son indéfectible bonne humeur auront été déterminants pour la réalisation de ces recherches.

Je remercie tous les anciens de l'aventure 3D Sound Labs, terminée trop tôt, pour les moments vécus ensemble. Une mention spéciale à Xavier Bonjour, qui aura su tenir la barre jusqu'au bout et nous communiquer sa foi inébranlable.

Je souhaite également remercier tout le personnel de CentraleSupélec qui, par son travail et son implication, y rend possible la Recherche. Karine Bernard, bien sûr, sans qui l'administration s'effondrerait. Tous les membres du service technique, toujours prêts à donner vie à nos projets parfois insensés. Ceux du 5050, là aussi toujours souriants et prêts à résoudre nos problèmes informatiques.

Je n'oublie pas non plus mes compagnons de galère, doctorants du niveau 3 de FAST et SCEE. C'est bientôt votre tour. ;)

Merci à ma famille et mes amis pour leur confiance et leur soutien tout au long de ces années. Merci à ma compagne pour sa présence, sa patience et son amour inestimable !

Merci enfin à tous ceux que j'oublie. Qu'ils se rassurent, c'est bien involontaire...

Table des matières

Remerciements **5**

Liste des abréviations, sigles et symboles **11**

1 Introduction au son spatialisé : de la morphologie à la psycho-acoustique **13**
1.1 Aux origines du son binaural . . . 14
1.1.1 Principe . . . 14
1.1.2 Théoriciens et précurseurs . . . 15
1.1.2.1 Genèse . . . 15
1.1.2.2 Lord Rayleigh et la *Duplex Theory* . . . 16
1.1.2.3 AT&T Bell Laboratories . . . 16
1.1.2.4 La formule de Woodworth . . . 17
1.1.2.5 La HRTF de Blauert . . . 18
1.2 Applications et limitations actuelles . . . 18
1.2.1 Intérêts pratiques . . . 18
1.2.1.1 Réalité Virtuelle . . . 18
1.2.1.2 Multimédia . . . 19
1.2.1.3 Santé humaine . . . 20
1.2.2 Verrous . . . 21

2 De la personnalisation des HRTF : État de l'Art **23**
2.1 L'indispensable HRTF . . . 24
2.1.1 Définitions . . . 24
2.1.1.1 HRTF ⇆ HRIR . . . 24
2.1.1.2 Dérivés . . . 25
2.1.2 Représentations mathématiques . . . 26
2.1.2.1 Approximation de l'acoustique linéaire . . . 26
2.1.2.2 Filtre à phase minimale . . . 27
2.1.2.3 Égalisations . . . 29
2.1.2.4 Métriques . . . 30
2.1.3 Indices de localisation . . . 31
2.1.3.1 ITD . . . 31
2.1.3.2 ILD . . . 36
2.1.3.3 Signature spectrale . . . 38

2.2 Stratégies d'obtention : circuits « directs » . . . 41
2.2.1 La mesure acoustique . . . 41
2.2.1.1 Des locaux adaptés . . . 41
2.2.1.2 Haut-parleurs . . . 42
2.2.1.3 Enregistrement . . . 42
2.2.1.4 Signal excitateur . . . 43
2.2.2 Les simulations numériques . . . 44
2.2.2.1 Formulation du problème . . . 44
2.2.2.2 Optimisations . . . 45
2.2.2.3 Moteurs de calcul . . . 45
2.3 Stratégies d'obtention : circuits « indirects » . . . 46
2.3.1 Choisir au sein d'une collection . . . 47
2.3.1.1 Critère morphologique . . . 47
2.3.1.2 Adaptation . . . 47
2.3.2 Synthèse à partir d'un ensemble . . . 48
2.3.2.1 Réseaux neuronaux . . . 48
2.3.2.2 Analyse statistique . . . 48
2.3.3 Fine-tuning : le sujet en première ligne . . . 49
2.3.3.1 Critère psycho-acoustique . . . 49
2.4 Bases de données . . . 50
2.4.1 Passage en revue . . . 50
2.4.2 Analyse comparée . . . 55
2.4.3 Application à notre cas d'utilisation . . . 57
2.5 Tests subjectifs . . . 58
2.6 Chaîne de personnalisation proposée . . . 60
2.6.1 Création des modèles déformables . . . 61
2.6.2 Bases de données et fonction de couplage . . . 62
2.6.3 Processus utilisateur . . . 63
2.6.4 Simulations numériques et impédance acoustique . . . 64
2.6.5 Limitations du champ d'étude . . . 65

3 Modélisations morphologiques 67
3.1 Base réelle et modèle d'oreille . . . 68
3.1.1 Collecte de données réelles . . . 68
3.1.1.1 Prérequis . . . 68
3.1.1.2 Acquisitions . . . 70
3.1.2 Modèle déformable statistique d'oreille . . . 71
3.1.2.1 Construction . . . 71
3.1.2.2 Correspondance . . . 72
3.1.2.3 Analyse statistique . . . 76
3.2 Modèles paramétriques complets . . . 77
3.2.1 Modèles synthétiques . . . 77
3.2.2 Modèle mixte . . . 80

3.3 De la 2D à la 3D : Optimisation sur photos 80

4 Production de HRTF 83
4.1 Focus sur le canal auditif 84
4.1.1 Représentation du canal auditif en simulation 84
4.1.2 Source ponctuelle vs éléments vibrants 87
4.1.2.1 Protocole expérimental 87
4.1.2.2 Résultats 88
4.1.2.3 Conclusions 94
4.1.3 Géométrie du canal 95
4.1.3.1 Protocole expérimental 95
4.1.3.2 Résultats 97
4.1.3.3 Conclusions 101
4.2 Optimisations numériques 102
4.2.1 Maillage adaptatif 102
4.2.2 Dépendance en fréquence du maillage 104
4.2.2.1 Définition 104
4.2.2.2 Impacts en simulation 104
4.2.2.3 Gains calculatoires 109
4.2.2.4 Conclusion 109
4.3 Impédance des matériaux 109
4.3.1 Problème des simulations 110
4.3.2 Existence d'une solution par l'impédance 113
4.3.2.1 Protocole expérimental 113
4.3.2.2 Recherche par simplex 114
4.3.2.3 Recherche quadrillée 116
4.3.3 Travail par zones 120
4.3.4 Évaluation et généralisation 125

5 D'un monde à l'autre : liens entre morphologie et HRTF 133
5.1 Bases de données 134
5.1.1 Base synthétique 134
5.1.1.1 Constitution 135
5.1.1.2 Analyse 137
5.1.2 Base aléatoire 141
5.1.2.1 Constitution 142
5.1.2.2 Analyse 142
5.1.3 Base mixte 145
5.1.3.1 Constitution 147
5.1.3.2 Analyse 148
5.2 Décomposition et couplage 149
5.2.1 Décomposition par ACP 150
5.2.1.1 Mise en place 150

5.2.1.2 Compression des données . . . 151
5.2.2 Couplage . . . 153
5.2.2.1 Couplage linéaire . . . 154
5.2.2.2 Réseaux de neurones . . . 154
5.2.2.3 Couplage barycentrique . . . 155
5.3 Évaluation subjective . . . 155
5.3.1 Procédure de test . . . 156
5.3.2 Résultats . . . 157
5.3.2.1 Base synthétique . . . 157
5.3.2.2 Base aléatoire . . . 160
5.3.2.3 Base mixte . . . 163
5.3.3 Discussion . . . 166

6 Conclusions et perspectives **169**

A SoundStage **175**

B Descriptif des tests **181**
B.1 Localisation . . . 182
B.2 Simulateur de test . . . 185

C Modèles ACP **187**
C.1 Modèles de la base synthétique . . . 188
C.1.1 Modèle morphologique . . . 188
C.1.2 Modèle binaural . . . 194
C.2 Modèles de la base aléatoire . . . 199
C.2.1 Modèle morphologique . . . 199
C.2.2 Modèle binaural . . . 205
C.3 Modèles de la base mixte . . . 210
C.3.1 Modèle morphologique . . . 210
C.3.2 Modèle binaural . . . 216

Bibliographie **221**

Liste des abréviations, sigles et symboles

ACI	Analyse en Composantes Indépendantes
ACP	Analyse en Composantes Principales
BEM	Boundary Element Method
CIPIC	Center for Image Processing and Integrated Computing
DTS	Digital Theater Sound
ES	Exponential Sweep
FFT	Fast Fourier Transform
FM-BEM	Fast-Multipole Boundary Element Method
FMM	Fast Multipole Method
HAT	Head-And-Torso
HPTF	HeadPhone Transfer Function
HRIR	Head-Related Impulse Response
HRTF	Head-Related Transfer Function
IACC	InterAural Cross-correlation
IID	Interaural Intensity Difference
ILD	Interaural Level Difference
IRGD	Integrated Relative Group Delay
ISSD	Inter-Subject Spectral Difference
ITD	Interaural Time Difference
JND	Just Noticeable Difference
KEMAR	Knowles' Electronics Manikin for Acoustic Research
LDDMM	Large Deformation Diffeomorphic Metric Mapping
LMS	Least Mean Square
LSCM	Least Squared Conformal Mapping
MESM	Multiple Exponential Sweep Method
MHD	Modified Hausdorff Distance

MLP	Multi Layer Perceptron
OATSP	Optimised Aoshima's Time-Stretched Pulse
SOFA	Spatially Oriented Format for Acoustics
SYMARE	Sydney York Morphological and Acoustic Recordings of Ears
TSP	Time-Stretch Pulse

Chapitre 1

Introduction au son spatialisé : de la morphologie à la psycho-acoustique

Johnnie Walken tapota une de ses bottes avec la badine noire qu'il tenait à la main. Un petit bruit sec emplit aussitôt la pièce et le chien dressa légèrement les oreilles.

- Haruki Murakami - *Kafka sur le rivage*

Sans forcément nous en rendre compte, nous faisons tous au quotidien l'expérience du *son spatialisé*. Qu'on nous appelle du 3^{e} étage et nous levons la tête. Qu'un bruit de moteur semble se rapprocher de nous et nous remontons rapidement sur le trottoir. Qu'un piaillement d'oiseaux se fasse entendre et nous scrutons un arbre plutôt qu'un autre à la recherche d'un spécimen.

Par son *spatialisé*, et ces quelques situations en sont autant d'exemples, on entend donc un son auquel est associée une information de *localisation*. Cette information est ensuite interprétée par le cerveau humain et traduite en sensation, elle-même utilisable en vue d'opérations de plus haut niveau. Ainsi, discuter avec un convive lors d'un banquet n'est possible qu'en parvenant à faire abstraction des autres conversations et bruits environnants. Sans cette faculté, le discours de notre interlocuteur se trouverait noyé sous les bruits de couverts et son écoute rendue des plus pénibles. Le premier bruit de fourchette, la première chaise qui glisse, le jappement du chien qui n'en peut plus d'attendre et nous voilà perdus ! De manière inconsciente, nous faisons donc tous, toutes, et de tout temps cet exercice psycho-acoustique d'interprétation des sons.

Pour autant, les systèmes actuels de reproduction sonore peinent à retranscrire cette sensation, livrant la plupart du temps un rendu intra-crânien – pour ce qui touche à l'écoute au casque – ou n'ayant d'autre origine que l'enceinte reproductrice. En d'autres termes, l'illusion que les sons perçus proviennent de telle ou telle direction de l'espace est rarement reproduite. Il existe donc un manque manifeste dans les théories classiques utilisées dans le monde de l'audio, manque que vise à combler par son application la théorie du *son binaural*.

1.1 Aux origines du son binaural

1.1.1 Principe

Le terme *binaural* – du latin *bini*, paire, et *auris*, l'oreille – est, étymologiquement parlant, applicable à tout ce qui se réfère aux deux oreilles et, selon le dictionnaire Larousse, « se dit des perceptions auditives engendrées par une stimulation simultanée des deux oreilles ». Avec le temps, son usage a peu à peu évolué et recouvre désormais un champ de recherche visant à comprendre et maîtriser les mécanismes permettant à l'être humain de percevoir l'origine spatiale des sons. Comme son étymologie le laisse présager, la présence chez l'homme de deux oreilles jouerait un rôle prépondérant dans ces mécanismes, et l'on comprend facilement pourquoi. Imaginons par exemple une source sonore placée à hauteur de tête, à quelques dizaines de centimètres et légèrement sur la droite. Le son s'en échappant arrivera immanquablement à l'oreille droite avant de parvenir à l'oreille gauche. Qui plus est, il y arrivera avec un niveau sonore sensiblement plus élevé. La différence des temps d'arrivée est plus classiquement appelée *Interaural Time Difference* (ITD), tandis que la différence des niveaux sonores est appelée *Interaural Level Difference* (ILD). La simple présence de deux oreilles, à des emplacements à l'évidence distincts, suffit donc à créer ces deux indices de localisation utilisables par le cerveau, permettant en l'occurrence de distinguer les sons en provenance de la droite de ceux en provenance de la gauche.

Ces indices de temps et de puissance sont notamment à la base de la théorie du son *stéréophonique* – ou *stéréo* – et de ses multiples variantes.

Néanmoins, imaginons la même source placée droit devant nous. Les sons nous parviendront en même temps aux deux oreilles et avec la même intensité. Faisons monter ou descendre cette source, les sons nous parviendront toujours en même temps – ITD nulle – et avec la même intensité – ILD nulle. Rien ne distingue donc ces différentes positions. Or, nous l'avons vu, notre cerveau les distingue ! En toute logique, il doit donc exister un autre indice de localisation que les deux mentionnés précédemment. Et celui-ci peut être cherché dans le contenu fréquentiel même de chacun des sons arrivant à nos oreilles. En effet, avant de parvenir aux tympans, les ondes sonores suivent un trajet potentiellement tortueux et semé d'embûches. Telle partie du spectre subira une absorption par les cheveux, telle réflexion sur l'épaule viendra créer une interférence constructive ou au contraire destructive. En fin de compte, le son perçu ne sera plus spectralement identique à celui émis. Et cette différence – on emploiera aussi le terme de *coloration* – étant fonction de la position de la source par rapport à l'auditeur, elle est en mesure de véhiculer l'information nécessaire au cerveau pour supprimer les ambiguïtés résiduelles.

Ainsi, le binaural stipule que la morphologie de chacun, en ce qu'elle agit comme filtre fréquentiel et directionnel sur les sons nous parvenant, figure parmi les fondements de notre faculté à percevoir le son de façon spatialisée [1].

1.1.2 Théoriciens et précurseurs

1.1.2.1 Genèse

Le monde de l'audio tel que nous le connaissons aujourd'hui et l'industrie qui l'accompagne trouvent leur racines dans la deuxième moitié du XIX^{e} siècle. Il faut remonter à 1857 et aux travaux d'Édouard-Léon Scott de Martinville pour trouver trace du premier enregistreur sonore, baptisé *phonautographe*. On date à 1860 son premier enregistrement d'une voix humaine. Cet appareil n'était toutefois capable que d'enregistrer des sons, et non de les reproduire.

Ce n'est que 20 ans plus tard, avec le *paléophone* de Charles Cros et surtout le *phonographe* de Thomas Edison, qu'il devient tout à la fois possible d'enregistrer et de reproduire. Commercialisé dans la foulée, le phonographe est un succès commercial qui marque la naissance de l'industrie du disque.

Quatre ans plus tard seulement, dans le cadre de l'Exposition Universelle de Paris de 1881, Clément Ader et la Société Générale des Téléphones mettent

FIGURE 1.1: *Théâtrophone – Affiche de Jules Chéret*

1. On parlera aussi parfois de son à 360° ou de son 3D.

en place la première expérience d'écoute 2-canaux proposée au public. Dénommée *théâtrophone*, elle permettait d'écouter au travers de deux récepteurs téléphoniques depuis le Palais de l'Industrie les airs joués à l'Opéra Garnier, sur la scène duquel était installés deux micros. Précurseur de la stéréophonie, ce dispositif sera étendu à d'autres salles de spectacle et commercialisé jusqu'en 1936.

1.1.2.2 Lord Rayleigh et la *Duplex Theory*

Si les prouesses technologiques s'enchaînent d'année en année, avec à leur suite un intérêt sans cesse grandissant du public, c'est en 1907 seulement qu'est pour la première fois théorisé [2] un processus de spatialisation du son : La *Duplex Theory* [131].

Son auteur, le physicien britannique Lord Rayleigh, mène à cette période de nombreuses expériences visant à décrire précisément les capacités de l'oreille humaine à estimer la direction de provenance d'un son. Analysant les résultats de tests subjectifs d'écoute réalisés dans le plan azimutal à l'aide de sons purs, de voix ou encore d'écoulement d'eau, il détermine que la distinction gauche / droite est très aisée à obtenir, que le sujet soit autorisé ou non à bouger la tête. Pour les sources situées devant ou derrière le sujet en revanche, la difficulté est d'autant plus grande que le son est pur, mais dans la grande majorité des cas, un léger mouvement de tête permet de lever l'incertitude de localisation. Il tire également de ses travaux la conclusion que la différence de phase – respectivement la différence d'intensité – ne peut expliquer à elle seule ses observations et que c'est dans l'association des variations de phase et d'intensité qu'il faut chercher une explication plus complète. Néanmoins, à chaque indice est associé une zone du spectre dans laquelle il prédomine. La phase en-dessous de 512 Hz, l'intensité au-dessus.

1.1.2.3 AT&T Bell Laboratories

Dans les années 1930, la stéréophonie se démocratise un peu plus avec la commercialisation des premiers récepteurs radio stéréo. De leur côté, les laboratoires Bell travaillent sur des systèmes multicanaux de perspective auditive. Leurs recherches incluent notamment une évaluation acoustique de la localisation des sons.

En 1933, à l'occasion de l'Exposition Universelle de Chicago, ils franchissent un pas supplémentaire en direction du son binaural tel que nous le concevons aujourd'hui avec la présentation d'*Oscar* – visible figure 1.2 –, premier mannequin humanoïde pourvu de microphones. Placés sur les joues, légèrement en avant des oreilles, ils permettaient l'enregistrement *in situ* de la scène sonore. Le mannequin, installé dans une pièce vitrée, donnait aux curieux attroupés à l'extérieur le loisir d'écouter ce qu'il entendait grâce à un casque relié à ses microphones. Bien que le rendu nous paraîtrait très moyen de nos jours, le caractère novateur de cette attraction en fit l'une des plus prisées de l'exposition. Il préfigure ce que sont les mannequins actuels tels que le KU-100 de Neumann ou le KEMAR de Knowles' Electronics.

2. Des travaux similaires ont été mené dès 1904 par le Prof. L. T. More [122] mais n'ayant publié ses résultats qu'en 1909, il fut oublié...

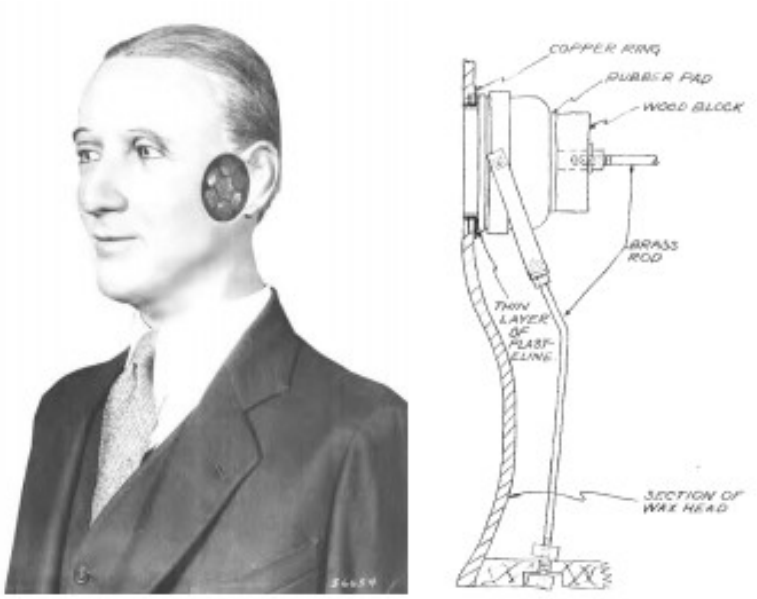


FIGURE 1.2: *Oscar, de la compagnie AT&T, ancêtre des mannequins d'enregistrement binaural.*

1.1.2.4 La formule de Woodworth

En 1938, dans son manuel *Experimental Psychology* [156] le professeur en psychologie *Robert S. Woodworth* reprend la *Duplex Theory* de Lord Rayleigh et l'applique au cas d'une boule, prise comme première approximation de la tête humaine.

S'intéressant au cas d'une source sonore située à l'infini et émettant en direction d'une boule affublée de deux oreilles – fictives – diamétralement opposées, il établit la formule portant aujourd'hui son nom exprimant l'ITD comme fonction du rayon a de la boule, de la vitesse de propagation c des ondes et de l'angle d'incidence θ.

$$ITD(\theta) = \frac{a}{c}(\theta + sin\ \theta) \tag{1.1}$$

Particularité intéressante de cette formulation du problème : la symétrie de révolution autour de l'axe interaural, i.e. l'axe joignant les deux oreilles. Celle-ci amène une dépendance selon un seul angle, donnée nécessaire et suffisante pour définir un cône de sommet le centre de la boule, d'axe de révolution l'axe interaural, de demi-angle au centre θ et matérialisant un ensemble de lieux d'ITD constants. En conséquence, toutes les directions situées sur ce cône ne peuvent être discriminées par

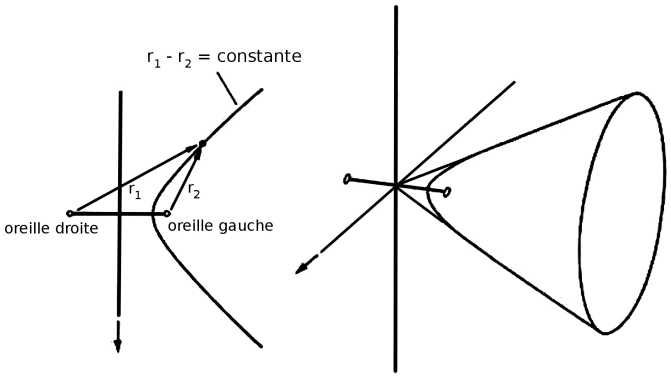


FIGURE 1.3: *Cône de confusion*

leur ITD et l'on parlera tout naturellement de *cône de confusion*.

Dans les cas extrêmes ce cône se voit réduit à une droite ($\theta = 0$) ou, au contraire, est totalement aplati ($\theta = \frac{\pi}{2}$) et coïncide avec le plan perpendiculaire à l'axe interaural, aussi appelé *plan médian*. Bien évidemment, il ne s'agit là que d'une approximation et le caractère purement conique disparaît dès lors que la symétrie du problème est rompue (par l'adjonction d'un torse, un autre positionnement des oreilles ou une forme plus ellipsoïdale donnée à la tête).

1.1.2.5 La HRTF de Blauert

Si les premiers liens entre la forme complexe des pavillons d'oreille et notre faculté de localisation auditive remontent aux travaux de Batteau [7] publiés en 1967, c'est en 1974 qu'est pour la première fois intronisée la notion fondamentale de l'écoute binaurale : la fonction de transfert reliée à la tête, ou *Head-Related Transfer Function* (HRTF). Le psycho-acousticien allemand Jens Peter Blauert [17] la définit pour chaque oreille sous forme de filtres fréquentiels et directionnels représentant l'ensemble des altérations subies par le signal sonore avant d'atteindre le tympan. Pour chaque individu peut donc se définir une HRTF droite et une HRTF gauche.

Pour une direction donnée, l'application des filtres droit et gauche à un signal mono permet de reconstituer le signal tel qu'il aurait été perçu par l'auditeur si la source avait été réellement présente dans ladite direction. Ainsi, deux opérations de filtrage suffisent à donner à l'auditeur l'illusion d'un signal spatialisé.

1.2 Applications et limitations actuelles

Comme on peut le constater, la spatialisation du son n'a eu de cesse de passionner et de rassembler les scientifiques, les industriels et le public. Il faut dire que les champs d'applications sont immenses, tout comme les obstacles qui subsistent encore.

1.2.1 Intérêts pratiques

Initialement présenté comme une curiosité, une attraction (cf. le théâtrophone, Oscar), le son binaural trouve de nos jours de plus en plus d'applications possibles. Cela tient en partie à la révolution numérique et au développement croissant de la puissance de calcul de l'informatique actuelle.

1.2.1.1 Réalité Virtuelle

L'anaglyphe, la stéréoscopie ou l'holographie sont autant de procédés développés pour reproduire l'illusion d'une réalité visuelle tridimensionnelle. Et si leurs résultats firent longtemps pâle figure devant l'imaginaire des œuvres de science-fiction [3], le dernier quart de siècle est en passe de renverser la tendance, tant les progrès y ont été nombreux et spectaculaires.

3. On pensera notamment à l'hologramme d'Obi Wan Kenobi, projeté par R2D2 dans *La Guerre des Étoiles*, sorti en 1977.

Le cinéma 3D, qui connut son premier véritable essor en 2009 pour la sortie d'Avatar, la télévision 3D, qui malgré un succès commercial en demi-teinte tient toujours une place dans les rayons d'électronique grand public, et désormais les casques de réalité virtuelle, de l'Oculus Rift au Gear VR en passant par le Google card board [4], sont autant de matérialisations des progrès en question. Néanmoins, ces illusions nécessitent d'être préparées avec minutie sans quoi le cerveau, se sentant à bon droit floué, rejettera la supercherie. Dans les faits, cela se traduit généralement par une sensation de nausée, des maux de tête et une incapacité à rester immergé bien longtemps dans ces univers virtuels.

Le manque de réalisme est donc un écueil fatal à éviter absolument par qui entend tromper nos sens. Et bien que d'énormes avancées aient été faites en matière visuelle, l'ouïe se voit généralement traitée avec plus de légèreté, alors même qu'elle fait partie des sens à ne pas négliger. En effet, on peux se laisser berner ou non par une illusion, mais il n'y a pas d'entre-deux. Et un monde virtuel nous apparaîtrait-il réaliste si, mettons, les voix des personnes que l'on pourrait y croiser ne nous semblaient pas provenir de leur bouches ? Ou si le bruit de nos pas ne nous arrivaient pas du sol ? Voilà des incohérences bien importunes qui se trouvent naturellement effacées par l'utilisation d'un moteur de son binaural ! Le son *et* la vision demeurant l'un et l'autre cohérents, l'illusion s'en trouve renforcée et perdure d'autant plus.

1.2.1.2 Multimédia

Dans le domaine du 7^{e} art et de la musique, on observe depuis des décennies une tendance de l'industrie à multiplier le nombre de canaux. Après de nombreuses années de règne quasi-hégémonique [5] de la stéréo, les laboratoires Dolby conçoivent en 1975 la première norme Dolby Stéréo, qui offrait quatre canaux distincts : avant droit et gauche, centre et arrière. En 1990, le *Dolby SR-D* [6] fait son apparition. Il est constitué de cinq canaux principaux et d'un dernier pour les basses. Six enceintes sont donc nécessaires pour en profiter au mieux, cinq d'entre elles – correspondant aux canaux principaux – devant être situées respectivement en face de l'auditeur, à $\pm 40°$ et à $\pm 110°$. Cette technologie est suivie trois ans plus tard par son concurrent, le *DTS* [7]. Ces procédés sont rapidement devenus des standards et il est rare de trouver un DVD dont elles soient toutes deux absentes. La liste n'a depuis eu de cesse de s'allonger, accueillant en son sein le 6.1, le 7.1, le 10.2 [8] ou encore le 22.2 [9], pour n'en citer que quelques unes.

Aussi nombreuses qu'elles puissent être, elles ont toutes en commun de limiter le champ sonore à un nombre réduit de sources, situées à des positions précises et fixes autour de l'auditeur. Par ailleurs, il est nécessaire d'investir dans un système d'écoute d'autant plus coûteux et complexe à installer qu'il y a de canaux à jouer.

4. Avec une pensée particulière pour la console *Virtual Boy* de Nintendo, sortie en 1995 et dont la technologie permettait déjà un rendu en relief.

5. On note quelques tentatives intéressantes, telle le *Chase Surround Sound*, plus connu comme *Fantasound*, ou le son multicanal de la société Todd-AO, mais rapidement avortées faute de succès commercial suffisant.

6. Dolby Spectral Recording Digital, ou système 5.1

7. Digital Theater Sound

8. Format multicanal développé par Tomlinson Holman qui apporte l'information de hauteur.

9. Format multicanal développé par la NHK pour accompagner le super high vision (7 680 x 4 320 pixels).

Il n'en faut pas plus au binaural pour apporter une plus-value. En effet, le multicanal n'est « in fine » qu'une liste de sources associées à des directions figées par convention. Or la technologie binaurale peut simuler la présence d'une source sonore dans n'importe quelle direction. On peut donc l'utiliser pour créer des haut-parleurs virtuels aux endroits voulus, chacun étant en charge de la restitution d'un canal.

Au-delà de la simplicité du procédé, le gain est multiple. Tout d'abord économique car il n'est plus besoin d'une lourde installation matérielle. Un simple casque stéréo suffit [10]. Pratique ensuite, aucune installation physique n'étant nécessaire. Pas d'encombrement à craindre non plus, ni de réglage hasardeux en fonction des dimensions de la pièce. La pièce elle-même est superflue, l'usage pouvant être désormais nomade. Enfin, l'essentiel du travail étant d'ordre logiciel, il est très facile de s'adapter à l'ensemble des formats existants – et il ne s'agit pas uniquement des formats multicanaux – ou à venir sans avoir à investir quoi que ce soit.

Dans les faits, n'importe quel contenu multicanal est donc déjà « compatible » avec le binaural et peut en tirer parti sans pré-traitement particulier.

1.2.1.3 Santé humaine

Loin des ces applications strictement divertissantes, le monde de la médecine peut lui aussi tirer parti des bénéfices apportés par le son binaural. En particulier en audiologie, la réponse apportée aux personnes souffrant de perte d'audition est bien souvent l'utilisation d'une prothèse auditive. Plusieurs centaines de milliers en sont vendues chaque année en France, chiffre en constante hausse et que l'allongement de la durée de vie ne fait que soutenir un peu plus.

Or ces prothèses sont loin d'être parfaites et nombre de leurs utilisateurs les laissent de côté dès que l'environnement sonore devient trop bruyant, comme lors du banquet évoqué ci-avant. Concrètement, les patients se plaignent de ne pouvoir distinguer les sons entre eux, que toutes les conversations ainsi que les bruits parasites sont portés au même niveau, nécessitant donc les plus grands efforts pour en extraire le contenu intelligible d'intérêt. Les autres convives – non appareillées – ne se plaignant pas d'une telle gêne, c'est donc bien la prothèse qu'il faut incriminer.

En effet, en ne tenant pas compte de l'audition binaurale naturelle de la personne, elle la prive de sa faculté de localisation. Or c'est cette faculté qui permet à notre cerveau de discriminer et traiter différemment des sons provenant de sources différentes. Il n'en faut pas plus pour comprendre que l'incorporation de la technologie binaurale au sein de ces appareils améliorerait grandement l'intelligibilité du rendu et diminuerait dans le même temps sa pénibilité.

Bien sûr, cela n'a pas échappé aux industriels du milieu et certains proposent d'ores-et-déjà des solutions dites « binaurales ». Toutefois, la prudence est de mise quant à savoir ce que recouvre réellement ce terme, surtout lorsqu'il provient d'une brochure promotionnelle.

10. Dans le cas d'une écoute sur hauts-parleurs, on parlera de technologie transaurale, dont la description sort du champ d'étude du présent manuscrit. On soulignera tout de même que le gain en confort lié à l'absence de casque est contrebalancé par de fortes contraintes techniques liées à la position de l'auditeur.

1.2.2 Verrous

Malgré ces nombreuses applications potentielles, leur donner corps demande à relever plusieurs défis, et non des moindres.

Le problème de consommation de ressources tout d'abord. « Binauraliser » une trame sonore revient à lui appliquer un filtre fréquentiel. Il s'agit donc, au choix, de passer dans le domaine fréquentiel pour effectuer la multiplication des spectres avant de revenir dans le domaine temporel, ou de demeurer dans ce dernier pour y réaliser une convolution. Quelle que soit la méthode adoptée, le coût calculatoire est très important. Cela a un impact direct sur la consommation CPU de l'électronique utilisée et, par ricochet, sur sa consommation énergétique. Dans le cas des prothèses auditives, l'esthétique étant de mise, les appareils doivent se faire les plus discrets possibles et l'encombrement laissé à la batterie est réduit au strict minimum. Il s'ensuit que cette dernière doit déjà être (trop ?) régulièrement changée. Tout ajout de charge de calcul ne fait donc qu'aggraver le problème, sauf à augmenter l'encombrement et rendre la prothèse plus apparente et pesante.

Vient ensuite le problème des outils de développement. Trop peu nombreux, aux fonctionnalités limitées ou trop chers, leurs manquements sont autant de freins à la production de contenus exploitant toutes les possibilités offertes par la spatialisation du son. Et sans contenu, pas de public pour l'écouter. Et sans public, pas d'industrie pour développer les outils. C'est le classique problème de l'œuf ou de la poule auquel se voit confronté tout marché naissant et dont la durée est d'autant plus courte que l'investissement initial est grand.

Cela étant, ce développement d'un écosystème tourné vers le son 3D souffre d'un manque de standardisation. Ce n'est en effet que tout récemment, en 2015, qu'a émergé un standard d'échange de données acoustiques spatialisées, le SOFA [11] [42]. S'il semble bien s'implanter dans la communauté, sa jeunesse laisse imaginer le travail à fournir avant que chaque outil de post-production cinématographique, chaque logiciel de mixage, chaque kit de développement audio l'intègre pleinement. Cela donne également une idée du travail de reprise de données requis pour assurer une certaine rétro-compatibilité avec l'existant, telles les HRTF que l'on peut trouver dans les différentes bases à travers le monde.

Par ailleurs, l'effet psycho-acoustique des HRTF est le plus convaincant lorsque celles-ci sont spécifiquement adaptées à l'auditeur. En l'absence d'individualisation cet effet est plus aléatoire, bien moins frappant, et ne permet en général pas au binaural de se démarquer véritablement des autres technologies de spatialisation du son. La personnalisation n'a malheureusement rien de trivial, surtout lorsque les contraintes du monde industriel viennent mettre leur grain de sel. Les exigences de l'utilisateur sont naturellement plus grandes vis-à-vis d'un produit payant, en particulier concernant sa fiabilité et sa simplicité d'usage.

Les travaux de cette thèse entendent s'attaquer à ce dernier point et faciliter la production de HRTF personnalisées en assurant tout à la fois la fiabilité du résultat, la simplicité de l'individualisation et la rapidité du processus.

11. Spatially Oriented Format for Acoustics

Chapitre 2

De la personnalisation des HRTF : État de l'Art

Il m'est arrivé de prêter l'oreille à un sourd.
Il n'entendait pas mieux.

- Raymond Devos -

Si les relations entre la morphologie et l'audition humaines ont de longue date été étudiées [157], on note depuis près d'un quart de siècle un intérêt croissant dans la communauté scientifique pour le problème de l'individualisation, c'est-à-dire de la prise en compte des spécificités propres à chacun.

En particulier, l'attention s'est portée sur l'individualisation des HRTF, représentations mathématiques de la coloration fréquentielle des sons que nous percevons.

2.1 L'indispensable HRTF

2.1.1 Définitions

2.1.1.1 HRTF ⇆ HRIR

En adoptant une approche système du problème de la propagation du son, on peut représenter l'ensemble des altérations subies par le signal entre une source ponctuelle et les tympans d'un auditeur comme l'effet d'un filtre spécifique. Ces altérations sont d'origines multiples – la pièce, le milieu de propagation, divers obstacles, etc. – mais il en est une qui pour être présente en permanence se particularise de manière naturelle : l'auditeur lui-même. C'est donc sans surprise qu'elle se voit associer une fonction de transfert propre, à savoir la fameuse HRTF [1], et son pendant temporel, la HRIR [2].

Plus concrètement, en notant x_s le signal émis par une source S et x_L – resp. x_R – le signal reçu au tympan gauche – resp. droit – de l'auditeur, on a :

$$x_{L/R} \;=\; h_{L/R} \;*\; x_s \tag{2.1}$$

où $*$ est l'opérateur de convolution et h_L – resp. h_R – représente la HRIR gauche – resp. droite. Par une simple transformée de Fourier, on obtient alors les HRTF correspondantes :

$$X_{L/R} \;=\; H_{L/R} \;\cdot\; X_s \tag{2.2}$$

Bien évidemment, ces filtres dépendent de la position relative de l'auditeur à la source et l'on parlera plus volontiers de jeux de HRIR – resp. HRTF –, chacun étant constitué de quelques centaines à plusieurs milliers de HRIR – resp. HRTF –, soit autant que de positions distinctes à l'étude. L'ensemble de ces positions est communément appelé *grille d'évaluation*. En pratique, l'origine du repère est placé sur l'axe interaural, c'est-à-dire celui joignant les deux oreilles, à équidistance de chacune d'elles. Un autre usage est de ne regrouper par jeu que des filtres pris à une même distance de l'origine du repère, c'est-à-dire sur une sphère centrée sur la tête de l'auditeur. Suivant que cette distance sera petite ou grande, les filtres seront dits en *champ proche* ou en *champ lointain*. La limite généralement admise les séparant est $r = 1\,\mathrm{m}$ [27].

1. *Head-Related Transfer Function*
2. *Head-Related Impulse Response*

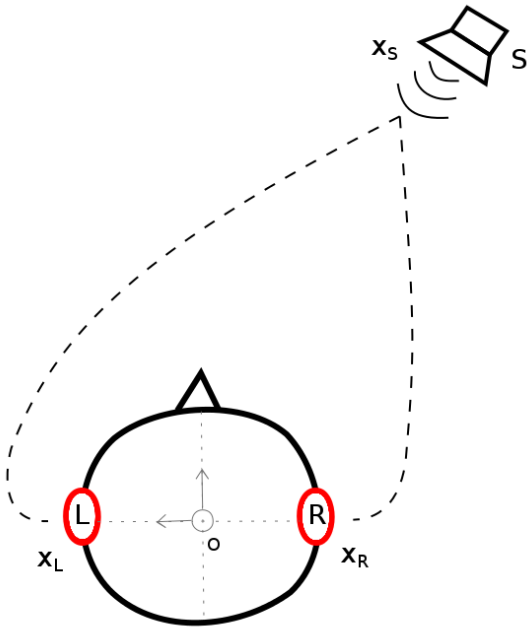


FIGURE 2.1: *Représentation schématique du principe de la HRTF.*

2.1.1.2 Dérivés

En parallèle du couple HRTF / HRIR, la littérature fait état d'un nombre important d'autres types de filtres, que l'on pourrait voir comme autant de proches cousins, avec en tête de file la *Directional Transfer Function* ou *DTF* définie par Middlebrooks [111]. Comme son nom peut le laisser présager, il s'agit de la HRTF privée de toute information non-directionnelle, cette dernière étant obtenue en moyennant toutes les HRTF du jeu. Bien qu'à proprement parler la notion de *Directional Impulse Response* n'existe pas, une transformée de Fourier inverse permet de repasser simplement toute DTF dans le domaine temporel. Il est donc tout à fait possible d'en utiliser un jeu en lieu et place du jeu de HRTF dont il est issu. Cette manipulation présentant d'ailleurs l'avantage de diminuer la coloration des sons binauralisés, elle est largement répandue, au point que par abus de langage les DTF sont elles-mêmes fréquemment appelées HRTF [3].

Très fréquemment employé également, le terme *BRIR* – pour *Binaural Room Impulse Response* – fait référence aux filtres binauraux liés à l'auditeur et à la pièce dans laquelle il se trouve. Concrètement, ils sont obtenus de la même manière que les HRIR à la différence près que la fenêtre temporelle d'enregistrement sonore est agrandie de façon à capter la réverbération due à la pièce. L'utilisation de BRIR permet donc de bénéficier des avantages d'une réverbération naturelle, connue pour apporter un réalisme supplémentaire, notamment en ce qui concerne l'externalisation des sons [41]. Cependant, c'est au prix d'une immobilisation de l'auditeur dans son environnement virtuel. En effet, sa position

3. Ce manuscrit ne fait pas exception et, sauf mention contraire, on parlera souvent de HRTF en lieu et place des DTF, notamment aux chapitres 4 et 5.

et son orientation dans la pièce ayant une influence sur la réverbération, ces informations sont par nature gravées dans ces filtres et ne sauraient être modifiées. Pour cette raison, l'emploi de BRIR ne conviendra pas aux usages autorisant l'auditeur à se mouvoir dans la scène sonore. Comme pour l'obtention des DTF, il est possible de supprimer la composante non-directionnelle des BRIR. Les nouveaux filtres sont appelés *Directional Room Impulse Response* ou *DRIR*.

Plus confidentielle, la notion de *PRTF* – pour *Pinna-Related Transfer Function* – est parfois mise en avant. Il s'agit là d'isoler, dans la HRTF, la contribution liée à l'oreille externe. Son intérêt principal réside dans le fait que la forme du pavillon est tenue pour responsable des principaux indices de localisation en élévation. Cette isolation ne pouvant se faire au moment de la capture, il est nécessaire de mettre en place un post-traitement dédié. Spagnol *et al.* [141] proposent ainsi d'appliquer un fenêtrage de Hann de 1 ms sur les HRTF de la base CIPIC – cf. section 2.4.1 – pour en retirer la contribution des épaules. Restent alors les contributions du pavillon mais aussi de la tête. Néanmoins, les auteurs assimilent cette dernière à une sphère, dont la réponse en fréquence est relativement plate dans le plan médian alors à l'étude. De cette manière, ils justifient le fait de ne pas avoir besoin de traitement supplémentaire pour considérer avoir extrait les PRTF. Toutefois, cette procédure laissant intacte l'ILD, elle-même principalement due au masquage de la tête, il est difficile d'y voir de pures PRTF. Une façon d'y remédier, notamment décrite par Geronazzo *et al.* [60], consiste en l'utilisation d'un modèle structurel de HRTF permettant d'estimer la contribution de la tête et du torse seuls. Dans le cas de Geronazzo, il s'agissait d'un modèle HAT [4], mais on trouve chez Duda *et al.* [50], Satarzadeh *et al.* [134] ou encore Aussal [5] d'autres représentations possibles.

Enfin, citons également les *HeadPhone Transfer Functions* ou *HPTF*, qui représentent l'effet de couplage d'un casque sur les oreilles. Leur prise en compte permet de supprimer une partie des perturbations liées au matériel, ce qui est un avantage certain. Néanmoins, elles dépendent tout à la fois du modèle de casque, de la géométrie des oreilles et du positionnement du casque sur la tête, ce qui peut être source de grande variabilité. Contrairement aux précédents filtres présentés, les HPTF ne sont pas directionnelles : il y a qu'un filtre gauche et un filtre droit.

2.1.2 Représentations mathématiques

2.1.2.1 Approximation de l'acoustique linéaire

La mécanique des fluides est une discipline intrinsèquement non-linéaire. Les ondes de choc, les écoulements turbulents, la viscosité des milieux et les effets thermoacoustiques en sont autant d'exemples. Très complexes à résoudre dans le cas général, les équations régissant les phénomènes de propagation ondulatoire peuvent sous certaines conditions se linéariser et offrir un cadre d'étude bien plus simple, parfois appelée *hypothèse acoustique* ou *approximation acoustique*.

4. *Head-And-Torso*

La première de ces conditions est l'hypothèse, classique, des « petits » mouvements. Une grandeur A telle que la pression, la masse volumique ou la vitesse est autorisée à varier de ΔA autour d'une valeur donnée tant qu'elle ne s'en écarte pas trop – i.e. $\Delta A << A$. À titre d'exemple, la pression P de l'air avoisine les 1 013 hPa au niveau de la mer et la surpression ΔP correspondant au seuil de douleur – et donc rare dans la vie courante – vaut à peine 100 Pa, soit 134 dB. L'application numérique nous donne : $\Delta A/A = 9,87 * 10^{-4} << 1$. L'approximation est ainsi largement valable. Pour ce qui est de la vitesse des particules, une surpression de 170 dB – près de 6 000 Pa – amène à manipuler des valeurs de l'ordre de 10 m.s^{-1}, soit près du trentième de la vitesse du son dans l'air. En fixant cette dernière à 340 m.s^{-1}, cela donne : $\Delta A/A = 2,94 * 10^{-2}$. La marge de manœuvre est donc plus réduite. Néanmoins, un dialogue d'intensité normal ne dépasse guère les 60 dB, soit 20 mPa. L'approximation considérée est alors une nouvelle fois tout à fait vérifiée pour notre cas d'étude.

Autre hypothèse simplificatrice, l'absence de forces volumiques. En particulier, l'effet de la pesanteur est supposée négligeable face aux autres forces en présence. À côté de cela, on émet également l'hypothèse de conservation de la masse, c'est-à-dire l'absence de création ou de destruction de matière durant l'expérience.

Enfin, les phénomènes dissipatifs – conduction thermique, viscosité – sont eux aussi négligés, ce qui revient à supposer le fluide comme étant parfait, rendant par suite son mouvement adiabatique. Dans le cas de l'air, cette hypothèse est elle aussi largement vérifiée dans notre cadre de travail, i.e. dans les conditions normales de température et de pression.

2.1.2.2 Filtre à phase minimale

Dans ce cadre d'approximation acoustique linéaire, les HRTF peuvent elles-mêmes être vues comme des filtres linéaires, stables et invariants dans le temps. Ces conditions réunies, elles admettent une décomposition en filtre à phase minimale et filtre à excès de phase :

$$H(f) \;=\; H_{min}(f) \cdot H_{exc}(f) \tag{2.3}$$

$$\text{avec} \begin{cases} H_{min}(f) \;=\; |H| \cdot e^{j\phi_{min}(f)} \\ H_{exc}(f) \;=\; e^{j\phi_{exc}(f)} \end{cases} \tag{2.4}$$

L'intérêt de cette représentation est multiple. Tout d'abord, pour un filtre H donné, H_{min} est unique et entièrement déterminé par le module de H. Plus exactement, sa transformée de Hilbert permet d'exprimer la phase :

$$\phi_{min} \;=\; \mathrm{Im}(\mathrm{Hilbert}(-\ln(|H|))) \tag{2.5}$$

En outre, une particularité intéressante des filtres à minimum de phase est qu'ils concentrent l'énergie du signal dans leur premiers coefficients. Mathématiquement, si l'on note h la réponse impulsionnelle du système, le choix du filtre à minimum de phase minimisera la quantité $\sum_{n=m}^{\infty} |h(n)|^2$, pour tout $m \in \mathbb{N}$, et cela a deux conséquences. La

première est que les filtres à minimum de phase seront de fait plus compacts, ce qui peut s'avérer utile pour répondre à des problématiques de transfert de données ou de stockage. La seconde est que ces filtres supprimeront toute composante linéaire de la phase, c'est-à-dire toute composante associée à un pur délai temporel. En vulgarisant davantage, on dira qu'ils « calent » les réponses impulsionnelles à gauche, faisant fi du temps de vol du signal de la source à l'oreille. En particulier, l'ITD ne peut se retrouver dans H_{min}. Mais de l'équation 2.3, on peut alors déduire que s'il est exclu de H_{min}, c'est donc qu'il est entièrement contenu dans H_{exc}, faisant de cette représentation un moyen très commode pour manipuler et analyser séparément les indices spectraux et l'ITD.

Par ailleurs, ainsi que le soulignent Mehrgardt *et al.* [108], le filtre à excès de phase des HRTF a la particularité d'être quasiment linéaire jusqu'à 10 kHz ou plus – cf. figure 2.2. Il peut donc être modélisé dans cette bande de fréquences par un retard pur. De plus, il a été observé qu'au-delà de 1 500 Hz la phase cessait rapidement d'être un indice de localisation. Par conséquent, étendre cette modélisation sur l'ensemble du domaine fréquentiel est une simplification à priori valide.

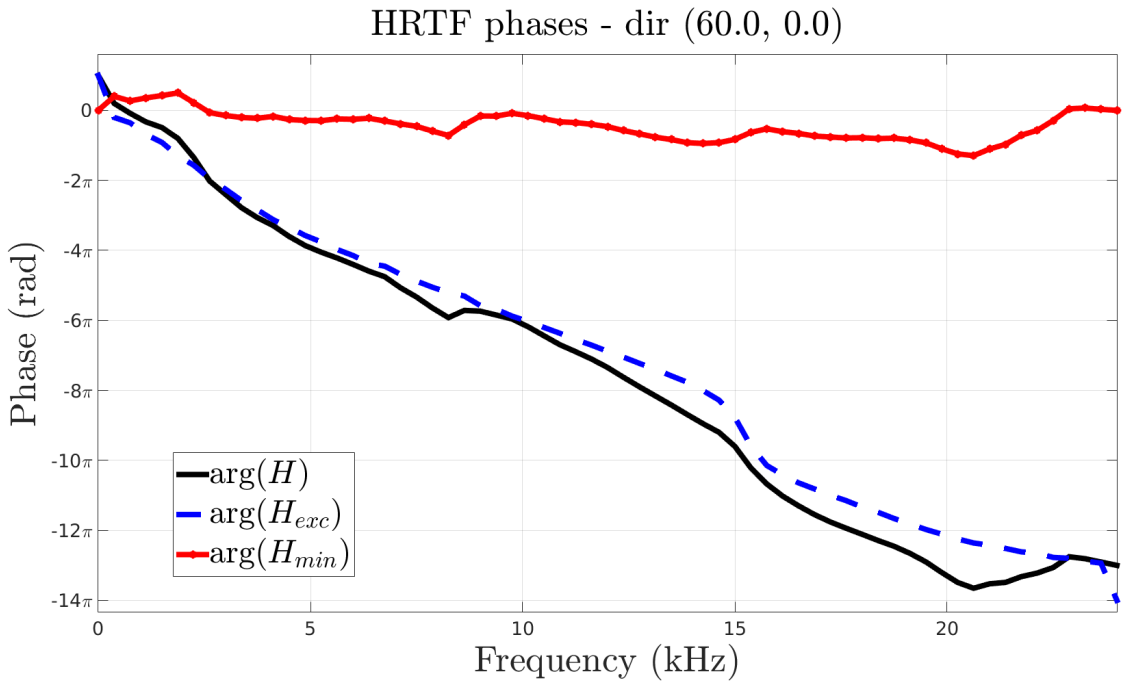


FIGURE 2.2: *Phase d'une HRTF acoustique du mannequin Kemar. Une fois dépliées, les phases de H et H_{exc} exhibent une décroissance quasi-linéaire imputable au temps de propagation des signaux et absente du filtre à minimum de phase.*

Néanmoins, l'impact perceptif de cette approximation dans la représentation des HRTF n'est au premier abord pas si trivial et a été abondamment étudié. Ainsi, Kistler & Wightman [96] ont mené une série de tests de localisations mettant notamment en concurrence des HRTF mesurées avec leur équivalent à minimum de phase auquel est adjoint un pur retard temporel. La similarité des résultats obtenus leur fait conclure en la validité de l'approximation. Kulkarni *et al.* [97] ont pour leur part effectué cette comparaison au sein d'un test à choix multiple dans lequel quatre sujets expérimentés se

voyaient attribuer la tâche de distinguer quel stimulus, parmi deux possibles, avait été généré à partir d'une HRTF à minimum de phase et retard pur en lieu et place de leur HRTF acoustique. Le test a de surcroît été réalisé pour quatre azimuts distincts : $\pm 90°$, $0°$ et $180°$. Il en sort que les HRTF à minimum de phase se sont révélées indistinguables des HRTF originales aux azimuts $0°$ et $180°$ mais pas aux positions latérales, sans que cette dernière observation ne puisse être expliquée. Un an plus tard, Plotsgies *et al.* [123] font le même constat et exhibent des HRTF à minimum de phase que certains auditeurs peuvent distinguer de leurs HRTF brutes aux positions latérales. Cette fois-ci, les auteurs poussent davantage leurs investigations et montrent qu'une estimation plus soignée de l'ITD permet de gommer les différences de perception entre HRTF avec et sans minimum de phase.

2.1.2.3 Égalisations

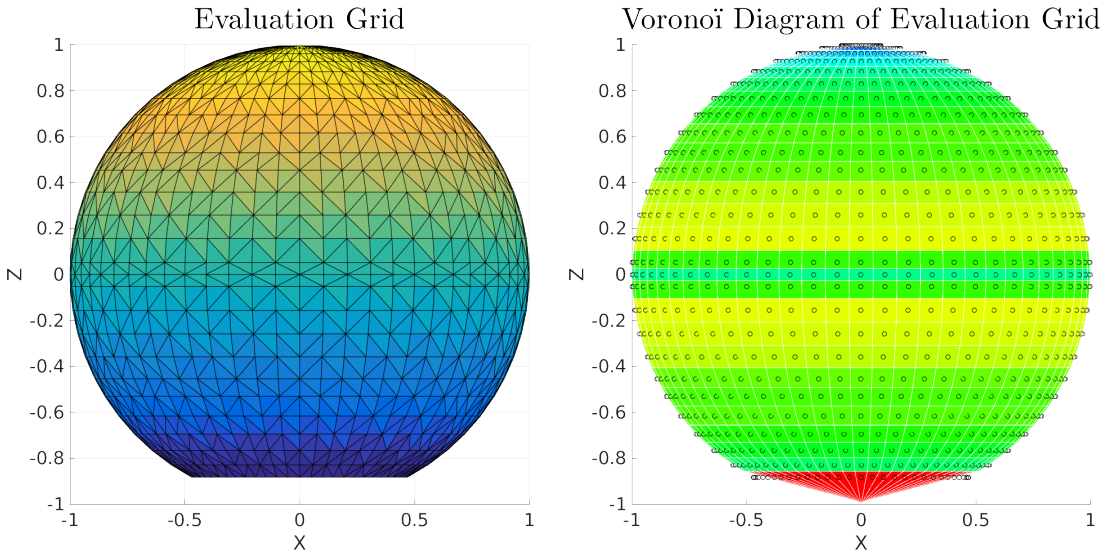
Un système de mesure de HRTF est communément constitué d'un haut-parleur, d'une paire de microphones et d'un auditeur placé entre les deux. Dans un monde idéal, les éléments électromécaniques de la chaîne ont tous des réponses plates et ne perturbent en rien la mesure. En pratique, ils introduisent de manière inévitable un ensemble d'artefacts qu'il convient de retirer après coup. C'est l'objet de l'*égalisation*.

Parmi les différentes stratégies possibles, la plus naturelle consiste à réitérer la mesure « à blanc » – c'est-à-dire sans auditeur – et à l'utiliser pour corriger la mesure réelle. Cette correction peut être effectuée en divisant cette dernière par la mesure à blanc dans le domaine fréquentiel. Cette méthode impose toutefois d'avoir à disposition la mesure corrective. Par ailleurs, les effets de couplage entre les microphones et les canaux auditifs ne sont pas concernés par cette égalisation. Plus précisément, l'effet de résonance du canal, souvent indésirable, n'est pas traité.

Alternativement, on peut faire choix d'une normalisation par rapport à la HRTF dans une direction de référence. On parle alors d'égalisation *en champ libre* [69, 44] et le jeu de HRTF à lui seul suffit à effectuer l'opération. Bien souvent, la direction $(\text{azimut}, \text{élévation}) = (0°, 0°)$ est utilisée. À la condition que tous les hauts-parleurs de mesure soient identiques, cette approche élimine leurs effets, ceux des microphones et l'éventuel couplage avec le canal. Néanmoins, elle présente l'inconvénient majeur de rendre parfaitement plate la HRTF dans la direction de référence, au mépris de tous les indices spectraux – cf. section 2.1.3.3 – qu'elle aurait pu contenir.

Ce problème peut toutefois se voir évacué en cessant de particulariser une direction de référence. C'est l'objet de l'égalisation *en champ diffus*. Dans celle-ci, la moyenne des HRTF dans toutes les directions est calculée puis retirée du jeu initial. De cette façon, les informations directionnelles sont préservées. Le calcul de cette moyenne doit cependant être réalisé avec soin. D'une part, il convient de prêter attention à la répartition spatiale de la grille d'évaluation car toute sur-représentation d'une direction ou d'une autre doit être prise en compte pour obtenir une moyenne correcte. À cette fin, la grille est pondérée par son diagramme de Voronoï – cf. figure 2.3.

D'autre part, la notion de *moyenne* est sujette à interprétation. Parmi les plus naturelles,



(a) (b)

FIGURE 2.3: *Côte à côte, une grille d'évaluation utilisée en mesures acoustiques (a) et son diagramme de Voronoï (b). Sur ce dernier, les cercles correspondent aux positions des sommets de la grille. Les couleurs du diagramme reflètent la taille relative de ses cellules, des plus petites en bleu au plus grandes en rouge. On observe que l'absence d'élévations basses, très fréquente pour les grilles d'évaluation acoustiques, force à donner un poids très important à la dernière couronne de sommets. Le haut de la sphère, en revanche, présente une forte densité de points qui oblige à en diminuer l'importance. Le plan azimutal, lui aussi légèrement plus dense que son voisinage immédiat, voit son importance relative réduite d'autant.*

bien sûr la moyenne quadratique :

$$H_{MQ} = \sqrt{\sum_i w_i \cdot |H_i|^2} \tag{2.6}$$

où les w_i désignent les poids issus du diagramme de Voronoï, mais les moyennes arithmétique ou géométrique sont également de possibles candidates.

2.1.2.4 Métriques

Dès lors que plusieurs jeux de HRTF sont disponibles, il est naturel de chercher à les comparer, à quantifier leurs ressemblances / dissemblances. À cette fin, plusieurs métriques ont à ce jour été proposées avec en premier lieu la famille des distances spectrales.

Très fréquemment utilisée [81, 78, 66], la distorsion spectrale est une mesure de la différence des amplitudes des HRTF exprimées en dB. Elle-même en décibel, elle est donnée par :

$$SD = \sqrt{\frac{1}{\Delta f} \int_{\Delta f} \left[|H_1|_{dB}(f) - |H_2|_{dB}(f) \right]^2 \cdot df} \tag{2.7}$$

où H_i, $i \in \{1, 2\}$ est une HRTF dans une direction donnée. La comparaison de jeux

entiers de HRTF nécessitera donc de calculer la moyenne des distorsions dans toutes les directions, pondérées par le diagramme de Voronoï, couvertes par ces jeux. Une version logarithmique de cette norme a été proposée par Huopaniemi [79]. Dans celle-ci, les fréquences sont échantillonnées selon une échelle logarithmique afin de s'approcher davantage de la perception humaine.

Autre métrique usuelle, l'*Inter-Subject Spectral Differences* – ou *ISSD* – initialement introduite par Middlebrooks [109] puis reprise avec de notables changements [69, 130]. Pour la définir, Middlebrooks part de deux jeux de DTF, dont il considère les amplitudes en dB, et d'un banc de 64 filtres. Ceux-ci sont triangulaires et équi-répartis selon une échelle logarithmique de fréquences couvrant la bande $[3\,700, 12\,900]$ Hz. Ces filtres sont ensuite appliqués à la différence des amplitudes des DTF, donnant ainsi 64 coefficients par direction. Pour chaque direction, la variance de ces coefficients est calculée et la moyenne de ces variances définit l'ISSD. Mathématiquement parlant, on peut l'écrire comme suit :

$$ISSD_{Middlebrooks} = \frac{1}{N}\sum_{k=0}^{N-1} \underset{i}{\mathrm{VAR}}(\mathrm{FILT}_i(|DTF_1|_{dB} - |DTF_2|_{dB})) \quad (2.8)$$

où $FILT_{i\in I}$ désigne la famille de filtres et N le nombre de directions. Son unité est le dB^2.

La version décrite par Guillon [69] diffère de celle de Middlebrooks en ce que le banc de filtres est laissé de côté. On a donc plus simplement :

$$ISSD_{Guillon} = \frac{1}{N}\sum_{k=0}^{N-1} \underset{D_f}{\mathrm{VAR}}(|DTF_1|_{dB} - |DTF_2|_{dB}) \quad (2.9)$$

où $D_f = [4\,000, 13\,000]$ Hz désigne le domaine d'étude fréquentiel. Une conséquence de la suppression de l'opération de filtrage est que le domaine d'échantillonnage fréquentiel redevient celui des DTF, c'est-à-dire une échelle linéaire dans la plupart des cas.

Ce dernier point n'ayant pas échappé à Rugeles [130], celui-ci propose à son tour une version de l'ISSD, baptisé *logISSD*, qui réintroduit l'évolution logarithmique de l'échelle fréquentielle dans la version de Guillon.

2.1.3 Indices de localisation

2.1.3.1 ITD

Déjà brièvement introduit section 1.1.1, l'ITD peut se décrire grossièrement comme la différence, pour un signal provenant d'une position (r, θ, ϕ) de l'espace, entre les temps d'arrivée à l'une et l'autre oreille. Indice de localisation fondamental s'il en est, la littérature fourmille d'études à son sujet. D'un fort impact pour la latéralisation des sources, il est vu comme l'indice principal dans le domaine basses fréquences. Le seuil fréquentiel généralement retenu pour admettre cette prédominance est $f_{max} = 1\,500$ Hz [155]. Au-delà, il sera plus fortement subordonné aux autres types d'indices. En sus de lui accorder un poids important, l'oreille humaine est aussi très sensible à ses variations. Ainsi, la *Just Noticeable Difference*[5]

5. JND

de l'ITD, c'est-à-dire sa plus petite variation détectable à l'oreille, est de l'ordre de quelques microsecondes. Carlile *et al.* [39] rapportent une JND de 6 μs, Mills [113] la mesure à 10 μs tandis que Plotgies *et al.* [123] parlent de 30 μs. Malgré tout, même dans ce dernier cas, cela ne représente qu'un décalage temporel d'à peine un échantillon et demi pour un signal capté à 48 kHz. Dans ces conditions, rien d'étonnant à ce que l'on puisse se pencher avec attention sur cet objet d'étude, à commencer par sa définition. Car si celle proposée en début de section peut se révéler utile pour la compréhension de la nature de l'ITD, elle laisse toutefois une large part d'interprétation au lecteur, notamment sur ce que doit être le moment d'arrivée d'un signal étendu dans le temps et comment l'évaluer au mieux. En conséquence, une multitude de méthodes de mesure, reposant sur autant d'acceptions différentes de la définition précédente, parsèment la littérature sans que l'une ou l'autre ne soit plus indiscutable que le reste.

Ainsi, la méthode de *seuillage* s'intéresse au premier dépassement d'une certaine valeur de référence par les valeurs absolues des HRIR à l'étude. Cette valeur peut s'exprimer tantôt en dB, tantôt en pourcentage du maximum des HRIR. Pour chaque direction, les temps d'arrivée à l'oreille droite et gauche sont mesurés et leur différence définit l'ITD dans cette direction. Un seuil trop haut risquerait de manquer le premier front montant – voire la HRIR dans son ensemble – tandis que trop bas, il risquerait d'être perturbé par l'éventuel bruit de mesure. À l'évidence, le domaine utile de valeurs seuil est donc déterminé par la dynamique et le rapport S / N des HRIR controlatérales. Raffinement supplémentaire, on peut suréchantillonner le signal pour accroître la précision de la mesure et / ou rendre le seuil dépendant de la direction pour se donner une plus grande marge de manœuvre, le principe sous-jacent restant inchangé. Dans ce dernier cas, on parlera de *seuillage adaptatif*. Pour se faire une idée des ordres de grandeur pratiqués, on peut se référer à l'étude de Sandvad & Hammershoï [133] portant sur les types de représentation des HRTF et dans laquelle les auteurs choisissent un seuil à 5 %, à l'article de Minnaar *et al.* [114] qui met en avant un seuil à 10 % mais aussi et surtout à l'étude des HRTF d'un modèle sphérique faite par Duda & Martens [51] dans laquelle, par comparaison de mesures expérimentales et de données théoriques, les auteurs déterminent qu'un seuil de 15 % permet d'obtenir de très bons résultats.

Sensiblement plus sophistiquées, les méthodes reposant sur le calcul de l'inter-corrélation sont également fréquemment utilisées [96, 103, 114]. On les verra souvent regroupées dans la famille *IACC*[6]. Mathématiquement, la IACC dans la direction (θ, ϕ), pour un décalage temporel τ, est donnée par :

$$IACC(\theta, \phi, \tau) \;=\; \frac{\int_{t_0}^{t_1} hl(\theta, \phi, t) \;.\; hr(\theta, \phi, t+\tau).dt}{\sqrt{\int_{t_0}^{t_1} hl^2(\theta, \phi, t).dt} \;.\; \sqrt{\int_{t_0}^{t_1} hr^2(\theta, \phi, t).dt}} \tag{2.10}$$

où hl – resp. hr – sigle pour la HRIR gauche – resp. droite. Des étapes additionnelles telles que le fenêtrage ou le suréchantillonnage sont autant d'options pour accroître la robustesse du calcul. Une première idée pour obtenir l'ITD est ici de retenir, pour chaque direction,

6. InterAural Cross-correlation

le décalage temporel assurant une inter-corrélation maximale des HRIR droite et gauche. Cette approche est connue sous le nom de *maxIACC* et définit l'ITD par :

$$ITD(\theta, \phi) = \underset{\tau}{\operatorname{argmax}}\ IACC(\theta, \phi, \tau) \tag{2.11}$$

Alternativement, il est possible d'effectuer le calcul de l'intercorrélation sur les enveloppes énergétiques des HRIR, c'est-à-dire sur les modules de leurs transformées de Hilbert. Cette procédure génère des signaux plus lisses, limitant ainsi l'obtention de discontinuités spatiales. En option supplémentaire, on peut faire choix de s'intéresser à la centroïde des signaux (*cenIACC*) plutôt qu'à leur maximum [47].

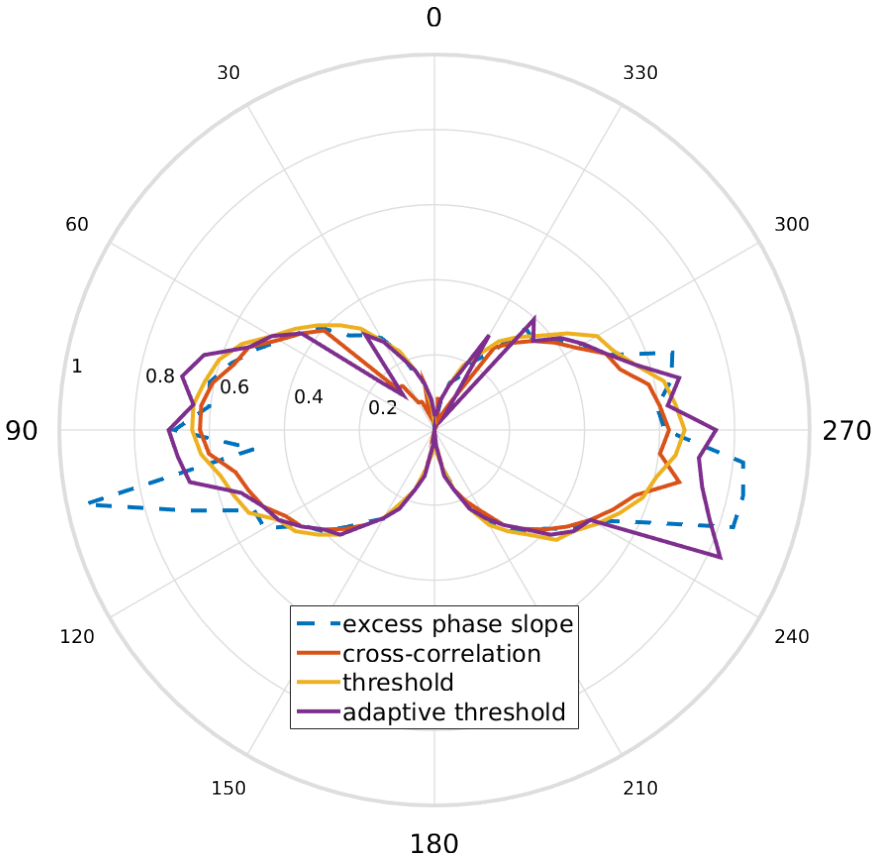


FIGURE 2.4: *Exemples d'estimations d'ITD, en* ms*, dans le plan azimutal pour un même jeu de HRTF acoustiques selon différentes méthodes. La distance au centre donne la mesure d'ITD, tandis que l'angle sur le pourtour représente l'angle d'incidence des ondes sonores. Malgré une cohérence générale (ainsi, l'ITD est bien nulle pour un angle de* 0° *ou* 180°*), un certain nombre de différences et d'artefacts apparaissent. Par ailleurs, ces variations vont généralement bien au-delà du seuil de détection de l'oreille humaine, ce qui les rend très audibles en pratique.*

En marge de ces méthodes entièrement fondées sur l'analyse temporelle, on en trouve d'autres reposant sur l'analyse des spectres fréquentiels, et plus particulièrement de leurs phases [87]. En effet, pour un signal en provenance du côté gauche d'un auditeur, le son se

propageant à l'oreille droite aura plus de distance à parcourir et verra sa phase augmenter à mesure qu'il chemine. En notant ϕ_l – resp. ϕ_r – la phase de la HRTF gauche H_l – resp. droite H_r –, on peut mathématiquement le traduire par :

$$ITD(f) = \frac{\phi_l(f) - \phi_r(f)}{2\pi f} \tag{2.12}$$

Cette formulation amène immédiatement à faire deux observations. La première est que l'ITD devient à priori dépendant de la fréquence. Il ne s'agit plus d'une valeur unique mais d'un ensemble variant selon notre position dans le spectre fréquentiel. Le cas particulier d'un ITD constant revient à représenter la différence de phase par une fonction affine. La seconde observation est que les phases étant définies modulo 2π, cette formule est à l'évidence inapplicable en l'état en hautes fréquences, un problème de repliement spectral apparaissant dès lors que la différence de marche entre les signaux droit et gauche excède la longueur d'onde. En notant Δx cette différence de marche, f la fréquence et c la célérité du milieu, ce domaine de validité fréquentielle est d'ailleurs régi par $\Delta x < \frac{c}{f}$. À titre d'exemple, pour un trajet aérien égal au demi-périmètre d'un cercle de diamètre 16 cm, on obtient $f < 1\,400\,\text{Hz}$[7].

Néanmoins, il est possible de lever cette limitation en s'intéressant non plus au *retard de phase* mais au *retard de groupe* :

$$ITD(f) = \frac{1}{2\pi} \cdot \frac{\partial\left(\phi_l - \phi_r\right)}{\partial f} = \frac{1}{2\pi} \cdot \frac{\partial \arg\left(H_l / H_r\right)}{\partial f} \tag{2.13}$$

Et de l'ITD instantané ainsi obtenu, on peut également tirer une valeur globale estimée à partir d'une bande de fréquences. Cette dernière est parfois appelée *Integrated Relative Group Delay* ou *IRGD*. C'est notamment ce à quoi se sont attachés Jot *et al.* [87] et Huopaniemi & Smith [80]. Dans chaque cas, le retard est déterminé en estimant la pente de la phase du filtre à excès de phase – cf. figure 2.2. Sur la bande $[1\,000, 5\,000]\,\text{Hz}$ pour Jot *et al.*, sur la bande $[500, 2\,000]\,\text{Hz}$ pour Huopaniemi & Smith. De leur côté, Plotgies *et al.* [123] adoptent une approche hybride des précédentes, liant seuillage et retard de groupe. Plus précisément, l'équivalent d'un seuillage est effectué sur les HRIR, donnant accès à un premier estimé de l'ITD. Ensuite, la représentation à minimum de phase est utilisée sur les HRIR résiduelles, c'est-à-dire les HRIR dont ont été retirés les échantillons précédant le dépassement du premier seuil. À partir des composantes à excès de phase ainsi obtenues, le retard de groupe à $0\,\text{Hz}$ est extrait et utilisé pour compenser le caractère approximatif de l'estimation par seuillage. Les auteurs ne précisent pas le ou les avantages de leur approche par rapport à celle de Jot ou Huopaniemi mais effectuent en revanche une série de tests subjectifs permettant de conclure en la validité de la représentation à minimum de phase, pour autant que l'ITD est correctement évalué.

Dans une étude comparative dédiée au sujet [92], Katz & Noisternig listent et traitent des performances respectives de neuf méthodes d'estimations d'ITD. Et bien qu'aucune

7. Il est d'ailleurs intéressant de remarquer que cette valeur correspond peu ou prou à la limite de prédominance de l'ITD comme facteur de localisation.

définition ne prévale véritablement sur une autre, et donc qu'aucune vérité terrain ne soit disponible en pratique, il demeure possible de comparer les variations obtenues en passant d'une méthode à une autre ou, au sein d'une même méthode, en passant d'un paramétrage à un autre. Leur observation majeure est que la variance mesurée de l'ITD dépasse parfois de plusieurs multiples l'éventail de valeurs de JND classiquement admis. Cette large variance implique que le choix de la méthode d'estimation n'est pas anodin, le passage de l'une à l'autre étant à priori nécessairement perçu par l'auditeur.

Parallèlement à ces approches expérimentales, nombre de chercheurs se sont penchés sur les modélisations possibles de l'ITD afin d'en fournir un modèle prédictif. En tête de file, on trouve bien sûr le fameux modèle de Woodworth, déjà évoque ci-avant – cf. équation 1.1 – et qui assimile la tête à une sphère sur laquelle les oreilles sont diamétralement opposées. Dans sa forme première, son ITD n'est donné que dans le plan azimutal, les valeurs dans les autres directions de l'espace devant s'en déduire par invariance par rotation selon l'axe interaural. Afin d'y remédier, Larcher & Jot [100] en proposèrent une formulation plus générale :

$$ITD_{Larcher}(\theta, \phi) = \frac{a}{c}\left(\arcsin(\sin\,\theta \cdot \cos\,\phi) + \sin\,\theta \cdot \cos\,\phi\right) \tag{2.14}$$

Celle-ci a ensuite été légèrement simplifiée par Savioja *et al.* [135] qui en présentent la variante suivante :

$$ITD_{Savioja}(\theta, \phi) = \frac{a}{c}\left(\sin\,\theta + \theta\right) \cdot \cos\,\phi \tag{2.15}$$

Dans chaque cas, pour $\phi = 0$, le modèle sphérique initial est bien retrouvé.

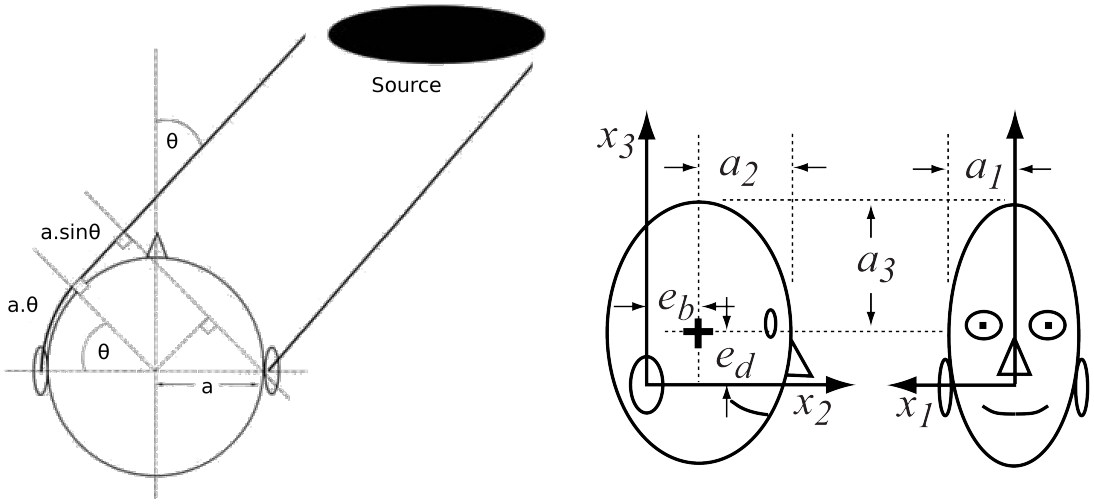
Néanmoins, ce modèle s'avère parfois trop fruste. Sur des mesures d'ITD réalisées à UC Davis, Larcher [99] montre en effet qu'il cesse d'être valable pour des sources latéralisées dès ±45° d'azimut. Plus particulièrement, une dépendance de l'ITD à l'angle d'élévation est observée, aplatissant sensiblement les cônes de confusion à mesure que l'on s'approche des oreilles. Et ce décalage entre prédictions et mesures atteint parfois 18 %, ce qui est considérable.

Afin de réduire l'erreur moyenne, Algazi *et al.* [1] proposent de remplacer le rayon de la sphère du modèle de Woodworth par un rayon optimal issu de différents paramètres morphologiques, à savoir les demies largeur, profondeur et hauteur de la tête du sujet – resp. a_l, a_p et a_h –, données en cm :

$$a_{opt} = 0,51 \cdot a_l + 0,019 \cdot a_p + 0,18 \cdot a_h + 3,2 \tag{2.16}$$

Dans cette expression, les différents coefficients sont obtenus par régression linéaire à partir des données anthropométriques et des HRIR de 25 sujets et l'on peut remarquer que la profondeur de la tête a_p semble revêtir une importance anecdotique vis-à-vis des autres dimensions. À noter également l'avertissement des auteurs, qui font le constat qu'utiliser la moyenne des a_i s'avère en pratique un très mauvais choix. Malgré tout, cette optimisation n'intègre pas de dépendance en élévation au modèle sous-jascent qui demeure sphérique.

Prenant à bras le corps ce problème, le modèle de Duda *et al.* [50] – cf. figure 2.5 – intègre le caractère plus ellipsoïdal que sphérique de la tête et le positionnement propre des oreilles, se posant ainsi en alternative significativement plus complète. Certes, le nombre de paramètres passe de un à cinq, mais l'erreur moyenne d'estimation par rapport aux ITDs obtenus par seuillage à 15 % passe de 30 μs à 10 μs. Toutefois, il est important de noter que ce gain n'a été rendu possible que par une optimisation des paramètres du modèle et non par leur simple mesure *in situ*. Il s'agit donc du meilleur résultat possible à tirer du modèle tout en connaissant la cible. En rapprochant cela de la difficulté à dégager une mesure d'ITD pouvant faire autorité, on comprend que les valeurs numériques en elles-mêmes sont à considérer avec précaution et sont plus à même de souligner une amélioration qualitative que quantitative.



(a) (b)

FIGURE 2.5: *Côte à côte, le modèle sphérique de Woodworth (a) et le modèle ellipsoïdal de Duda (b).*

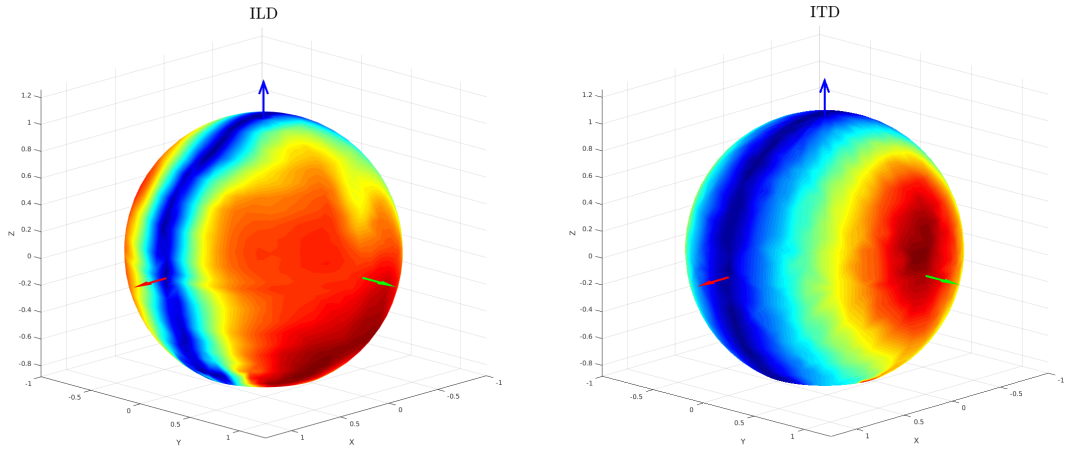
Cependant, même sans avoir de lien parfait avec la morphologie, la donnée d'un modèle d'ITD représentant de manière fiable les variations directionnelles peut être mise à profit. En particulier, dès qu'il s'agit de pallier le manque de robustesse de certaines méthodes de mesure directe, la possibilité de remplacer les discontinuités et valeurs aberrantes par les prédictions du modèle s'avère très utile.

2.1.3.2 ILD

Autre indice majeur de localisation, la différence des niveaux sonores arrivant jusqu'aux oreilles. Mieux connu sous l'acronyme d'ILD ou parfois IID, pour *Interaural Intensity Difference*, cet indice est déjà présent dans la *Duplex Theory* de Lord Rayleigh, qui lui donnait une importance équivalente à l'ITD.

Cet ILD, tout comme l'ITD avant lui, évoluera en fonction de la position relative de la source par rapport au couple d'oreilles. Un son émis à proximité immédiate de l'oreille gauche y sera naturellement perçu plus fort qu'à l'oreille droite. Non seulement celle-ci

est plus éloignée mais elle est également masquée par la présence de la tête. Les mêmes phénomènes étant à l'œuvre pour l'ILD que pour l'ITD, rien d'anormal à ce que leurs variations suivent des évolutions semblables – cf. figure 2.6. En particulier, il est important de noter que le plan médian est à la fois un plan iso-ITD et iso-ILD. De plus, les valeurs absolues des mesures augmentent avec la latéralisation de la source.



(a) (b)

FIGURE 2.6: *Représentations des valeurs absolues des ILDs (a) et ITDs (b) acoustiques d'un sujet mesurées en champ lointain. Le plan médian constitue dans les deux cas une zone de minimum (en bleue), les valeurs maximales (en rouge) étant quant à elles situées à proximité des oreilles. Et si l'ITD croit de manière très concentrique à mesure que l'on s'en approche, donnant tout son sens à l'appellation « cône de confusion », cette régularité se perd quelque peu dans le cas de l'ILD.*

Cet indice voit son domaine de prédominance commencer après celui de l'ITD, c'est-à-dire au-delà de $f_{max} = 1\,500\,\text{Hz}$, ce que deux phénomènes physiques principaux permettent de bien comprendre. Tout d'abord, l'ITD doit faire face à d'inévitables problèmes de recouvrement fréquentiel et perd donc immanquablement en poids au-delà de f_{max}. Mais en parallèle, plus la fréquence augmente et plus la longueur d'onde devient du même ordre ou plus petite que les dimensions caractéristiques de la tête. Celle-ci représente alors un véritable obstacle à la propagation des ondes sonores moyennes et hautes fréquences. La différence d'intensité perçue entre les deux oreilles augmente donc mécaniquement avec la fréquence.

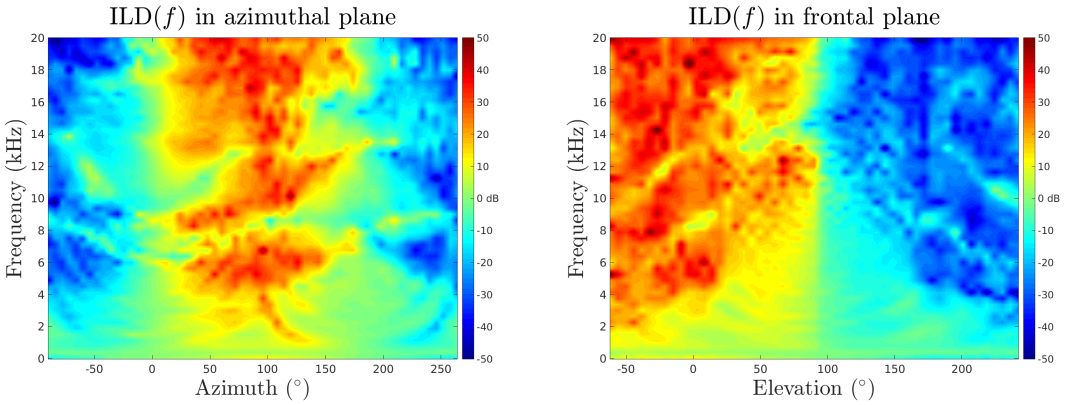
Du point de vue mathématique, l'ILD dans une direction et à une distance données peut aisément s'exprimer à partir du rapport des HRTF gauche et droite :

$$ILD(f) \;=\; 20 \cdot \log\left(\left|\frac{H_l(f)}{H_r(f)}\right|\right) \tag{2.17}$$

La figure 2.7 présente l'ILD fréquentiel dans les plans azimutal et frontal [8] d'une HRTF

8. C'est-à-dire le plan vertical passant par les azimuts $\pm 90°$

acoustique. On constate que les variations d'amplitude peuvent atteindre de très larges valeurs, mais uniquement à partir d'un ou deux kilohertz. En basses fréquences, l'ILD est à peu près le même dans toutes les directions et n'y représente à l'évidence pas un indice de localisation. En moyennes et hautes fréquences, de nombreux motifs apparaissent et la mesure de l'ILD dans une direction est de fait rendue délicate.



(a) (b)

FIGURE 2.7: *ILDs fréquentiels d'un sujet mesurés dans les plans azimutal (a) et frontal (b).*

À cette définition d'un ILD variant en fréquence, on peut donc préférer celle d'un ILD moyenné sur une bande de fréquences, mesure plus synthétique et moins impactée par ces motifs fréquentiels :

$$ILD \;=\; 10 \cdot \log\left(\frac{\int |H_l(f)|^2 \cdot df}{\int |H_r(f)|^2 \cdot df}\right) \tag{2.18}$$

Néanmoins, l'ILD et l'ITD seuls ne suffisent à expliquer la localisation auditive dans toutes les directions de l'espace. En témoigne le plan médian, sur lequel ILD et ITD sont tout deux nuls, sans que cela ne nous empêche de nous y repérer. Certes, les performances y sont moins bonnes – Damaske [46] relève une incertitude de localisation allant jusqu'à 20° pour une source placée au zénith alors que les expériences menées par Mills [113] montrent une précision de l'ordre de 2° pour une source placée face à l'auditeur, dans le plan azimutal – mais tout de même significatives. Pour cette raison, l'attention s'est ensuite portée sur les variations de spectre en elles-mêmes.

2.1.3.3 Signature spectrale

Il faut remonter jusqu'en 1959, et les travaux de McLean [107] pour trouver la première mention du lien entre la morphologie du pavillon d'oreille et la faculté de localisation. Plus précisément, il y montra qu'une simple torsion des oreilles perturbait la sensation de localisation des sons.

La décennie suivante connut alors une forte activité scientifique autour de cette thématique. Ainsi, Batteau [7] théorisa un processus de transformations et d'analyse des

signaux auditifs par le cerveau dans lequel les pavillons tiennent un rôle de premier ordre et sont à l'origine d'une signature directionnelle. En parallèle, Teranishi & Shaw [144] étudièrent également le lien entre les variations de spectre et la forme des oreilles. Sans pour autant émettre d'hypothèse quant au lien entre ces variations et les capacités humaines de localisation, ils proposèrent un modèle simple d'oreille externe donnant un spectre réaliste jusqu'à 7 kHz.

À la fin de années 1960, Roffler & Butler [128] et Blauert [18] menèrent plusieurs séries d'expériences visant à étudier les facultés humaines de localisation dans le plan médian, c'est-à-dire dans des conditions d'iso-ITD et d'iso-ILD. En jouant à un panel de volontaires des bruits blancs d'une largeur de bande d'un tiers d'octave, Blauert constata ainsi que l'origine ressentie des sons était fortement liée à la fréquence centrale du signal. Plus précisément, ces sons pouvaient être perçus, de façon répétable et statistiquement significative, à une position très différente de celle du haut-parleur les jouant. Celui-ci pouvait être placé soit directement en face de l'auditeur, soit à la position diamétralement opposée. La figure 2.8 décrit les zones spatiales retenues dans ces expériences et illustre le type de résultats obtenus. De ces résultats ressort l'importance du contenu spectral du signal dans la perception que l'on peut en avoir, ouvrant le champ à l'étude d'un nouveau type d'indice de localisation.

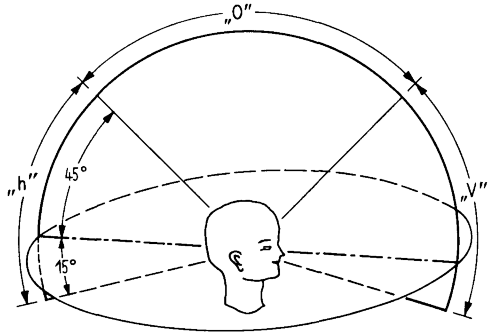


Fig. 3. Nominal scale of the observers to describe the direction of the sound sensations.

(a)

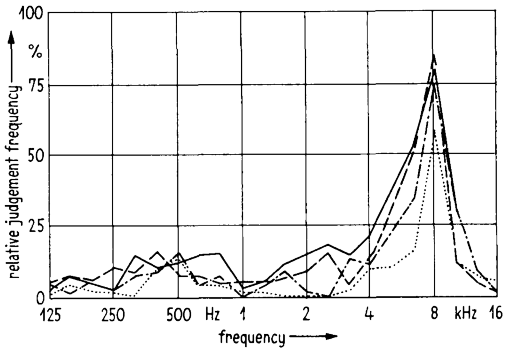


Fig. 6. Relative frequency of "o"-judgements (20 observers, each stimulated once from the front, once from the rear, in each of the 1/3 octave bands).
——— 60 dB,
– – – 50 dB,
–·—·– 40 dB,
····· 30 dB.

(b)

FIGURE 2.8: *Extraits de l'expérience de localisation de Blauert. La figure (a) présente les différentes zones de localisation possible, à savoir « devant », « en haut » ou « derrière ». La figure (b) présente les proportions d'apparition du jugement « en haut » relativement à la fréquence centrale du signal. On observe une nette prédominance de ce type de ressenti pour des sons de fréquence proche de 8 kHz.*

Ces idées, maintes fois reprises depuis lors [74, 119, 36, 90], ont cristallisé l'importance de la morphologie de l'auditeur et amené la communauté à se pencher davantage sur la partie haute du spectre.

De ces travaux, plusieurs éléments d'information essentiels peuvent être retenus. Tout d'abord, la présence de l'oreille se traduit par une signature fréquentielle et directionnelle propre à chacun. On évoquera fréquemment les « peaks » et les « notchs » des HRTF, c'est-à-dire les zones de résonances et d'anti-résonances du spectre, comme une signature propre de la morphologie. La figure 2.9 en présente des exemples dans plusieurs plans de coupe. De plus, cette influence de l'oreille n'apparaît qu'au-delà d'un certain seuil fréquentiel. Par utilisation de la simulation numérique, Katz *et al.* [90] ont mis en évidence que la contribution de l'oreille aux variations d'amplitude du spectre commençait à se faire sentir à partir de 3,5 kHz. En outre, ces signatures spectrales sont un indice monaural. Les expériences de localisation de Butler *et al.* [35, 36] en condition monaural – ie avec une oreille bouchée – ou celles, plus parlantes encore, de Hausler *et al.* [73] sur des populations naturellement sourdes d'une oreille, montrent qu'une seule suffit pour disposer d'une capacité de localisation normale en élévation. Dans cette dernière étude, il est également intéressant de noter que les résultats en azimut étaient, pour leur part, comparables à des sorties de tirage aléatoire.

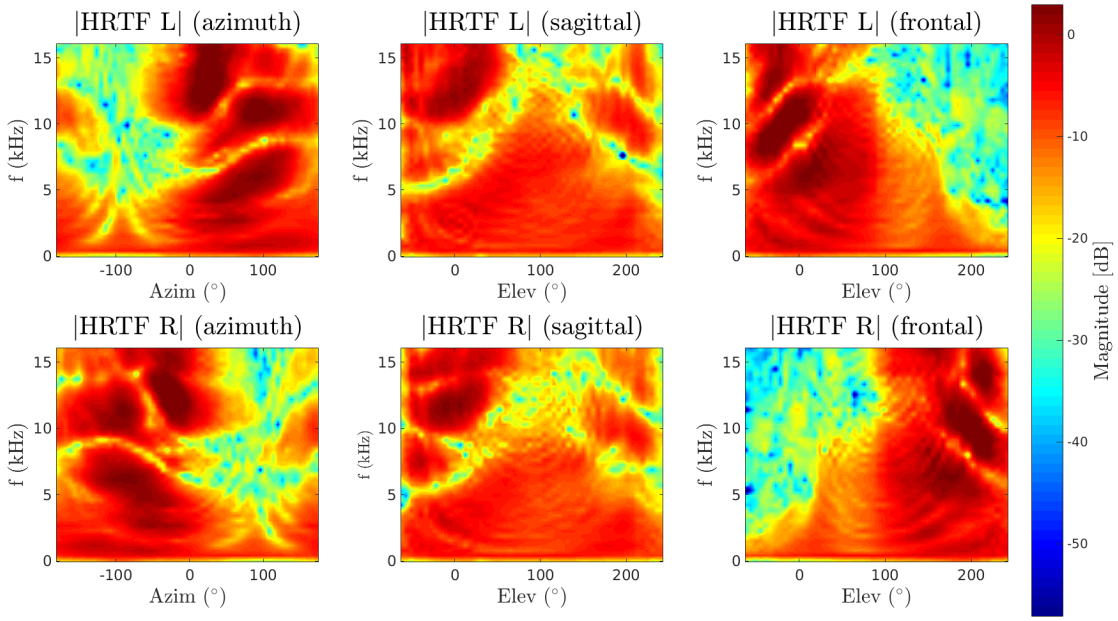


FIGURE 2.9: *De gauche à droite, coupes dans les plans azimutal, sagittal et frontal des HRTF acoustiques gauche (en haut) et droite (en bas) d'un même sujet. Parmi les observations les plus immédiates, la symétrie manifeste des deux jeux – par rapport au plan sagittal – et la présence, en position controlatérale, d'une zone légèrement plus énergétique, assimilable au* bright spot.

Par ailleurs, parler ainsi de « signature spectrale » souligne implicitement la répartition de l'information sur un ensemble de fréquences, ce que confirme la difficulté à localiser des sons de contenu spectral pauvre. Il a de nombreuses fois été rapporté [128, 59, 35] la nécessité de disposer de cette richesse spectrale pour des performances optimales, surtout en élévation. King & Oldfield [95] rapportent même une largeur de bande de 13 kHz afin d'obtenir des performances de localisation optimales en élévation. De leur côté, Hebrank & Wright [158]

concluent que les indices spectraux en élévation appartiennent à la bande $[4, 16]$ kHz. En azimut, la présence de l'ITD et de l'ILD rend cette signature moins indispensable mais toujours utile. Elle est alors plus une aide à la réduction du nombre de confusions avant-arrière.

Enfin, ce tour d'horizon des indices de localisation amène à une réflexion sur leurs représentations, en particulier fréquentielle. En effet, la sensibilité de l'oreille humaine suit une progression logarithmique, accordant plus d'importance et une meilleure sensibilité aux basses fréquences qu'aux hautes fréquences et il n'est ainsi pas rare de retrouver des représentations de HRTF selon une échelle fréquentielle également logarithmique. Ce choix est néanmoins discutable si l'on se réfère aux domaines de prédominance de chaque indice. Plus précisément, la localisation en basses fréquences est principalement gérée par l'ITD, lui-même souvent résumé à une valeur par direction. La signature spectrale n'étant en revanche un indice qu'aux moyennes et hautes fréquences, l'adoption d'une échelle log crée un effet de loupe sur une zone dépourvue d'information critique tout en comprimant les autres. Pour cette raison, et sauf mention contraire, les figures de HRTF présentées dans ce manuscrit viennent avec une échelle linéaire.

2.2 Stratégies d'obtention : circuits « directs »

Les caractéristiques et usages des HRTF personnalisées étant établies, il reste à savoir comment en obtenir en pratique. Et en premier lieu de façon directe. Pour cela, deux approches s'affrontent et se complètent : la mesure, qui cherche l'information *in situ*, et la simulation numérique, qui s'attache à modéliser l'expérience de mesure elle-même.

2.2.1 La mesure acoustique

L'idée de base de la mesure acoustique est assez simple : il s'agit d'aller capter le signal à l'endroit où il se trouve, c'est-à-dire au niveau du canal auditif du sujet. La réalisation pratique, en revanche, nécessite de l'investissement matériel et beaucoup de précautions.

2.2.1.1 Des locaux adaptés

Tout d'abord, il faut pouvoir réaliser les mesures dans un lieu exempt de toute pollution sonore. Concrètement, une chambre anéchoïque acoustique [9], également appelée chambre sourde, est un bon candidat. Comme son nom l'indique, celle-ci ne doit pas produire d'échos acoustiques car son objectif est de reproduire les conditions d'écoute en champ libre. À cette fin, ces pièces sont tapissées de matériaux poreux (mousse polymère, fibres de verre) absorbant les ondes sonores. Si le sol n'en est pas pourvu, on parlera de chambre semi-anéchoïque. Elle doit être suffisamment grande pour contenir le système de haut-parleurs destiné à produire les signaux à mesurer, mais aussi pour assurer la validité des mesures en basses fréquences. En effet, à chaque pièce est associée une fréquence de coupure au-delà de laquelle le coefficient d'absorption des parois est supérieur à 99 %. En-deçà, l'approximation

9. À noter que le même type de dispositif existe à l'attention des ondes électromagnétique. On parle alors de chambre anéchoïque... électromagnétique.

« champ libre » n'est plus valide. Cette fréquence dépendra essentiellement des dimensions mais aussi du revêtement utilisé et sa valeur doit donc être mesurée au cas-par-cas. L'ordre de grandeur classique en est de quelques dizaines de hertz pour une chambre cubique de 10m de côté.

2.2.1.2 Haut-parleurs

Vient ensuite le système de production sonore. Il doit être capable d'émettre depuis une large variété de positions et sur une large gamme de fréquences. Dans un design souvent observé, de nombreuses enceintes se trouvent fixées à un arceau. Ce dernier est éventuellement rendu mobile, selon que l'installation dispose ou non d'un plateau rotatif capable de faire tourner le sujet. Les enceintes acoustiques venant avec une caisse de résonance d'autant plus grande qu'elle peuvent produire de basses fréquences, il faut un arceau d'autant plus grand que l'on souhaite diminuer l'écart angulaire entre deux points de mesure adjacents.

2.2.1.3 Enregistrement

Concernant le système de captation, il consiste en une paire de microphones installés au niveau des canaux auditifs du sujet, lui-même situé au centre du système d'enceintes. Selon les spécificités de la méthode, on cherchera ou non à boucher préalablement le canal auditif. Dans les faits, la mesure du champ sonore à l'entrée d'un canal ouvert est entaché d'un effet de couplage dû à l'onde réfléchie sur le tympan. Cela se traduit par une baisse de la puissance sonore omnidirectionnelle aux alentours de 3 kHz, ce qui est cohérent avec la profondeur usuelle du canal, de l'ordre de 2,5 cm. Dans le cas de dispositifs de mesure placés dans le canal [124, 76], au plus près du tympan, cette précaution n'est plus requise. Toutefois, cette proximité avec le tympan rend cette méthode plus risquée pour le sujet et d'autres précautions deviennent alors nécessaires, appelant parfois le concours d'un professionnel de santé. Par ailleurs, l'encombrement du système n'est pas négligeable dans un espace aussi confiné et se trouve de nature à perturber les mesures. Ainsi, en réalisant des mesures d'HPTF [10] avec et sans microphone intra-auriculaire, Pralong *et al.* [124] montrent, sans s'attarder dessus, une atténuation du spectre entre 11 kHz et 14 kHz pouvant aller jusqu'à 20 dB. Pour toutes ces raisons, la mesure de HRTF canal bouché est souvent considérée comme le meilleur choix possible [72, 126].

Suivant l'installation et les techniques adoptées, le processus de mesure acoustique dure de quelques dizaines de minutes et plus de 2 h. Dans ces conditions, on ne peut attendre du sujet qu'il reste debout, immobile. Pour autant, même assis, il n'est pas à l'abri d'effectuer des mouvements de tête, conscients ou non, qui invalideraient la prise de son. Ce cas de figure est facilement décelable, notamment par la présence de discontinuités dans l'ITD. C'est pourquoi on pourra parfois adjoindre un système de fixation de la tête et/ou un système de « tracking » en charge de la détection et de la correction des éventuelles dérives.

10. HeadPhone Transfer Function

2.2.1.4 Signal excitateur

Pour ce qui est du signal d'entrée à choisir, plusieurs options sont possibles. Si le Dirac est celle qui permettrait un temps de mesure minimal pour un post-traitement réduit, l'impossibilité à la réaliser physiquement force à se tourner vers des solutions alternatives.

Parmi les nombreuses voies envisagées, citons en premier lieu l'excitation impulsionnelle périodique – ou *Periodic Impulse Excitation* (PIE) – proposée par Berman *et al.* [14]. Peu coûteux à mettre en place, il utilise une série d'impulsions comme signal d'entrée. Le principe est d'approximer au mieux un Dirac puis d'améliorer le rapport signal à bruit de la réponse en multipliant les mesures et en moyennant les réponses. Néanmoins, toute non-linéarité dans la chaîne de mesure [11] induit des distorsions dans la réponse que le calcul d'une moyenne ne permet pas de corriger.

Dans une autre catégorie, on trouve la famille des séquences binaires périodiques pseudo-aléatoires et son représentant le plus connu, le *Maximum-Length Sequence* (MLS) [127]. L'idée de base est de présenter cette séquence en entrée du système linéaire à étudier, en échantillonner la réponse et étudier l'inter-corrélation de cette dernière avec la séquence d'entrée. On obtient ainsi la réponse impulsionnelle du système. Proposant un ratio S / N plus avantageux, elle n'est pas forcément la plus adaptée au problème. Demandant par exemple d'utiliser une trame de $2^n - 1$ échantillons, elle est une « mauvaise » candidate pour la très classique FFT (Fast Fourier Transform). Permettant de pallier ce dernier point, on trouve les codes Golay, qui présentent par nature une taille de 2^n échantillons. Ceci étant, il a été montré que leur utilisation dans le cadre de mesures de HRTF pouvait facilement introduire des artefacts en cas de mouvement de tête [164].

À côté de cette famille s'en trouve une autre, celle des *Time-Stretch Pulse* (TSP). Remontant au moins au début des années 1980 et aux travaux de Berkhout [13] et d'Aoshima [4], elle rassemble une série de méthodes dont le point commun est de tenter « d'étirer » temporellement un Dirac, d'effectuer la mesure acoustique voulue, puis de « compresser » correctement la réponse pour retrouver le résultat qu'aurait produit un véritable Dirac. Le plus souvent, le signal excitateur prend la forme d'une rampe de fréquence et la phase de compression s'opère par simple convolution de la mesure avec la rampe inverse. De nombreuses variantes en ont été proposées [57, 118, 142, 55, 104]. Parmi les membres qu'elle compte en son sein, citons l'Optimised Aoshima's TSP (OATSP) de Suzuki *et al.* [142], utilisée pour la constitution de la base de données RIEC (cf. section 2.4.1) et la rampe de fréquence exponentielle ou *Exponential Sweep* (ES), décrite par Farina [55]. De fort rapport S / N, cette dernière technique présente l'avantage de pouvoir être parallélisée, donnant naissance à la MESM – *Multiple Exponential Sweep Method* –, proposée par Majdak *et al.* [104] puis elle-même optimisée par Dietrich [48]. L'idée est de tirer parti de la robustesse au bruit de l'ES pour commencer à jouer une nouvelle rampe avant la fin de la dernière, diminuant de fait la durée de l'opération.

Pour finir ce tour d'horizon des différentes possibilités, citons encore le filtre adaptatif *Least Mean Square* (LMS) et ses dérivés (NLMS, VSNLMS, MVSS...). Filtres de la famille des descentes de gradient et notamment présents dans les algorithmes d'annulation

11. Telles celles inhérentes aux haut-parleurs.

d'échos [12], Enzer a le premier proposé leur utilisation dans le cadre des mesures de HRTF [54, 53]. La particularité du système est qu'il demande à jouer un signal, tel un bruit blanc, de façon continue pendant que l'enceinte tourne tout autour du sujet. De cette manière, la prise de mesure voit sa durée considérablement amoindrie. Une comparaison détaillée des différentes variantes est présentée par Correa *et al.* [45]

2.2.2 Les simulations numériques

Développée dans le cadre du binaural depuis près d'un quart de siècle [30, 91, 88, 32], la famille des méthodes aux éléments finis vise à modéliser puis résoudre le problème aux dérivées partielles posé par la propagation du son de la source aux tympans du sujet, résumé en l'équation d'Helmholtz, la condition de Sommerfeld et les conditions aux limites choisies.

Elles offrent de nombreux intérêts pratiques comparativement aux méthodes de mesure, à commencer par un contrôle complet des conditions expérimentales. Il n'est pas rare en effet qu'une mesure acoustique soit invalidée par un mouvement indésirable du sujet, un micro ne tenant pas en place, un bruit intempestif... Autant de problèmes qui n'en sont plus en simulation. De la même manière, le sujet ne risque pas d'être blessé par la pose du micro, les enceintes ne présentent plus de non-linéarité, leur nombre peut être arbitrairement grand et leurs positions modifiées à l'envie.

2.2.2.1 Formulation du problème

Mais avant d'aller plus avant, il convient de rappeler la base du problème à résoudre. Supposons donc donné un individu dont on note S la surface extérieure (peau, chevelure, vêtements...) et un repère orthonormal centré en sa tête. Supposons également présente une source sonore générant un champ acoustique incident Φ^i. Le calcul de la HRTF de l'individu dans la direction de la source et pour la fréquence $f = \frac{\omega}{2\pi}$ demande à déterminer la valeur du champ acoustique $\Phi = \Phi^i + \Phi^{ref}$ au niveau des oreilles de l'individu, où Φ^{ref} représente le champ acoustique réfléchi.

Φ^i étant laissé au choix de l'expérimentateur [12], il ne reste plus qu'à déterminer Φ^{ref}, ce que permet la résolution de l'équation d'Helmholtz :

$$\nabla^2\Phi^{ref} + k^2\Phi^{ref} = 0, \qquad k > 0 \tag{2.19}$$

où k représente le nombre d'onde. Il s'agit d'un cas particulier de l'équation d'onde de d'Alembert apparaissant lorsque l'on en recherche les solutions stationnaires, ou modes propres.

À celle-ci, il convient d'associer la condition de rayonnement de Sommerfeld :

$$\lim_{r\to\infty}\left[r\left(\frac{\partial\Phi^{ref}}{\partial r} - ik\Phi^{ref}\right)\right] = 0 \tag{2.20}$$

qui exprime en substance le fait que l'énergie rayonnée par les sources se disperse à l'infini.

12. Traditionnellement une onde plane ou sphérique.

Enfin, restent à définir les conditions aux limites, typiquement :

$$\left.\frac{\partial \Phi^{ref}}{\partial n}\right|_S = -\left.\frac{\partial \Phi^{i}}{\partial n}\right|_S \tag{2.21}$$

pour une surface S rigide.

Bien que n'ayant pas de solution analytique dans le cas général, le problème précédent est toutefois assuré d'avoir une solution, et qui plus est unique. En répétant l'opération pour toutes les fréquences et directions souhaitées, on obtient le jeu de HRTF complet de l'individu.

2.2.2.2 Optimisations

Un point noir de toutes les méthodes numériques réside dans le temps de calcul. Il est en effet dépendant de la résolution fréquentielle choisie, devenant rapidement non-négligeable à mesure que l'on s'aventure dans les hautes fréquences, mais aussi de la résolution spatiale voulue, c'est-à-dire du nombre de directions mesurées. Il faut dire que la précédente approche est très inefficiente. Pour s'en convaincre, on peut remarquer qu'elle apporte une solution pour tous les points du domaine d'étude alors que deux seulement – les tympans – nous intéressent, laissant augurer d'un nombre important de calculs superflus. Pour cette raison, diverses reformulations du problème ont été proposées, livrant tout autant de méthodes, telles la *Direct Boundary Element Method* (DBEM), la *Indirect Boundary Element Method* (IBEM), la *Infinite-Finite Element Method* (IFEM) ou encore la *Fast-Multipole Boundary Element Method* (FM-BEM).

Parmi les optimisations notables, citons également le *principe de réciprocité*, qui permet d'intervertir les rôles des microphones et des enceintes, constituant un moyen élégant de s'affranchir de la dépendance spatiale. On passe ainsi de plusieurs centaines – voire milliers – de problèmes à résoudre, à seulement deux. Valable aussi bien lors de mesures de HRTF que lors de calculs, ce n'est que dans ce dernier cas qu'il peut être pleinement utilisé. À l'évidence, placer des enceintes à l'intérieur de canaux auditifs réels pose un problème de taille, l'utilisation de haut-parleurs miniatures forçant à se passer de la mesure de la partie basse du spectre, qu'il faut donc combler par d'autres moyens [170]. Par ailleurs, la puissance maximale émise doit être drastiquement contrôlée de sorte que l'ouïe du sujet ne s'en trouve pas détériorée, diminuant d'autant le rapport S / N en sortie. En simulation en revanche, ces considérations disparaissent.

2.2.2.3 Moteurs de calcul

Enfin, la mise en pratique des méthodes numériques étant un travail complexe à lui seul, cette introduction ne saurait s'achever sans un tour d'horizon des solutions logicielles actuellement disponibles. Divers programmes, libres ou propriétaires, dédiés au calcul de HRTF ou non, ont en effet été développés au fil des années et sont aujourd'hui disponibles.

Parmi les solutions open-source, citons tout d'abord *OpenBEM*, qui se définit comme une implémentation à visée acoustique de la BEM. Développée au Danemark par une équipe de l'université DTU, elle est écrite en Matlab et traite aussi bien les problèmes 2D que 3D.

Toutefois, elle n'est documentée que par le code lui-même, les exemples afférents [13] et les publications de ses auteurs [75]. Par ailleurs, de l'aveu même de ces derniers, OpenBEM n'est pas « user-friendly » car en constante évolution.

Plus récemment, les membres de l'*Acoustics Research Institute* (ARI) ont mis à disposition une autre alternative : *mesh2hrtf* [167, 168]. Elle entend proposer un pack logiciel simple d'utilisation pour le calcul de HRTF. Écrite en Python, C et Matlab, elle propose de résoudre par BEM la formulation de Burton-Miller [31] du problème à l'étude. En particulier, elle intègre les méthodes d'accélération telles que la *Fast Multipole Method* (FMM) et le principe de réciprocité et permet également de sortir des HRTF directement au format SOFA. De plus, un wiki détaillant la structure du projet, les écueils les plus fréquents et des tutoriels est disponible.

Parmi les solutions payantes, et sans chercher à en établir une liste exhaustive, on trouvera certains produits spécifiquement conçus pour la résolution de problèmes acoustiques, tel *Coustyx* – édité par ANSOL – ou *VA One* – édité par ESI. Chacun propose une résolution par BEM pouvant intégrer la FMM. Néanmoins, seul VA One supporte également la résolution par FEM, laissant la place à des modélisations plus complexes des propriétés physiques des éléments à simuler. Enfin, comme il est de rigueur pour une solution commerciale, un support client est disponible.

Pour finir, nombre de logiciels offrent de résoudre, de manière plus générale, des problèmes par BEM et / ou FEM, les rendant à la fois plus flexibles mais potentiellement plus ardus à prendre en main. *COMSOL multiphysics* – édité par COMSOL –, *MD Nastran* – édité par mscsoftware – ou *MyBEM* – issu des travaux du laboratoire CMAP de l'école Polytechnique [5] – en sont quelques exemples.

La comparaison de tous ces produits dépassant bien évidemment, et de loin, le cadre de cette recherche, nous nous en tiendrons là. Cependant, on pourra observer et regretter qu'en dépit de la multiplicité des offres, les études comparatives sur le sujet, pourtant riches d'enseignement, sont rares [14].

2.3 Stratégies d'obtention : circuits « indirects »

Parmi les stratégies indirectes de personnalisation, la littérature a vu émerger de nombreuses méthodes que l'on peut classer en deux grandes familles : Les *méthodes de synthèse*, qui s'attachent à calculer ou recréer des jeux de HRTF, et les *méthodes adaptatives*, qui cherchent à découvrir, parmi un ensemble donné et au prix éventuel de transformations mineures, les jeux les plus adaptés à un individu.

13. Dont des cas acoustiques simples venant avec une solution analytique.
14. Le lecteur intéressé pourra se pencher sur l'étude menée par Molares *et al.* [116] et comparant COMSOL, VA One et MD Nastran.

2.3.1 Choisir au sein d'une collection

2.3.1.1 Critère morphologique

Remettant au premier plan l'importance de la morphologie propre à chacun, Zotkin *et al.* [171] décrivent l'oreille au travers de 7 paramètres morphologiques mesurables sur une vue de profil de l'oreille. Ces paramètres permettent alors de définir une distance entre les individus qui est utilisée pour sélectionner le plus proche voisin dans la base CIPIC d'un sujet donné. On remarquera que les HRTF ainsi sélectionnées on ensuite fait l'objet d'une modification pour les fréquences inférieures à 3 kHz. En effet, pour les basses fréquences – $f \leq 500\,\text{Hz}$ –, un modèle HAT est utilisé pour synthétiser les HRTF. Entre 500 Hz et 3 kHz, un recollement affine est opéré pour passer progressivement des HRTF de synthèse aux HRTF sélectionnées.

En 2001, la société Arkamys a déposé un brevet [94] sur une méthode de sélection morphologique. L'idée est de constituer trois BDD. La première contient les HRTF d'un ensemble d'individus, la seconde un jeu de paramètres morphologiques de ces individus et la troisième les préférences d'écoute de ces individus, i.e. pour chaque sujet, la classification qu'il fait des HRTF de la première base. Une fois cela posé, une étude des corrélations entre la 2$^{\text{e}}$ et la 3$^{\text{e}}$ BDD est réalisée pour classer les paramètres morphologiques par ordre d'importance. Du côté des HRTF, une analyse dimensionnelle de l'espace est menée – par exemple une *Analyse en Composantes Principales* (ACP) – pour en obtenir une base dans laquelle elles deviennent représentables. Chaque HRTF est alors donnée par son jeu de coordonnées dans cette base. Les liens entre les K paramètres morphologiques les plus importants et les coordonnées des HRTF dans l'espace précité sont alors calculés, établissant un pont entre morphologie et HRTF. Pour un nouvel individu, la mesure des K paramètres morphologiques mis en lumière précédemment permet de se positionner dans l'espace des HRTF. Là, le plus proche voisin présent en base est recherché et constitue le résultat de la personnalisation.

On retrouve ici le problème rencontré par les précédentes méthodes utilisant des paramètres morphologiques, à savoir : comment définir le nombre et la localisation de ceux-ci ? Par ailleurs, se pose la question des critères de définition de la distance mathématique utilisée car le résultat de la sélection dépend de cette dernière.

2.3.1.2 Adaptation

Enfin viennent les méthodes de sélection adaptée, dont le représentant le plus parlant est sans doute le *Frequency Scaling*.

Introduite par Middlebrook *et al.* [110], cette opération repose sur l'idée que l'interaction d'une onde sonore de fréquence donnée avec un solide dépendra des dimensions de ce dernier. En particulier, toute homothétie opérée sur l'objet doit s'accompagner, si l'on souhaite toujours observer la même interaction, d'une homothétie de rapport inverse sur la fréquence. Appliquée à l'individualisation, cette idée revient à dire qu'en connaissant les HRTF d'un individu de référence – ou même d'un mannequin – et le rapport d'échelle – *scaling factor* – entre la morphologie de cette référence et celle d'un sujet à individualiser,

on peut améliorer la sensation de localisation apportée à celui-ci par les HRTF de référence en leur appliquant une mise à l'échelle de rapport inverse.

En parallèle du *frequency scaling*, Maki & Furukawa [105] ont montré que, partant de la donnée de l'angle entre un pavillon d'oreille de référence et un pavillon test, une rotation du système de coordonnées donnant la direction des HRTF permet de réduire significativement les différences inter-individu. En d'autres termes, ce procédé utilise le fait, en le restreignant au pavillon d'oreille, qu'une rotation du sujet induit la même rotation au niveau des HRTF mesurées.

Ces approches, si utiles soient-elles, ne sauraient néanmoins constituer à elles seules des procédés complets de personnalisation. Cela reviendrait à réduire la variabilité des HRTF à un ou deux paramètres seulement. Toutefois, elles peuvent être vues comme de bons compléments à d'autres méthodes.

2.3.2 Synthèse à partir d'un ensemble

Une approche alternative au calcul direct des HRTF consiste à analyser un ensemble représentatif de HRTF réelles pour en faire émerger les principaux modes de variation.

2.3.2.1 Réseaux neuronaux

C'est notamment ce que réalisent les travaux de Busson [32] sur les réseaux de neurones artificiels (RNA). L'idée développée est de réaliser une prédiction des HRTF à partir de la mesure d'un nombre restreint d'entre elles. Cela passe en particulier par l'utilisation conjointe d'une carte de Kohonen et d'une *Classification Hiérarchique Ascendante* (CHA) comme préliminaires à l'élection de HRTF représentatives (jusqu'à 100 dans l'étude). Par la suite, un réseau de neurones de type *Multi Layer Perceptron* (MLP) à trois couches est construit et les HRTF représentatives de 44 sujets de la base CIPIC utilisées comme ensemble d'apprentissage. Bien que prometteuse, cette étude ne parvient pas à dégager de représentants universels, i.e. communs à tous les individus, ni ne présente de validation psycho-acoustique des résultats. De plus, il est également nécessaire de disposer d'un moyen d'accès auxdits représentants.

2.3.2.2 Analyse statistique

L'autre branche des méthodes alternatives de synthèse de HRTF se fonde, elle, sur l'ACP.

Kistler & Wightman [96] furent les premiers à proposer de décomposer les HRTF selon cette méthode. L'ensemble des HRTF est alors vue comme un sous-espace vectoriel de l'espace des mesures. La connaissance d'une base de ce sous-espace permet ensuite d'en atteindre n'importe quel représentant, i.e. n'importe quelle HRTF, par simple combinaison linéaire des vecteurs de base. C'est ce que permet l'ACP en fournissant une base orthonormale de l'espace engendré par les HRTF d'apprentissage. La dernière étape de la résolution du problème d'individualisation consiste alors à faire le lien entre les paramètres

morphologiques des individus et les coefficients de reconstruction par les vecteurs propres. Pour cela, des régressions linéaires multiples sont habituellement utilisées.

Partant des travaux de Kistler & Wightman, Xu *et al.* [161] ont proposé de grouper les HRTF des différents individus mesurés selon la direction (azimut, élévation) pointée avant d'effectuer l'ACP – une par groupe –, espérant ainsi réduire l'erreur d'estimation.

Zhang *et al.* [166] ont quant à eux proposé une méthode statistique d'estimation des paramètres anthropomorphiques les plus pertinents pour réaliser l'étape de régression.

En 2007, Vast Audio Pty Ltd. a de son côté déposé un brevet [86, 85] inspiré par ces idées. En pratique, ce dernier décrit tout d'abord la création d'une base de HRTF et d'une autre de paramètres morphologiques. Est ensuite invoquée l'utilisation d'une méthode d'analyse statistique pour décomposer en composantes élémentaires les espaces de paramètres et de HRTF, à la manière de ce que permet l'ACP. Par la suite, à l'aide d'une autre méthode d'analyse statistique, les liens entre les coefficients de reconstruction des paramètres morphologiques et ceux des HRTF sont déterminés. Par ailleurs, conjointement à la description, le brevet documente également l'utilisation d'un appareil de mesure intra-auriculaire des paramètres morphologiques caractérisant l'oreille.

Chaque variante proposée jusqu'à maintenant a généralement permis d'améliorer les résultats des méthodes antérieures sans toutefois offrir de rendu satisfaisant du point de vue psycho-acoustique, i.e. en conditions réelles. En particulier, le nombre et la localisation des paramètres morphologiques nécessaires sont très imprécis. De plus, dans le cas d'analyse simultanée de la morphologie et des HRTF, la découverte des liens entre les coefficients des deux mondes est d'autant plus complexe que les données sont laissées brutes.

2.3.3 Fine-tuning : le sujet en première ligne

2.3.3.1 Critère psycho-acoustique

Parmi les critères psycho-acoustiques, il convient en premier lieu de citer les travaux de Shimada *et al.* [138]. Partant d'une base conséquente de HRTF, ces derniers entendent réaliser des regroupements entre HRTF similaires. Pour ce faire, ils opèrent une décomposition cepstrale de 16 coefficients. La distance euclidienne naturellement associée à cet espace à 16 dimensions permet alors le regroupement des HRTF en *clusters* (au nombre de 8). Des jeux de HRTF sont ensuite choisis aléatoirement au sein des clusters et les sujets invités à élire le ou les clusters qui leur offrent la meilleure impression d'externalisation et de directivité.

Plus récemment, on pourra se référer aux travaux de Tame *et al.* [143] ou encore ceux de Xie *et al.* [159] qui utilisent respectivement des mixtures de gaussiennes et une décomposition en ondelettes pour réaliser le regroupement des HRTF.

Une fois le cluster sélectionné, une autre étape de sélection peut être ajoutée pour élire un jeu bien précis. Là encore, de multiples méthodes ont été publiées. Ainsi, Iwaya [83] décrit une procédure de sélection d'un jeu de HRTF parmi 32 disponibles en reprenant le principe des tournois d'échec. Une trajectoire sonore dans le plan horizontal est simulée par convolution d'un bruit rose avec les jeux de HRTF. 32 trajectoires sont donc obtenues

et mises en compétitions. À chaque rencontre, le sujet déclare vainqueur l'une des deux trajectoires selon qu'elle ressemble le plus à la trajectoire de consigne ou non. Le jeu sortant vainqueur du tournoi est déclaré le plus adapté au sujet.

Autre approche, celle de Seeber *et al.* [136], qui présente une sélection en deux étapes d'un jeu parmi douze. L'objectif affiché est d'être rapide sans entraînement préalable tout en offrant un résultat minimisant l'impression de son intra-crânien. La première étape consiste à désigner les cinq jeux présentant un meilleur rendu en terme de spatialisation dans la zone frontale. La seconde consiste à en éliminer quatre selon qu'ils pèchent à reproduire différents comportements tels que le déplacement d'une source sonore à vitesse constante, à élévation constante ou encore à distance constante. Une dizaine de minutes est nécessaire à la réalisation de la procédure.

Enfin, on citera également les travaux de Martens *et al.* [106] définissant le *bisection scaling*. L'idée est ici de créer à l'aide d'un test psycho-acoustique une table de correspondance entre les directions réelles associées à un jeu de HRTF et les directions perçues par le sujet. En pratique, pour un azimut donné, la tâche est de trouver la HRTF correspondant le mieux à la sensation d'une élévation à 45°. Les élévations extrémales – à 0° et 90° – étant supposées correctement perçues, une interpolation polynomiale du 2^{nd} ordre est ensuite opérée pour construire la table évoquée précédemment.

D'autres protocoles encore ont été proposés par la communauté scientifique mais aucun ne parvient à se défaire des inconvénients inhérents à ce type de méthodologie. En effet, même si l'on garde en mémoire que l'objectif n'est pas ici de trouver les HRTF exactes du sujet – il faudrait faire appel aux méthodes de synthèse – mais de sélectionner ou de s'adapter au mieux à l'existant, il n'en reste pas moins que la qualité de la solution optimale que l'on pourra proposer sera toujours limitée par la variabilité des jeux de HRTF ouverts à la sélection. Ainsi, pour un protocole donné, les résultats seront d'autant meilleurs que la base de donnée d'entrée sera importante. Or l'augmentation de cette dernière allonge de fait la durée de l'expérimentation, ce qui est d'autant plus gênant qu'elle repose sur la participation active du sujet.

2.4 Bases de données

2.4.1 Passage en revue

Ainsi que le détaillent les sections précédentes, de nombreuses méthodes de personnalisation – et la nôtre ne fera pas exception – reposent sur l'utilisation de données disponibles en grandes quantités. Par ailleurs, nombre de bases de données ont d'ores-et-déjà été constituées à travers le monde et, parfois, laissées ouvertes à la communauté. Dans ce dernier cas, elles ont l'avantage d'offrir une base de travail commune à tous, pouvant servir de référence et de point de comparaison indiscutable. Cependant, l'intérêt pratique de telle ou telle base est directement lié au cahier des charges de la méthode utilisée. Avant toute chose, il convient donc de s'intéresser d'un peu plus près aux bases morphologiques et / ou de HRTF existantes. La table 2.4.2 opère la synthèse des descriptions qui suivent.

CIPIC Parmi les plus connues – et les plus anciennes –, la base *CIPIC*[15] de l'université de Californie. Constituée en 2001 par Algazi *et al.* [2], elle rassemble des HRIR acoustiques et des valeurs de paramètres morphologiques pour 90 sujets. 45 d'entre eux ont été rendus publiques, dont deux variantes du mannequin KEMAR.

La grille de mesure utilisée pour l'acquisition des HRIR, c'est-à-dire l'ensemble des positions des haut-parleurs, se présente sous la forme d'un nuage de 1 250 points placés sur une sphère d'un mètre de rayon. Ces points se répartissent selon 25 azimuts et 50 élévations différents. L'écart angulaire moyen entre directions voisines est d'approximativement 5°. Elles ont été acquises canaux auditifs bouchés et sont de plus enregistrées à une fréquence de 44,1 kHz, longues de 200 échantillons et d'une durée proche de 4,5 ms. Les auteurs rapportent la présence fréquente de sauts d'ITD liés à de petits mouvements de tête, ce qui n'est pas forcément étonnant, des codes Golay ayant été utilisés comme signaux d'entrée – cf section 2.2.1. Bien que les jeux de données comportant de trop fortes discontinuités spectrales aient été écartés, certains autres peuvent tout de même en présenter des résidus.

Pour ce qui est des paramètres morphologiques, 10 se rapportent à l'oreille et 17 au reste du corps, pour un total de 27. Ils sont définis sur des représentations 2D de l'oreille et du corps. Des valeurs brutes, les auteurs tirent quelques données statistiques, comme la moyenne et l'écart-type associés à chaque paramètre. Néanmoins, il n'est pas clairement précisé quel jeu de données sert de base au calcul de ces statistiques. Les 90 sujets mesurés ? Les 45 rendus publiques ? KEMAR en est-il exclu ? Par ailleurs, sur les 45, seuls 37 présentent des jeux de mesure complets.

Cela étant, malgré ces failles, elle aura jusqu'à ce jour servi de base de travail à de nombreuses publications et fait encore figure de référence.

LISTEN Fruit d'une collaboration de 2003 entre la France, l'Autriche et l'Allemagne, la base LISTEN regroupe en accès libre les jeux de données morphologiques et acoustiques de 51 participants.

Les HRIR sont échantillonnées à 44,1 kHz et longues de 8 192 points. Elles sont prises canaux bouchés, dans une chambre anéchoïque, à 1,95 m de distance et dans 187 directions. L'espacement moyen de ces dernières est d'une quinzaine de degrés et les élévations ne descendent pas en-dessous de $-45°$. Des signaux de type ES ont été utilisé pour les mesures. Le résultat est disponible sous deux formes : brute ou égalisée. Dans la seconde, un fenêtrage à 512 points et une égalisation en champ diffus on notamment été appliqués.

Les données morphologiques, elles, sont largement inspirées du formalisme de CIPIC. Il s'agit plus précisément d'un sous-ensemble de 22 distances et angles, certains mesurés *in situ*, d'autres sur photos. Il est à noter que ces données sont parfois lacunaires ou manquantes.

FIU À la même époque, les laboratoires de la *Florida International University* réalisent eux aussi une série d'acquisition, consistant en des jeux de HRIR et des scans 3D de 15 personnes. Longtemps restreinte à un usage privé, cette base n'a été présentée publiquement [71] et rendue accessible sur demande qu'en 2010.

15. Center for Image Processing and Integrated Computing

Pour chaque sujet, 72 HRIR – 12 azimuts et 6 élévations – échantillonnées à 96 kHz sont enregistrées. La prise de son a été effectuée canaux bloqués et des séquences de type code Golay ont servi de signaux d'entrée. Par ailleurs, l'environnement de mesure est une simple pièce tapissée de mousse pour atténuer les réflexions. Celles-ci sont de toute façon à priori absentes des données du fait du fenêtrage temporel qui leur est appliqué.

Si les caractéristiques de ces HRIR font pâle figure face à celles des concurrentes de l'époque, l'originalité de FIU est plutôt à chercher dans la partie morphologique des données. Celles-ci sont en effet constituées de scans 3D de la tête et des oreilles des participants et sont accompagnées de certaines mesures anthropométriques. Ces dernières reprennent quelques définitions de CIPIC comme la hauteur et la largeur de l'oreille et y adjoignent de nouveaux paramètres, propres à la 3D, tels l'aire et le volume de la conque. Ces données 3D étant la plupart du temps très lacunaires, il est malheureusement difficile d'en sortir quoi que ce soit d'autre.

Takeda Au milieu des années 2000, l'université de Nagoya mène une série d'expérimentations visant à l'individualisation des HRTF dans le plan azimutal. À cette fin, ils réalisent en 2005-2006 une campagne d'acquisition de HRIR acoustiques totalisant 86 participants [120]. Ce nombre a par la suite été poussé jusqu'à 111, et des données anthropométriques ont été prises pour 80 d'entre eux. Laissée en libre accès, les liens initiaux de téléchargement ne sont toutefois aujourd'hui plus valides.

Pour chaque sujet, 72 HRIR échantillonnées à 48 kHz ont été mesurées tous les 5° dans le plan azimutal. Le système de mesure se constitue d'un haut-parleur installé dans une pièce non-insonorisée, à 1,52 m de distance de l'axe de rotation de la chaise sur laquelle sont positionnés les sujets. Ces derniers sont équipés de microphones placés à l'entrée de leur canaux auditifs sans que ceux-ci soient pour autant hermétiquement bouchés.

Les données anthropométriques rassemblent les mesures de 9 paramètres de tête et d'oreille. Leur définition est issue des spécifications du mannequin KEMAR [29].

RIEC Entre 2003 et 2008, un ensemble d'universités et instituts de recherche japonais, réunis autour du *Research Institute of Electrical Communications*, ont constitué une nouvelle base de HRTF [151]. Rendue publique en 2014 et accessible librement pour usage académique, elle rassemble les HRIR acoustiques au format SOFA de 105 sujets – dont deux mannequins – et des scans 3D de 39 d'entre eux.

En utilisant des signaux de type OATSP, ses auteurs ont procédé à l'acquisition des HRIR de chaque sujet dans 865 directions – 72 azimuts et 13 élévations – à 1,5 m de distance. La grille de mesure présente une précision de 5° en azimut et 10° en élévation. De plus, les valeurs des élévations s'échelonnent de $-30°$ à 90°. La fréquence d'échantillonage est de 48 kHz et la longueur finale des trames – après fenêtrage et padding – de 512. Enfin, la prise de son s'est effectuée en chambre anéchoïque, canaux auditifs bouchés. Parmi les remarques faites par les auteurs après analyse de leurs données, la présence résiduelles de discontinuités spatiales, qu'ils expliquent par les non-linéarités propres à chaque haut-parleur.

Les scans, s'ils jouissent d'une précision appréciable de l'ordre du millimètre, présentent cependant de nombreuses discontinuités et trous, en particulier au niveau des oreilles. Acceptables pour des mesures de caractéristiques anthropométriques de type CIPIC, ils sont en revanche inutilisables dans le cadre de simulations numériques.

ARI Initiée en 2009 par l'*Acoustic Research Institute*, la base ARI [6] rassemble à ce jour les HRTF acoustiques d'environ 150 personnes et des données anthropométriques pour 60 d'entre elles. En accès libre, cette base voit régulièrement sont nombre de sujets augmenter.

Les mesures acoustiques ont été faites à 48 kHz, dans une chambre semi-anéchoïque et se répartissent sur une grille de mesure de 1 550 points. Les élévations sont comprises entre $-30°$ et $80°$ et l'écart angulaire moyen avoisine les $5°$, sauf à l'avant, pour les azimuts compris entre $-45°$ et $45°$, où le pas de mesure azimutal a été divisé par 2. Les microphones ont pour leur part été placés à l'entrée des canaux auditifs. Ceux-ci n'ont toutefois pas été hermétiquement bouchés. Les résultats, disponibles sous forme de HRTF et de DTF, sont compilés au format SOFA.

Les données anthropométriques reprennent en grande partie le formalisme de CIPIC. Toutefois, d'autres choix de distances caractéristiques d'oreilles sont également définis, offrant une alternative intéressante.

SYMARE Une autre collaboration, cette fois entre les universités de York et de Sydney a permis la réalisation en 2012 de la base SYMARE [70, 84]. Celle-ci se compose de données morphologiques et de HRIR – acoustiques comme numériques – pour 62 sujets dont 10 sont accessibles au public.

Les données acoustiques se présentent sous la formes de HRIR acquises canaux bouchés, à une fréquence de 48 kHz et longues de 256 échantillons. Un haut-parleur positionnable sur une sphère de 1 m de rayon a servi à l'émission de codes Golay en entrée. Chaque jeu de HRIR comprend 393 directions. Le bras robotisé sur lequel a été monté le haut-parleur ne pouvant toutefois pas atteindre les élévations inférieures à $-45°$, les directions correspondantes ne font pas partie des mesures.

Les données morphologiques sont quant à elles des maillages tridimensionnels des oreilles, de la tête et du buste des participants. Réalisés à partir d'IRM, ces maillages se déclinent en plusieurs versions : oreilles seules, tête + oreilles, buste + tête + oreilles ou buste + tête. Les plus denses totalisent environ 450 000 sommets pour une résolution minimale de l'ordre du dixième de millimètre [16]. Afin de s'adapter aux contraintes de la FM-BEM, divers remaillages et sous-échantillonnages ont également été opérés, aboutissant à la génération d'autant de variantes supplémentaires des données initiales.

Car une des spécificités de la base SYMARE est de proposer en sus des mesures acoustiques des jeux de HRIR obtenus par simulations numériques. Les caractéristiques générales de directions et longueur de trames sont identiques à celles des mesures. Le nombre et l'espacement des fréquences simulées n'est pas explicité. Pour chaque sujet, deux jeux

16. À ne pas confondre avec la précision vis-à-vis de la vérité terrain. Le réglage des IRM est en effet particulièrement délicat et des écarts de plusieurs millimètres ont été observés sur au moins un sujet de cette base.

ont été calculés à l'aide du logiciel COUSTYX. Le premier à partir d'un maillage de tête et d'oreilles et allant jusqu'à 16 kHz. Le second à partir d'un maillage incluant également le torse mais s'arrêtant à 5 kHz. Ces limites fréquentielles, toutes regrettables qu'elles soient, ne font que traduire le coût computationnel important des méthodes numériques, et ce, en dépit de l'utilisation d'une solution professionnelle récente de simulation. Cela étant, les simulations ayant le bon goût de pouvoir être rejouées à loisir, il n'est pas dit que de nouvelles HRIR, plus complètes, ne seront pas ajoutées à l'avenir.

FABIAN Peu après, en 2013, est publiée la base FABIAN [22]. Sa réalisation s'insère dans le projet SEACEN et a été réalisée à l'université de Berlin *TUB*[17]. Elle entend étudier l'impact, sur les HRIR, de l'orientation de la tête par rapport aux épaules. Initialement composée de 11 mesures acoustiques de HRTF prises en chambre anéchoïque, elle est complétée en 2017 par leurs pendants numériques [23].

Plus en détail, un buste artificiel muni d'une tête orientable – le mannequin FABIAN – a été utilisé pour la réalisation des mesures. L'angle dont est tournée la tête vis-à-vis du torse croît de $-50°$ à $+50°$ par pas de $10°$. La grille de mesure est très dense, avec 11 345 directions réparties sur une sphère de 1,7 m de rayon et allant jusqu'à $-64°$ d'élévation basse. La fréquence d'échantillonage est de 44,1 kHz et la bande de fréquence utile est [100, 21 000] Hz. Les maillages 3D sont forts de 82 228 sommets et réputés permettre un calcul de HRTF jusqu'à 22 kHz. Cette base est publiquement accessible.

BiLi-Ircam Déjà partie prenante dans la collecte et le traitement des données de la base LISTEN, les laboratoires français de l'Ircam ont une nouvelle fois mis leur savoir-faire à l'oeuvre et leurs infrastructures à profit au sein du projet collaboratif BiLi [18], dont les résultats ont été présenté en 2014 [40]. Ceux-ci sont librement accessibles au format SOFA à des fins de recherche et d'enseignement.

Il s'agit cette fois-ci d'un ensemble de HRIR échantillonnées à 96 kHz pour 54 sujets et trois têtes artificielles. Réalisées en chambre anéchoïque, canaux auditifs bouchés, en utilisant des ES comme signaux d'entrée, ces acquisitions totalisent 1 680 directions différentes, d'écart angulaire moyen $6°$. Les élévations couvrent un champ allant de $-51°$ à $86°$. En particulier, cette grille de mesure a été spécialement conçue pour faciliter la décomposition des HRTF en harmoniques sphériques.

L'ajout des maillages 3D des participants et des mesures anthropométriques associées fait partie des travaux futurs et aucune donnée morphologique n'est pour l'heure disponible.

ITA Entrée suivante dans la liste, la base de HRIR acoustiques et de scans 3D de l'institut *ITA*[19], présentée pour la première fois en 2016 [20]. Forte de 48 sujets, elle est disponible sur demande et utilisable à toute fin non commerciale.

Les HRIR qu'elle contient sont longues de 256 échantillons, acquises à 48 kHz et se répartissent selon 2 304 directions, pour un écart angulaire moyen avoisinant les $5°$. Les

17. Technische Universität Berlin
18. Binaural Listening
19. Institu für Technische Akustik

participants mesurés l'ont été canaux auditifs bouchés, dans une chambre semi-anéchoïque. Les haut-parleurs sont quant à eux supportés par un arceau de 1,2m de rayon et ont servi à l'acquisition selon la méthode *optimized MESM*. Notamment disponibles au format SOFA, elles couvrent les fréquences allant de 200 Hz à 18 kHz.

Autre point important : l'ensemble des données morphologiques. Celui-ci consiste en effet en divers paramètres liés à la tête – et empruntés à CIPIC – mais aussi et surtout en des scans 3D des oreilles reconstitués à partir de résultats d'IRM. Bien que n'allant pas aussi loin que SYMARE, dont les scans 3D incluent la tête et le buste, la base ITA vient avec des données 3D de première importance.

HUTUBS Enfin, dernier ajout à cette liste, la base HUTUBS [24, 21] de HRIR acoustiques, de HRIR calculées, de scans 3D – de la tête – et de paramètres anthropométriques de la *TUB*.

Publiée en 2019, elle offre cet ensemble très fourni de mesures pour 96 sujets. Toutefois, seuls 58 d'entre eux ont consenti à la mise à disposition de toutes leurs données, diminuant d'autant le nombre de jeux complets disponibles. Par ailleurs, deux entrées sont en réalités des mesures du mannequin FABIAN. Cette base fait donc également un écho intéressant à la base du même nom.

Les HRIR ont été acquises – resp. simulées – canaux bouchés, à une fréquence de 44,1 kHz et sont longues de 256 échantillons. Chaque jeu mesuré – resp. simulé – comprend 440 – resp. 1 730 – directions réparties sur une sphère de diamètre 1,47 m.

Les maillages disponibles ont entre 140 000 et 240 000 sommets et sont plus fins au niveau des oreilles. Cette taille, très importante, et le passage parfois abrupte entre zones de résolutions différentes laisse supposer qu'il s'agit là de maillages obtenus après recollement des oreilles sur la tête et non de ceux en entrée de simulation.

Les paramètres anthropométriques, quant à eux, se composent de 25 mesures largement inspirées de CIPIC mais retravaillées pour la 3D.

2.4.2 Analyse comparée

Sans être exhaustive – les bases totalement privées ou celles ne comportant pas de données morphologiques ayant été passées sous silence –, la liste établie ci-avant offre une vue d'ensemble des données et techniques utilisées ces quinze dernières années. On peut ainsi constater que la base CIPIC, probablement grâce à son statut de précurseur, dispose encore d'une énorme influence sur les travaux actuels. Sans même chercher à dénombrer la pléthore d'articles utilisant leurs données pour construire et présenter de nouveaux travaux, le seul nombre de bases ultérieures reprenant tout ou partie des définitions des paramètres morphologiques 2D est déjà en soi remarquable.

À souligner également, la difficulté à dépasser la centaine de sujets. Seules les bases Takeda, RIEC et ARI ont jusqu'ici franchi ce seuil. La constitution de telles collections de données est bien sûr un processus coûteux, complexe et de longue haleine. Il n'est donc pas étonnant que leur taille habituelle soit de l'ordre de quelques dizaines d'entrées. Pour

Base	Date d'acquisition	Type	Taille	Échantillonage	Distance	Disponibilité
CIPIC	2001	HRIR acoustiques et données anthropométriques	90	44,1 kHz	1 m	semi-publique
LISTEN	2003	HRIR acoustiques et données anthropométriques	51	44,1 kHz	1,95 m	publique
FIU	$<$ 2004	HRIR acoustiques, scans 3D et données anthropométriques	15	96 kHz	-	publique
Takeda	2005 —- 2006	HRIR acoustiques et données anthropométriques	111	48 kHz	1,52 m	publique
RIEC	2003 —- 2008	HRIR acoustiques et scans 3D	105	48 kHz	1,5 m	publique
ARI	2009 —- 2017	HRIR acoustiques et données anthropométriques	~150	48 kHz	1,2 m	publique
SYMARE	2012	HRIR acoustiques et simulées et scans 3D	62	48 kHz	1 m	semi-publique
FABIAN	2013 —- 2017	HRIR acoustiques et simulées et scans 3D	11	44,1 kHz	1,7 m	publique
BiLi-Ircam	2014	HRIR acoustiques scans 3D à venir	57	96 kHz	1,95 m	publique
ITA	2016	HRIR acoustiques et scans 3D	48	44,1 kHz	1,2 m	publique
HUTUBS	2019	HRIR acoustiques et simulées, scans 3D et données anthropométriques	96	44,1 kHz	1,47 m	publique

TABLE 2.1: *Bases de HRTF et de données morphologiques à travers le monde.*

autant, la variabilité manifeste des HRTF et des morphologies incite plutôt à voir ce seuil, tout symbolique qu'il soit, comme loin du minimum nécessaire.

Autre point notable, la montée en puissance des acquisitions 3D. Encore balbutiantes au milieu des années 2000, elles ont franchi un cap lors de la publication de SYMARE et sont appelées à se développer, comme en témoignent les base BiLi-Ircam, FABIAN, ITA et HUTUBS. Ce développement, bien évidemment lié à de plus grandes facilités d'acquisitions, traduit également la difficulté à traiter les liens entre morphologie et HRTF à l'aide de représentations schématiques. Il est l'expression de l'insuffisance de l'approche paramétrique, initialement mise en place faute de mieux, à traiter efficacement ce problème.

En outre, SYMARE, FABIAN et HUTUBS se particularisent aussi par la fourniture de HRIR simulées, qu'elles sont les seules à proposer. Et si la nécessité d'avoir des scans 3D de qualité en est un premier verrou, la capacité de calcul en est un autre. Là encore, les récents développements scientifiques et technologiques expliquent la levée progressive de ce dernier. En effet, l'apparition de méthodes de résolutions numériques de plus en plus performantes associée au développement d'ordinateurs de plus en plus puissants – à coût constant – fait depuis quelques années des simulations une solution viable.

2.4.3 Application à notre cas d'utilisation

Replaçant ces observations dans le contexte du procédé de personnalisation proposé, les enseignements à en tirer sont multiples. Tout d'abord, la production de HRIR acoustiques est une technique répandue mais au coût d'infrastructure et d'organisation trop important pour se démocratiser véritablement. Toutes les bases listées en proposent, mais les campagnes s'étendent généralement sur plusieurs années et ne dépassent jamais quelques dizaines de jeux. Et si l'envie prenait quelqu'un de les fusionner pour en accroître la taille, il faudrait alors résoudre les problèmes posés par l'hétérogénéité des données, tant elles diffèrent entre elles aussi bien en terme de fréquence d'échantillonage, qu'en distance de prise de son ou en grille d'évaluation. Par ailleurs, on l'a vu, chacune est susceptible de venir avec ses biais propres. Qu'ils soient liés à la salle de mesures, au type de signaux d'entrée ou à tout autre chose, ils amènent nécessairement à s'interroger sur les qualités relatives de chacune d'entre elles. Ce point est d'autant plus important qu'aucune évaluation subjective n'accompagne jamais les HRIR.

En outre, si des données anthropométriques parviennent à décrire schématiquement la tête et le buste d'un individu, elles sont loin de suffire à décrire la complexité de l'oreille humaine. Déjà plus facile à définir qu'à mesurer, la perte inévitable d'information qu'elles induisent en faisaient dès l'origine un pis-aller. L'acquisition 3D, bien que reposant uniquement sur l'IRM dans les exemples que sont SYMARE et ITA, peut en réalité être menée avec de bien plus modestes moyens. Le marché foisonne en effet de scanners portatifs adaptés à une multitude d'usage, dont le scan de tête ou d'oreille. Moyennant un investissement réduit, il est donc réaliste d'envisager la constitution d'une base de scans 3D de qualité de taille suffisante pour y appliquer les méthodes d'analyses statistiques les plus courantes. Ces scans étant de surcroît le point d'entrée vers l'obtention de HRTF

calculées, aujourd'hui accessibles, ils sont la donnée nécessaire et suffisante à la résolution du problème de personnalisation de HRTF par modèle déformable d'oreille.

N'étant néanmoins pas encore disponibles – à tout le moins en quantités suffisantes –, ces scans doivent faire l'objet d'une phase de collecte. En d'autres termes, les bases actuelles ne répondant pas à nos besoins mais le coût nécessaire d'acquisition étant raisonnable, l'une des étapes à prévoir est la constitution d'une base de données de scans 3D et dont sera dérivée une base de données de HRTF correspondantes.

2.5 Tests subjectifs

Le principal intéressé par les HRTF étant le sujet lui-même, leur personnalisation ne saurait aller sans une phase d'évaluation perceptive. Celle-ci revêt de nombreuses formes et il est rare de voir deux équipes distinctes utiliser rigoureusement le même protocole. Néanmoins, plusieurs caractéristiques récurrentes se dégagent.

Qualitatif vs quantitatif Et tout d'abord, le type de test, qui peut être tantôt *qualitatif*, tantôt *quantitatif*. Le premier sera privilégié pour évaluer le ressenti lié à l'écoute. Cela recouvre les concepts portant dans leur définition même une part inaliénable de subjectif tels que la préférence globale, l'émotion, le timbre ou encore la coloration d'un son. Le choix est alors laissé à l'auditeur pour classer les stimuli proposés les uns par rapport aux autres ou selon une échelle définie *ad hoc*. Catégorisation et classification sont ici les maîtres-mots.

Le second est adapté aux notions pour lesquelles des métriques indiscutables existent. La localisation et l'externalisation en sont deux exemples typiques. Dans la mesure où la spatialisation fidèle des sons est en pratique l'un des objectifs premiers de l'usage du binaural, c'est cette seconde catégorie qui compte le plus grand nombre de représentants. L'aspect quantitatif permet aussi de produire des résultats plus aisément comparables à ceux issus d'autres études. Néanmoins, rien n'empêche de traiter des notions à priori quantitatives sous l'angle qualitatif.

Auditeur L'expérience de l'auditeur en matière de tests d'écoute est un autre point fortement discriminant. On parlera d'auditeur *naïf* s'il en a peu ou pas et d'*expert* dans le cas contraire. Il est fréquent qu'un panel de sujets mixte, mélangeant experts et auditeurs naïfs, amène à différencier les résultats selon ce critère dans la discussion [56]. L'homogénéité d'un panel est donc également notable et se voit précisée dans la description du test [9, 11, 136].

Stimuli Vient ensuite le type de son reproduit. On trouve une large utilisation de bruits blancs [158, 98, 136, 77, 115, 93, 160] ou rose [132, 83, 38, 162, 148]. Ils ont l'avantage de permettre un test de la HRTF dans son ensemble et donnent une égale importance à toutes les fréquences [20]. Aisément implémentables, ils sont privilégiés lors des tests quantitatifs.

20. selon une échelle linéaire pour le bruit blanc, logarithmique pour le bruit rose.

Cela étant dit, l'écoute d'un bruit, qu'il soit blanc ou coloré, est assez peu adapté pour exprimer une préférence ou générer une émotion... Ceci est d'autant plus vrai pour les auditeurs naïfs, qui seront souvent déroutés par le contenu proposé avant de pouvoir à nouveau se concentrer sur les consignes initiales. Bien évidemment, il peut aussi être inadapté si l'objectif même du test est d'étudier une caractéristique précise d'un type de signal comme par exemple l'intelligibilité de la voix [37]. En alternative fréquente se retrouvent donc tous les sons dits « familiers », généralement la voix [11, 10, 117, 49] ou la musique [56, 165]. Cela étant, le contenu fréquentiel proposé n'est alors plus aussi bien connu et contrôlé. Or, il est démontré que la richesse spectrale d'un stimulus est un facteur de bonne localisation [34, 58, 28]. Les stimuli familiers sont donc moins adaptés pour des tests de performance pure mais préférables lorsqu'il est important de s'approcher d'un cas d'utilisation réel. Par ailleurs, dans un tel cas, Blauert [19] a montré que de meilleurs résultats sont à attendre d'une voix connue que d'une voix étrangère.

Réverbération Le besoin d'approximer du mieux possible le réel peut d'ailleurs amener à en reproduire une autre caractéristique : la réverbération.

Les HRTF étant le plus souvent sèches [21], c'est-à-dire sans effet de réverbération, rien n'empêche de leur ajouter un effet de salle [22]. Une telle manipulation est connue pour augmenter la sensation d'externalisation des sons [9, 52, 139]. Dans un environnement anéchoïque, celle-ci est en effet peu fiable, tant elle dépend de l'éloignement de la source et de son intensité propre. Les réflexions sont l'ensemble des chemins indirects empruntés par le son pour atteindre le tympan. Le rapport de l'énergie reçue directement sur celle reçue indirectement offre un indice d'estimation de l'éloignement, un fort ratio étant généralement signe de grande distance. Il peut donc être opportun d'inclure ce facteur dans le protocole de test pour en optimiser le réalisme général. Cela étant, sa contribution exacte n'est pas clairement établie. L'étude menée par Begault *et al.* [9] en est à ce titre un bon exemple. S'ils relèvent une augmentation de l'externalisation liée à l'ajout d'un effet de réverbération, ils notent aussi une dégradation des performances de localisation. Toutefois, ils leur faut préciser que les HRTF utilisées n'étaient pas individuelles mais génériques et que des stimuli de voix, et donc à spectre réduit, avaient été joués. Cette interdépendance des facteurs mis en jeu a d'ailleurs amené certaines équipes à étudier l'impact de la réverbération en conjonction avec d'autres éléments tels que le head-tracking [11] ou la présence d'indices visuels [149].

Mouvements du sujet Parmi ces autres éléments, la possibilité laissée au sujet de bouger la tête, voire de se déplacer, trône en bonne place [11, 115, 25, 154]. L'une de ses caractéristiques additionnelles mais non systématique est l'utilisation d'un capteur de mouvement. Ce dernier cas définit la sous-famille des tests *avec head-tracking.*

Une telle liberté d'action du sujet permet notamment de diminuer le taux de confusions avant-arrière dans les tests de localisation. En effet, si l'on prend l'exemple d'un auditeur

21. soit qu'elles auront été mesurées en environnement anéchoïque, soit qu'elles auront été calculées à partir d'un modèle physique ne prenant pas en compte cet aspect.

22. Il s'agit même d'une de leurs forces, un jeu de HRTF pouvant alors être employé quel que soit l'environnement.

ayant une source sonore située face à lui, une rotation de la tête vers la gauche augmentera l'intensité du signal perçu à l'oreille droite et la diminuera à la gauche alors qu'une source située à l'arrière aurait provoqué le phénomène inverse. Déjà mis en évidence il y a près de 70 ans par Wallach [150], ce mécanisme justifie l'utilisation d'un capteur de mouvements lors de tests de localisation. Dans une version plus récente des expériences de Wallach, Wightman & Kistler [154] confirment d'ailleurs les bénéfices observés. D'une part, ces derniers observent qu'autoriser les sujets à bouger la tête pour localiser une source diminue de manière drastique lesdites confusions par rapport au cas où les sujets n'y sont pas autorisés – toutes choses étant égales par ailleurs –, mettant ainsi en avant le rôle du système vestibulaire dans la représentation sonore. D'autre part, et de manière très intéressante, ils observent que lorsque les sujets sont immobiles, si la source se déplace indépendamment de leur volonté – par le biais de l'expérimentateur, par exemple – alors les confusions demeurent mais que s'ils sont à l'origine des déplacements de la source alors les confusions disparaissent de la même manière que s'ils avaient pu bouger la tête. Laisser la main à l'auditeur lui permet donc de formuler des hypothèses quant à la position de la source qui seront confirmées / infirmées à l'écoute. Cette boucle rétroactive ne fait plus effet si l'auditeur n'est plus impliqué dans les déplacements.

D'une autre série d'expériences, Bronkhorst [25] conclut que des sources virtuelles créées à l'aide de HRTF individualisées peuvent être localisées presque aussi précisément que des sources équivalentes réelles pour peu que les mouvements de tête soient autorisés et que les stimuli soient suffisamment longs. L'utilisation de HRTF non-individualisées se traduit le plus souvent par une dégradation des performances, tout comme l'utilisation de stimuli courts couplée à une interdiction des mouvements de tête.

Du point de vue de l'externalisation, s'il est généralement rapporté que la présence de head-tracking en améliore la sensation [52, 152], il ne s'agit toutefois pas d'un consensus unanime [11].

2.6 Chaîne de personnalisation proposée

Comme on peut le constater, les idées ne manquent pas pour parvenir à une personnalisation des HRTF. Malgré cela, aucune d'entre elles ne propose pour l'heure de personnalisation à l'échelle industrielle, c'est-à-dire réalisable à peu de frais et de n'importe où. On l'a vu, les mesures nécessitent de grandes installations. Elles sont de fait peu nombreuses et d'accès restreint à un public professionnel. Les simulations numériques demandent, elles, à fournir un maillage 3D précis de la personne et de ses oreilles, restreignant là aussi les débouchés à un petit cercle. Quant aux méthodes indirectes, on peut leur reprocher d'être trop fatiguantes pour le sujet, de reposer sur des données complexes à obtenir – les HRTF dans quelques directions dans le cas de Busson [32], les paramètres morphologiques pour Kistler & Wightman [96] – ou les deux à la fois. Ainsi, indépendamment de leurs résultats effectifs, ces méthodes demandent beaucoup au sujet, que ce soit de manière active en le faisant choisir selon ses préférences ou passive en le mesurant ou en acquérant son scan 3D.

La chaîne de personnalisation ici proposée entend éviter ces écueils et répondre aux impératifs de simplicité, de rapidité et d'accessibilité précédemment évoqués. Pour y arriver,

le cheminement général peut se décomposer en trois sous-parties. Les deux premières sont préparatoires et réalisées en amont – cf. figure 2.10 et 2.11 ci-après – tandis que la dernière est dédiée à l'utilisateur final – cf. figure 2.12.

Par ailleurs, cette chaîne de personnalisation a fait l'objet d'un *engineering brief* [64] à la 141e convention de l'AES [23] et d'un brevet [63].

2.6.1 Création des modèles déformables

En ce qu'il va permettre la constitution de nos bases de données morphologiques et de HRTF, le modèle déformable 3D de tête, buste et oreilles (TBO) est à n'en pas douter un élément clef de notre procédé. Dans sa version la plus complète, il est issu de la fusion entre un modèle générique de tête et de buste et un modèle déformable spécifique à l'oreille. Son processus de création est schématisé figure 2.10.

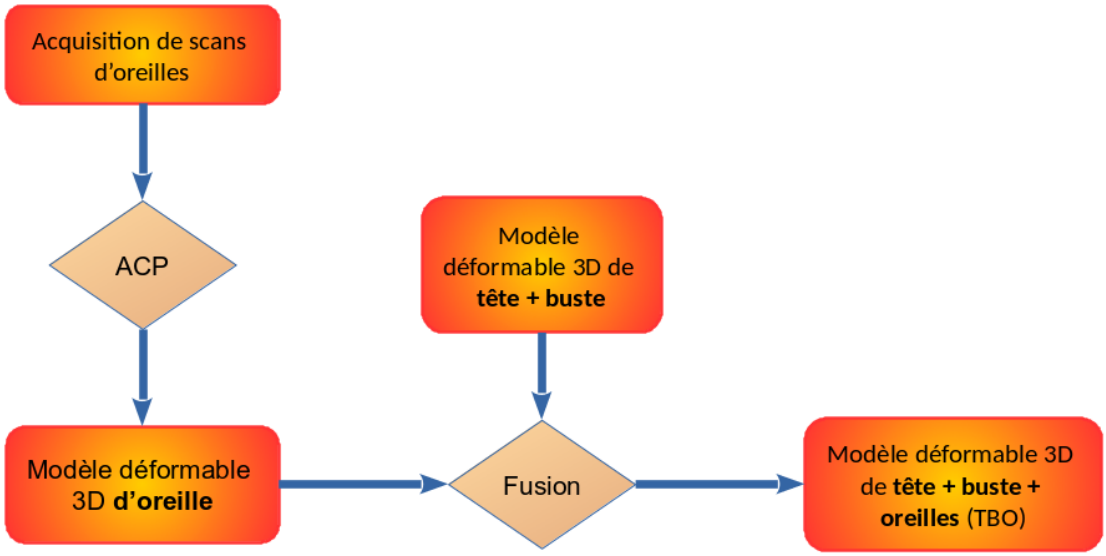


FIGURE 2.10: *Étapes de création des modèles déformables.*

Ici, le modèle de tête et buste a été réalisé grâce au concours de la société DynamiXYZ et peut être vu comme un apport extérieur à la thèse. Toutes les autres parties, en revanche, ont fait l'objet d'un travail spécifique avec en premier lieu l'acquisition de scans d'oreilles. Ce travail au long cours est détaillé à la section 3.1.1.

Une fois les données acquises, de multiples étapes de nettoyage et surtout une mise en correspondance des différentes oreilles ont été nécessaires en amont de l'ACP, prélude à la réalisation du modèle déformable 3D d'oreille. Un tel modèle n'ayant jamais été créé – ou rendu public – à ce jour, sa réalisation est une contribution majeure des présents travaux et se trouve détaillée à la section 3.1.2 du manuscrit. Elle a également fait l'objet du dépôt du brevet [62].

Enfin, une version simplifiée de la fusion est présentée section 3.2.2 et donnera lieu à ce que nous nommerons par la suite le *modèle mixte*.

23. *Audio Engineering Society*

2.6.2 Bases de données et fonction de couplage

Une fois le modèle TBO créé, ce dernier est utilisé pour constituer une première base de données, uniquement morphologique. Comme nous le verrons chapitre 5, dans les faits, plusieurs variantes du modèle ont été utilisées. Néanmoins, toutes suivent le même schéma global – visible figure 2.11.

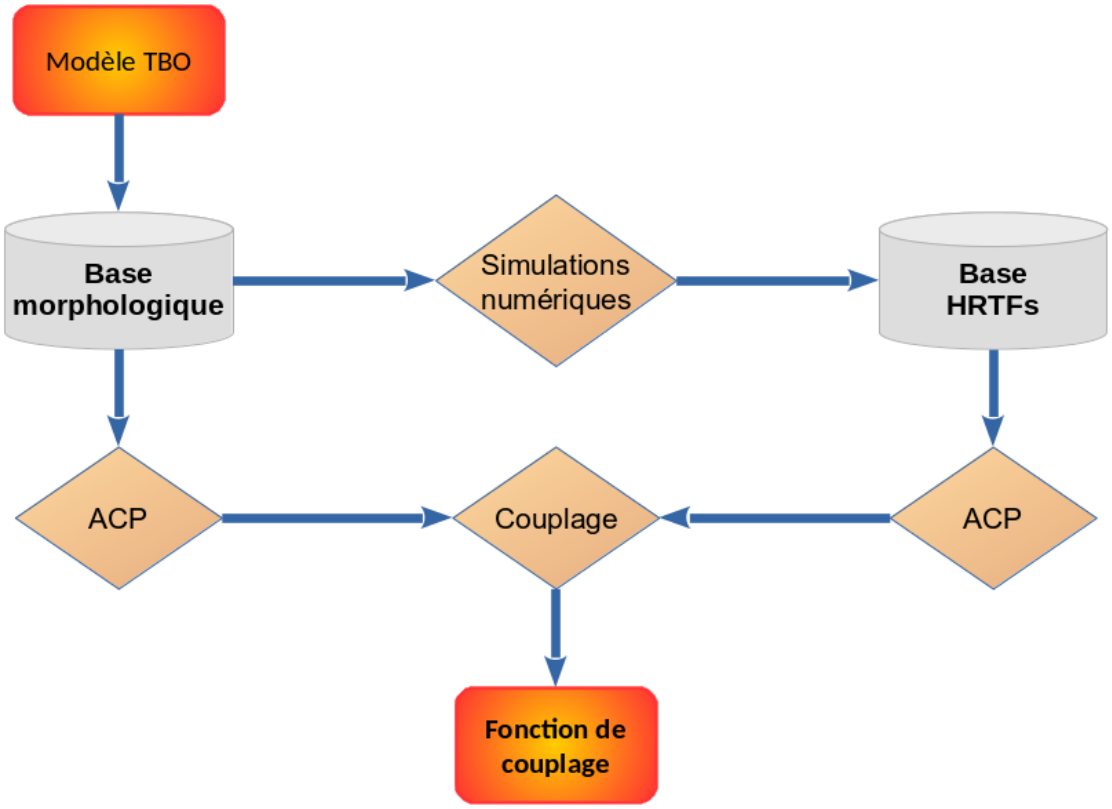


FIGURE 2.11: *Étapes de création des bases de données et de la fonction de couplage les liant.*

Partant de là, nous utilisons ensuite cette base pour :

1. créer par ACP un espace représentatif des morphologies,
2. construire par simulations numériques la base des HRTF correspondantes.

Cette dernière faisant elle aussi l'objet d'une ACP, nous gagnons accès à un autre espace représentatif : celui des HRTF. La constitution de ces bases et des espaces qui leur sont liés se voit détaillée section 5.1. C'est là encore une contribution majeure de ces travaux, l'un des modèles présentés ayant d'ailleurs mené à la création et la mise en ligne de la base CHEDAR, disponible sur `www.sofaconventions.org` ainsi qu'à une publication académique lors de la 148^{e} convention de l'AES [61].

Une fois les deux espaces représentatifs disponibles, la dernière étape, détaillée et évaluée section 5.2.2 consiste en la création d'une fonction dite *de couplage* liant les deux mondes. De cette façon, le passage de l'un à l'autre devient immédiat et généralisé à une

infinité de morphologies et de HRTF. C'est là aussi une contribution inédite et majeure de ces travaux.

2.6.3 Processus utilisateur

Vient enfin le processus de personnalisation proposé à l'utilisateur final – cf. figure 2.12. Points d'entrée du système, le type et la quantité de données à acquérir sont à définir avec attention. Ils sont en effet le résultat d'un compromis entre simplicité et précision. Simplicité d'accès à la personnalisation pour l'utilisateur, tant elle conditionnera sa volonté de tester le procédé. Précision – et donc qualité – du résultat final, sans quoi il est inutile de se donner tant de mal.

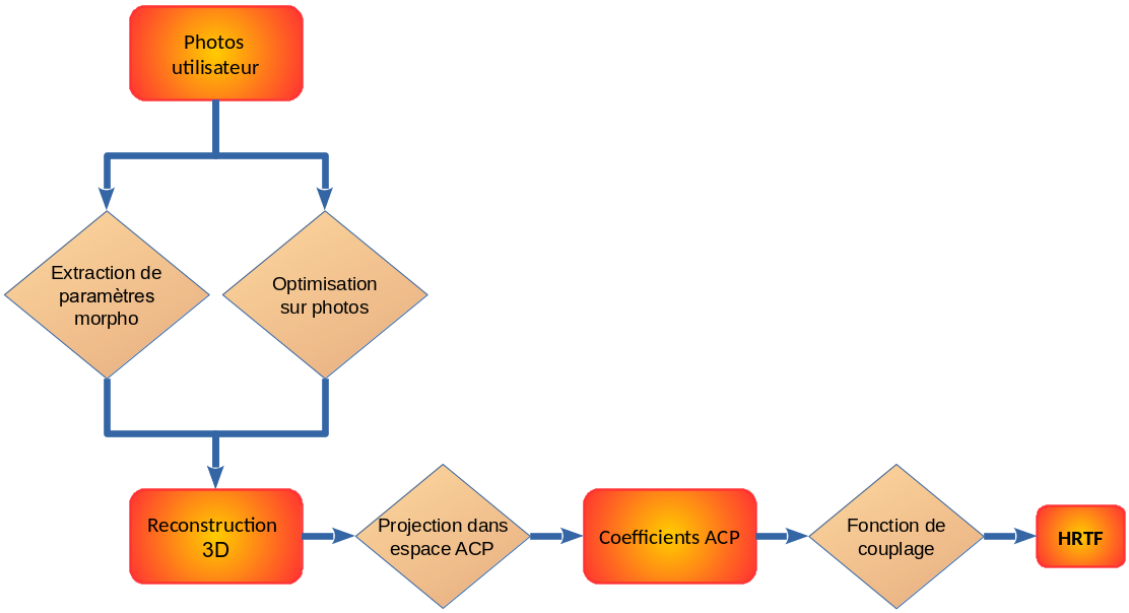


FIGURE 2.12: *Chaîne de personnalisation proposée à l'utilisateur final.*

En raisonnant par condition nécessaire, l'acquisition doit porter sur chaque oreille, la tête et le torse. De plus, les appareils photos et caméras étant largement répandus de nos jours – constat d'autant plus vrai depuis l'apparition des smartphones –, il est tout à fait naturel de les utiliser comme vecteur d'acquisition. Ainsi, un jeu de photos correctement prises de l'utilisateur devrait répondre au cahier des charges sus-mentionné. Idéalement, trois photos sont suffisantes : une de l'oreille droite, une autre de la gauche et une dernière du buste[24]. Les oreilles jouant un rôle plus critique dans la perception sonore que le reste du corps, elles peuvent faire l'objet de prises de vue supplémentaires sans que cela devienne impératif. Typiquement, une résolution donnant accès à des détails de dimension inférieure au millimètre est suffisante.

À ce stade, il est utile de noter qu'avec la poursuite des développements technologiques actuels, on peut s'attendre à voir apparaître des caméras de profondeur ou stéréoscopiques

24. la tête et les épaules.

équiper les prochaines générations de smartphone. Si tel était le cas, cela pourrait constituer une alternative intéressante à la photo 2D.

Plusieurs informations sont extraites de ces quelques vues. Tout d'abord, les dimensions générales de la personne telles que la distance interaurale – à minima – ou la largeur d'épaules sont récupérées de la photo du buste. Suivant le cas d'usage, il est envisageable d'extraire encore d'autres données : hauteur de tête, hauteur du cou, orientation des oreilles, etc.

Ensuite et surtout, les photos d'oreilles. Celles-ci servent à en reconstituer la forme tridimensionnelle. Pour cela le modèle 3D d'oreille vu précédemment est mis à contribution et donne lieu à l'étape d'optimisation sur photos. Il permet après une opération de convergence de trouver la forme 3D et l'angle de prise de vue les plus adaptés ayant pu générer les photos de l'utilisateur.
Réalisée au sein de la société 3D Sound Labs, cette brique d'entrée du système apparaît section 3.3.

Une fois ces opérations effectuées plusieurs variantes du procédé sont possibles mais dans le cas le plus poussé les valeurs des paramètres morphologiques extraites précédemment sont utilisées afin de paramétrer au mieux la forme de la tête et du buste du maillage 3D. Par la suite, ce maillage est projeté dans l'espace ACP représentatif des morphologies et ses coordonnées sont alors utilisées par la fonction de couplage évoquée à la section précédente pour donner naissance aux HRTF personnalisées de l'utilisateur.

2.6.4 Simulations numériques et impédance acoustique

Grâce aux trois sections précédentes nous disposons à ce stade d'un descriptif général de la chaîne de personnalisation envisagée. Toutefois, en marge des travaux déjà évoqués, d'importants efforts ont également été déployés pour sécuriser et optimiser l'étape de simulation numérique (cf. figure 2.11). Présentées chapitre 4, ces recherches ont aussi bien porté sur des considérations anatomiques (cf. section 4.1.3), s'attachant à déterminer la sensibilité des simulations à une mauvaise connaissance de la forme du canal auditif ou à différents placements possibles des microphones, que sur des considérations plus numériques (cf. section 4.2), cherchant à optimiser le temps de calcul ainsi que les ressources informatiques nécessaires.

Par ailleurs, nous avons constaté durant nos recherches que des différences substantielles existent entre la personnalisation de HRTF par mesures acoustiques et la personnalisation par simulations numériques, telles que l'état de l'art les préconisent. Nous nous sommes donc également attaché à investiguer sur l'origine de ces différences et présentons, section 4.3, nos travaux et leurs conclusions quant à l'influence de l'impédance acoustique choisie lors de simulations numériques. Nous y montrons en particulier qu'un choix judicieux d'impédance de la peau peut permettre à des HRTF simulées d'exhiber un rendu subjectif aussi bon que leur équivalent acoustique. Ce résultat constitue là aussi une contribution majeure de cette thèse et a également fait l'objet d'une contribution académique lors de la 148^{e} convention de l'AES.

2.6.5 Limitations du champ d'étude

Néanmoins, plusieurs obstacles de taille n'ont pour l'heure pas été surmontés et s'opposent au bon fonctionnement de notre chaîne de personnalisation, telle que nous l'avons mise en place. En effet, bien que nous ayons pu montrer qu'une impédance correctement choisie pouvait grandement améliorer les HRTF obtenues par simulation, nous n'avons pas pu passer l'étape de généralisation. En ce sens, nous ne pouvons pas prétendre pouvoir générer à coup sûr par simulation des HRTF au bon rendu subjectif. Or ce rendu est un point primordial de notre base sur lequel on ne peut transiger si l'on souhaite offrir une personnalisation par photos au plus grand nombre.

Par ailleurs, nous avons aussi dû reconnaître que de nombreuses HRTF acoustiques ne tenaient pas non plus leurs promesses en termes de rendu, ne se démarquant nullement de HRTF génériques lors de tests subjectifs. Il ne serait donc pas suffisant de remplacer nos bases simulées par des bases issues de mesures [25], il faudrait aussi et surtout s'assurer que les HRTF de remplacement soient subjectivement pertinentes, ce qui ne va pas de soi.

La conséquence générale est qu'en l'état actuel des choses il serait illusoire d'espérer obtenir des résultats convaincants sur l'ensemble de la chaîne présentée figure 2.12, c'est-à-dire avec des utilisateurs réels. Néanmoins, le tableau n'est pas aussi sombre qu'il peut sembler. En effet, comme nous le verrons à la section 5.2.2, dédiée à l'évaluation du couplage, l'utilisation du simulateur de tests subjectifs permet d'agir comme si les bases de HRTF utilisées proposaient un rendu subjectif impeccable ! Ainsi, nous pouvons rendre temporairement indépendante notre chaîne du problème du rendu subjectif de la base de HRTF utilisée. Au final, le seul véritable impact de tout cela est que la partie « Photos utilisateurs » vers « Reconstruction 3D » n'a été que partiellement mise au point, car bien moins prioritaire que le reste dans le contexte que nous venons d'évoquer, et que le lecteur lui trouvera donc peu de développements dédiés dans les pages qui suivent.

25. si tant est qu'il en existe de suffisamment grandes...

Chapitre 3

Modélisations morphologiques

Un secret a toujours la forme d'une oreille.

- Jean Cocteau -

Ce chapitre se consacre aux modélisations morphologiques dans leur ensemble, depuis l'acquisition des données brutes jusqu'à la création de modèles déformables complets en passant par toutes les étapes et traitements intermédiaires.

Nous allons voir dans quelle mesure les données de synthèse et les données réelles peuvent se compléter mutuellement et les contraintes à prendre en compte lors de leur génération / acquisition, surtout en prévisions des simulations numériques de HRTF.

Puis nous passerons à l'étape de création du modèle déformable d'oreille, en prêtant une attention toute particulière à la mise en correspondance des données. Celle-ci est en effet un point épineux et crucial du procédé.

Par la suite, nous expliquerons comment passer du simple modèle d'oreille à un modèle plus complet comprenant une tête et un buste. Cette partie comportera notamment les descriptions d'un modèle purement synthétique et d'un autre dit « mixte », utilisés ultérieurement pour la constitution de bases morphologiques 3D d'envergure. Elle est également fondamentale car la grande majorité des simulations des chapitres suivants utiliseront un maillage dérivé de ces modèles.

Enfin, nous ferons un dernier détour par l'étape d'optimisation sur photo et verrons comment en tirer des données tridimensionnelles.

3.1 Base réelle et modèle d'oreille

3.1.1 Collecte de données réelles

3.1.1.1 Prérequis

Comme explicité précédemment, la première étape à traiter est la constitution d'une base de données apte à servir en simulations numériques. Les contraintes qui en découlent sont de plusieurs ordres. En premier lieu, la zone à scanner doit être aussi grande que possible, quitte à retirer par la suite le superflu. Cela comprend bien sûr les oreilles et la tête mais aussi une grande partie du buste. De plus, le centre de l'audition étant situé dans l'oreille, la criticité de l'acquisition croît avec la proximité à l'oreille. Présenté de façon plus triviale, une petite bosse n'aura pas le même impact sur le trajet des ondes sonores selon qu'elle sera située à l'arrière du crâne ou à l'entrée du canal auditif. Suivant ce principe, on peut s'autoriser une précision moindre pour la tête et le buste que pour les oreilles. Ceci est par ailleurs renforcé par les nombreuses occlusions et anfractuosités que peut présenter l'oreille.

Deuxièmement, les méthodes numériques venant avec la fameuse règle des six éléments par longueur d'onde, il est impératif de s'intéresser à la précision minimale à assurer. Établie empiriquement, cette règle lie la taille l des arêtes du maillage à la fréquence maximale f_{max} qui pourra être simulée de façon fiable :

$$l < \frac{c}{6 \times f_{max}} \tag{3.1}$$

À titre d'application numérique, pour $f_{max} = 24\,\text{kHz}$ et $c = 343\,\text{m.s}^{-1}$, il sort une

taille d'arête maximale $l_{max} = 2,3\,\text{mm}$. Bien sûr, si un maillage venait à être trop grossier, il serait toujours possible de le suréchantillonner pour le faire passer sous cette limite. Néanmoins, il est aussi toujours plus reposant de savoir que les données brutes satisfont naturellement ce critère. Ceci est d'autant plus vrai que les méthodes numériques sont également sensibles à la qualité du maillage, en particulier à sa triangulation. Concrètement, plus les faces se rapprocheront de triangles équilatéraux et plus simple sera la résolution des systèmes d'équations.

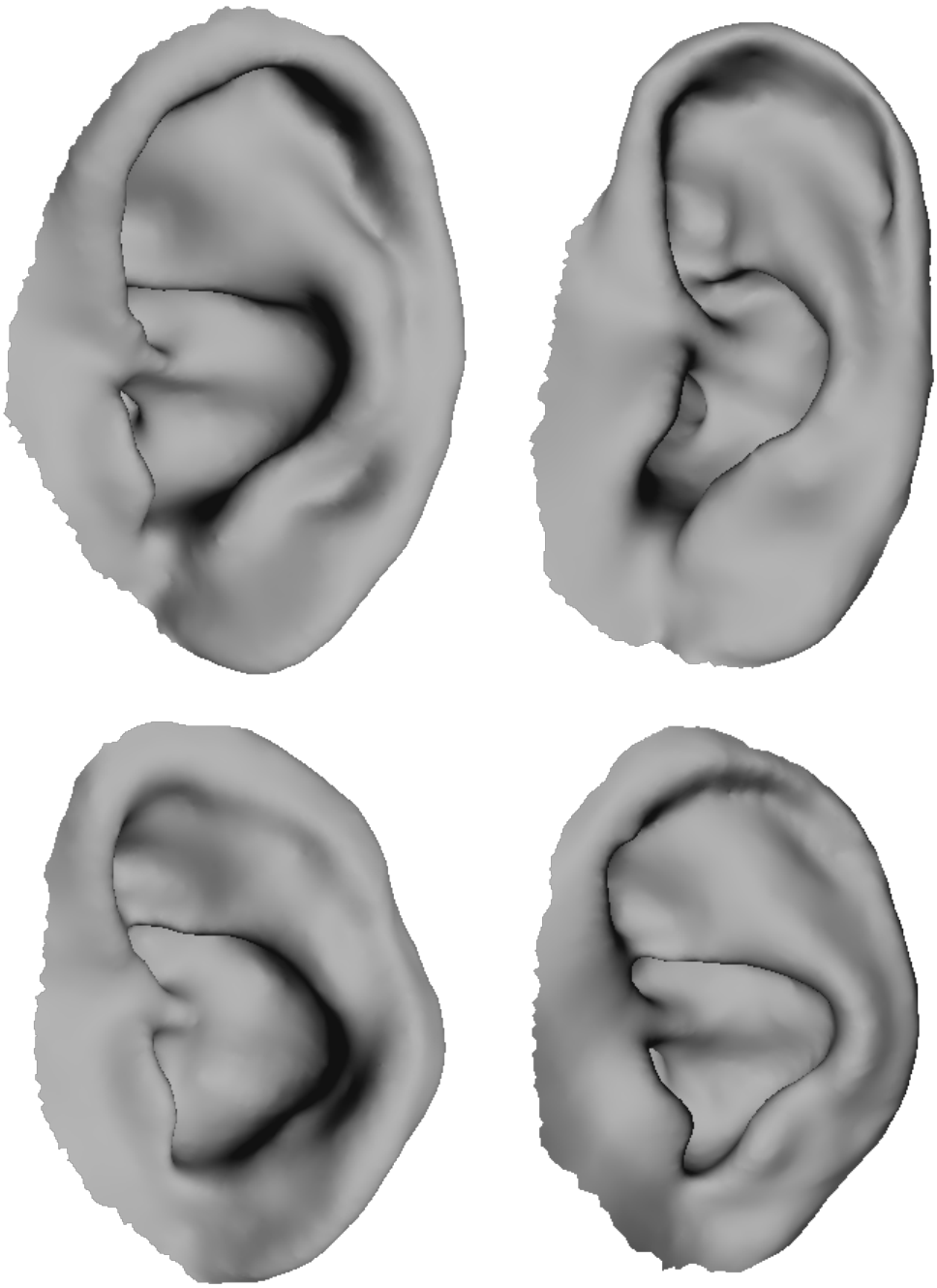


FIGURE 3.1: *Exemples de scans d'oreille.*

Dernier point : la taille globale du maillage. Selon la méthode d'acquisition, les dimen-

sions seront absolues ou seulement relatives, c'est-à-dire valables à un facteur d'échelle près. Ce dernier cas aboutissant à l'introduction d'un décalage fréquentiel dans les HRTF, il n'est pas gênant en soi dès lors que l'on connaît la valeur dudit facteur. Néanmoins, s'il est connu à l'avance, force est de reconnaître qu'il vaut mieux corriger les maillages avant la simulation plutôt que d'en retravailler les sorties.

3.1.1.2 Acquisitions

Ces différents prérequis étant établis, il a été décidé de réaliser deux types d'acquisitions par sujet. L'une consacrée à la tête et au buste, l'autre dédiée aux oreilles. Un scanner optique permettant de gérer les occlusions et de capter une partie du canal auditif – jusqu'à 2 cm – a été choisi. D'une grande précision, il renvoie un scan aux dimensions réelles de longueur d'arête moyenne 0,5 mm. Il est à noter que l'arrière de l'oreille est bien souvent difficile d'accès ou aboutit à l'obtention d'artefacts. Pour cette raison, seul l'avant du pavillon a été pris en considération. La figure 3.1 en présente quelques exemples. En ce qui concerne la tête et le buste, deux méthodes différentes, correspondant à deux campagnes d'acquisitions distinctes, ont été employées. La première a consisté en la prise d'une multitude de photos des sujets, sous une large variété d'angles – jusqu'à 80 différents –, afin d'alimenter des algorithmes de photogrammétrie. Les maillages résultants sont fidèles aux sujets d'origine mais n'ont que des dimensions relatives. Pour en retrouver la taille réelle, un alignement sur les scans d'oreille est réalisé. La seconde méthode a été d'utiliser une Kinect pour effectuer l'opération. Moins contraignante à l'usage et diminuant les efforts de nettoyage post-acquisition, elle a aussi l'avantage appréciable de produire des maillages aux dimensions réelles. La figure 3.2 présente les résultats obtenus par chacune des méthodes.

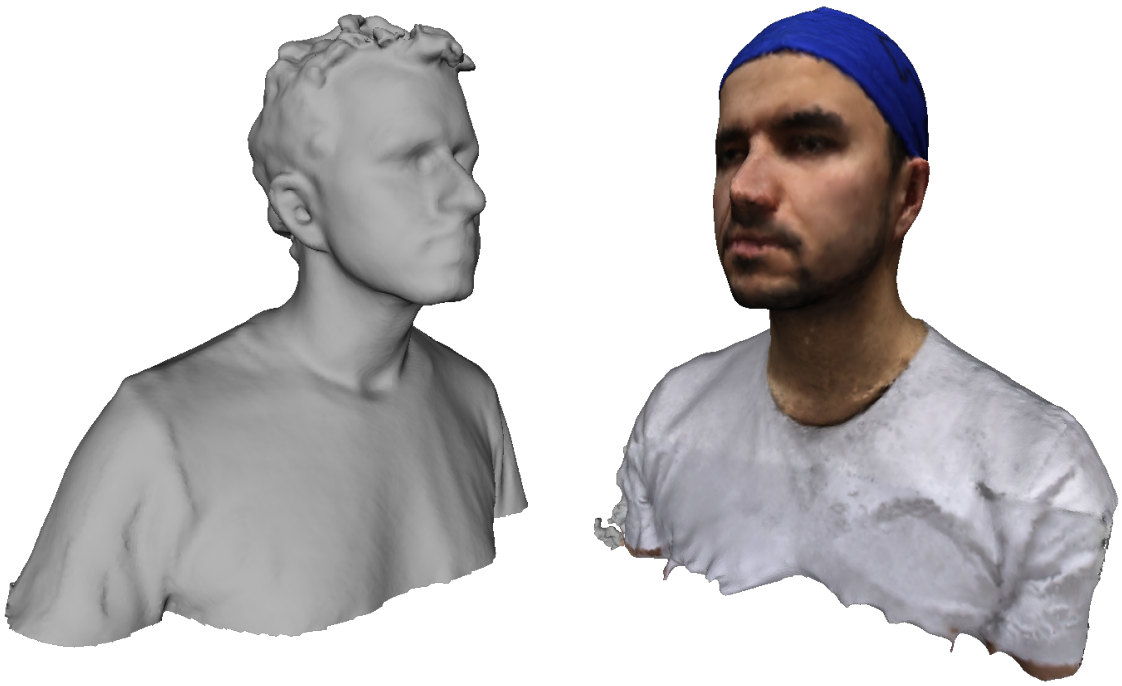


FIGURE 3.2: *Acquisitions de tête et torse. Par Kinect (à gauche) et par photogrammétrie (à droite).*

Pour ces travaux, 140 sujets se sont prêtés à ces opérations de scan entre 2015 et 2017. Cela représente donc 2×140 formes d'oreille et autant de jeux de HRTF à étudier. Forts de cette quantité de données spécifiquement destinées à notre procédé de personnalisation, nous pouvons entamer la création du modèle déformable statistique d'oreille.

3.1.2 Modèle déformable statistique d'oreille

3.1.2.1 Construction

Introduits pour la première fois par Blanz & Vetter [15] en 1999 dans le cadre de la modélisation des visages, les modèles déformables n'ont depuis eu de cesse de gagner en popularité. Aussi bien utilisés en animation 3D [163, 129] qu'à des fins de reconnaissance ou de vérification d'identité [16, 121, 26, 82], ces modèles ont été peu à peu transposés à de multiples autres sujets tels les oreilles [102, 33], le corps humain dans son ensemble [3] ou encore les squelettes animaliers [125].

Néanmoins, quels que soient les sujets d'études, les étapes de construction demeurent peu ou prou les mêmes, à savoir :

1. Acquisition de données – ici, des oreilles 3D – servant d'exemples d'apprentissage statistique.
2. Mise en correspondance dense – aussi dénommée *dense registration* – desdits exemples.
3. Création d'un espace vectoriel propre au sujet d'étude par l'utilisation d'une méthode d'analyse statistique telle que l'ACP, l'*Analyse en Composantes Indépendantes* (ACI) ou leurs dérivées.

Car l'idée de base de chacune de cette famille de modèles est d'étudier la corrélation d'un ensemble d'exemples d'apprentissage donné et de s'en resservir pour la création d'un espace représentatif plus adapté. Plus précisément, si l'on s'attache au cas de l'ACP, chaque exemple est à considérer comme la réalisation d'un jeu de variables aléatoires indépendantes. On peut donc en calculer la moyenne et la matrice de corrélation. Cette dernière ayant l'avantage d'être symétrique – dans le cas de variables à valeur dans $\mathbb{R}$, hermitienne si l'on travaille dans $\mathbb{C}$ –, le théorème spectral nous assure l'existence d'une base orthonormale de diagonalisation, et donc d'une nouvelle façon de décrire l'espace de travail. Cette base venant par ailleurs avec un ensemble de valeurs propres réelles, ses vecteurs peuvent être classés par ordre d'importance. Par conséquent, certaines parties de l'espace vont naturellement charrier plus d'information pertinente que d'autres pour la représentation des données. Mais sans même aborder les possibilités de réduction dimensionnelle offertes par la méthode, la simple décomposition précédente permet déjà de reconstruire nos données comme la somme de leur moyenne et d'une combinaison linéaire desdits vecteurs propres, que l'on peut voir comme autant de *modes de déformation*. Par extension du procédé il devient possible, en jouant sur les coefficients de la combinaison linéaire, de créer de nouveaux jeux de données de façon aussi simple qu'efficace.

Toutefois, malgré l'apparente simplicité de la méthode, sa mise en pratique doit au préalable résoudre deux problèmes de taille : déterminer quels points pourront être mis en correspondance dans chaque exemple d'apprentissage et effectuer cette association sur

un nombre suffisant de points [1]. En effet, afin de pouvoir considérer les oreilles de la base comme un même objet ayant simplement subi des déformations différentes, il est nécessaire de leur associer une certaine sémantique et de rendre leurs représentations vectorielles initiales cohérentes entre elles. Plus précisément, si les trois premières valeurs du vecteur codant pour la première oreille représentent les coordonnées (X, Y, Z) de la pointe du lobe alors il doit en être de même pour toutes les autres oreilles. C'est à cette condition seulement que le calcul du vecteur moyen pourra aboutir à une forme moyenne du sujet d'étude.

3.1.2.2 Correspondance

Malheureusement, à l'issue de la phase d'acquisition, les données n'ont à priori aucune raison de satisfaire ce type de condition et il est nécessaire de les réordonner pour y remédier. C'est l'objet de la deuxième étape de construction évoquée précédemment.

Historiquement, Blanz & Vetter ont utilisé les techniques de *flux optique* pour y parvenir [15]. Celles-ci sont adaptées au traitement de vidéos, dans lesquelles elles repèrent et quantifient les déplacements d'objets en s'attachant aux variations de pixels d'une image à l'autre. Blanz & Vetter, cherchant à construire un modèle 3D de visage, s'en trouvaient quelque peu éloignés. Toutefois, les visages ne présentant que peu d'occlusions, il leur a été possible d'en acquérir des représentations en coordonnées cylindriques et donc de travailler avec des projections 2D des visages. Bien sûr, nous ne sommes pas encore dans le cadre standard du flux optique mais il s'agit d'un premier pas essentiel. Par ailleurs, les résultats présentés témoignent de la faisabilité de la méthode.

Plus près de nous, Li *& al.* ont proposé une méthode intitulée *Triangle Mesh Hierarchy Growth* [102] pour répondre à cette problématique et l'ont appliqué au cas spécifique de la création d'un modèle 3D d'oreille. Avant toute chose, il convient de préciser que les données utilisées proviennent de la collection J2 de l'Université de Notre-Dame et que celle-ci est constituée de photos de profil de 415 sujets et de cartes de profondeurs prises simultanément et sous le même angle. L'oreille, du fait de sa géométrie est donc incomplètement scannée et le modèle final sera inévitablement lacunaire. Cette précision apportée, revenons à la méthode. Cette dernière implique d'annoter les photos d'oreille selon une procédure systématique reposant sur la détection de contours. Une fois quelques positions initialisées à la main, une série de règles géométriques simples permettent d'en augmenter le nombre. Le processus est itératif et chaque étape double le nombre de points annotés sur les contours – contour extérieur de l'oreille et contour de la conque. Une triangulation de l'ensemble est ensuite effectuée avant suréchantillonnage. De cette façon, plusieurs milliers de points peuvent être mis en correspondance sur les images 2D. Celles-ci étant prises sous le même angle que leurs pendants tridimensionnels, c'est sans difficulté que peuvent être également mis en correspondance les points 3D.

Une autre approche possible est celle adoptée par Zolfaghari *& al.* [169] – déjà à l'origine de la base SYMARE – et qu'ils appliquent aux oreilles 3D contenues dans cette dernière. Leur idée est d'utiliser la méthode *Large Deformation Diffeomorphic Metric*

1. Plusieurs milliers en règle générale.

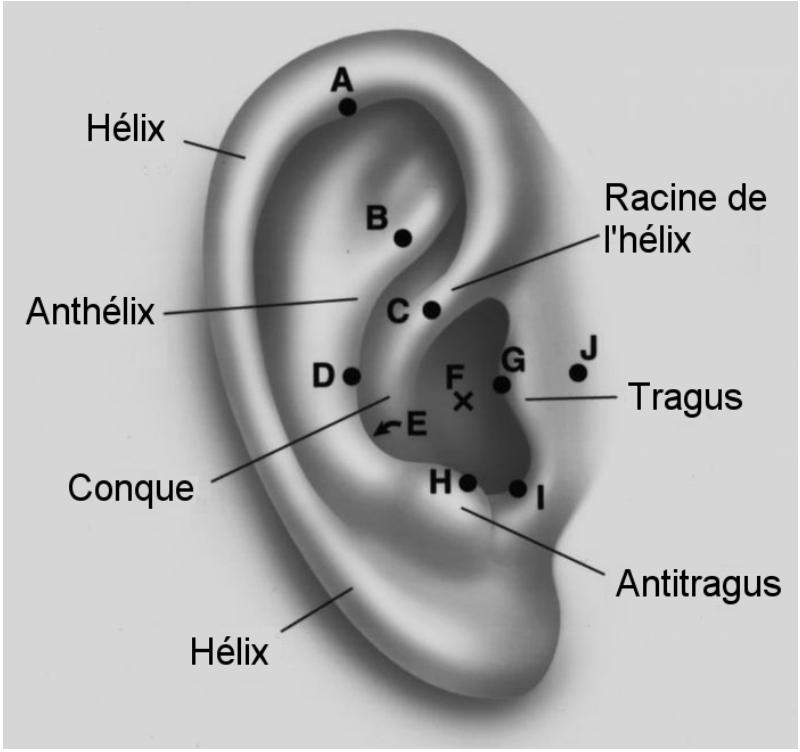


FIGURE 3.3: *Représentation schématique de l'oreille avec ses principales régions anatomiques et des exemples de marqueurs (points noirs).*

Mapping au problème de transformation d'une oreille en une autre. D'un point de vue strict, cette approche s'affranchit de la notion même de correspondance entre points et entend travailler sur les surfaces en tant qu'objets à part entière. Toutefois, dès lors qu'une surface est transformée en une autre, le chemin est court jusqu'à l'association des sommets qui constituent chacune d'entre elles. Très intéressante pour le degré d'abstraction qu'elle porte en elle, cette méthode perd toutefois la sémantique des objets traités, au sens où il n'est pas possible de définir de points repères et donc de reprendre la main sur la transformation en imposant tel ou tel comportement. En d'autres termes, rien dans cette méthode ne permet d'être sûr que le lobe d'un sujet sera bien associé aux lobes des autres, et ce qui vaut pour le lobe vaut bien évidemment pour toutes les autres parties caractéristiques de l'oreille.

La méthode mise en place dans les présents travaux navigue entre ces différentes approches et en constitue une contribution majeure. Le point de départ est la définition de marqueurs morphologiques sur les oreilles 3D de notre base, à la façon d'une mise en correspondance grossière, puis de s'en servir pour trianguläriser les surfaces à l'étude pour enfin effectuer, dans chacun des triangles, une mise en correspondance dense fondée sur les coordonnées barycentriques des points s'y trouvant. Plus concrètement, nous cherchons tout d'abord à nous ramener à un problème 2D. Malheureusement, compte tenu du type de scans obtenus, la simple projection qu'utilisaient Blanz & Vetter n'est d'aucun secours.

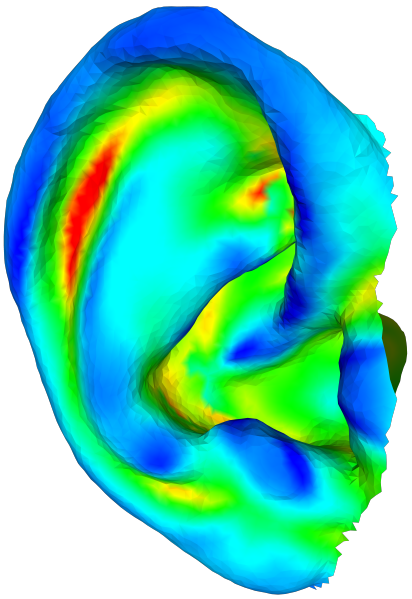


FIGURE 3.4: *Oreille colorisée selon sa courbure locale.*

Ceci étant, ces scans ne sont rien d'autre que des nappes de $\mathbb{R}^3$ et peuvent en tout état de cause être dépliés avant annotation en 2D. Néanmoins, il est impératif de pouvoir continuer à repérer les structures constitutives de l'oreille une fois le dépliement effectué. Pour cela, la courbure locale de chaque oreille est calculée en chacun de ses points puis rattachée au maillage à la manière d'une texture. La méthode *Algebraic Point Set Surface* [68, 67] est utilisée pour l'occasion. Un exemple de résultat est présenté figure 3.4.

Une fois les oreilles colorisées, l'algorithme *Least Squared Conformal Mapping* [101] est utilisé pour le dépliement proprement dit. Ne reste plus alors qu'à réaliser les associations de points. Pour simplifier la gestion de l'opération, une oreille dite de référence est choisie, sur laquelle sont positionnés 38 marqueurs spécifiques. Pour faciliter leur positionnement, les lignes d'iso-courbure sont surimposées aux oreilles 2D. Concernant les localisations choisies en elles-mêmes, les structures les plus caractéristiques de l'oreille sont retenues. Dans les faits, le tragus, l'anti-tragus, l'hélix et l'antihélix – entre autres – sont repérés.

Ensuite, ces points sont utilisés pour effectuer une triangulation de Delaunay – cf. figure 3.5 – , ayant pour effet de subdiviser notre oreille de référence. Les mêmes marqueurs sont alors repérés sur l'ensemble des autres oreilles, permettant de leur transférer la triangulation de référence et leur conférant par là même une topologie commune. Un tel transfert est présenté figure 3.6.

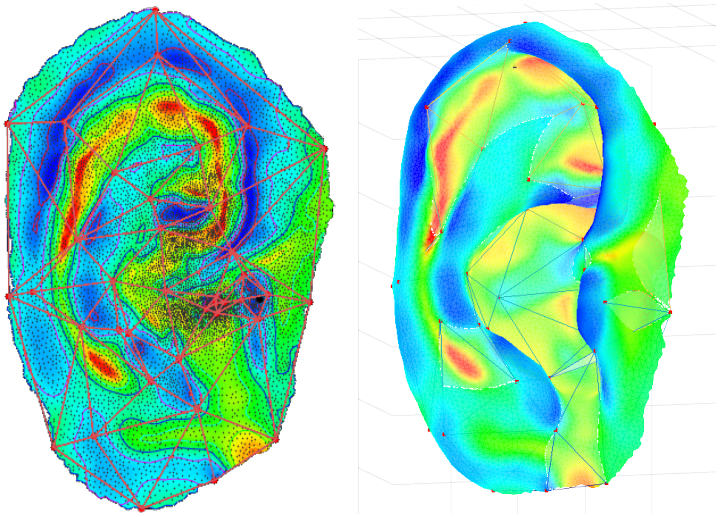


FIGURE 3.5: *Triangulation de l'oreille de référence, en 2D (à gauche) et en 3D (à droite). Les points en rouge sont les marqueurs annotés manuellement. Ceux en noir – oreille 2D uniquement – sont les points restants du maillage.*

À ce stade, il serait déjà possible de s'atteler à l'étape d'analyse statistique. Cependant, le modèle obtenu ne comporterait alors que 38 points, trop peu pour assurer une bonne résolution spatiale. Afin

d'augmenter le nombre de marqueurs, une étape de mise en correspondance automatique est ajoutée. L'idée est d'utiliser l'unicité de la topologie introduite précédemment et de travailler à l'échelle du triangle. En pratique, cette unicité nous permet de numéroter chaque triangle et d'avoir l'assurance que le n-ième d'entre eux couvrira une zone similaire sur chaque oreille.

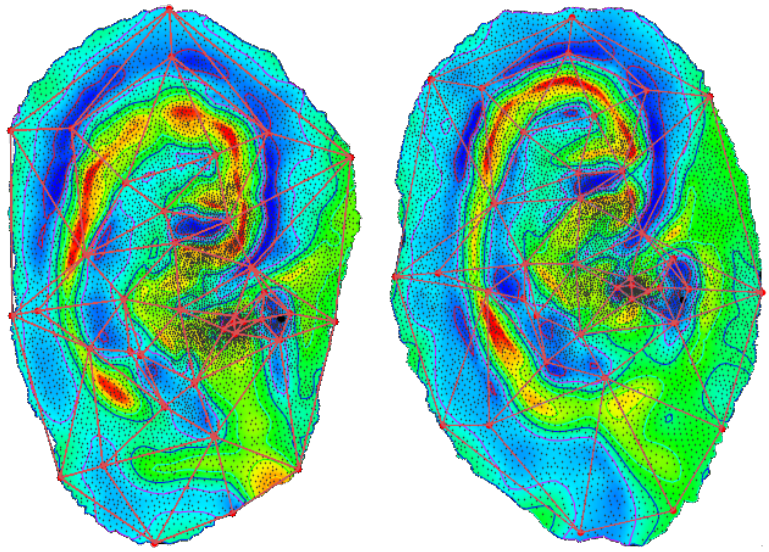


FIGURE 3.6: *Triangulation de l'oreille de référence (à gauche) et la même triangulation transférée sur une autre oreille de la base (à droite).*

Pour mieux fixer les idées, plaçons-nous désormais dans ce n-ième triangle et dotons-le de son système de coordonnées barycentriques. Les scans originaux comprenant plusieurs milliers de points, notre triangle en contient aisément entre plusieurs dizaines et quelques centaines, chacun d'entre eux étant repérable par ses coordonnées. Ceci étant valable quelle que soit l'oreille, ça l'est à fortiori pour l'oreille de référence. Par conséquent, nous pouvons parcourir l'ensemble des points intérieurs au n-ième triangle de l'oreille de référence, noter leurs coordonnées barycentriques et pour chacun, l'associer aux points des autres oreilles dont les coordonnées barycentriques – dans leur n-ième triangle – sont les plus proches de celles du point à l'étude. Ainsi, une association plus dense de points peut être obtenue. Toutefois, c'est au prix d'une certaine part de perte. En effet, le remplissage d'un triangle dépend aussi de l'oreille considérée, certains étant plus dotés en points que l'oreille de référence, d'autres l'étant moins. En conséquence, tous les points ne peuvent être mis en correspondance et ceci est d'autant plus vrai que le nombre d'entrées de la base est grand. La figure 3.7 illustre le résultat de la procédure.

Afin de remédier à cette perte, plutôt que de sélectionner des points existants des autres oreilles, nous en créons de nouveaux. Pour un point P_{ref} de coordonnées (u, v) donné de l'oreille référence, nous sélectionnons les trois points les plus proches entourant le point P_i de mêmes coordonnées dans l'oreille candidate à ajouter au modèle. Cela est très simplement fait en effectuant une « sous-triangularisation » à partir des point contenus dans le n-ième triangle de l'oreille candidate puis en sélectionnant le « sous-triangle » dans lequel tombe le point de coordonnées (u, v). Ce sous-triangle définit naturellement un nouveau repère barycentrique. Nous notons (u_i, v_i) les nouvelles coordonnées de P_i dans ce repère.

Puisque nous connaissons par ailleurs les coordonnées (x, y, z) des points définissant notre nouveau repère, il est aisé de remonter aux coordonnées (x_i, y_i, z_i) de P_i, ce qui achève la création de ce point. En ré-échantillonnant de la sorte toutes les oreilles, on aboutit à une mise en correspondance quasi-totale de la référence avec les autres entrées, et ce, indépendamment de la taille de la base. Le résultat est un ensemble d'oreilles correctement ordonnées pour alimenter l'analyse statistique qui suit.

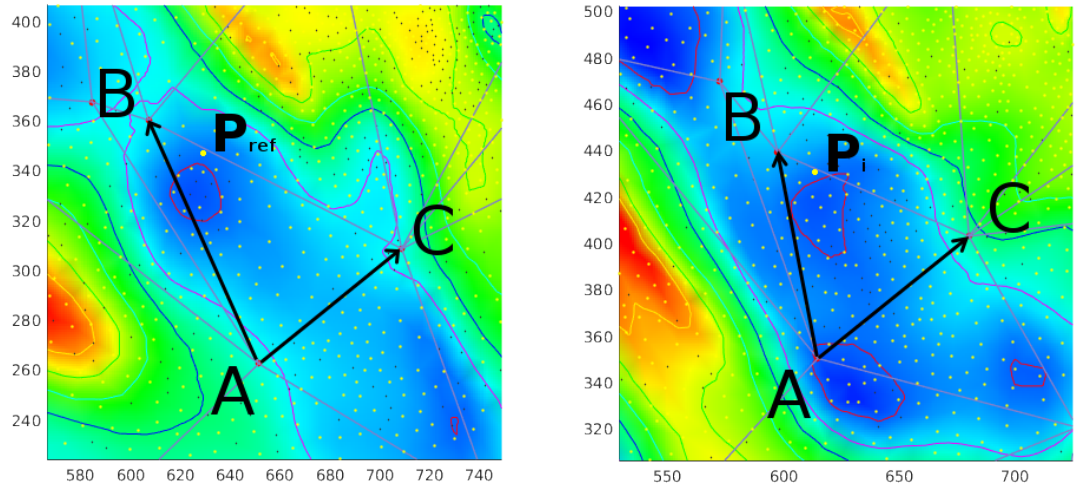


FIGURE 3.7: *Mise en correspondance barycentrique par recherche du plus proche voisin des points d'un même triangle. À gauche, l'oreille de référence. À droite, une autre oreille. En jaune, les points ayant été mis en correspondance. En noir, les points laissés de côté à l'issue de l'opération. On peut constater la présence de points noirs aussi bien dans le triangle de référence que dans l'autre. Les coordonnées barycentriques prises en compte sont celles des repères* $(A, \overrightarrow{AB}, \overrightarrow{AC})$ *de chaque oreille. À titre d'exemple, la correspondance entre* P_i *et* P_{ref} *est mise en exergue.*

3.1.2.3 Analyse statistique

Comme indiqué en début de section, l'étape finale de création du modèle statistique est l'application d'une ACP sur les entrées de la base. Celle-ci peut sans difficulté être effectuée sur les données précédemment mises en correspondance. Néanmoins, les variations de positions, d'orientation et d'échelle des différents exemples d'apprentissage seront alors intégrées dans l'analyse. Afin de s'affranchir de ce que l'on peut qualifier de manière générique de *paramètres de pose*, une étape de normalisation est introduite. Elle consiste aligner chaque oreille sur la référence. Pour cela, une transformation procrustéenne est effectuée. Seules subsistent par la suite les différences intrinsèquement dues aux formes des oreilles. On notera dès à présent que les paramètres de pose ainsi pris en compte sont à garder en mémoire pour les futurs traitements, en particulier lors de la modélisation de HRTF. La figure 3.8 illustre cette étape.

Ces préparatifs achevés, l'ACP permet enfin d'obtenir le modèle tant attendu. Celui-ci consiste donc en un vecteur moyen – une forme moyenne d'oreille – et des déformations portées par les vecteurs propres.

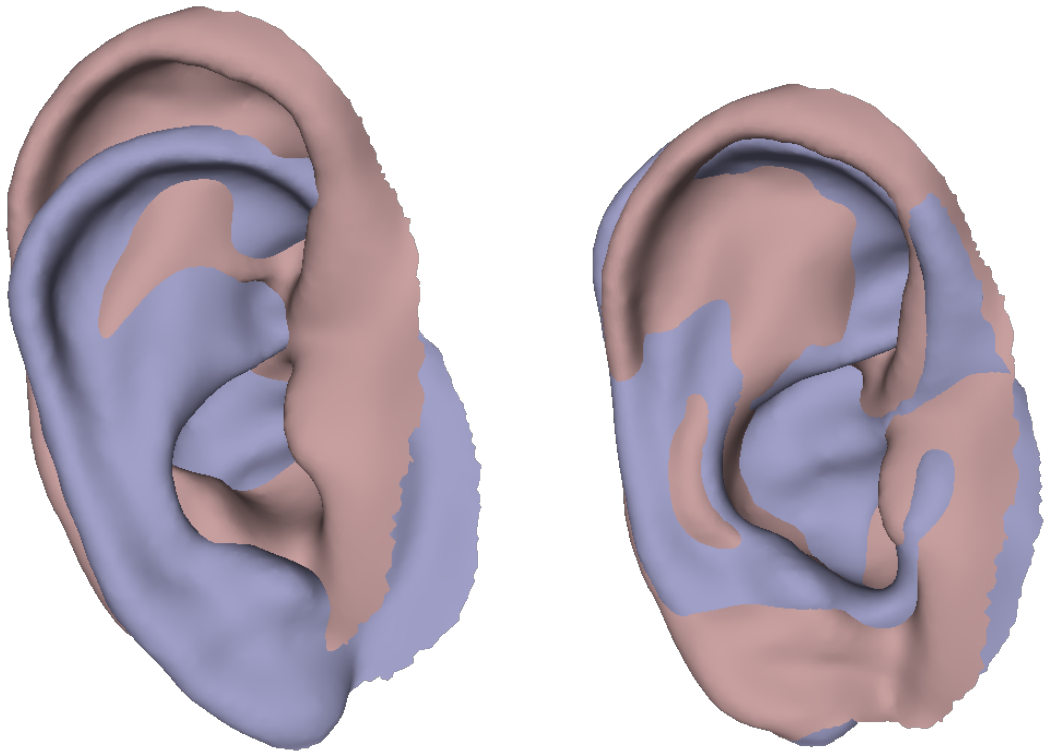


FIGURE 3.8: *Effets de l'analyse procrustéenne sur une oreille de la base. État originel à gauche. Après transformations à droite. L'oreille de référence en rose, une autre oreille de la base en violet.*

3.2 Modèles paramétriques complets

3.2.1 Modèles synthétiques

Puisqu'il est construit à partir de données réelles, le modèle d'oreille précédent est à la fois assuré d'être réaliste et de porter en lui un maximum de variabilité. Toutefois, aucun contrôle ne peut être posé sur les modes de déformation qui le composent. Ceux-ci sont entièrement déterminés par les données sous-jacentes. Or certaines applications, comme l'étude de l'influence de tel ou tel paramètre anthropométrique sur les HRTF, peuvent requérir une capacité à superviser finement de quelle manière peuvent se déformer certaines parties spécifiques du maillage. Est-ce la forme de la conque qui fait la HRTF ? Le lobe d'oreille a-t-il son mot à dire ? N'a-t-on vraiment aucun besoin d'information sur le canal auditif ? Voilà autant de questions que l'on est en droit de se poser dès lors que l'on traite du problème de la personnalisation de HRTF et dont les réponses permettraient de séparer le superflu du nécessaire.

Pour toutes ces raisons, un modèle paramétrique de torse, de tête et d'oreilles, doté d'un fort réalisme général, a été réalisé. Il a par la suite donné lieu à la constitution d'une base de données de HRTF synthétiques.

L'une des premières contraintes à laquelle ce modèle doit se conformer est de pouvoir déplacer / agrandir / rétrécir une zone précise indépendamment du reste du maillage. Pour

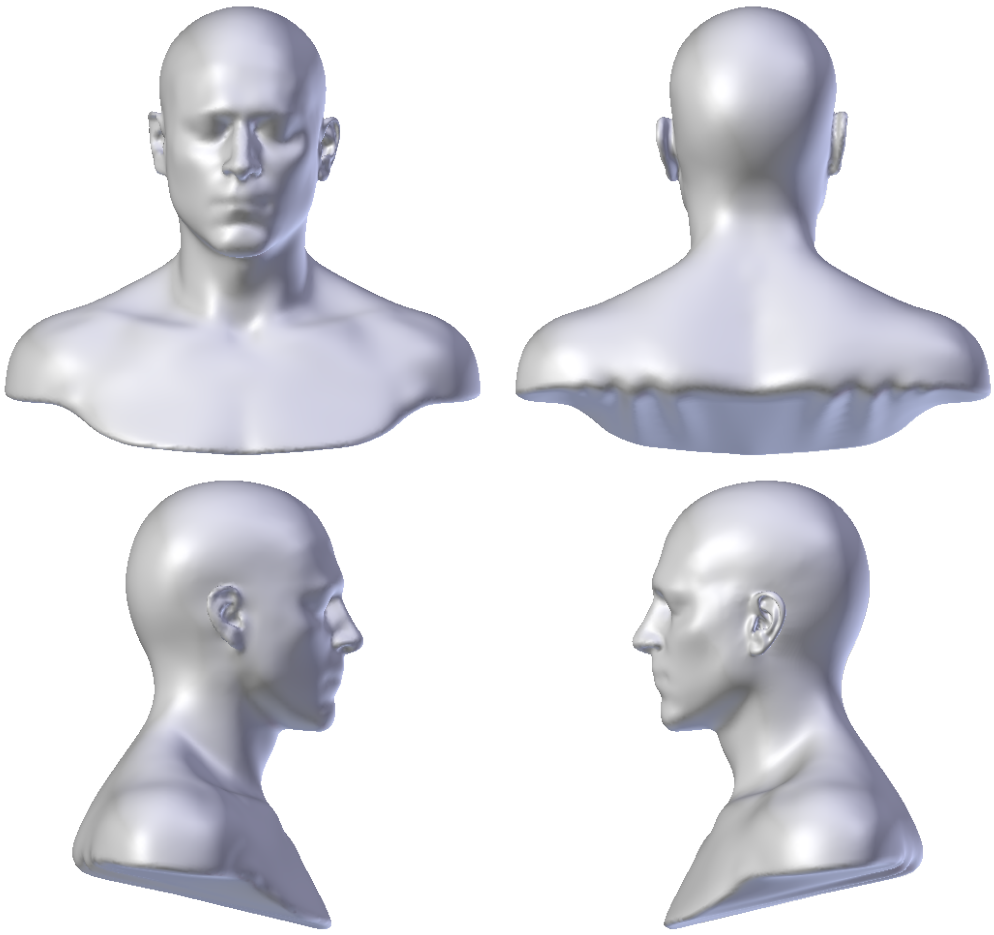


FIGURE 3.9: *Vues d'ensemble du modèle synthétique.*

ce faire, la première étape est de définir lesdites zones d'intérêt. Le modèle doit en effet rester simple d'usage, ce qui passe par un paramétrage aussi épuré que possible. Dans cette optique, seule une oreille du modèle – celle de gauche – est ici rendue déformable. Au besoin, une symétrie par rapport au plan médian permet d'obtenir un maillage avec une oreille droite déformée. De plus, pour se comparer facilement à l'état de l'art, il est utile d'y incorporer les déformations les plus souvent citées dans la littérature.

Les dimensions θ_1, θ_2 et d_1 à d_6 telles que définies par CIPIC ont été reprises pour les déformations d'oreille. Un paramètre de taille permet d'agrandir ou de réduire l'échelle de l'oreille par rapport à la tête. Les modifications possibles de la tête, du cou et du torse sont dans chacun des cas la hauteur, la largeur et la profondeur de l'élément considéré. L'emplacement des oreilles sur la tête fait également partie des éléments modifiables, tout comme la largeur d'épaules. La tête peut ainsi être agrandie ou réduite selon l'axe X, Y ou Z. Les rotations du cou ont été intégrées pour de futures recherches. En ce qui concerne l'oreille, un attention particulière de modélisation a été portée au canal auditif, capable d'élongation, contraction et d'une torsion sur le côté. De plus, deux déformations spécifiques de forme ont été définies : la taille de la racine de l'hélix, qui parcourt la conque de part en part et se trouve être très marquée chez certains individus et un « effet parabole »,

qui entend en quelque sorte « déplier l'oreille ». Dans la suite du manuscrit, elles seront respectivement appelées paramètres R et P. La figure 3.10 présente l'oreille de base et illustre ces deux nouvelles déformations.

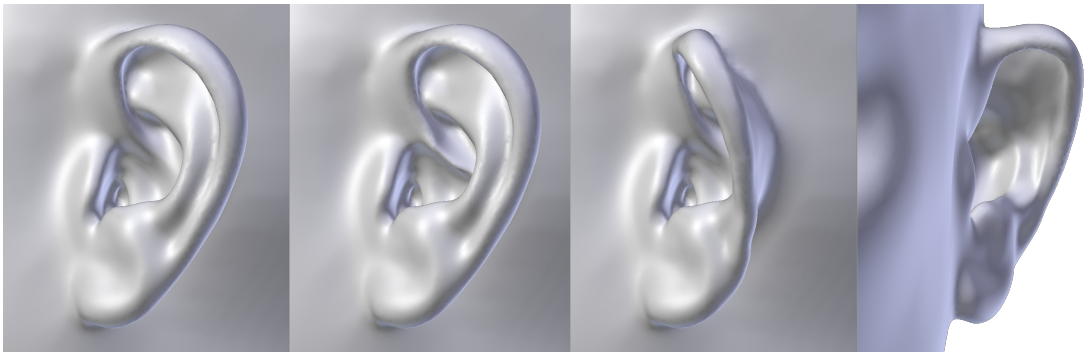


FIGURE 3.10: *Oreille du modèle synthétique. De gauche à droite : la forme de base, la déformation « racine de l'hélix » (paramètre R), deux vues de la déformation « parabole » (paramètre P).*

À l'occasion de la conception de ce modèle, quelques problèmes pratiques sont apparus, sur lesquels il peut être utile de faire un bref retour d'expérience. En premier lieu, la difficulté à mesurer dans $\mathbb{R}^3$ des distances définies dans $\mathbb{R}^2$, c'est-à-dire dans un plan de projection. Car, dans le cas des paramètres CIPIC, ce plan n'est pas précisément défini. S'agit-il du plan orthogonal à l'axe interaural ? Ou est-ce celui maximisant la surface de l'oreille projetée ? Dans le premier cas, cela implique que deux oreilles de formes identiques mais présentant deux angles de décollement différents par rapport à la tête auront des largeurs apparentes distinctes et donc des valeurs également différentes pour les paramètres d_3 et d_6, ce qui est illogique. Nous devons donc écarter cette solution. Dans le second, la tâche est alors extrêmement ardue et, bien qu'encore possible en simulation, elle est irréaliste dans des conditions réelles de prises de photos tant le biais apporté par le choix du photographe est alors important. Il est donc préférable de l'abandonner également. Et plus largement encore, il nous faut abandonner l'idée de paramètres 2D et tirer parti du fait que nos données sont tridimensionnelles. Autant que nécessaire, les définitions des paramètres à mesurer ont donc été adaptées au monde 3D. Les logiciels classiques de visualisation 3D offrant tous des outils de mesure de distance entre points, la récupération de leurs valeurs est à porter de main.

Vient ensuite le problème des collisions, c'est-à-dire des intersections du maillage avec lui-même. Celles-ci sont à proscrire en simulations numériques mais constituent un risque à priori inévitable dès lors que l'on applique une ou plusieurs déformations dont les zones d'influences se recoupent. Par exemple, ce cas est susceptible de se rencontrer si l'on applique une déformation réduisant la largeur globale de l'oreille en conjonction avec une déformation augmentant la largeur de la conque. De petites variations ne poseront à priori pas de problème mais on comprend aisément qu'il est risqué de les augmenter inconsidérément. Certaines précautions peuvent cependant être prises lors de la construction du modèle pour en limiter le nombre. En particulier, la définition soigneuse des zones d'influence, à savoir

l'ensemble des sommets concernés par une déformation, a un impact significatif à l'usage. Pour une déformation donnée, cela se traduit par la définition d'une transition douce entre zone déformable et zone rigide. À l'échelle du modèle en lui-même, cela passe par des déformations aussi indépendantes que possible, i.e. dont les zones d'influence se recoupent aussi peu que possible. En parallèle, l'utilisation d'une triangulation très régulière est un plus. Ces précautions, ajoutées à une étape de détection et de nettoyage éventuel des maillages en entrée de simulation, permettent en pratique de réduire considérablement les rejets.

Dernier point mais non des moindres, ce modèle vient avec un ensemble de points caractéristiques servant de référence de mesure. De cette manière, le processus d'extraction de paramètres anthropométriques est systématisé, reproductible et indépendant de tout opérateur. Inversement, la construction d'un maillage à partir d'un ensemble de valeurs de tels paramètres devient également possible, et de manière industrielle, ce qui permet de programmer la génération à grande échelle d'une base de maillages 3D couvrant une large variété de morphologies.

3.2.2 Modèle mixte

Dorénavant, nous disposons donc d'un modèle déformable d'oreilles réelles et d'un modèle déformable synthétique de tête, torse et oreille. Une situation idéale serait de pouvoir fusionner les deux et d'allier ainsi la variabilité naturelle de la base réelle à la complétude des données synthétiques. Toutefois, une telle opération est d'une grande complexité infographique et n'a pu être réalisée. Néanmoins, une solution intermédiaire a été mise en place, dans laquelle le modèle déformable d'oreille est fusionné avec un torse fixe, donnant naissance à ce que l'on appellera par la suite le modèle *mixte*.

En pratique, un morceau de tête en forme de médaillon et une partie arrière d'oreille ont été ajoutés à la référence du modèle d'oreille (cf. figure 3.11). Puis cet ajout a été transféré aux autres sujets de la base. L'algorithme des *Thin Plate Splines* a été utilisé pour cela. Il a permis d'adapter l'arrière à la morphologie de chaque oreille tout en gardant celle-ci parfaitement intacte. La frontière du médaillon s'est quant à elle vu imposer un positionnement fixe. De cette façon, la fusion des médaillons à la tête du modèle a été grandement facilitée. Il est à déplorer qu'au cours de ce processus, 10 maillages sur les 140 initiaux ont exhibés des intersections de faces, les rendant impropres au calcul de HRTF. Pour tous les autres, nous avons grâce à cette procédure pu transférer une géométrie réelle d'oreille sur un buste de référence. Ces maillages étant encore en correspondance, une simple ACP permet de donner naissance au modèle mixte, fusion d'un buste et d'une tête synthétiques à un modèle déformable d'oreilles scannées. La forme moyenne ainsi que les dix premières composantes en sont disponibles à l'annexe C.

3.3 De la 2D à la 3D : Optimisation sur photos

Dernier élément de ce chapitre consacré aux modélisations morphologiques mais surtout point d'entrée du système de personnalisation tel qu'il se présente à l'utilisateur final : la

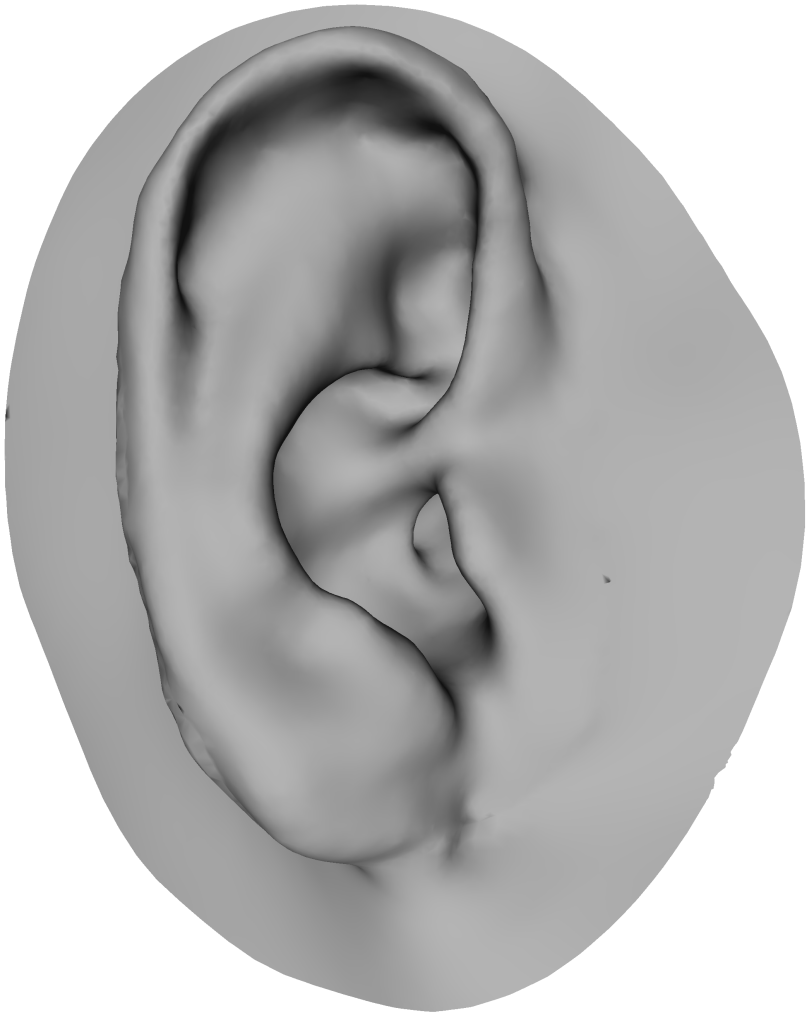
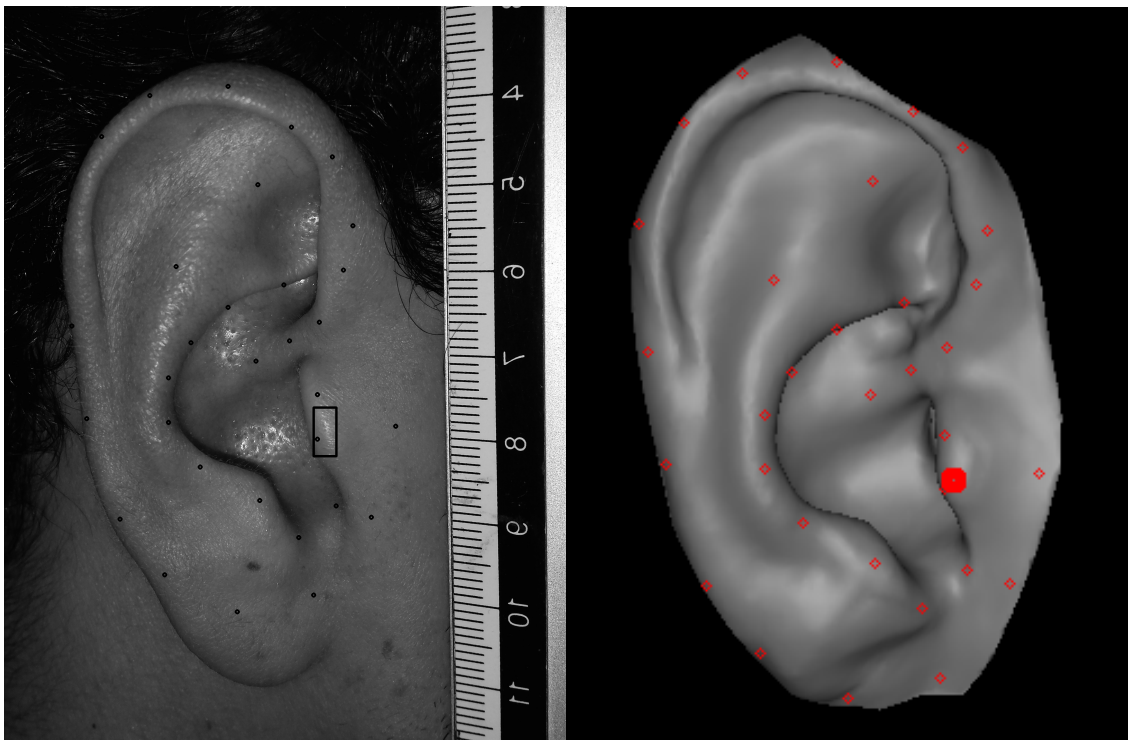


FIGURE 3.11: *Oreille augmentée.*

transformation de photos d'oreilles en leurs maillages 3D.

En effet, qu'elles soient mesurées ou calculées, les HRTF sont par nature dérivées de données morphologiques 3D. Or la personnalisation proposée section 2.6 repose entièrement sur l'utilisation de données 2D, nommément des photos des utilisateurs. Notre objectif – l'obtention de HRTF à partir de photos – requiert donc que l'on s'attache à reconstruire les données 3D à partir de celles en 2D. La prise de photos équivalant peu ou prou à une projection mathématique, et celle-ci étant notoirement non bijective, le problème majeur d'une telle entreprise est donc de reconstituer l'information perdue. Le modèle statistique d'oreille va pour cela se révéler être un atout majeur.

Sans surprise, les outils infographiques de modélisation actuels permettent tout à fait de créer de toutes pièces des scènes d'un réalisme saisissant. Cet état de fait est ici mis à



(a) (b)

FIGURE 3.12: *Photo d'utilisateur (a) et l'oreille synthétique induite par le modèle (b).*

profit pour créer des oreilles réalistes à partir du modèle 3D. Celles-ci sont ensuite prises en photo – de manière numérique ; pas de réel appareil ici –, sous des angles choisis et dans des conditions de luminosité contrôlées. Ces photos synthétiques peuvent alors être comparées à de réelles photos d'oreilles, et une mesure d'erreur peut en être extraite. Ainsi, les bases d'une procédure d'optimisation de la forme que doit prendre le modèle d'oreille sont posées. Les paramètres de cette procédure sont essentiellement les paramètres du modèle et la minimisation du critère d'erreur, c'est-à-dire l'obtention d'une photo synthétique aussi proche que possible de la photo réelle, aboutit à la forme d'oreille 3D la plus proche de celle de l'utilisateur. La figure 3.12 en présente un exemple de résultat.

Cette dernière brique, développée par les équipes de la société 3D Sound Labs, constitue l'ultime maillon de cette première partie de personnalisation, centrée sur la morphologie de l'individu. Il nous reste alors à nous atteler à l'étude des HRTF proprement dites et aux liens que l'on peut établir entre le monde morphologique et le monde auditif.

Chapitre 4

Production de HRTF

C'est difficile, de bien voir, de bien entendre, de tout sentir, sans filtre.

- Justine Lévy - *Rien de grave*

Maintenant que nous disposons d'un modèle 3D complet, nous pouvons envisager la production de HRTF et la mise en place de la brique « Simulations numériques » figurant sur le schéma de fonctionnement 2.11. Toutefois, étant donnée la position critique que cette brique occupe, un travail préliminaire de vérification et d'optimisation est indispensable. Dans ce chapitre, nous nous attacherons donc à vérifier par la simulation que la bonne connaissance du canal auditif est superflue dans notre cas. Nous nous intéresserons également à l'impact que peut avoir la représentation des haut-parleurs sur les HRTF et les DTF, de façon à en définir le meilleur paramétrage possible. Nous présenterons ensuite deux optimisations numériques conduisant à des gains substantiels, tant en durée de calcul qu'en puissance nécessaire. Enfin et surtout, nous étudierons avec attention l'impact des conditions aux limites choisies – c'est-à-dire de l'impédance acoustique des matériaux mis en jeu – sur les HRTF résultantes. Nous verrons que ce point à lui seul mérite d'importantes recherches.

Il convient à ce stade de l'étude de préciser que l'ensemble des simulations numériques ont été réalisées à l'aide du solveur *mesh2hrtf*. Celui-ci présentant par ailleurs une forte intégration avec les logiciels d'infographie *Blender* et *OpenFlipper*, ces derniers ont également été privilégiés pour réaliser les préparatifs de simulation. Le premier nous permet de manipuler simplement les modèles déformables et de générer de nouveaux maillages tandis que le second nous donne les moyens de les remailler convenablement.

4.1 Focus sur le canal auditif

Joignant le tympan au pavillon, le canal auditif est un passage obligé des ondes sonores. D'une longueur moyenne de 2,5 cm et présentant une forme en « S » caractéristique, son positionnement stratégique tout comme les questions qu'il a déjà soulevées dans la littérature demandent à ce qu'on lui prête une attention particulière. En effet, son emplacement central fait qu'une mauvaise appréciation de son influence aurait des conséquences dramatiques sur l'ensemble des travaux ultérieurs.

Et l'enjeu s'en trouve encore accru si l'on considère son extrême difficulté d'accès ! Une personnalisation nécessitant une connaissance fine du canal auditif du sujet ne saurait à priori se passer du concours d'un professionnel pour en prendre les mesures, donnant par là même un sérieux coup de frein à la démocratisation de l'écoute binaurale pour chacun. Et même ainsi la partie ne serait pas gagnée tant il est vrai que sa géométrie n'est pas unique mais qu'il voit son tracé et son diamètre modifiés par nos mouvements de mâchoire.

Du rôle que l'on doit donner ou non au canal auditif dans le cadre de la spatialisation du son dépendent donc deux épées de Damoclès, l'une menaçant la validité des expériences en vue, l'autre la viabilité de tout procédé de personnalisation grand public reposant sur la morphologie.

4.1.1 Représentation du canal auditif en simulation

Une représentation possible du canal est celle d'un guide d'onde sonore. De la même manière qu'un tube métallique peut acheminer sans encombres une onde électromagnétique

de l'une à l'autre de ses extrémités ou que la fibre optique fait de même avec des ondes lumineuses, le canal auditif permettrait de véhiculer sans altération majeure les ondes sonores de son entrée, située à la jonction avec le pavillon, jusqu'au tympan.

On voit poindre ici l'intérêt évident d'une telle modélisation : elle évacue purement et simplement tout impact sur la localisation sonore. Le rôle tenu par le canal est alors restreint à un travail de protection du tympan, qui, par son intermédiaire, peut être à l'écoute du monde extérieur sans pour autant devoir s'exposer aux multiples agressions qui s'y trouvent. La conséquence concrète pour les HRTF est qu'elles pourraient être captées – ou simulées – en n'importe quel point [1] du canal auditif.

La littérature fait mention de plusieurs expériences visant à répondre à ce questionnement. Dès 1946, Weiner & Ross [153] réalisent des mesures de pression acoustique au niveau du canal auditif de plusieurs sujets. Une sonde flexible pouvant aller jusqu'au tympan est utilisée pour y capter le champ acoustique. La même captation est également effectuée au milieu et à l'entrée du canal. Le stimulus est généré par un haut-parleur successivement placé à différentes positions du plan azimutal – 0°, 45° et 90° – et couvre la bande de fréquences [200, 8 000] Hz. L'ensemble des mesures est effectué dans une chambre semi-anéchoïque. À l'issue de l'expérience, il est observé que le champ au niveau du tympan présente une résonance allant jusqu'à 20 dB aux alentours de 3,5 kHz et une anti-résonance vers 7 kHz, et ce, indépendamment de la position de la source. Ces mesures sont en accord avec les valeurs théoriques de résonance d'un tube rigide bouché d'une longueur de 2,5 cm, elle-même en accord avec la taille typique d'un canal auditif. Les auteurs précisent également que des mesures effectuées avec et sans la présence d'une seconde sonde à l'intérieur du canal ont montré des différences de niveaux n'excédant pas 1 dB, leur permettant ainsi de conclure au caractère négligeable des perturbations induites par la présence du microphone.

En 1972, Shaw [137] prolonge cette expérience en s'intéressant cette fois à des variations d'angle d'incidence en élévation de $-15°$ à $135°$ par rapport au plan azimutal pour des fréquences allant de 300 Hz à 20 kHz. Un oreille artificielle est utilisée pour l'occasion. Sa conclusion est que la transmission du son par le canal est indépendante de l'angle d'incidence pour les fréquences allant jusqu'à 14 kHz. Il note néanmoins que pour des positions légèrement décalées de l'axe central, cette indépendance ne peut être établie que jusqu'à une limite légèrement inférieure et attribue cette variation à la présence de modes transverses de résonance.

En 1989, Middlebrooks *et al.* [112] mènent une étude sur six sujets incluant 356 directions différentes. L'une des idées y est de mesurer et de comparer les fonctions de transfert – jusqu'à 16 kHz – en deux points du canal auditif, espacés de 9 mm l'un de l'autre. L'analyse statistique des résultats leur fait observer une très forte correspondance des fonctions de transfert, signe d'une indépendance de la mesure vis-à-vis du point choisi. À noter cependant qu'ils rapportent la présence de variations substantielles des spectres liés à la présence des microphones, prenant ainsi le contre-pied de Weiner & Ross.

Quelques années plus tard, en 1996, Hammershoï & Moller [72] apportent également leur pierre à l'édifice en se penchant sur les mesures de HRTF de plusieurs sujets, captées en

1. N'importe lequel, donc le plus pratique.

différents points à l'intérieur et à l'extérieur du canal. Trois hauts-parleurs ont été utilisés : l'un situé en face du sujet, un autre à l'arrière et le dernier face à l'oreille. Cette étude se limite donc à trois directions du plan azimutal. Pour sa part, la bande de fréquences va de 200 Hz à 25 kHz. Leur conclusion est qu'une concordance existe entre la mesure prise dans le canal, celle prise à son entrée et celle prise à 6 mm à l'extérieur du canal. Au-delà, des variations importantes apparaissent et les points de mesures ne peuvent plus être traités comme équivalents du point de vue de la directionalité. À la lecture des tracés des HRTF présentées dans l'article, cette conclusion n'est cependant pas si évidente, le champ diffus [2] n'ayant pas été soustrait des spectres. À noter également une remise en cause du modèle du tube rigide fermé à une extrémité, le second mode de résonance n'étant généralement pas à la fréquence théorique attendue par rapport au premier.

Ainsi, malgré le faisceau concordant d'indices nous amenant à considérer acquise cette précieuse indépendance vis-à-vis du point de mesure, pour autant qu'il soit dans le canal, certaines oppositions et certains biais de mesure surgissent néanmoins çà et là, entretenant de fait un flou épineux autour de ce sujet crucial. Parmi les éléments les plus notables :

1. les différences extrêmes entre les jeux de directions mesurées, allant de trois pour Weiner & Ross et Hammershoï & Moller à 356 chez Middlebrooks *et al.*
2. les domaines de validité des conclusions, directement liés aux bandes de fréquences utilisées, et variant donc de [200, 8 000] Hz chez Weiner & Ross à [300, 25000] Hz pour Hammershoï & Moller.
3. l'impact du système de captation, négligeable pour Weiner & Ross mais observable voire nuisible – bien que non invalidant – pour Middlebrooks *et al.*

Or chacun de ces sujets est aisément pris en charge par le calcul numérique de HRTF. En effet, le nombre de points de mesure est quasi-illimité et l'on peut ainsi avoir une cartographie complète des éventuelles variations. Les fréquences simulées peuvent sans problème couvrir le spectre audible. Le système de captation a un volume nul et ne risque donc pas de perturber la mesure.

Il est à noter que ces avantages inhérents à la simulation numérique n'ont pas échappés à Ziegelwanger *et al.*, du laboratoire ARI, à l'origine de *mesh2hrtf*. Afin de qualifier la validité des HRTF produites par leur moteur de calcul, ils effectuent en 2015 une série d'expérimentations [168] dont certaines recoupent les expériences qui suivent et pourront offrir un bon point de comparaison.

Bien évidemment, il n'est pas perdu de vue que la validité des conclusions tirées de l'analyse de résultats de simulations est à jauger à l'aune de la validité des hypothèses simplificatrices de la modélisation physique adoptée. À titre d'exemples, on peut questionner la non-modélisation de la pilosité de l'oreille ou de la présence de cérumen.

Cette précaution d'usage étant prise, nous pouvons passer à l'expérimentation proprement dite et rappeler en premier lieu son objectif dans le cadre des présentes recherches. Il s'agit ici de déterminer l'impact d'une connaissance imparfaite de la forme du canal

2. Même s'il est peut-être un peu abusif de parler de champ diffus lorsque seules trois directions ont été mesurées.

auditif ainsi que l'impact du choix du point de mesure [3]. En particulier, le but n'est donc pas de vérifier la validité de la modélisation du tube rigide bouché ni de déterminer la distance exacte à partir de laquelle on se sera trop éloigné de l'entrée du canal pour estimer correctes les HRTF, mais de déterminer les écueils à éviter dans *notre* cas d'utilisation. C'est ce que proposent de déterminer les deux séries de simulations qui suivent.

4.1.2 Source ponctuelle vs éléments vibrants

4.1.2.1 Protocole expérimental

La première de ces deux séries a consisté en un calcul de HRTF pour différentes positions de la source sonore. Le maillage utilisé est commun à toutes les simulations. Nous avons sélectionné la morphologie de base du modèle synthétique et modifié la forme du canal auditif de façon à obtenir le résultat visible figure 4.1. La source voit sa position varier du fond du canal jusqu'à son entrée. Elle prend tantôt la place d'un triangle – ou d'un ensemble de triangles – dont la condition limite est celle d'un élément vibrant, ou bien la forme d'une source ponctuelle, alors réellement localisée dans le canal et non plus sur le maillage.

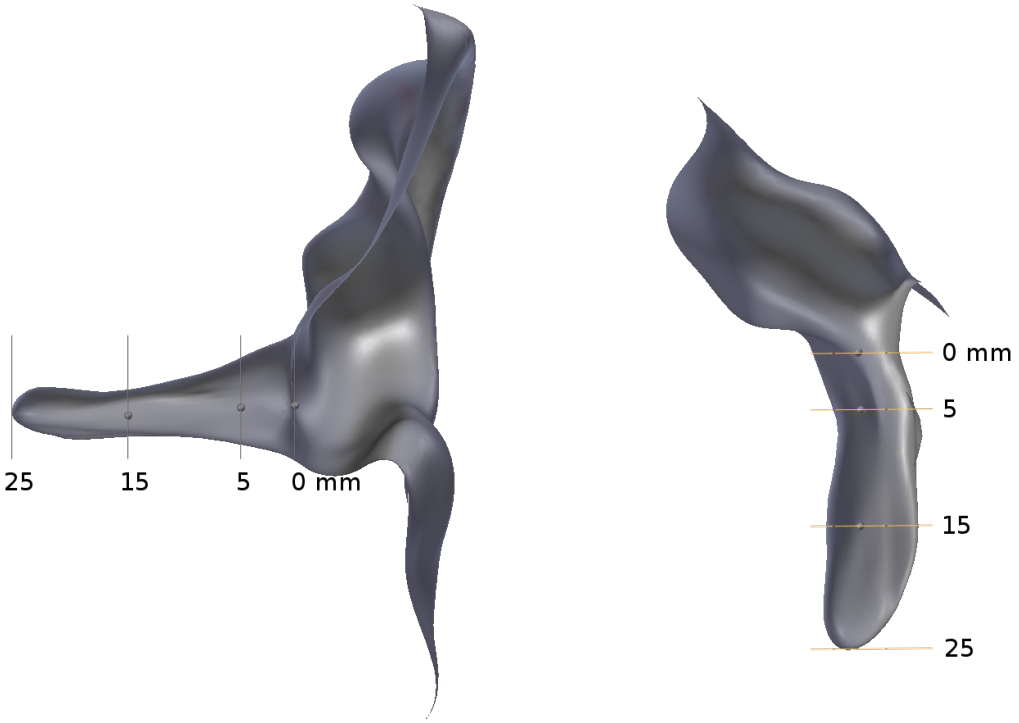


FIGURE 4.1: *Coupes frontale et azimutale de l'oreille utilisée dans la première série de simulations avec, en surimposition, les différentes distances retenues. Les petites sphères représentent les positions des sources ponctuelles.*

Dans le premier cas de figure, c'est-à-dire lorsque l'on utilise un élément du maillage comme émetteur, les trois profondeurs sélectionnées sont de 25 mm, 15 mm et 5 mm. Par

3. Ou plus exactement, du point d'émission, principe de réciprocité oblige.

défaut, un unique élément vibrant, arbitrairement choisi sur la paroi du canal, est utilisé. Ces simulations seront notées T_{25}, T_{15} et T_5. Une variante de T_{25} est également testée, dans laquelle c'est une petite zone de 6 mm^2 et comprenant 37 triangles qui fait office d'émetteur sonore. On la note TZ_{25}.

Dans le second cas de figure, c'est-à-dire lorsque l'on positionne une source ponctuelle comme émetteur, les trois profondeurs sélectionnées sont de 15 mm, 5 mm et 0 mm. Ces simulations seront notées S_{15}, S_5 et S_0. La figure 4.2 décrit ces différentes configurations.

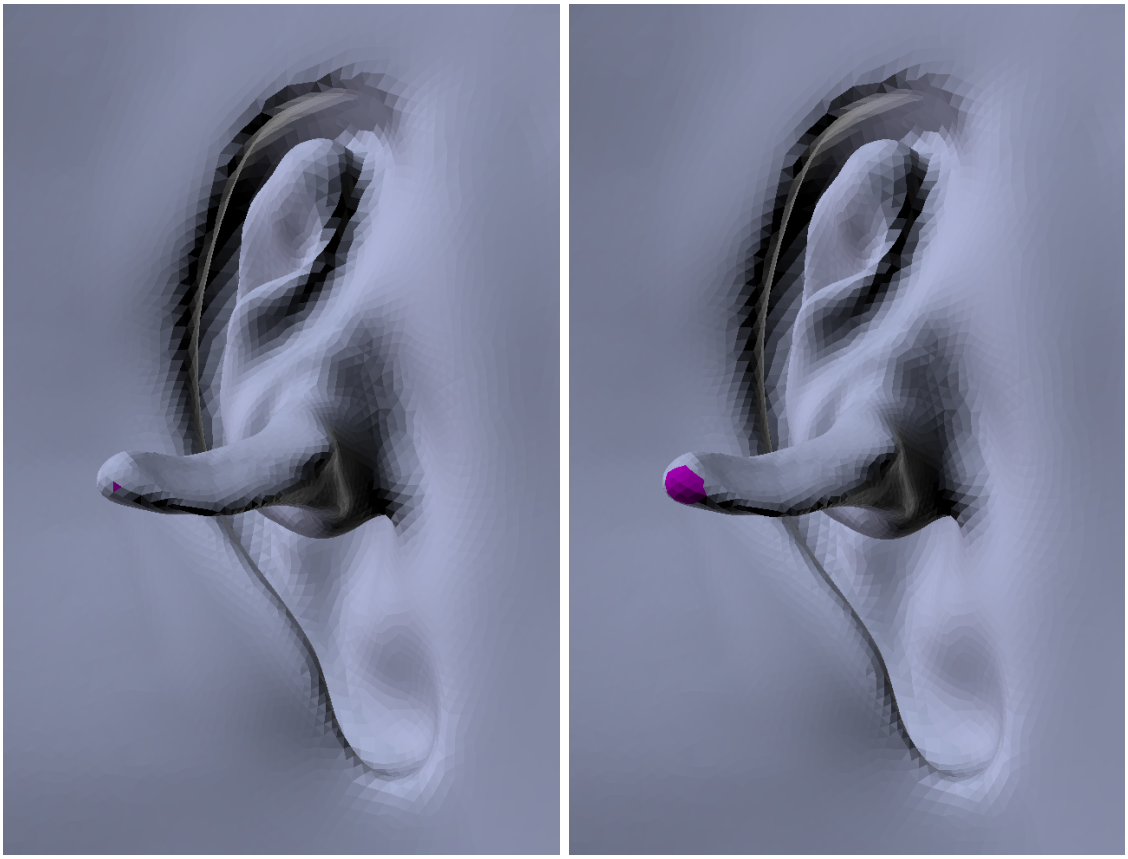


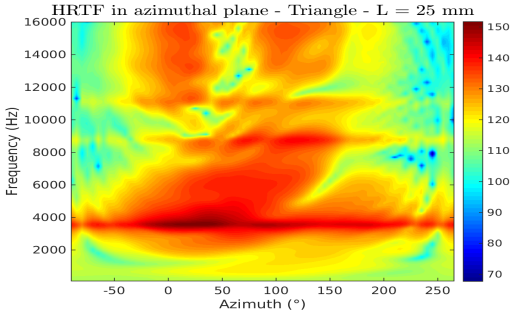
FIGURE 4.2: *Vue de l'intérieur du maillage. Les triangles colorés sont ceux pouvant servir d'éléments vibrants. À gauche, un seul élément (T_{25}). À droite, un ensemble d'éléments d'une surface de quelques millimètres carrés (TZ_{25}).*

4.1.2.2 Résultats

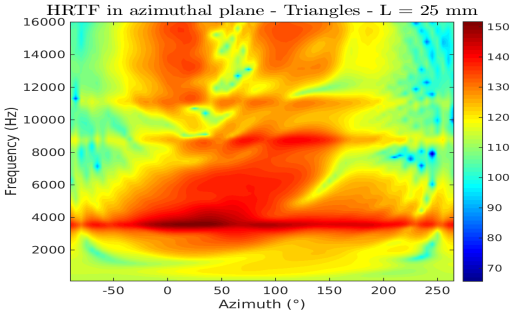
Les figures 4.3 à 4.7 compilent les résultats expérimentaux de cette première expérience. Les figures 4.3 et 4.4 présentent respectivement les coupes azimutale et sagittale des magnitudes en dB des HRTF brutes, c'est-à-dire sans post-traitement.

À toute fin utile, il est rappelé que pour des raisons d'optimisation du temps de calcul des simulations celles-ci vont jusqu'à 16 kHz et englobent donc déjà le spectre audible de la

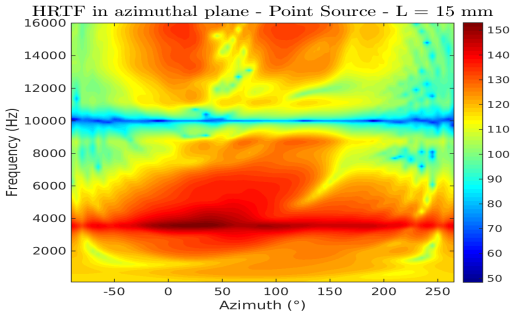
majorité de la population.



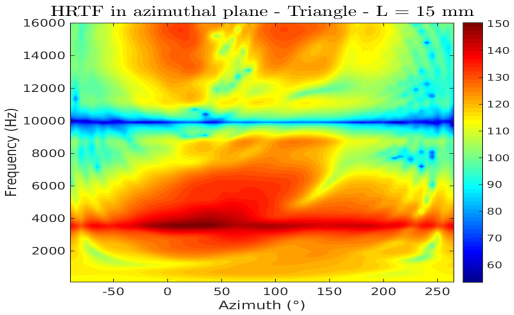
(a) T_{25}



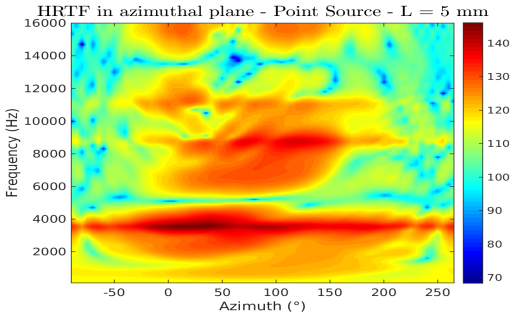
(b) TZ_{25}



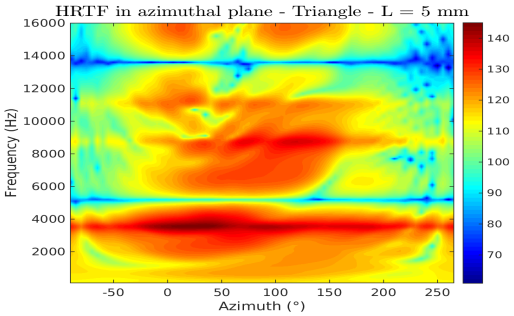
(c) S_{15}



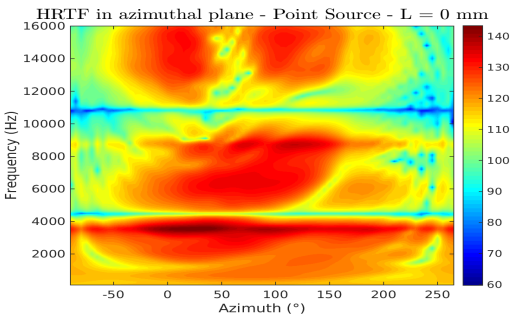
(d) T_{15}



(e) S_5

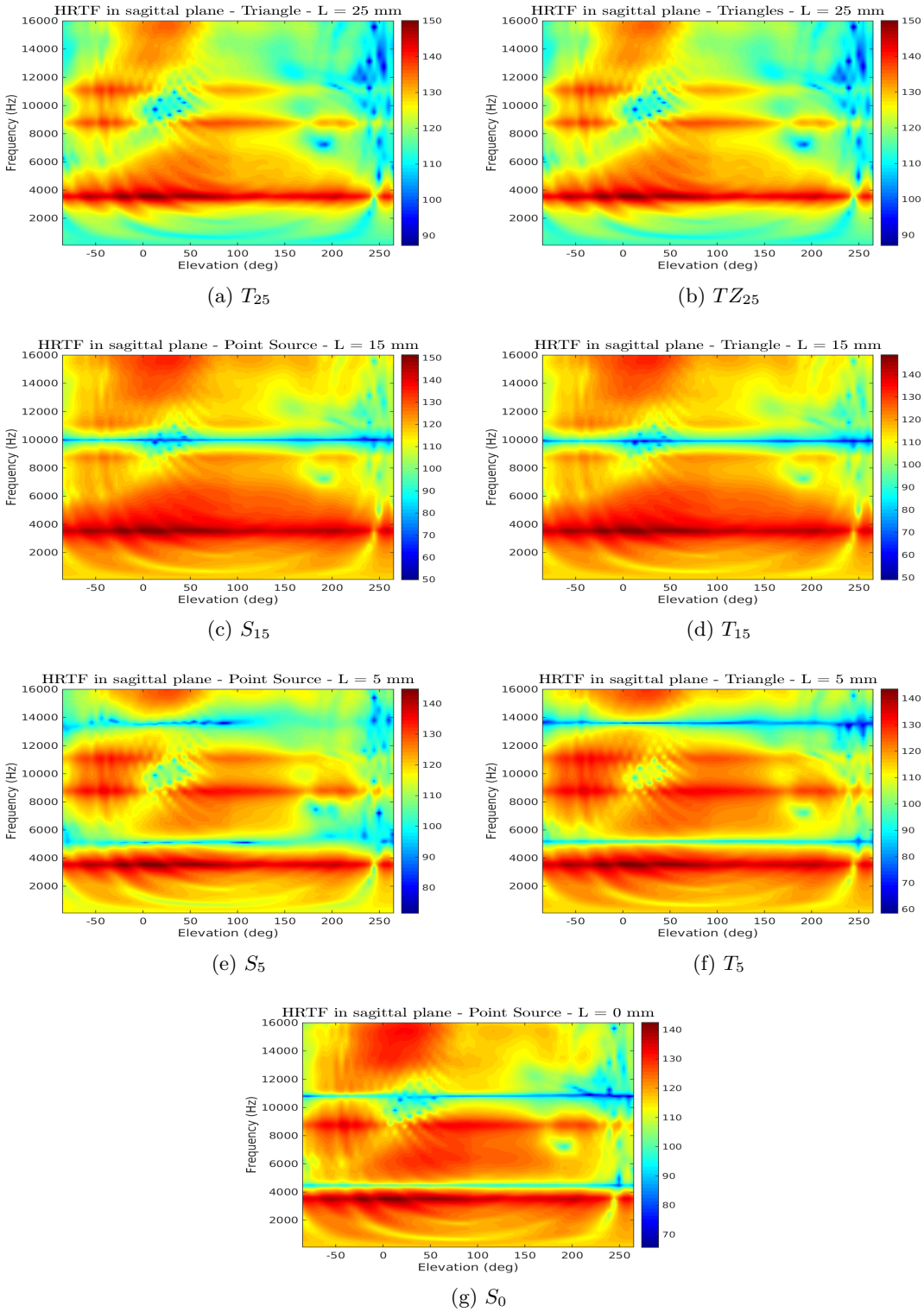


(f) T_5



(g) S_0

FIGURE 4.3: *De (a) à (g), les coupes azimutales des HRTF issues de la première série de simulations.*



(a) T_{25} (b) TZ_{25}

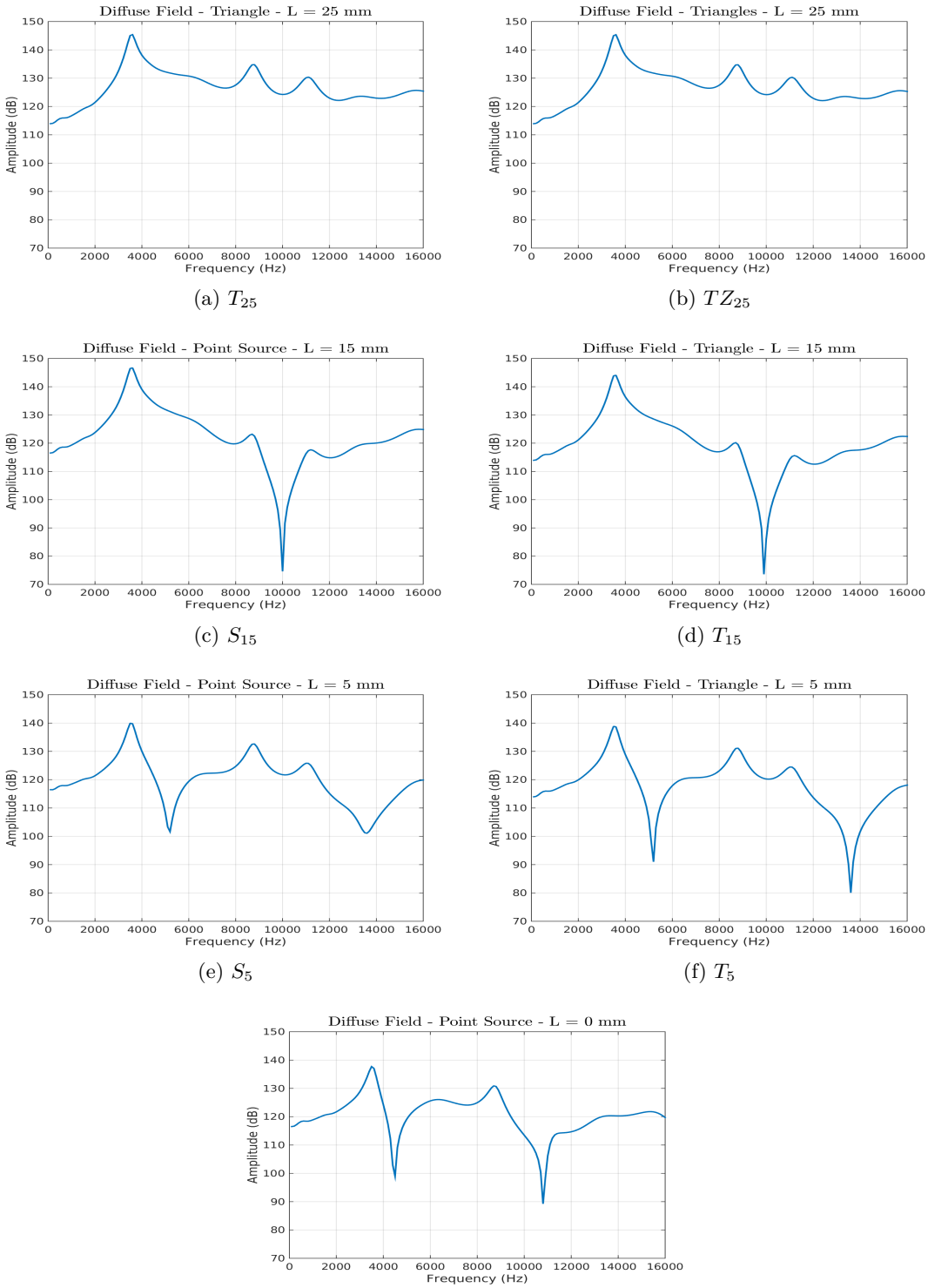
(c) S_{15} (d) T_{15}

(e) S_5 (f) T_5

(g) S_0

FIGURE 4.4: *De (a) à (g), les coupes sagittales des mêmes HRTF.*

La figure 4.5 présente les champs diffus de ces HRTF.



(a) T_{25} (b) TZ_{25}

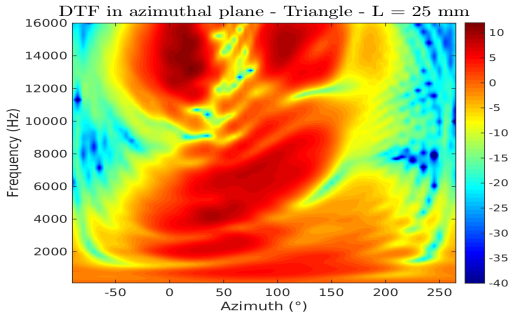
(c) S_{15} (d) T_{15}

(e) S_5 (f) T_5

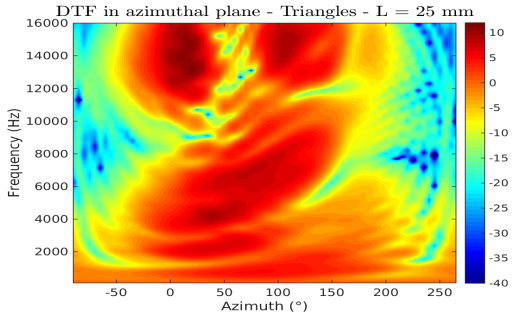
(g) S_0

FIGURE 4.5: *De (a) à (g), les champs diffus correspondants.*

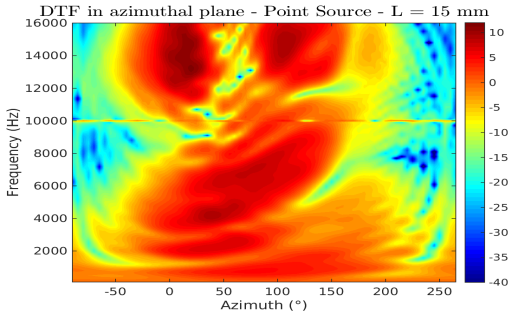
Les 4.6 et 4.7 présentent quant à elles les coupes azimutales et sagittales des DTF, c'est-à-dire des HRTF auxquelles ont été soustraits les champs diffus respectifs.



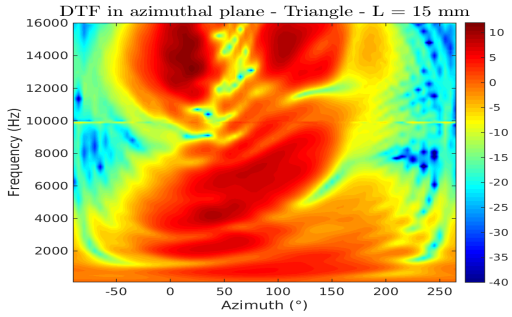
(a) T_{25}



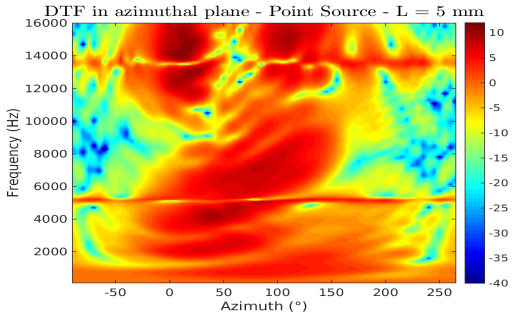
(b) TZ_{25}



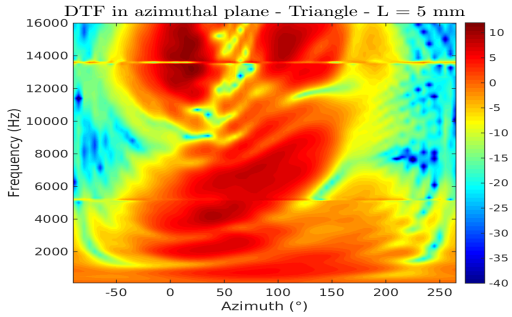
(c) S_{15}



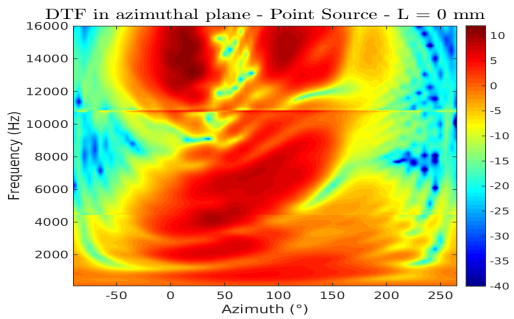
(d) T_{15}



(e) S_5

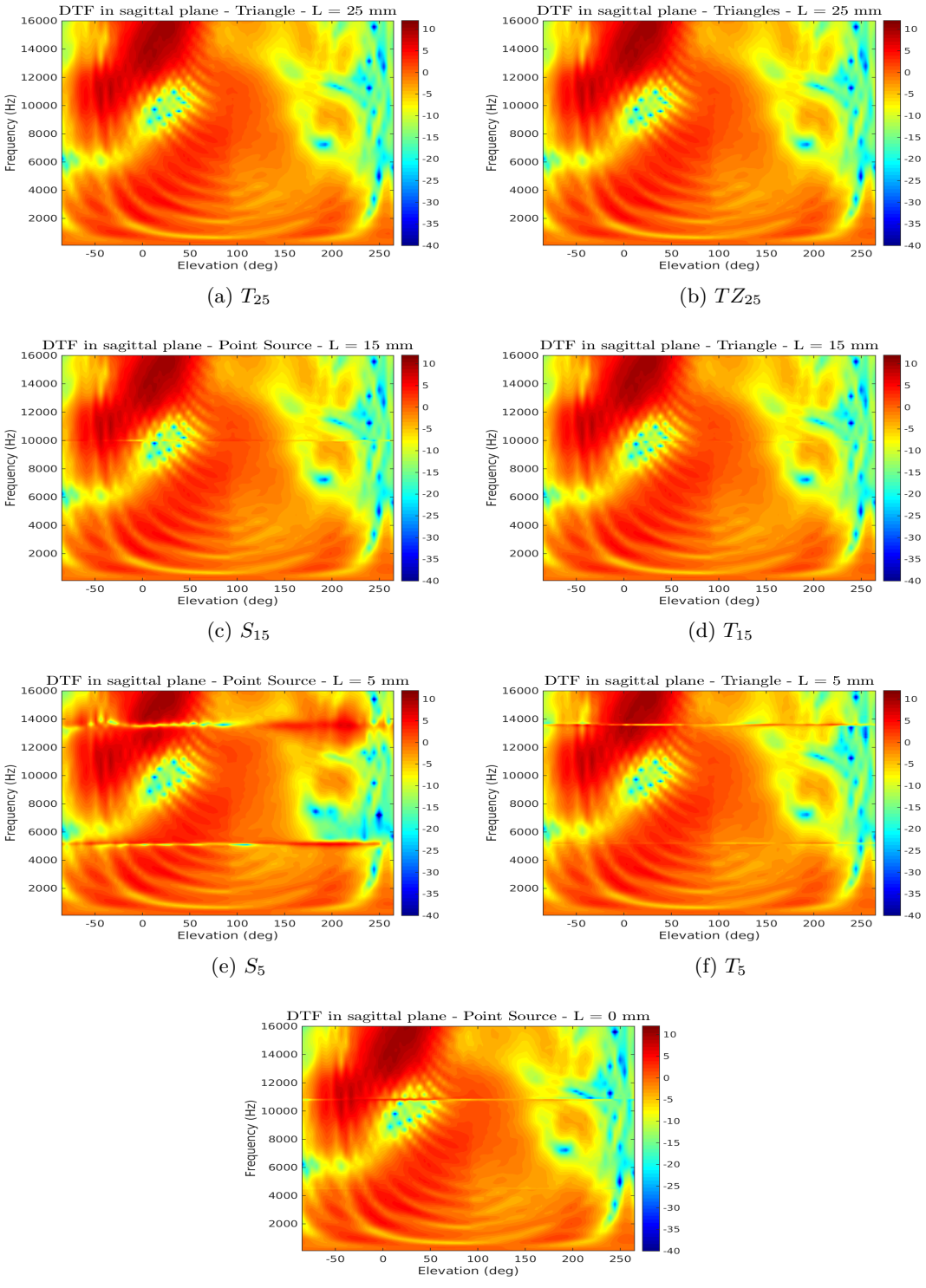


(f) T_5



(g) S_0

FIGURE 4.6: *De (a) à (g), les coupes azimutales des DTF correspondantes.*



(a) T_{25} (b) TZ_{25}

(c) S_{15} (d) T_{15}

(e) S_5 (f) T_5

(g) S_0

FIGURE 4.7: *De (a) à (g), les coupes sagittales des mêmes DTF.*

Enfin, le tableau 4.1 compile les normes SD des couples de DTF à l'étude. Celles-ci sont calculées sur tout le spectre fréquentiel disponible et moyennées (avec une pondération de Voronoï) sur toutes les directions de la grille d'évaluation.

	S_0	S_5	S_{15}	T_5	T_{15}	T_{25}	TZ_{25}
S_0	0	3.7160	1.0663	1.4389	0.9355	0.8887	0.8862
S_5	3.7160	0	3.6288	3.6119	3.5878	3.5717	3.5711
S_{15}	1.0663	3.6288	0	1.2786	0.5773	0.6363	0.6303
T_5	1.4389	3.6119	1.2786	0	1.1541	1.0940	1.1004
T_{15}	0.9355	3.5878	0.5773	1.1541	0	0.4281	0.4221
T_{25}	0.8887	3.5717	0.6363	1.0940	0.4281	0	0.0870
TZ_{25}	0.8862	3.5711	0.6303	1.1004	0.4221	0.0870	0

TABLE 4.1: *Normes SD calculées sur tous les couples de DTF.*

4.1.2.3 Conclusions

À la vue des HRTF brutes, le premier constat que l'on peut faire est que la seule variable qui importe véritablement ici est la distance de la source vis-à-vis du fond du canal. Et dès lors que l'émetteur n'est pas situé parfaitement au fond, on voit apparaître de fortes résonances et antirésonances fréquentielles. Le calcul des champs diffus montre par ailleurs clairement le caractère omnidirectionnel de ces dernières. Un fois les champs diffus retirés, le calcul des normes SD sur les couples de DTF vient confirmer l'impression de grande proximité qui se dégage des coupes azimutales et sagittales. De plus, à mesure que la source s'éloigne du fond, la première annulation descend en fréquence et l'on peut voir apparaître la seconde pour les sources placées à 5 mm et à l'origine. Toutefois, si l'analyse qualitative des phénomènes observés est en accord avec la théorie de la résonance du tube bouché, des divergences surgissent lors des applications numériques. En effet, si l'on se place à une distance x du fond, les fréquences d'annulation sont attendues aux valeurs théoriques suivantes :

$$f_n(x) = \frac{c}{4x}(1+2n), n \in \mathbb{N} \tag{4.1}$$

En prenant $c = 346.18\,\mathrm{m.s^{-1}}$ et $x = 10\,\mathrm{mm}$, on obtient alors $f_0 = 8.65\,\mathrm{kHz}$. Or cette valeur de x correspond ici à une source placée à 15 mm de l'entrée du canal et l'on observe manifestement – figure 4.5 (c) et (d) – une fréquence différente d'annulation, à savoir 10 kHz. Et des observations analogues peuvent être faites pour les simulations 5 mm et 0 mm. Par ailleurs, toujours selon le cadre théorique, le ratio des deux premières fréquences d'annulation $\frac{f_1}{f_0}$ doit valoir 3. Or il vaut approximativement 2,6 dans les simulations à 5 mm et 2,4 dans celle à 0 mm.

Causes possibles de tels écarts, le fait que la forme du canal n'est ici pas à proprement parler un tube bouché et que l'on ne simule pas la propagation d'ondes planes mais d'ondes sphériques. Cela étant dit, eu égard aux travaux de Hammershoï & Moller [72], ces différences sont en réalité les bienvenues car elles confortent leurs observations faites sur des HRTF mesurées.

Parmi les autres enseignements de cette première expérience, le fait que les DTF

partagent le même spectre quel que soit l'emplacement de la source, à ceci près que pour certaines valeurs des artefacts liés aux fréquences de résonances persistent. Cela se comprend bien car, pour obtenir les DTF, il faut d'abord calculer le champ diffus puis le soustraire aux HRTF. Or, au niveau des fréquences d'annulation, on manipule des valeurs extrêmement basses et donc entachées d'imprécisions. Lorsque la source sonore n'est pas placée au fond du canal, il est donc naturel que ces artefacts omnidirectionnels apparaissent. Ziegelwanger *et al.* n'avaient pas noté cette influence du positionnement de la source mais, dans leur cas, le canal auditif ne mesurait pas plus de 5 mm de profondeur et n'était pas en mesure de faire apparaître les résonances.

Le choix d'un triangle plutôt que d'une source ponctuelle est lui aussi sans incidence. Les différences résiduelles que l'on peut observer sont bien plus le fruit d'un inévitable écart de localisation entre les deux types de source (l'une étant située sur la paroi, l'autre au milieu du canal) qu'à tout autre chose.

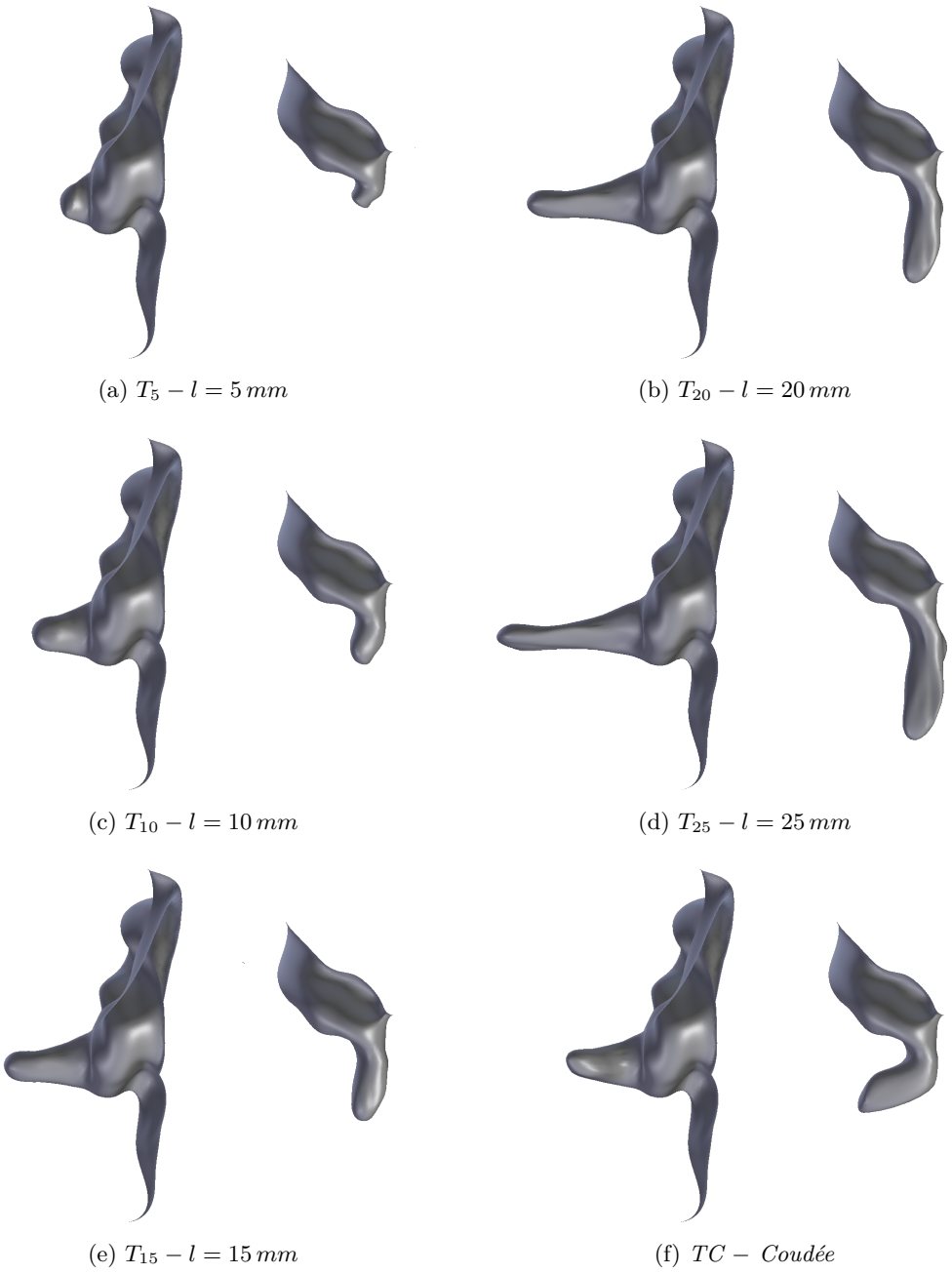
Enfin, utiliser un seul triangle – cas 1 – ou tous ceux d'une zone donnée – cas 2 – n'altère pas le résultat. Ceci est en accord avec les résultats de l'équipe ARI qui a comparé, entre autres, les DTF obtenues par un premier micro virtuel fait d'un seul élément et par un deuxième fait d'un ensemble couvrant un disque de 3 mm de rayon.

Néanmoins, quelques constats annexes ont pu être faits selon que l'on a choisi l'une ou l'autre de ces options. Tout d'abord, le temps de calcul se trouve légèrement accru dans le cas 1, passant de 44 h d'utilisation CPU à 48 h. De plus, les valeurs minimales atteintes dans les directions controlatérales y sont aussi légèrement plus grandes. Cela équivaut en quelque sorte à un léger lissage de ces régions. Cet effet est également rapporté par Ziegelwanger *et al.* Enfin, et il s'agit là d'un retour d'expérience impliquant un plus grand nombre de simulations, se placer dans le cas numéro 2 aboutit généralement à présenter à l'outil de calcul numérique des systèmes d'équations mieux conditionnés, ce qui en facilite la résolution. Pour toutes ces raisons, et sauf mention contraire, les HRTF calculées dans la suite des présents travaux l'auront été en plaçant la source sonore au fond du canal auditif, sous la forme d'un ensemble de triangles du maillage.

4.1.3 Géométrie du canal

4.1.3.1 Protocole expérimental

Dans la seconde série de simulations, l'émetteur a donc été fixé au fond du canal sous la forme d'un triangle, i.e. à la position attendue du tympan, et c'est la forme du canal qui a été modifiée, le reste du maillage demeurant inchangé. Plusieurs longueurs ont été testées ainsi qu'un changement d'orientation. L'ensemble des géométries étudiées est visible figure 4.8.



(a) $T_5 - l = 5\,mm$ (b) $T_{20} - l = 20\,mm$

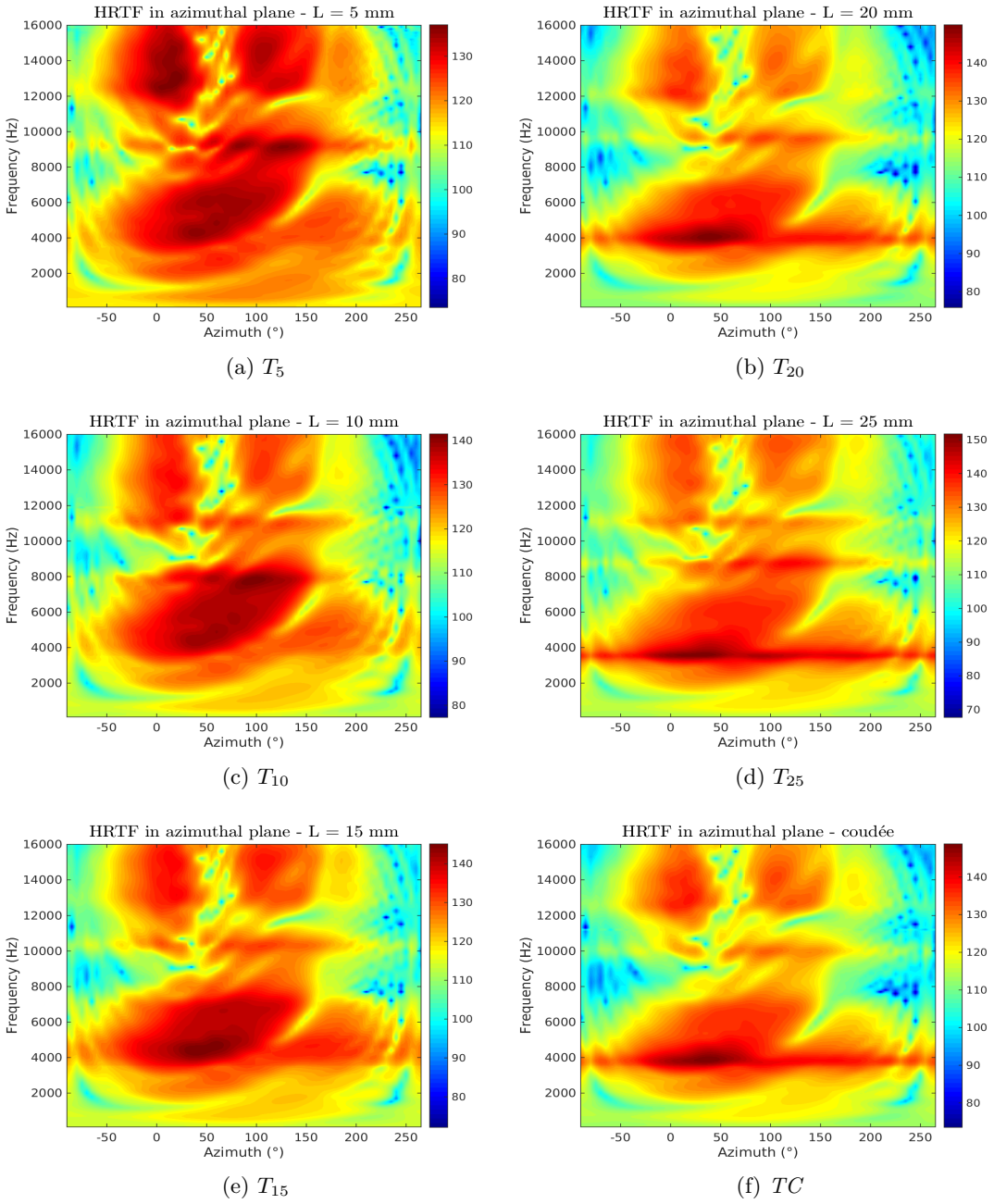
(c) $T_{10} - l = 10\,mm$ (d) $T_{25} - l = 25\,mm$

(e) $T_{15} - l = 15\,mm$ (f) $TC -$ *Coudée*

FIGURE 4.8: *De (a) à (f), les coupes de profil et de dessus des différentes oreilles utilisées dans la seconde série de simulations. L'oreille (f) présente un coude fortement marqué. Cette oreille mise à part - notée TC par la suite -, la longueur du canal croît de 5 mm à 25 mm par pas de 5 mm. Dans chacun des cas, la coupe frontale est à gauche et l'azimutale à droite. Ces cinq autres géométries sont naturellement notées* T_5, T_{10}, T_{15}, T_{20} *et* T_{25}.

4.1.3.2 Résultats

Les figures 4.9 à 4.13 compilent les résultats expérimentaux de la seconde expérience. Les figures 4.9 et 4.10 présentent respectivement les coupes azimutales et sagittales des magnitudes en dB des HRTF brutes, c'est-à-dire sans post-traitement. La figure 4.11 présente les champs diffus de ces HRTF. Les figures 4.12 et 4.13 présentent quant à elles les coupes azimutales et sagittales des DTF.

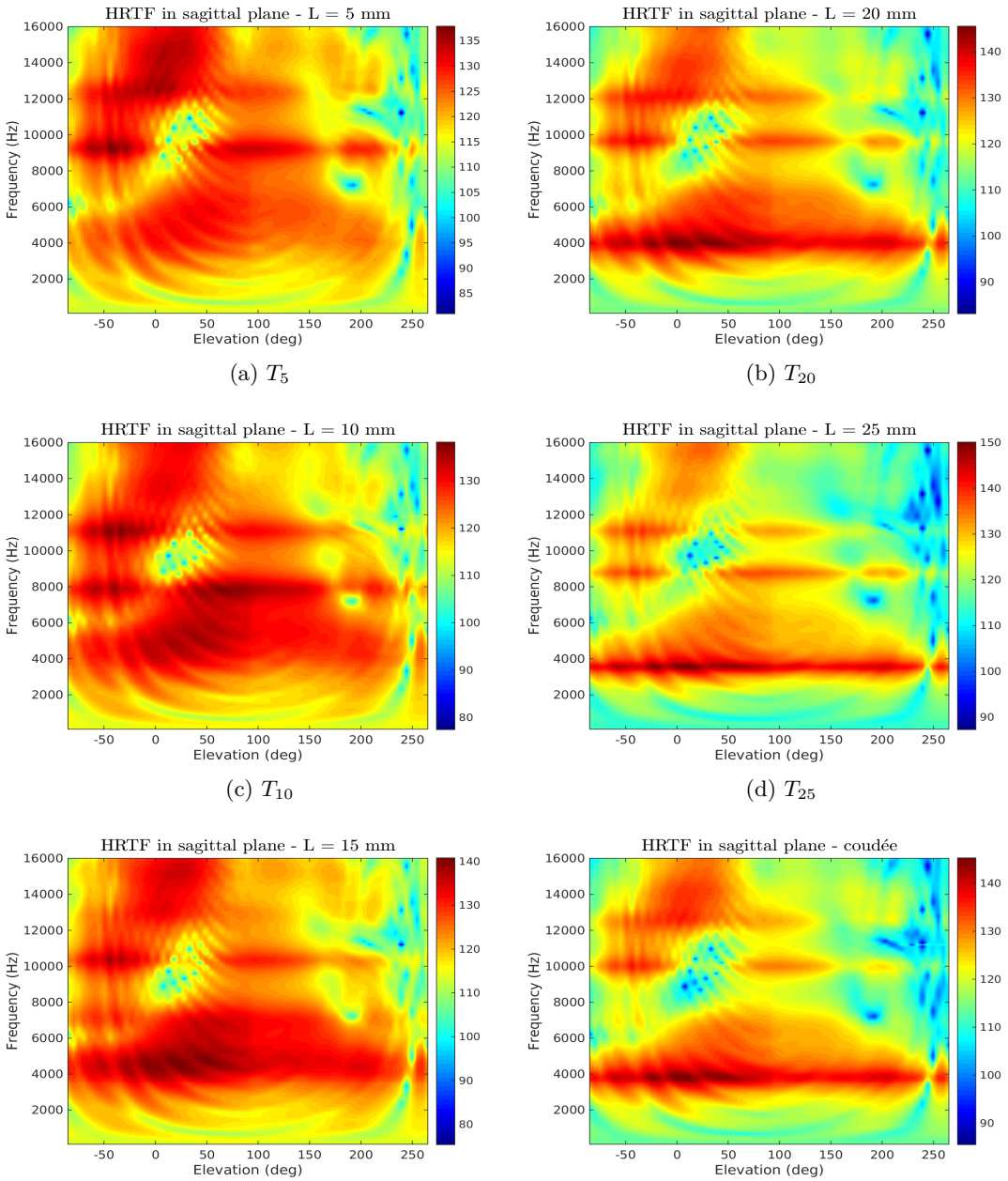


(a) T_5 (b) T_{20}

(c) T_{10} (d) T_{25}

(e) T_{15} (f) TC

FIGURE 4.9: *De (a) à (f), les coupes azimutales des HRTF issues de la seconde série de simulations.*

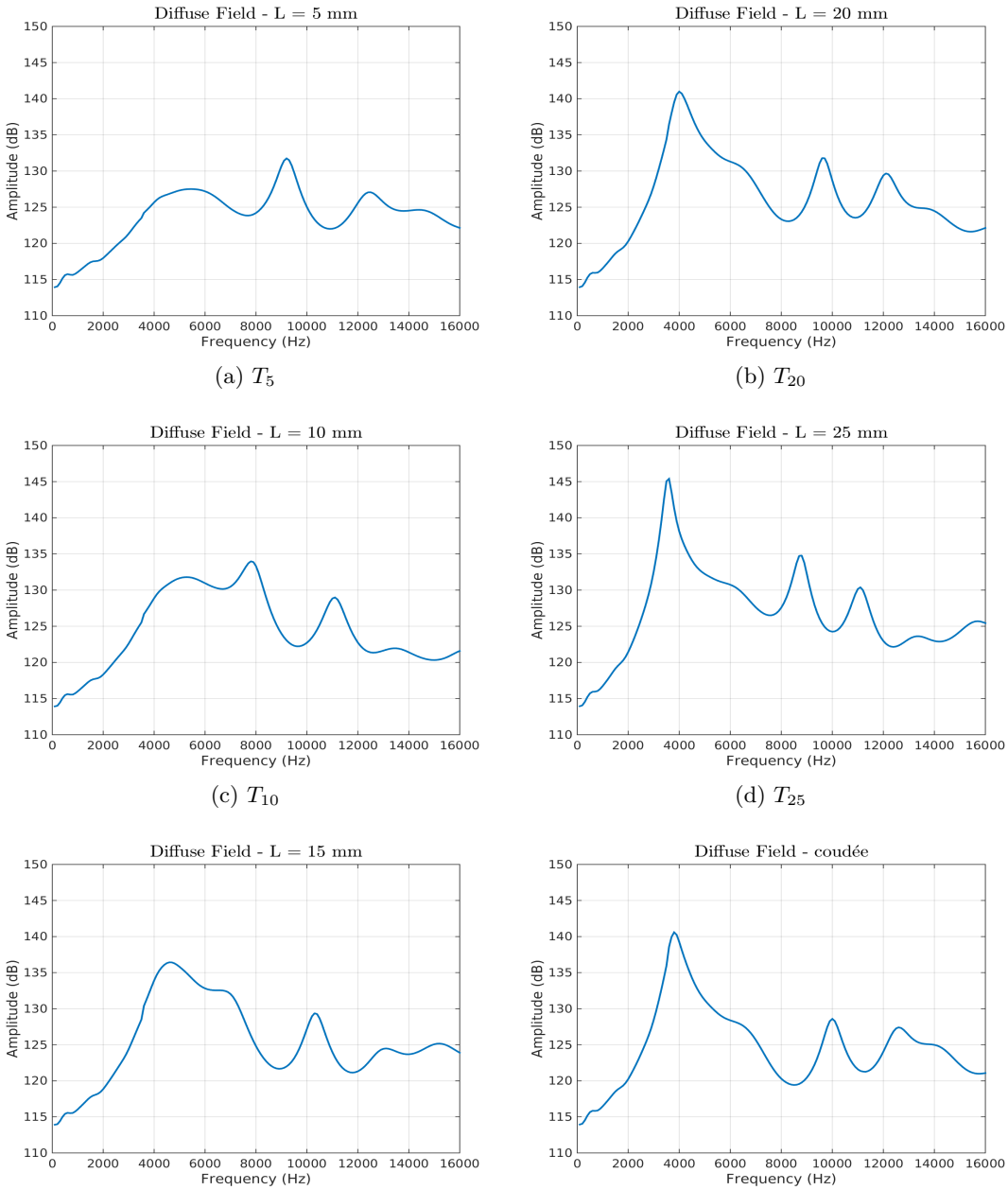


(a) T_5 (b) T_{20}

(c) T_{10} (d) T_{25}

(e) T_{15} (f) TC

FIGURE 4.10: *De (a) à (f), les coupes sagittales des mêmes HRTF.*

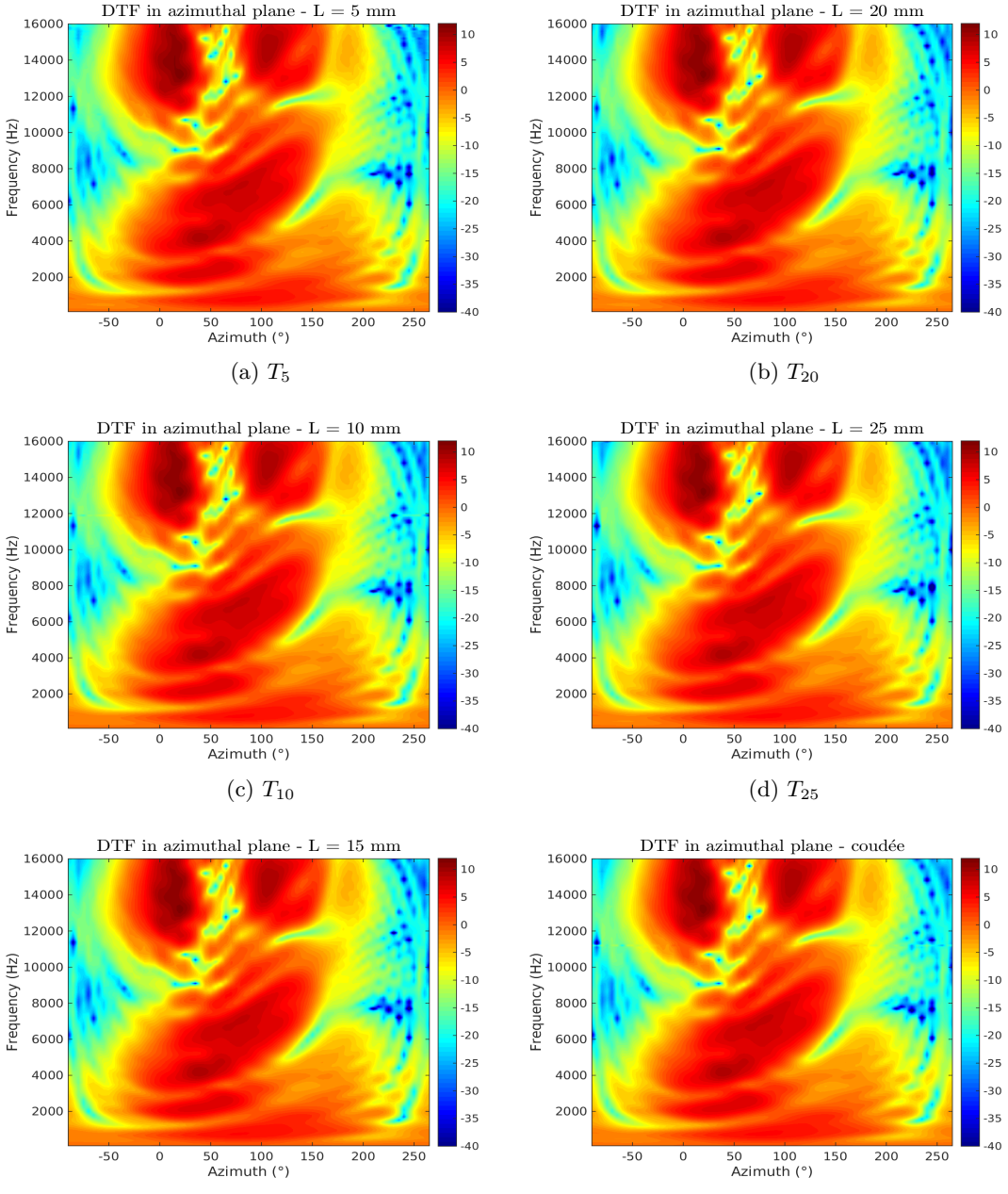


(a) T_5 (b) T_{20}

(c) T_{10} (d) T_{25}

(e) T_{15} (f) TC

FIGURE 4.11: *De (a) à (f), les champs diffus correspondants.*

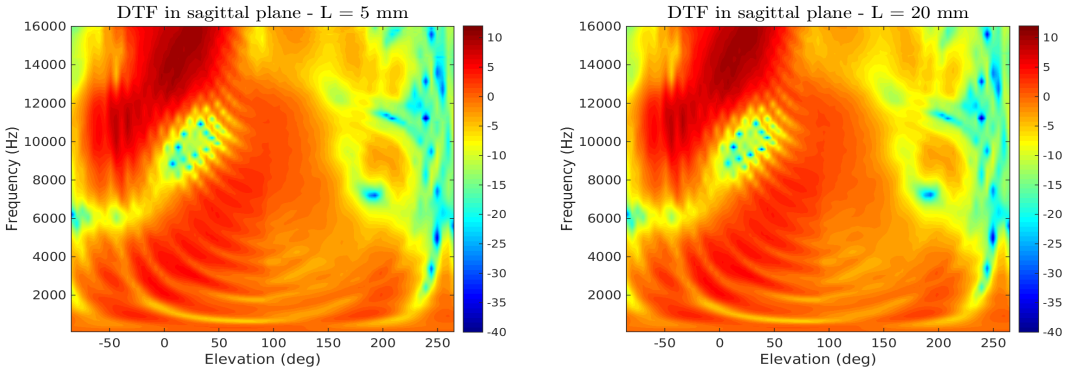


(a) T_5 (b) T_{20}

(c) T_{10} (d) T_{25}

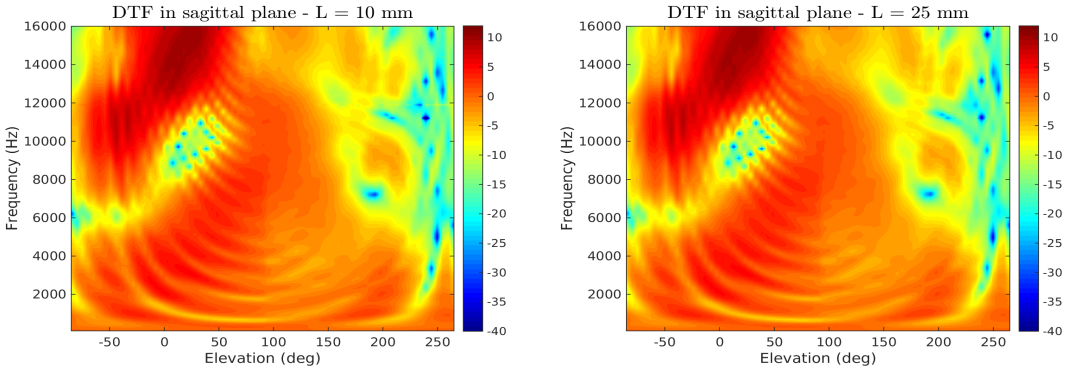
(e) T_{15} (f) TC

FIGURE 4.12: *De (a) à (f), les coupes azimutales des DTF correspondantes. Au contraire des HRTF brutes, les DTF présentent une extrême similarité.*



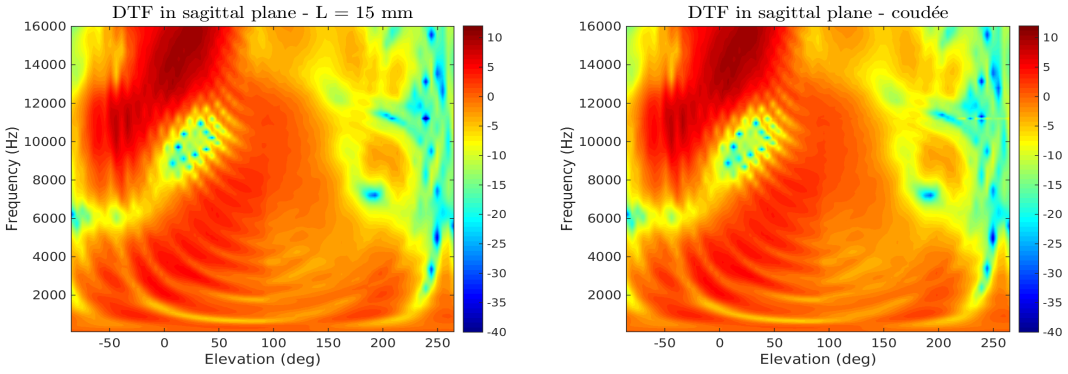
(a) T_5

(b) T_{20}



(c) T_{10}

(d) T_{25}



(e) T_{15}

(f) TC

FIGURE 4.13: *De (a) à (f), les coupes sagittales des mêmes DTF.*

Enfin, le tableau 4.2 compile les normes SD des couples de DTF à l'étude. Celles-ci sont calculées sur tout le spectre fréquentiel disponible et moyennées (avec une pondération de Voronoï) sur toutes les directions de la grille d'évaluation.

4.1.3.3 Conclusions

Comme on peut le constater, les HRTF brutes sont bien différentes les unes des autres et on ne saurait les confondre. En particulier, les champs diffus calculés – cf. figure 4.11 – font clairement apparaître de fortes résonances et annulations omnidirectionnelles. La seule

	T_5	T_{10}	T_{15}	T_{20}	T_{25}	TC
T_5	0	0.4031	0.4725	0.5970	0.6866	0.4724
T_{10}	0.4031	0	0.3316	0.4891	0.5339	0.3907
T_{15}	0.4725	0.3316	0	0.3306	0.4212	0.3136
T_{20}	0.5970	0.4891	0.3306	0	0.3184	0.4127
T_{25}	0.6866	0.5339	0.4212	0.3184	0	0.5131
TC	0.4724	0.3907	0.3136	0.4127	0.5131	0

TABLE 4.2: *Normes SD calculées sur tous les couples de DTF.*

variable de cette expérience étant la géométrie du canal, les fréquences auxquelles elles se manifestent en sont donc fonction.

Une fois ces champs diffus soustraits à leurs HRTF d'origine, il apparaît une correspondance extrêmement forte entre toutes les DTF, les différences résiduelles (et marginales - cf. table 4.2), pouvant sans risque être attribuées au calcul numérique.

Le dépouillement des résultats de simulation montre donc que la longueur du canal auditif a bien un impact sur la HRTF et que cet impact est en première approximation – mais en première approximation seulement – semblable à l'effet de résonance acoustique observé en mesurant la pression à l'extrémité fermée d'un tube semi-ouvert. Comparativement, la présence d'un coude, c'est-à-dire d'une déformation géométrique majeure du tube, n'a que très peu d'influence sur le résultat et constitue donc une variable mineure. Dans tous les cas, les différences observées sont omnidirectionnelles et ne se retrouvent pas dans les DTF. Pour autant que l'on ne travaille qu'avec ces dernières, ce qui est le cas présentement, il n'est donc pas nécessaire de disposer d'informations sur la géométrie du canal auditif pour les calculer. Dans tous ce qui suit, ne nous intéressant qu'aux DTF, nous les appellerons à nouveau, et par abus de langage, HRTF sauf en cas d'ambiguïté manifeste.

4.2 Optimisations numériques

Cette vérification d'usage sur la géométrie du canal étant achevée, il est désormais temps de se pencher sur la simulation numérique en elle-même. Étant donnée la quantité d'opérations à effectuer pour obtenir ne serait-ce qu'une HRTF, toutes les optimisations permettant de réduire les temps de calcul sont les bienvenues. Suivant cette ligne directrice, nous avons déjà préféré la *BEM* à la *FEM*. Plus encore, il s'agit de *FM-BEM*, notablement plus efficace que la version classique. Mais pour aller encore plus loin et parvenir à un coût marginal de simulation autorisant la constitution de nos bases, deux améliorations ont été mises en place : le *maillage adaptatif* et la *dépendance en fréquence*.

4.2.1 Maillage adaptatif

Très simplement, ce que l'on appelle *maillage adaptatif*, ou parfois *mesh grading*, est le fait de retravailler un scan de manière à ce qu'il soit plus finement maillé sur certaines

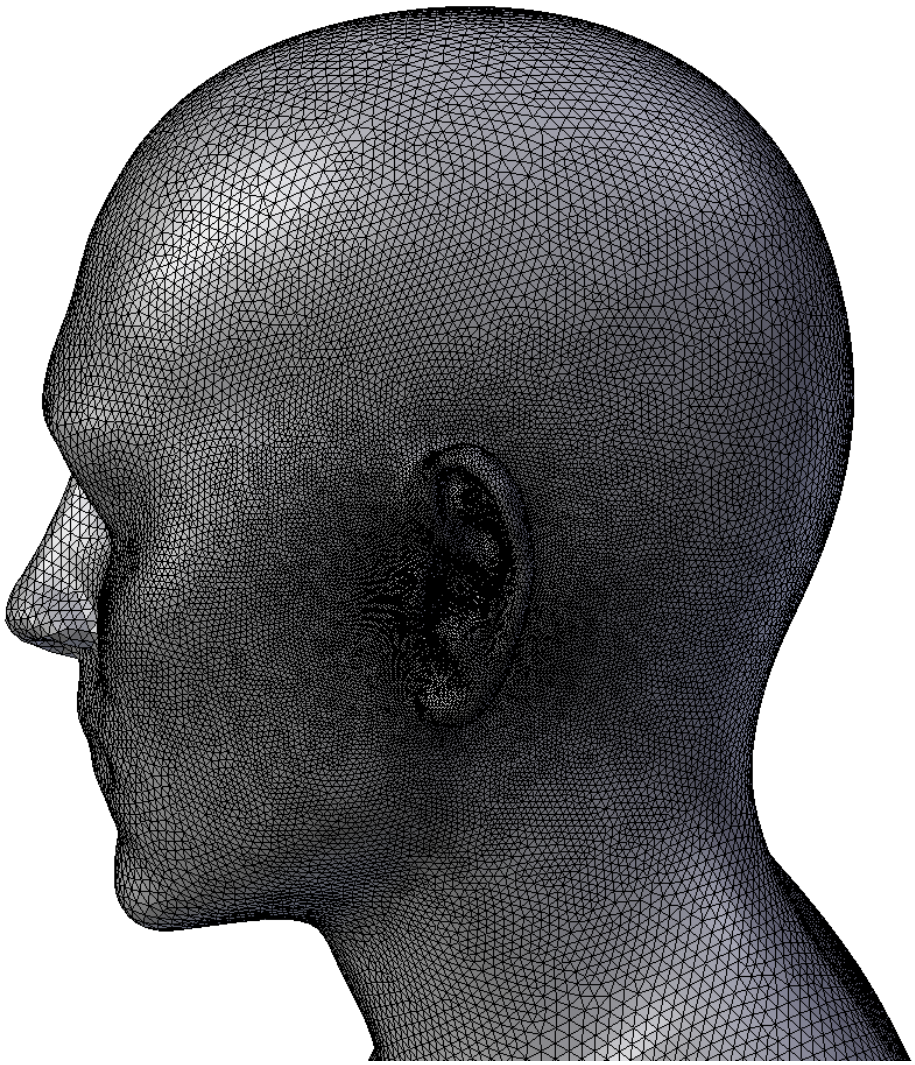


FIGURE 4.14: *Maillage utilisé en simulation. Les arêtes voient leur longueur varier de 0,7 mm, au niveau de l'oreille, à 5 mm sur la tête et le torse.*

zones que sur d'autres. Cette idée n'est pas nouvelle et a déjà été expérimentée et validée par Ziegelwanger *et al.* [168]. Pour mémoire, ils avaient estimé qu'un maillage caractérisé par une longueur d'arête moyenne de 1 mm au niveau du pavillon et de 2,5 mm ailleurs – avec une zone de transition de quatre rangées de triangles d'arête moyenne à 1,5 mm – remplissait tous les critères pour obtenir des simulations satisfaisantes jusqu'à 24 kHz. Plus précisément, la règle empirique des 6 éléments par longueur d'onde amène à une fréquence critique de $f_{crit} = 22,9\,\text{kHz}$ pour une moyenne d'arête de 2,5 mm et à $f_{crit} = 57,2\,\text{kHz}$ pour une moyenne de 1 mm.

Dans notre cas, nous avons fait le choix de nous limiter à la bande de fréquence [100, 16 000] Hz. Pour $f_{crit} = 16\,\text{kHz}$, la règle des 6 éléments nous donne alors une valeur moyenne d'arête de 3,57 mm. Cette donnée en tête, plusieurs simulations ont été menées pour définir la meilleure topologie pour nos maillages. Afin de préserver les détails physionomiques de l'oreille, une arête minimale de 0,7 mm a été choisie et concerne une zone de 10 cm de

diamètre centrée sur la position du micro. Au-delà, la consigne a été fixée à 5 mm. La transition entre les deux zones est quant à elle assez douce et s'effectue de façon continue. Au final, la taille des maillages obtenus en suivant ces principes oscille entre 50 000 et 60 000 sommets. La figure 4.14 illustre un tel remaillage. À titre de comparaison, un maillage uniforme de longueur d'arête 0,7 mm dépasse les 2×10^6 sommets. La puissance de calcul alors nécessaire devient prohibitive, ce que le maillage adaptatif permet d'éviter.

4.2.2 Dépendance en fréquence du maillage

4.2.2.1 Définition

Autre procédé d'optimisation important s'il en est : la dépendance en fréquence du maillage. Concrètement, il s'agit d'adapter celui-ci en fonction de la fréquence à simuler. En effet, les calculs étant faits de manière indépendante fréquence par fréquence, il est tout à fait possible de modifier le maillage en fonction des besoins. Nous tirons en l'occurrence parti de cette possibilité pour privilégier l'utilisation de scans plus grossiers en basses fréquences.

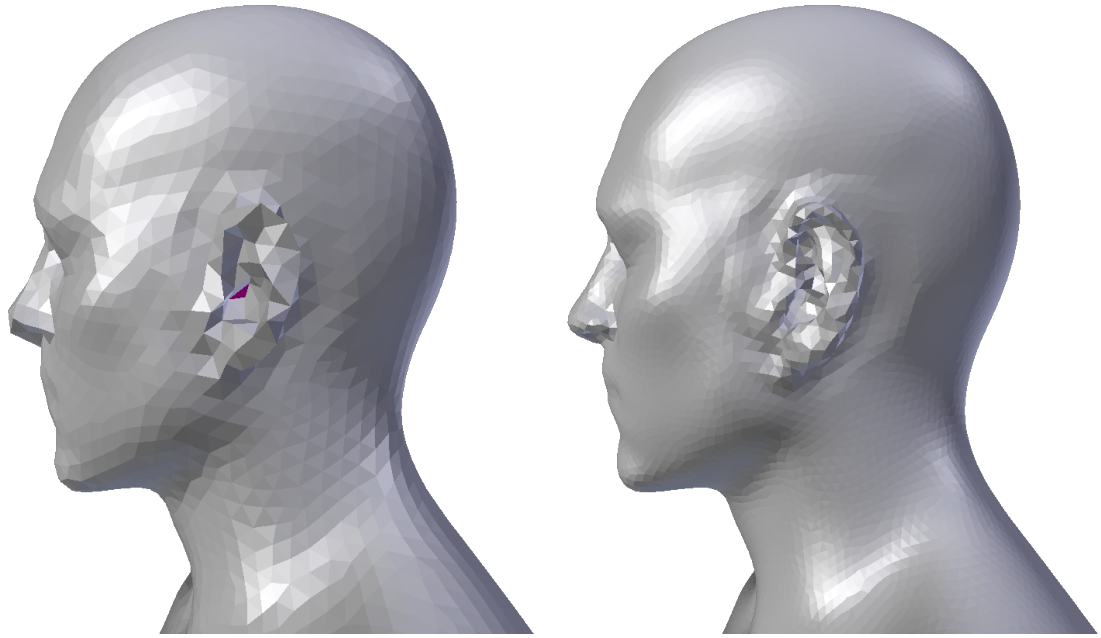
Ainsi, la plage de fréquence à l'étude a été divisée en quatre sous-plages et chacune d'entre elles s'est vue attribuée un maillage spécifique :

1. De 100 Hz à 400 Hz, un maillage uniforme d'arête moyenne 1 cm, noté *A*.
2. De 500 Hz à 2 000 Hz, un maillage uniforme d'arête moyenne 5 mm, noté *B*.
3. De 2 100 Hz à 3 500 Hz, un maillage progressif d'arêtes comprises entre 2 mm et 5 mm, noté *C*.
4. De 3 600 Hz à 16 000 Hz, un maillage progressif d'arêtes comprises entre 0,7 mm et 5 mm, noté *D*.

4.2.2.2 Impacts en simulation

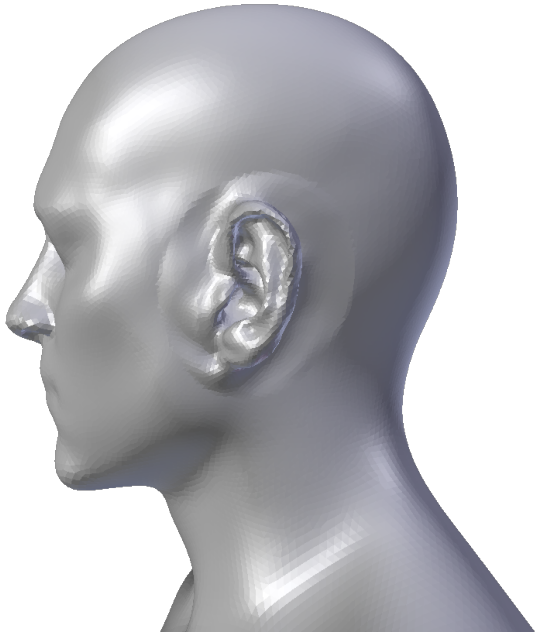
Bien entendu, un changement de maillage impliquant un changement de topologie, de position du micro, de taille du micro, etc., ce procédé induit inévitablement, dans les données, des discontinuités dont il faut tenir compte. Pour cela, une étude comparative des HRTF obtenues à partir de chacun des maillages a été réalisée au préalable sur la plage [100, 5 000] Hz. Un aperçu des données d'entrée est présenté figure 4.15. La figure 4.16 présente coupes azimutales des HRTF et DTF obtenues. Enfin, précisons que les simulations ont été réalisées avec une grille d'évaluation icosaédrique de rayon 2 m et que seules les HRTF

gauches ont été calculées.

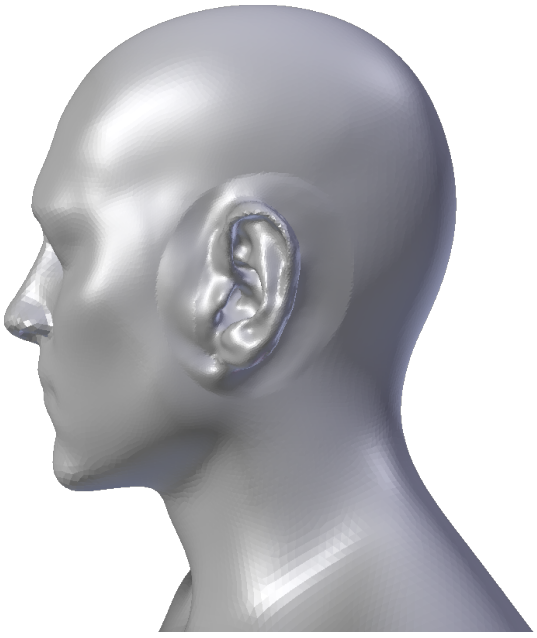


(a) Maillage A

(b) Maillage B

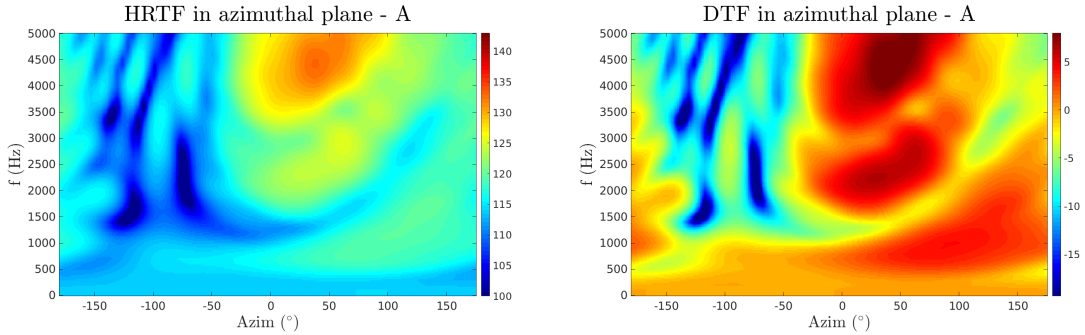


(c) Maillage C

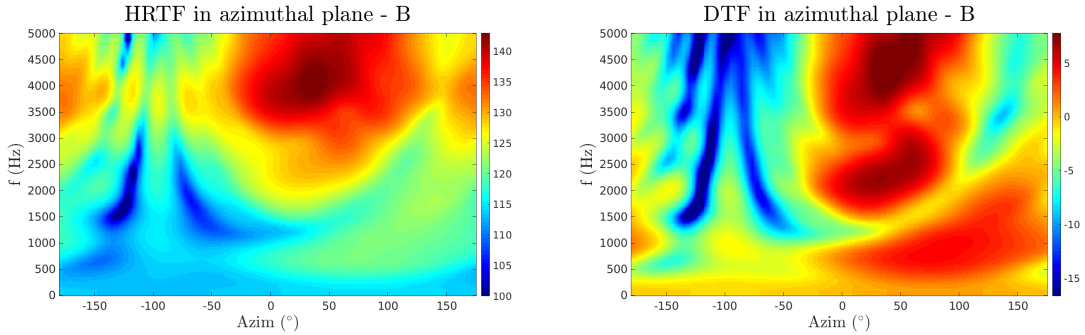


(d) Maillage D

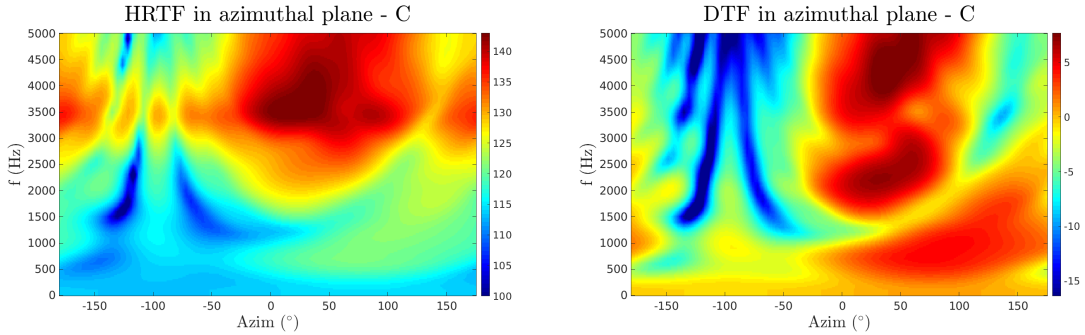
FIGURE 4.15: *De (a) à (d), les différents remaillages utilisés. Le triangle mauve indique la position du micro. Censé se trouver au fond du canal auditif, il n'est ici visible que sur le maillage (a).*



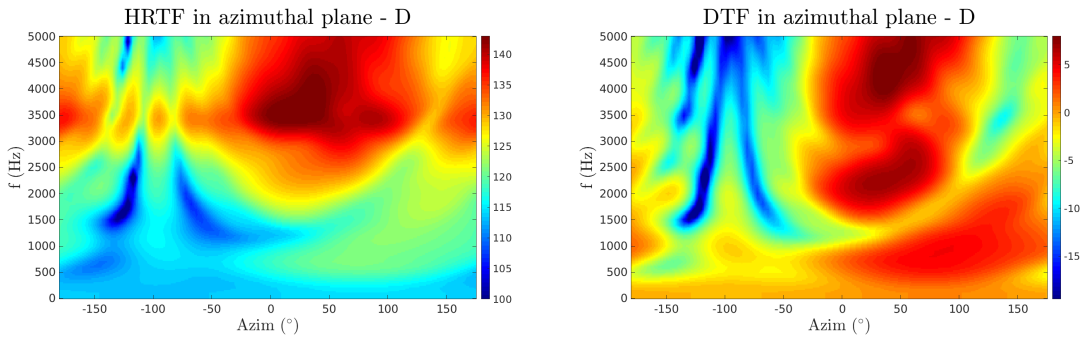
(a) HRTF et DTF liées au maillage A



(b) HRTF et DTF liées au maillage B



(c) HRTF et DTF liées au maillage C



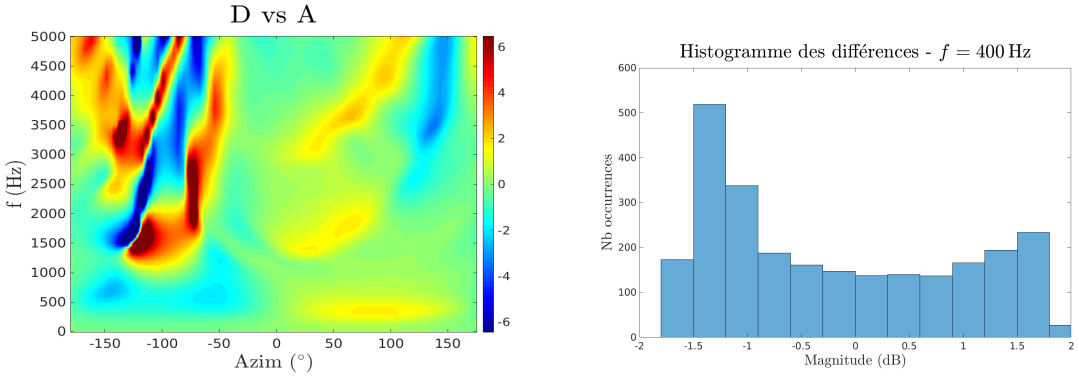
(d) HRTF et DTF liées au maillage D

FIGURE 4.16: *Pour chaque maillage, les HRTF (à gauche) et DTF (à droite) prises dans le plan azimutal.*

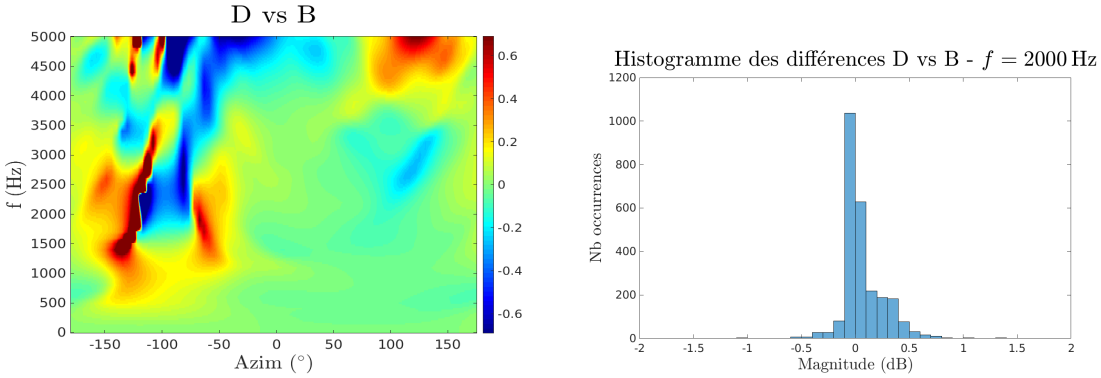
Ainsi qu'on peut le constater, les HRTF en elles-mêmes diffèrent bien plus que les DTF. Cela est particulièrement vrai lorsque l'on compare le maillage le plus grossier au maillage le plus fin. Nous attribuons cela aux changements inévitables de géométrie. En effet, plus

le maillage est grossier et moins le canal est profond. Or nous l'avons observé section 4.1.3, ses variations ont de fortes répercussions sur le champ diffus.

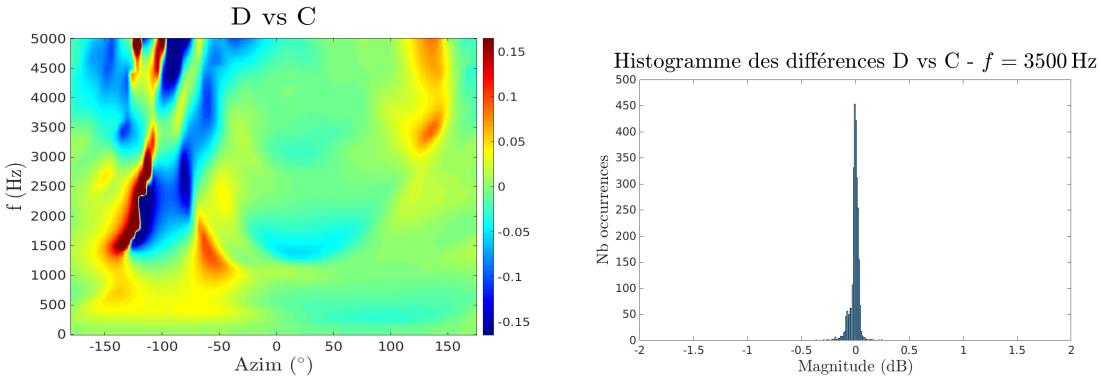
Pour ce qui est des DTF, on constate également des écarts, et ceux-ci s'accroissent à mesure que l'on monte en fréquence. La figure 4.17, représentant les cartes de différences vis-à-vis des DTF du maillage le plus fin et les histogrammes des écarts aux fréquences limites, permet de mieux appréhender ce phénomène.



(a) Maillage D VS Maillage A



(b) Maillage D VS Maillage B



(c) Maillage D VS Maillage C

FIGURE 4.17: *À gauche, les cartes des écarts – dans le plan azimutal – entre les DTF des maillages grossiers et la DTF du maillage le plus fin. À droite, les histogrammes des différences – sur toute la grille d'évaluation – prises aux fréquences limites (400 Hz, 2 000 Hz et 3 500 Hz).*

On note que les écarts sont bien moins prononcés aux fréquences 2 000 Hz et 3 500 Hz

qu'à 400 Hz, mais même là, la valeur moyenne des différences n'est que de -0,2 dB, pour un écart-type de 1,12 dB.

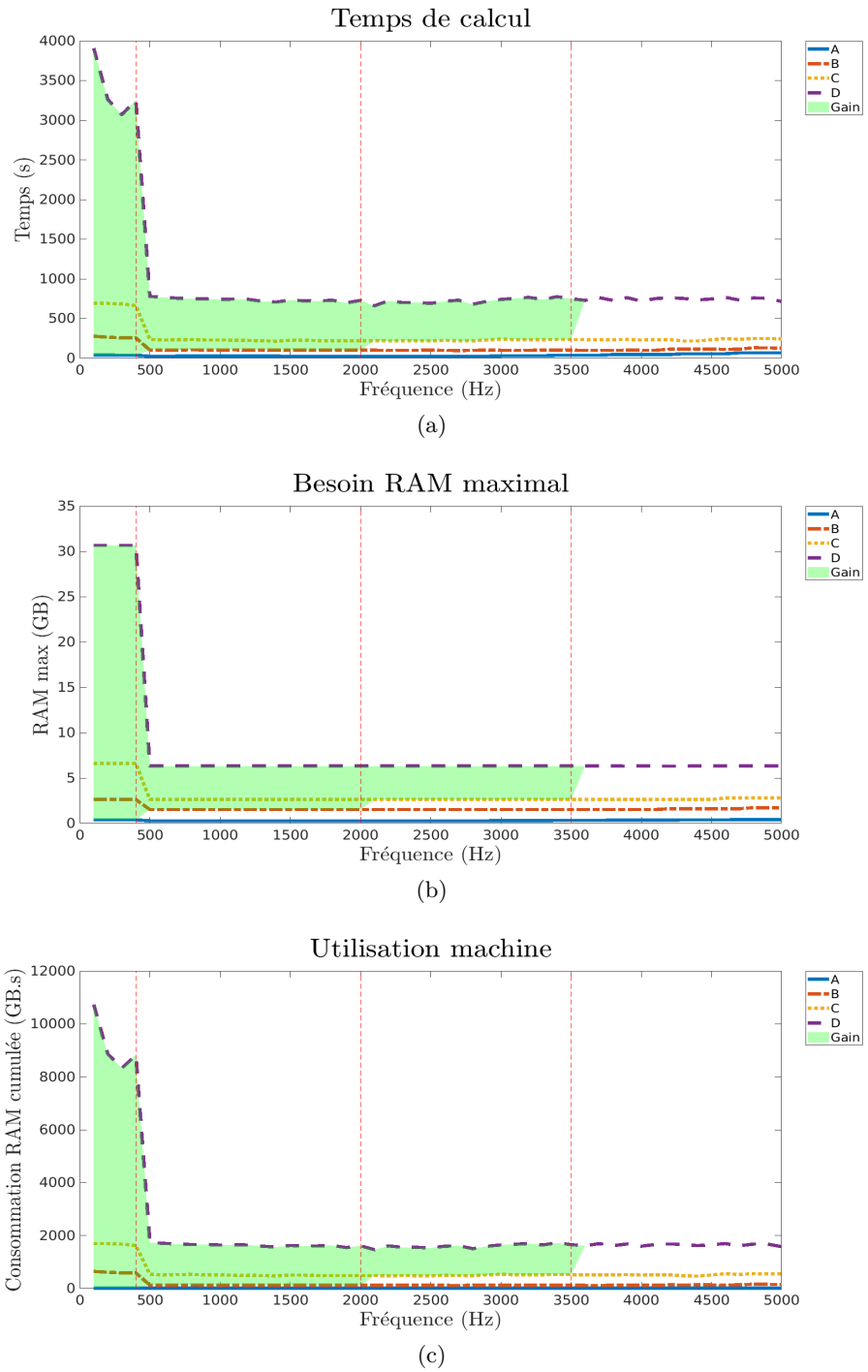


FIGURE 4.18: *De haut en bas, le temps de calcul (a), le pic de RAM (b) et l'utilisation machine (c) pour chacun des maillages. La zone verte correspond aux gains obtenus par l'utilisation des différents maillages au lieu du plus fin uniquement.*

4.2.2.3 Gains calculatoires

Côté performances, l'utilisation de la dépendance en fréquence apporte des gains manifestes. Ceux-ci sont clairement visibles sur les graphiques de temps de calcul, de besoin en RAM et d'utilisation machine – cf. figure 4.18. En cumulé, sur la plage [100, 3 500] Hz, elle représente une économie de 85,6 % du temps de calcul, les ordinateurs voient leurs capacités maximales en RAM diminuées de 79,4 % et cela s'en ressent sur l'utilisation $RAM \times temps\,CPU$ des machines, qui est réduite de 89,1 %.

Il est à noter que la tranche [100, 400] Hz a la particularité d'être beaucoup plus gourmande en RAM que le reste du spectre. Ceci est dû au fonctionnement même de mesh2hrtf qui sélectionne le type de formulation du problème le plus adapté à la fréquence. En l'occurrence, pour ces basses fréquences, la BEM simple est considérée plus fiable que la FM-BEM. Bien que cela soit coûteux en puissance de calcul, nous avons choisi de ne pas outrepasser ce paramétrage, issu de l'expertise des concepteurs du moteur numérique.

4.2.2.4 Conclusion

En ramenant ces résultats à l'échelle d'une simulation entière, i.e. sur [100, 16 000] Hz, l'utilisation conjointe du maillage adaptatif et de la dépendance en fréquence nous a permis de réduire le temps de calcul de plus de 10 % et de faire tourner les simulations sur des ordinateurs à la puissance plus raisonnable. Cela s'est fait sans altérer outre mesure la précision des résultats ni, plus important encore, toucher à la plage de fréquences traditionnellement associée à l'oreille.

4.3 Impédance des matériaux

Tout au long de cette thèse, de nombreuses expérimentations et tests subjectifs ont été compilés. Certains se sont révélés conformes aux attentes, d'autres beaucoup moins. En particulier, certaines différences entre les HRTF simulées et les acoustiques – différences que l'on pourrait considérer, à première vue, comme des dérives fréquentielles – ainsi que leurs conséquences en terme de performances aux test subjectifs de localisation nous ont amené à revoir nos positions sur le paramétrage généralement admis en simulation. Plus précisément, après avoir éliminé les suspects les plus évidents tels que la présence de bugs dans le moteur de calcul, une vitesse du son dans l'air mal paramétrée, un point virgule manquant ou tout autre erreur malencontreuse du même type, nous en sommes arrivé à émettre l'hypothèse que l'approximation consistant à assimiler la peau à un matériau acoustique totalement réfléchissant est trop grossière et ne permet pas, dans le cas général, d'aboutir à des HRTF au rendu subjectif significativement amélioré.

La section qui suit revient sur l'analyse du problème, la formulation de l'hypothèse, les expérimentations menées pour la tester ainsi que les résultats obtenues et les zones d'ombre restantes.

4.3.1 Problème des simulations

Tout d'abord, il est à noter que le problème que nous qualifions ici – et par abus de langage – de *dérive fréquentielle* a déjà été observé par le passé. En effet, l'équipe à l'origine de la base SYMARE a déjà rapporté des différences du même ordre entre les HRTF acoustiques des sujets qu'ils avaient scannés et les HRTF calculées associées [84]. En pratique, après alignement optimal des jeux de HRTF, ils ont mesuré les facteurs d'échelle optimaux faisant correspondre les jeux acoustiques et simulés. Alors que la valeur attendue vaut tout simplement 1, les valeurs obtenues se sont échelonnées entre 1 et 1,15. L'explication alors avancée a été de supposer une différence de température – et donc de vitesse de propagation du son – significative entre l'acquisition et la simulation.

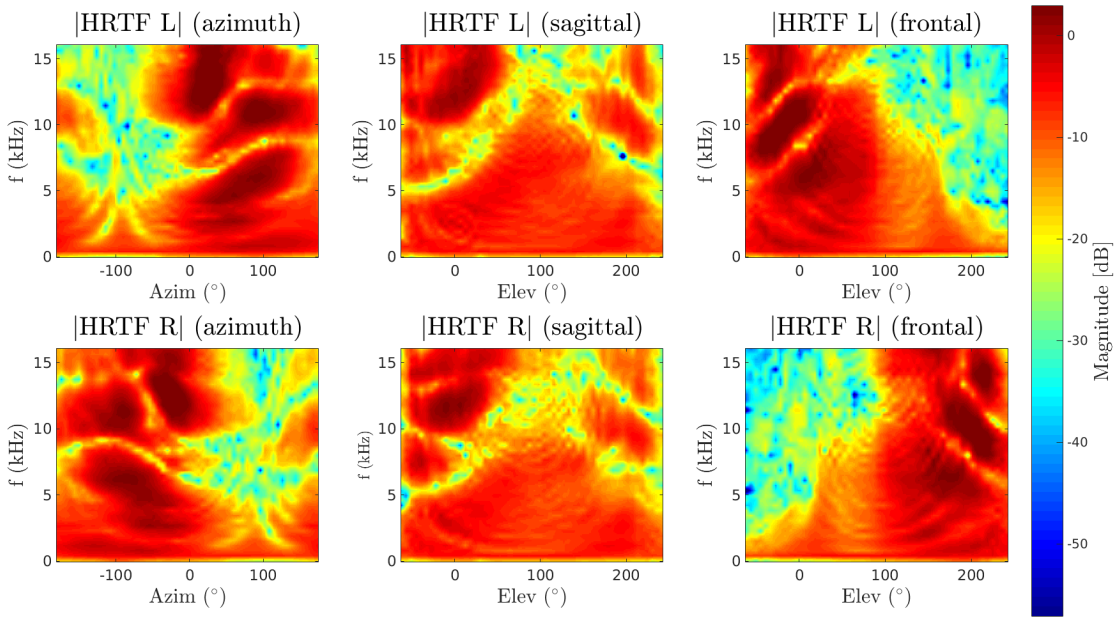


FIGURE 4.19: *De gauche à droite, les coupes dans les plans azimutal, sagittal et frontal des HRTF de référence du sujet 006. Mesure à l'oreille gauche en haut, à l'oreille droite en bas.*

Dans notre cas, nous avons mis à profit les quelques HRTF acquises à Orange Labs pour évaluer nos sorties de simulations. Les figure 4.19 et 4.20 présentent ces dernières et les HRTF calculées correspondantes pour l'un des sujets disponibles. Comme on peut le voir, quel que soit le plan de coupe, les motifs sont très semblables. Dans le plan azimutal, par exemple, les lieux de la fréquence de première annulation forment, côté ipsilatéral, une sorte de protubérance. Cela concerne aussi bien l'oreille gauche que l'oreille droite. Toutefois, des différences certaines sont également présentes. Pour n'en citer que trois, nous pouvons remarquer que :

1. ce motif se situe dans la zone [8, 9] kHz dans le cas Orange alors qu'il tourne autour des 10 kHz dans les simulations.
2. la zone du plan médian située entre 60° et 140° montre une plus grande atténuation au-delà de 10 kHz sur les mesures que par le calcul.

3. si les deux jeux de HRTF montrent bien des *ripples*, c'est-à-dire des sortes de vaguelettes sur chaque plan de coupe, leur amplitude est visiblement plus importante sur les jeux calculés que dans les HRTF de référence.

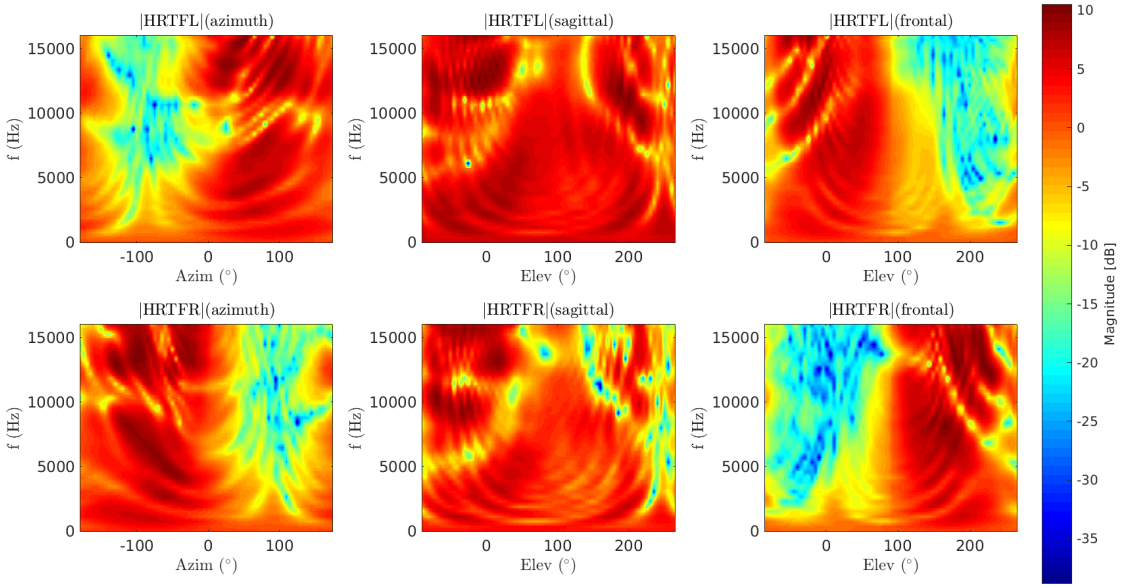


FIGURE 4.20: *De gauche à droite, les coupes dans les plans azimutal, sagittal et frontal des HRTF simulées du sujet 006. Calcul à l'oreille gauche en haut, à l'oreille droite en bas.*

En calculant le facteur d'échelle nécessaire pour faire correspondre les données acoustiques en notre possession aux simulations, nous observons un panel de valeurs allant de 1 à 1,25. Dans le cas présenté en exemple, elles sont de 1,18 pour l'oreille gauche et 1,25 pour l'oreille droite. Les coupes des HRTF résultantes sont visibles figure 4.21. La confirmation de ces écarts par des équipes, en des lieux et avec des calculateurs différents laisse peu de place au simple hasard ou à l'erreur malheureuse de manipulation.

Toutefois, afin de s'en assurer, de multiples tests ont été effectués. Tout d'abord, plusieurs vitesses de propagation du son, synonymes de températures différentes, ont été essayées sans véritable succès. Cela est quelque part rassurant car rien n'indique jusqu'ici que l'être humain se trouve dépourvu de ses capacités de localisation en-dessous de, mettons, 10°C ni au-dessus de 30°C par exemple. Ensuite, des simulations issues de Coustyx ont été comparées avec leurs équivalents calculés par mesh2hrtf. Dans les deux cas, la même dérive a été observée. De leur côté, les relectures de code et la vérification des dimensions des scans par rapport aux sujets réels n'ont rien donné non plus. Enfin, la mise à plat des hypothèses sous-jacentes de l'acoustique linéaire – cf. section 2.1.2 – n'a fait que confirmer la validité de ce cadre théorique.

Mais alors, sachant que lorsque vous avez éliminé l'impossible, ce qui reste, aussi improbable soit-il, est nécessairement la vérité [43], nous en sommes venus à suspecter les conditions aux limites de nos simulations, et plus précisément l'impédance attribuée à la peau humaine dans le spectre audible.

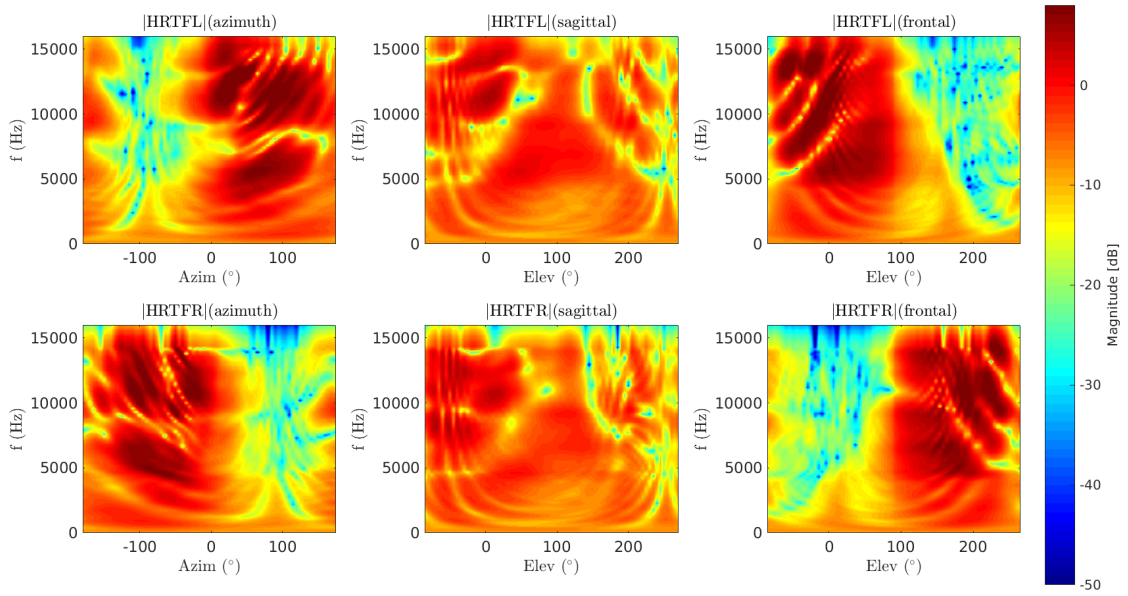


FIGURE 4.21: *De gauche à droite, les coupes dans les plans azimutal, sagittal et frontal des HRTF simulées en condition rigide, avec mise à l'échelle en fréquence, du sujet 006. Calcul à l'oreille gauche en haut, à l'oreille droite en bas.*

À notre connaissance, assez peu d'études se sont penchées jusqu'à présent sur le sujet. En 2000, Katz [89] étudie les impédances des cheveux et de la peau entre 1 et 6 kHz. Il observe que la peau semble très proche d'un matériau rigide sur cette plage de fréquences. Les cheveux en revanche présentent un coefficient d'absorption notable et dépendant de la fréquence. Il souligne également, étant donnée la méthode de mesure employée, l'impossibilité de discuter de l'impédance des matériaux en cas d'incidence oblique. En 2001, dans la continuité de ses travaux, il s'intéresse donc à l'influence des cheveux sur les HRTF grâce à la BEM [90] et observe déjà des variations dues à la présence de cheveux pouvant aller jusqu'à 6 dB.

En 2007, Treeby *et al.* publient une série d'expériences visant à étudier l'effet de l'impédance des cheveux sur les HRTF et les indices de localisation qu'elles véhiculent [146, 147, 145]. Ces travaux comportent des mesures d'impédance dans différentes conditions et des mesures de HRTF d'une sphère de rayon 12,4 cm munie ou non de cheveux. Là encore le spectre fréquentiel est limité aux basses fréquences, en l'occurrence [375, 3 000] Hz. Ils effectuent dans cette plage la mesure de l'impédance d'un matériau équivalant aux cheveux pour des angles d'incidences variant de 40° à 90°. Les auteurs ne discutent pas l'importance des variations en tant que telles mais font tout de même l'observation que ces variations sont de plus en plus faibles à mesure que l'on monte en fréquence. En ce qui concerne les indices de localisations, les auteurs se limitent à l'étude de l'ITD et de l'ILD. Ils présentent néanmoins des variations significatives. Ainsi l'ITD mesuré peut croître de 100 μs si l'on affuble la sphère rigide d'un matériau très absorbant. L'ILD quant à lui peut varier de 10 dB à 3 000 Hz selon les impédances choisies.

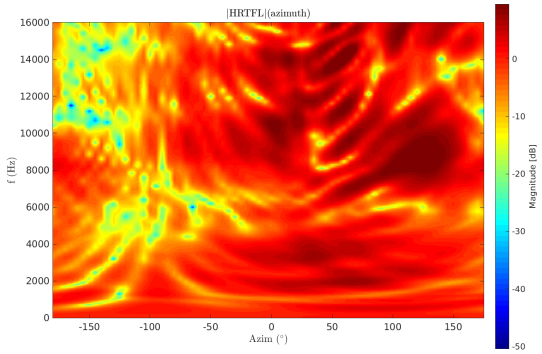
Indéniablement, l'enseignement que l'on peut en tirer est que l'impédance du matériau a une influence significative sur la HRTF. En outre, notre savoir en la matière est très

parcellaire, le spectre au-delà de 6 kHz étant *terra incognita*. Il est donc véritablement un sujet à creuser et que nous proposons d'attaquer sous l'angle de la simulation numérique.

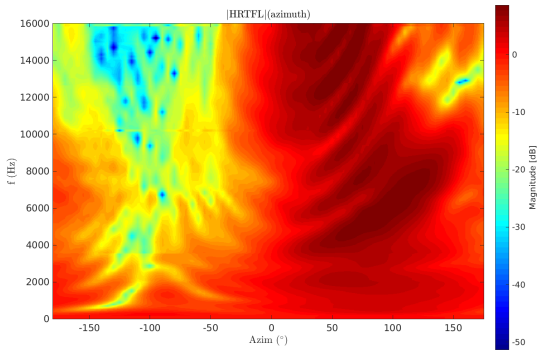
4.3.2 Existence d'une solution par l'impédance

4.3.2.1 Protocole expérimental

Afin de tester la validité de notre hypothèse, à savoir que les différences entre HRTF acoustiques et simulées peuvent s'expliquer par les valeurs données à l'impédance des matériaux (peau, cheveux, etc.), nous avons commencé par sélectionner le meilleur jeu de HRTF [4] acoustiques dont nous disposions. Il s'agit de HRTF acquises à Orange Labs et qui ont donné de très bons résultats subjectifs lors des tests de localisation. Ce jeu constitue notre référence.



(a) $R = e^{0.58j}$



(b) $R = 0.55 * e^{-0.009j}$

FIGURE 4.22: *Coupes azimutales de HRTF du sujet 006 obtenues pour deux valeurs extrêmes du coefficient de réflexion. Si les motifs sont encore assez semblables en basses fréquences, ils diffèrent de plus en plus à mesure que l'on monte en fréquence.*

En parallèle, le scan complet (tête, torse, oreilles) le plus propre du sujet dont il est question a été choisi et a servi pour les simulations. Lors de celles-ci, les microphones ont été placés au fond des canaux auditifs du maillage, lui-même situé au centre d'une grille d'évaluation de rayon 2 m similaire à la grille d'Orange. Les HRTF droite *et* gauche sont donc à l'étude. Comme dans le reste de ces travaux, ces dernières sont calculées sur la plage [100, 16 000] Hz avec un pas de 100 Hz. L'impédance sur le maillage est quant à elle uniforme et prend une valeur complexe dépendante de la fréquence.

Afin de mieux contrôler le domaine de recherche, nous choisissons le coefficient de réflexion, de valeur complexe, comme variable d'optimisation. Contrairement à l'impédance – qui peut être arbitrairement grande – il a l'avantage d'avoir un module compris entre 0 et 1. Cela n'est toutefois qu'un jeu d'écriture et, en notant Z l'impédance réduite [5] et R le coefficient de réflexion, la relation entre l'un et l'autre à la fréquence f est donnée par l'équation 4.2. Le cas totalement réflexif $R = 1$ correspond à une impédance infinie.

4. ou plus exactement « de DTF »

5. L'impédance absolue s'en déduit en la multipliant ensuite par l'impédance du milieu de propagation, ici l'air, $Z_0 = 409,84 \frac{kg}{m^2.s}$

$$Z(f) = \frac{1 + R(f)}{1 - R(f)} \tag{4.2}$$

Pour mieux appréhender l'impact possible de ce type de manipulation, la figure 4.22 en montre les effets dans le plan azimutal sur les simulations du sujet 006 pour deux valeurs extrêmes du coefficient de réflexion appliquées à l'ensemble du spectre. Comme on peut le constater, l'impédance apparaît comme un puissant levier d'ajustement.

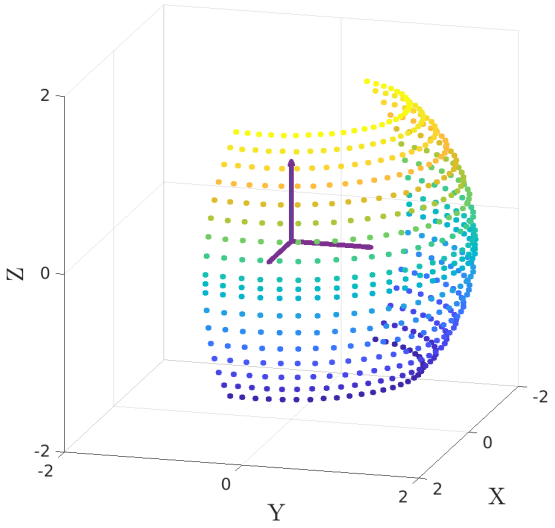


FIGURE 4.23: *Sous-grille d'évaluation utilisée pour les calculs d'optimisation de l'impédance.*

En sortie, la moyenne quadratique des écarts entre la simulation et la référence est calculée et nous informe sur notre distance vis-à-vis de la vérité terrain. Cette moyenne, calculée fréquence par fréquence, est pondérée par le diagramme de Voronoï mais, pour ne pas être affectée par les valeurs très faibles des magnitudes des HRTF controlatérales, elle est également limitée à une portion de l'espace les excluant. De manière analogue, la grille d'évaluation d'Orange Labs ne comprenant pas les élévations inférieures à $-62°$, une limite est là aussi à poser. Et en prenant en compte le fait que la présence de cheveux risque d'impacter négativement le résultat, une limite haute en élévation est également la bienvenue. C'est en définitive une sous-grille restreinte aux directions vérifiant $-15° < az < 160°$ et $-45° < el < 55°$ et représentée figure 4.23 qui a été retenue. Celle-ci n'est bien sûr valable que pour l'oreille gauche et donne naissance, par symétrie selon le plan médian, à une seconde sous-grille, destinée à l'oreille droite. Bien évidemment, l'objectif de l'opération est de trouver la courbe d'impédance minimisant l'écart entre simulations et mesures.

4.3.2.2 Recherche par simplex

Selon une première approche du problème, un algorithme du simplex est lancé pour chaque fréquence. En plus d'être facile d'utilisation, il peut sans peine traiter plusieurs variables et est assuré de converger. Cette convergence demande en contrepartie de nombreuses itérations, ce qui est essentiellement une gêne pour l'optimisation des hautes fréquences. D'autre part, le simplex n'est pas assuré de trouver un minimum global.

Les HRTF de référence, celles avant optimisation et celles après optimisation de l'impédance sont visibles figure 4.19, 4.20 et 4.24. Les courbes des coefficients de réflexion obtenus sont quant à elles présentées figure 4.25.

On peut noter que les deux premières différences que l'on a soulignées précédemment se trouvent amoindries. Les indices spectraux sont en effet un peu mieux placés – bien que

la « protubérance » de l'oreille droite ait quasiment disparu – et la zone du plan médian, à l'origine trop forte, voit son amplitude générale nettement réduite.

De manière inattendue, le coefficient de réflexion varie beaucoup entre 1 et 6 kHz alors que la littérature laissait à penser que l'on se serait maintenu proche du cas rigide. Cela doit néanmoins être nuancé car très peu de motifs sont présents dans les basses fréquences et l'expérience nous a montré – cf. figure 4.22 – que les changements d'impédance y avaient également moins d'impact.

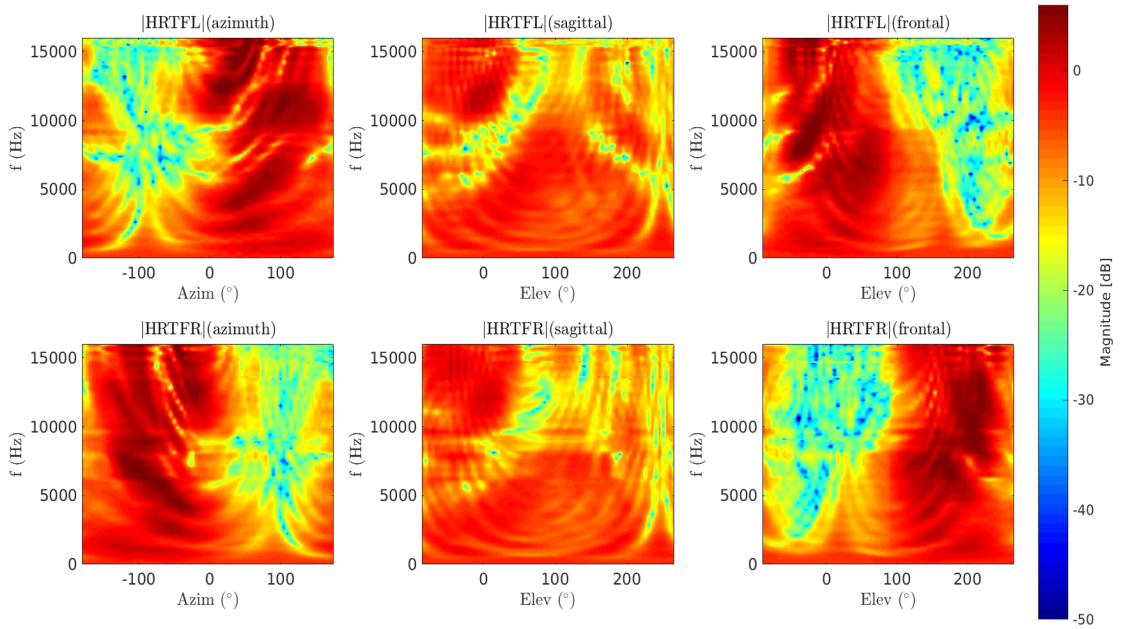
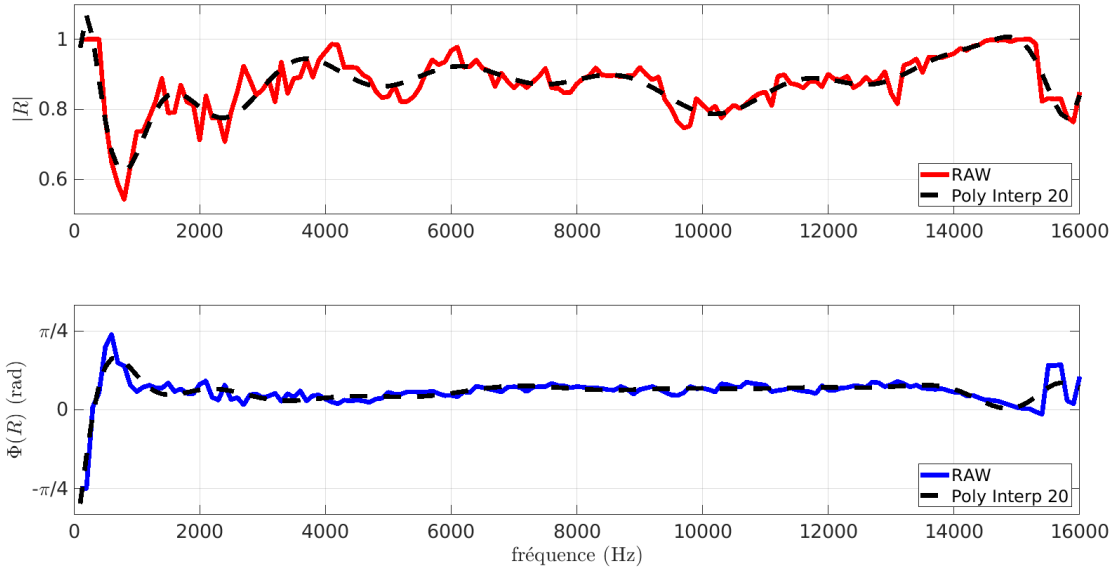


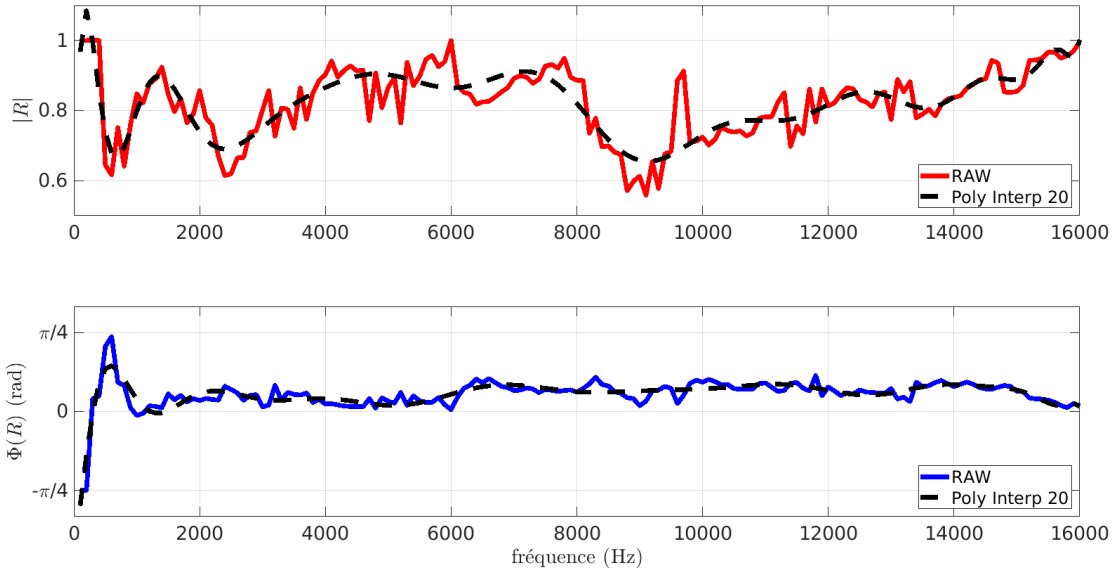
FIGURE 4.24: *De gauche à droite, les coupes dans les plans azimutal, sagittal et frontal des HRTF gauche (en haut) et droite (en bas) obtenues après optimisation de l'impédance.*

La phase, quant à elle, est assez stable et proche de 0,17 rad sur la quasi-totalité du spectre. Bien que cette stabilité ne soit pas pour déplaire, le signe, positif, de la phase laisse quelque peu perplexe. L'intuition physique aurait en effet voulu qu'une réflexion induise un retard de phase – donc de signe négatif – et non une avance de phase. Est-ce simplement dû à une erreur d'interprétation ? S'agit-il d'un effet normal du principe de réciprocité ? Notre impédance numérique, obtenue par optimisation sur des données réelles, vient-elle également compenser les différences de forme et d'orientation entre le sujet et son maillage ? Nous n'avons pour l'heure pas d'explication indiscutable à cette observation et préférons laisser cette question ouverte tant qu'elle n'entrave pas la poursuite des expérimentations.

Ces remarques sont valables aussi bien pour l'oreille droite que pour la gauche. Une observation supplémentaire à faire est que les modules optimisés pour une oreille sont relativement différents de ceux de l'autre oreille. Or si l'on manipule effectivement la valeur du coefficient de la peau, il n'est pas très intuitif que celui-ci doive dépendre du côté du sujet. Parmi les explications possibles, on peut encore une fois avancer l'idée que le paramètre « impédance » que nous manipulons voit d'une certaine manière son optimisation être « polluée » par des différences géométriques. Par ailleurs, comme nous l'avons rappelé



(a) Oreille gauche



(b) Oreille droite

FIGURE 4.25: *Courbes de module et de phase des coefficients de réflexion optimaux pour l'oreille gauche (a) et l'oreille droite (b). En pointillés, les résultats obtenus après interpolation par un polynôme de degré 20.*

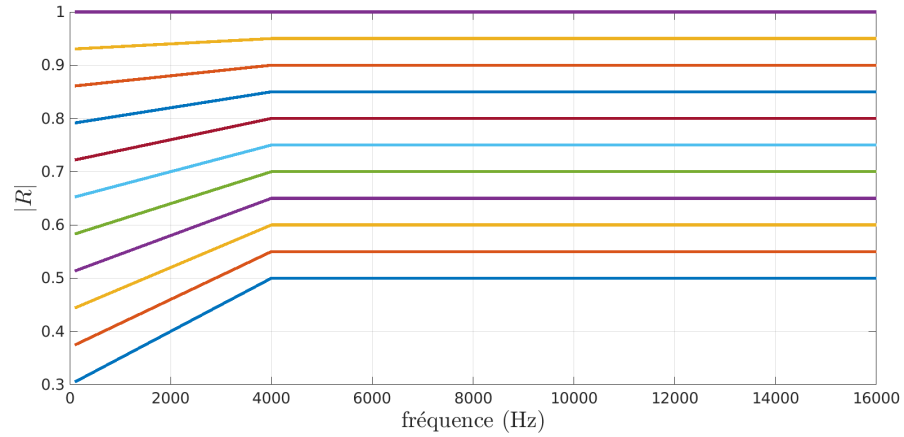
en début de paragraphe, rien n'assure que le simplex ait trouvé la meilleure solution au problème posé et nous pouvons tout à fait être en présence de deux minima locaux.

4.3.2.3 Recherche quadrillée

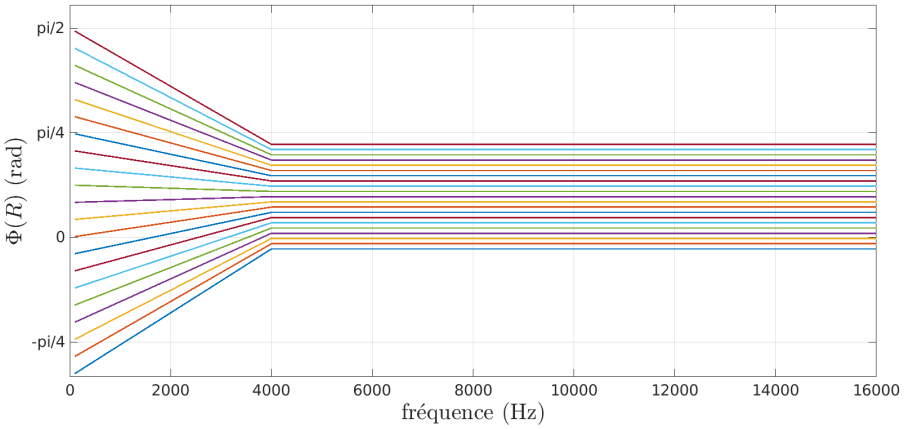
Un second mode d'optimisation a été mis en œuvre afin d'éclaircir ce point et dans lequel la zone de recherche couverte par le simplex est échantillonnée de façon régulière. Pour garder un maximum de clarté dans l'analyse et les figures, seule l'oreille gauche est ici étudiée. Toutefois, l'analyse des données issues des simulations de l'oreille droite mène à

des conclusions similaires.

De façon plus concrète, pour chaque fréquence, 11 valeurs distinctes ont été retenues pour $|R|$ et 21 pour $\arg(R)$ (comme le montre la figure 4.26), le tout pour un total de 231 combinaisons possibles [6]. De cette façon, la carte des erreurs nous est révélée et nous donne accès à une meilleure vue d'ensemble de ses minima.



(a)



(b)

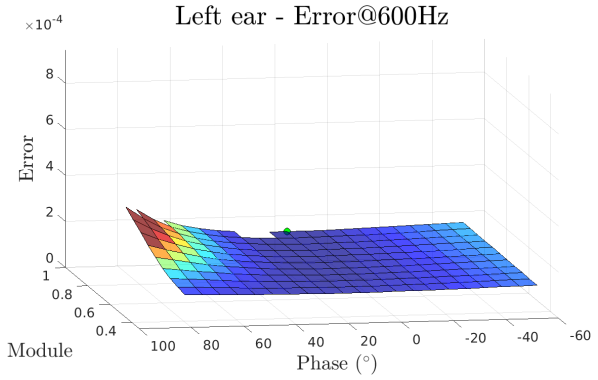
FIGURE 4.26: *Courbes d'échantillonage des modules (a) et phases (b) des coefficients de réflexion.*

Une fois ces cartes établies, le minimum, à chaque fréquence, en est extrait et la courbe complète reconstruite sur l'ensemble du spectre. La figure 4.28 présente le résultat de l'optimisation et le compare au simplex tandis que la figure 4.27 montre quelques exemples des cartes obtenues pour l'oreille gauche.

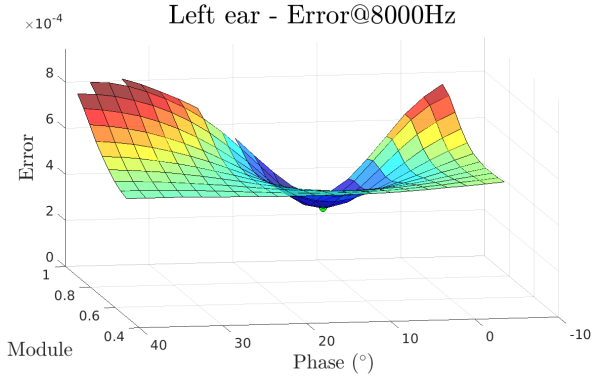
Parmi les enseignements à tirer de cette visualisation, il y a la confirmation que les basses fréquences sont peu altérées par les modifications d'impédance. Ainsi, à 600 Hz, l'erreur est très faible et ne varie quasiment pas avec R. Le fait que le simplex y ait convergé vers une valeur très éloignée du cas rigide, en dépit de ce que la littérature nous laissait attendre, n'est donc pas un véritable problème. C'est en montant en fréquence, à partir de 5 000 Hz, que le niveau général de l'erreur augmente significativement et que de véritables

6. Le choix de ces valeurs s'appuie sur la recherche par simplex.

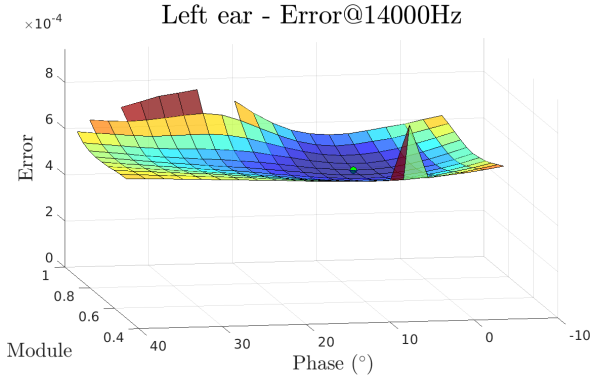
vallées, aux minima locaux marqués, commencent à apparaître. Cela se constate d'ailleurs très bien en visualisant les courbes des minima, figure 4.29. Là, les valeurs d'erreurs obtenues grâce à l'interpolation du simplex et à la recherche quadrillée sont comparées aux erreurs initiales.



(a)



(b)



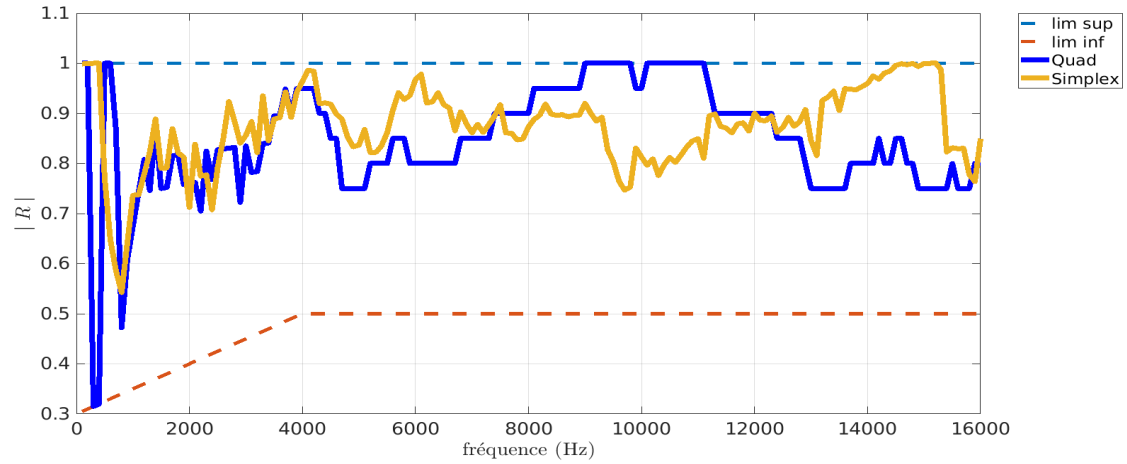
(c)

FIGURE 4.27: *Cartes des erreurs de l'oreille gauche pour $f = 600\,Hz$, en (a), $f = 8\,000\,Hz$, en (b) et $f = 14\,000\,Hz$, en (c). En rouge, les plus fortes valeurs, en bleues les plus faibles. Quelques simulations en échec expliquent les données manquantes. Le point vert symbolise le minimum du quadrillage, le point jaune celui obtenu grâce au simplex.*

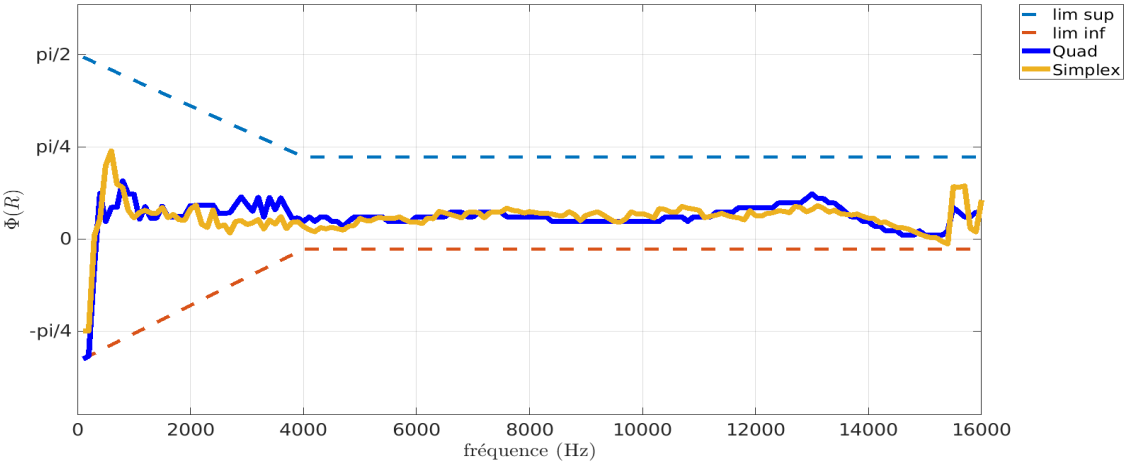
Si le simplex parvient par moment à trouver de meilleures solutions que le quadrillage systématique, il apparaît aussi bien souvent pris dans des minima locaux. L'interpolation polynomiale n'est bien sûr pas tout à fait étrangère aux moindres performances du simplex, mais même dans les cas où cette interpolation demeure anecdotique cette première méthode n'améliore qu'à la marge les résultats obtenus par quadrillage. Il ne semble donc pas utile de raffiner davantage les pas d'échantillonnage, qu'il s'agisse de celui du module ou de la phase de R.

Parmi les autres constats se trouve le fait que ces deux optimisations ont drastiquement réduit l'écart entre les HRTF acoustique et simulée, faisant passer l'ISSD sur la plage $[100,\ 16\,000]$ Hz de $37{,}95\,\mathrm{dB}^2$ à $15{,}79\,\mathrm{dB}^2$ ou $18{,}26\,\mathrm{dB}^2$, selon l'optimisation retenue. On doit également noter que les erreurs sont basses à très basses en-deçà de 5 kHz et grimpent plus fortement au-delà mais que les gains liés à l'optimisation, eux, n'apparaissent qu'à partir d'environ 6 kHz et se réduisent à nouveau à partir de 14 kHz En d'autres termes, l'apparente contradiction soulevée plus haut entre nos résultats de simulation et les données issues de la littérature se voit grandement mise à mal. Certes, nos optimisations nous ont conduit à privilégier des impédances relativement éloignées des impédances jusque là mesurées expérimentalement mais ces impédances n'ayant pas été mesurées au-delà de 6 kHz, elles demeurent dans une zone fréquentielle où l'optimi-

sation montre un très faible rendement. Or, nous l'avons déjà souligné, notre approche ne permet pas de distinguer les erreurs provenant d'une impédance mal calibrée de celles provenant de différences dans la géométrie du maillage. On peut donc tout à fait concilier la vision quasi-rigide de la peau, jusqu'à un peu moins de 6 kHz avec nos résultats d'expérience en considérant que nous avons essentiellement corrigé les écarts géométriques. Cela n'est en revanche plus possible au-delà, tant la réduction de l'erreur est importante. En suivant cette logique, il est aussi permis d'envisager que l'impédance réelle redevienne quasi-rigide au-delà de 15 kHz, expliquant ainsi la réduction du gain d'optimisation.



(a)



(b)

FIGURE 4.28: *Modules (a) et phases (b) des coefficients de réflexion obtenus grâce aux deux méthodes d'optimisation. En bleu, le résultat de la recherche quadrillée. En jaune, le résultat du simplex.*

En marge de ces résultats, il convient également de s'attarder sur l'aspect en « dents de scie » de l'erreur de la simulation rigide. En effet, la fréquence de ces variations n'est pas sans rappeler celle des vaguelettes évoquées en début de section et qui comptait parmi les problèmes manifestes de ce type de simulations. Jusqu'alors, il ne s'agissait que d'une observation qualitative. Désormais, elle semble quantifiable et apparaît comme un obstacle potentiel à une optimisation correcte. En effet, l'idée généralement admise pour expliquer

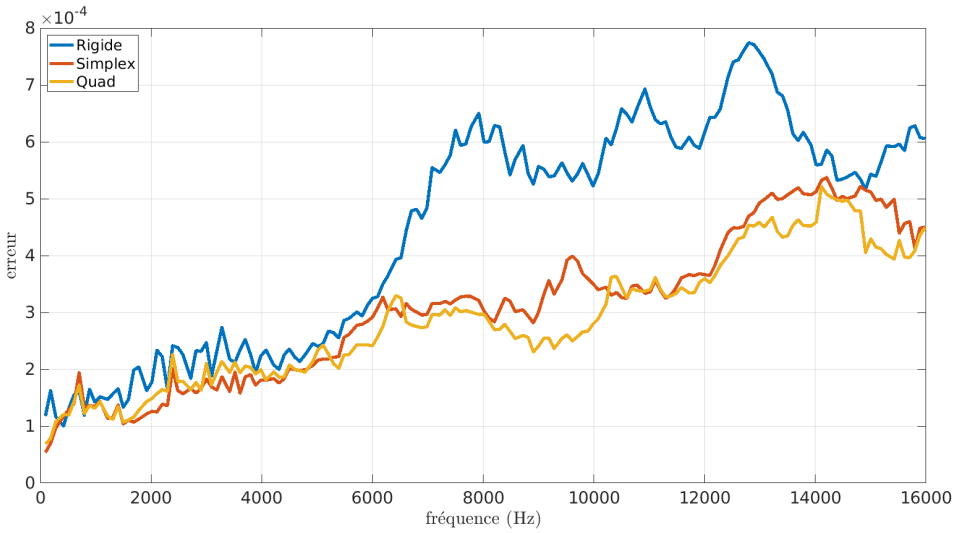


FIGURE 4.29: *Courbes d'erreurs obtenues selon chacune des approches du problème. En bleu, sans optimisation (simulation en condition rigide). En orange, l'optimisation par simplex. En jaune, celle par quadrillage.*

ce phénomène est que les réflexions sur les épaules et le torse viennent créer ce schéma d'interférences. Or, ce schéma d'amplitudes nous semble notablement plus marqué en simulations que dans le cas réel. Une étude différenciée des réflexions du buste et de la tête s'impose dès lors comme une suite naturelle.

4.3.3 Travail par zones

Pour ce faire, les simulations ont été modifiées pour attribuer au buste (cf. figure 4.30) une courbe de coefficient de réflexion propre. Celle-ci, visible figure 4.31, résulte des observations faites sur les simulations en notre possession avec en ligne de mire la correction des symptômes. Elle est donc résolument caricaturale et pourrait faire l'objet d'une étude future propre.

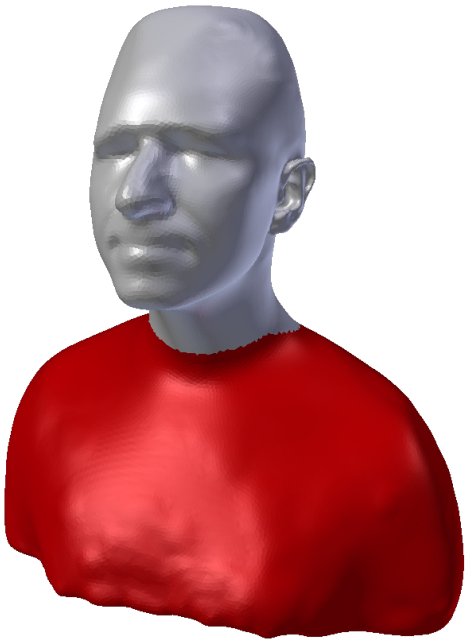


FIGURE 4.30: *Maillage utilisé en simulation avec, en rouge, le buste.*

Une simple simulation avec la peau en condition rigide et le torse absorbant est dans un premier temps réalisée. La HRTF obtenue – cf. figure 4.33 – ne laisse pas de place au doute : les réflexions acoustiques issues du torse viennent parasiter les HRTF en simulation et font partie des éléments à l'origine de la présence d'interférences très marquées en hautes fréquences.

L'élément le plus frappant est certainement la « protubérance » du plan azimutal, bien plus ressemblante à la vérité terrain – cf. figure 4.19 – que dans le cas entièrement rigide. Il est

aussi intéressant de constater que les interférences n'ont pas totalement disparu. On aurait en effet pu s'attendre à ce qu'elles soient à peine visibles tant le coefficient de réflexion a perdu en valeur. Pour être tout à fait précis, diminuer encore ce dernier aboutit à des simulations convergeant très difficilement.

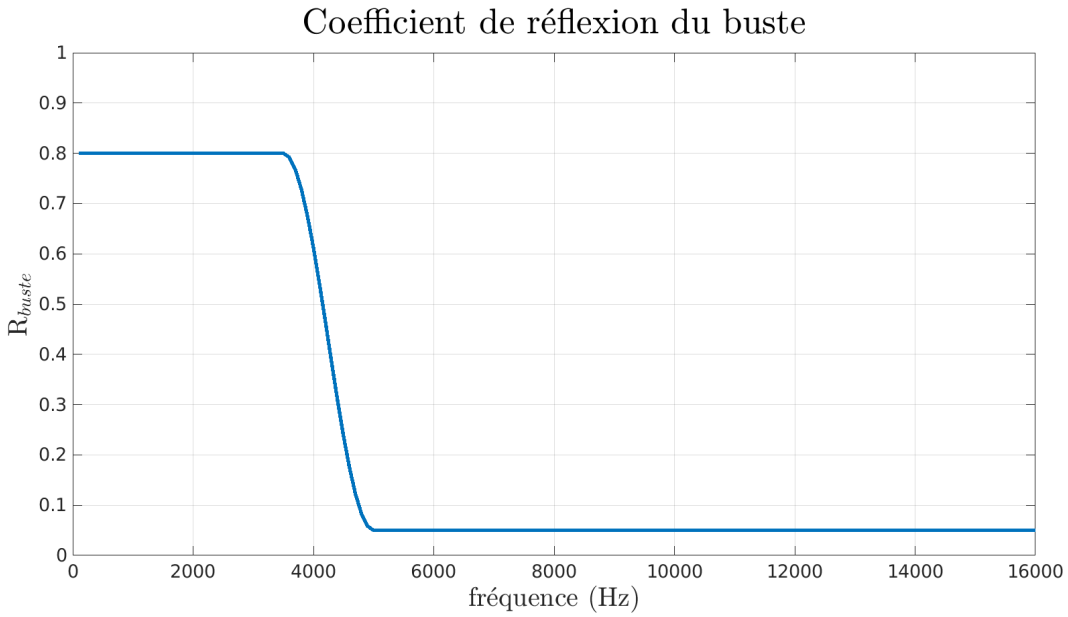


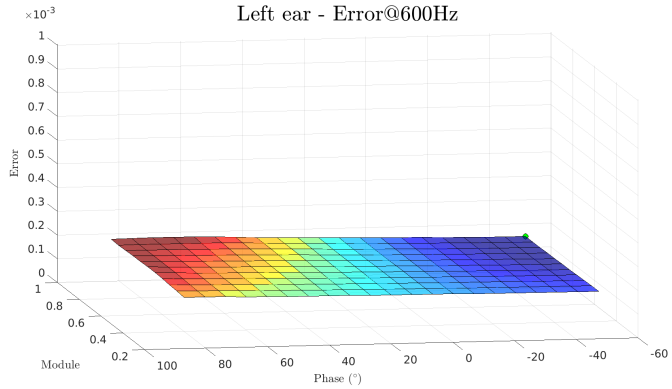
FIGURE 4.31: *Courbe du coefficient de réflexion appliqué au buste.*

Mais il n'en est rien et l'on distingue sans problème un schéma d'interférence du même type que celui de la vérité terrain sur l'ensemble du spectre. De deux choses l'une, soit le schéma restant est bien le résidu des interférences vues dans la simulation rigide, auquel cas il faut s'interroger sur le sens physique à donner aux valeurs – à priori fort peu réalistes – du coefficient de réflexion du torse, soit il ne l'est pas, auquel cas il y avait à l'origine superposition de deux schémas d'interférences, l'un lié au torse et l'autre à un élément différent. Dans ce dernier cas se posent les questions de la nature de cet élément, de l'utilité véritable du torse en simulation et des raisons de l'absence du schéma qui lui est propre. Ces interrogations dépassant le cadre de nos travaux actuels, nous les laissons ouvertes et sujets potentiels de futures recherches.

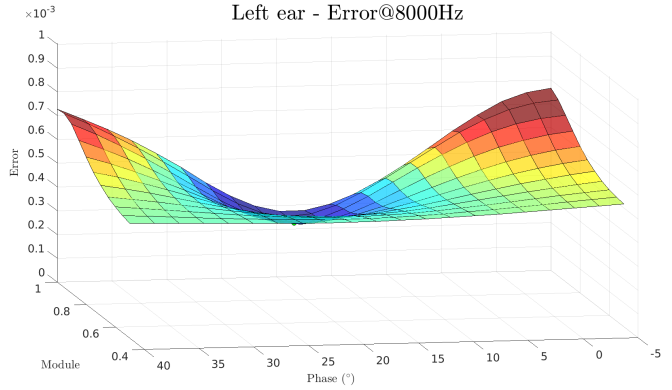
Pour l'heure, nous considérons comme acquis que l'ajout d'un profil d'impédance propre au torse permet d'éliminer de manière efficace un artefact des HRTF issues de simulations. Et si l'on reprend les expériences précédentes, cela devrait en toute logique se traduire par une meilleure optimisation. Pour le vérifier, un nouveau jeu de HRTF est simulé selon ces critères, c'est-à-dire 231 simulations avec torse absorbant et une impédance variable sur le reste du maillage.

La figure 4.34 présente le résultat de la nouvelle optimisation et le compare à la précédente tandis que la figure 4.32 montre quelques exemples des cartes obtenues pour l'oreille gauche. À nouveau, nous observons que les basses fréquences sont peu impactées par les modifications d'impédance. Ainsi, à 600 Hz l'erreur est encore très faible et ne varie quasiment pas avec R. En plus hautes fréquences se trouvent en revanche des vallées bien

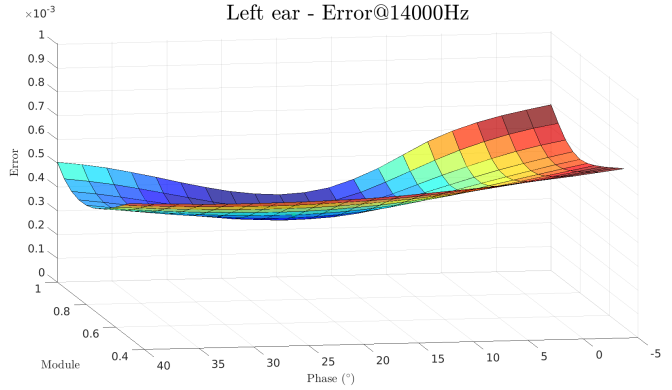
plus marquées et plus repérables.



(a)



(b)



(c)

FIGURE 4.32: *Cartes des erreurs de l'oreille gauche pour $f = 600\,Hz$, en (a), $f = 8\,000\,Hz$, en (b), et $f = 14\,000\,Hz$, en (c). En rouge les plus fortes valeurs, en bleu les plus faibles. Le point vert symbolise le minimum du quadrillage.*

Pour ce qui est des valeurs prises par R, beaucoup de variations figurent en basses et moyennes fréquences sans que cela soit un réel problème (comme nous l'avons expliqué précédemment en analysant les résultats à la fréquence $f = 600\,\text{Hz}$). À partir de 3 kHz, $|R|$ cesse de descendre en-dessous de 0,8, ce qui n'est pas un mal pour qui souhaite éviter de décrire la peau comme un matériau excessivement absorbant. De son côté, la phase est également assez chaotique en basses fréquences sans que cela soit là non plus un problème. Au-delà de 3 kHz elle se montre beaucoup plus stable sans pour autant être constante. La visualisation des courbes des minima d'erreur, figure 4.35, nous montre un progrès sensible sur les plages $[4,\,7]\,\text{kHz}$ et $[12,\,15]\,\text{kHz}$ – courbe *Quad + torse*.

En inspectant la HRTF résultante – visible figure 4.36 – on constate que celle-ci présente une amélioration par rapport à sa version sans optimisation, certains indices spectraux étant mieux positionnés. Néanmoins, la magnitude des directions controlatérales est parfois assez élevée et la trop faible atténuation au-dessus du sujet au-delà de 10 kHz demeure toujours présente. Il y a donc une amélioration mais ce n'est pas sans contreparties.

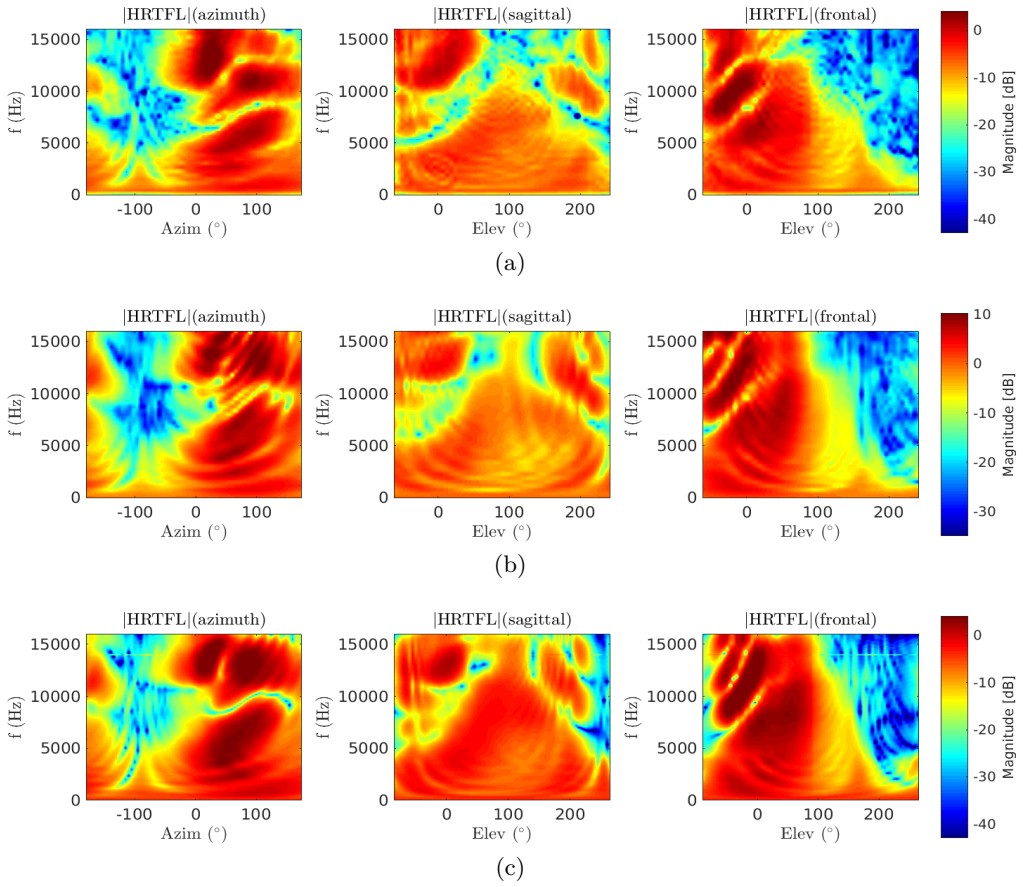


FIGURE 4.33: *En (a), coupes de la HRTF gauche du sujet 006. En (b), sa version simulée* sans *torse absorbant. En (c), sa version simulée* avec *torse absorbant.*

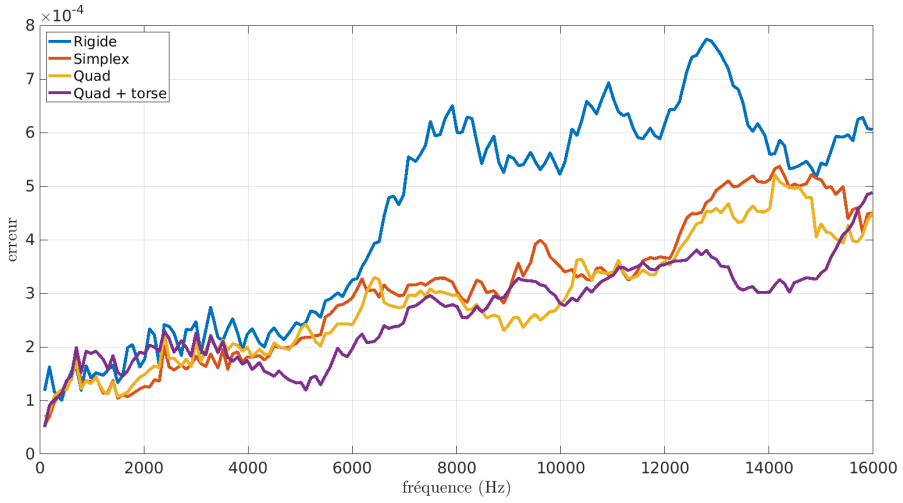
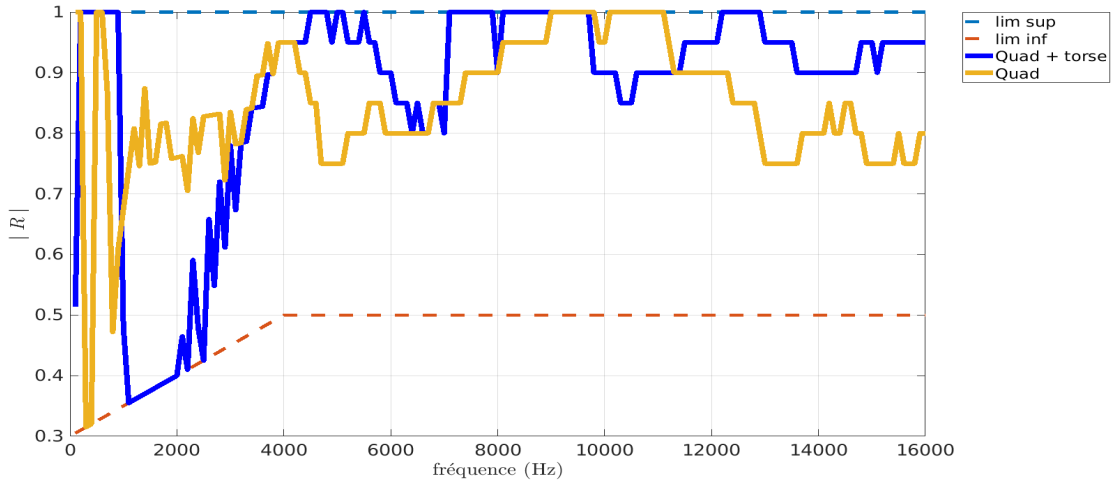
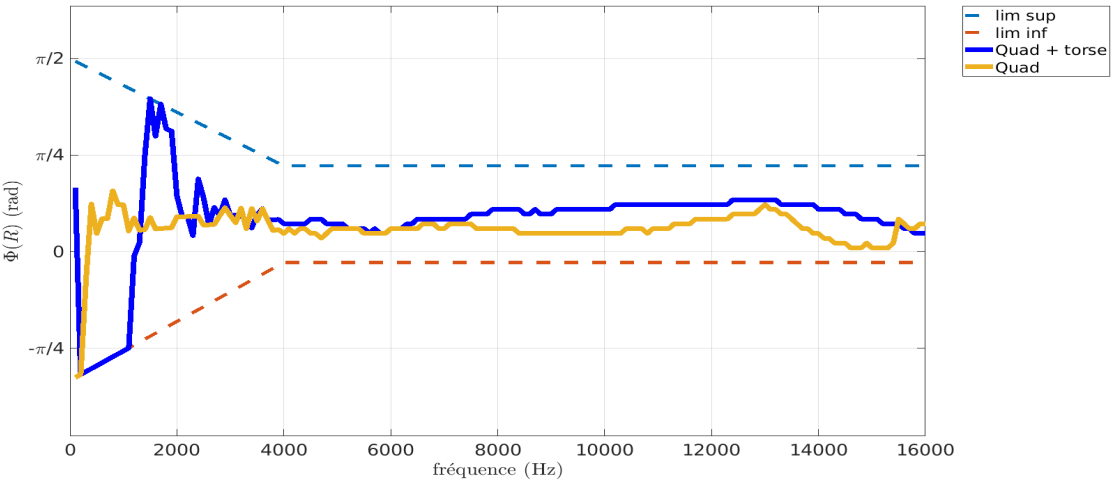


FIGURE 4.35: *Courbes d'erreurs obtenues selon chacune des approches du problème. En bleu, sans optimisation (simulation en condition rigide). En orange, l'optimisation par simplex. En jaune, celle par quadrillage. En violet, celle par quadrillage avec torse absorbant.*

Afin de répondre à ces problèmes, la tentation est grande de reprendre l'approche



(a)



(b)

FIGURE 4.34: *Modules (a) et phases (b) des coefficients de réflexion obtenues grâce aux deux recherches quadrillées. En jaune, le résultat sans torse absorbant. En bleu, le résultat avec torse absorbant.*

précédente et de l'appliquer aux cheveux. On les imagine en effet assez bien à même d'absorber une certaine quantité d'énergie, sans compter qu'ils figurent sur le trajet des HRTF controlatérales. Cette complexité toujours croissante du modèle force toutefois à la prudence, d'autant qu'un certain nombre de questions ont été laissées sans réponse. Pour l'heure, nous estimons donc préférable de s'en tenir au fait que l'application d'un traitement distinct au torse permet l'élimination de parasites dans les HRTF. Le traitement utilisé ici est très certainement exagéré et manque de sens physique pour convaincre totalement. Cependant, tout imparfait qu'il soit, il a permis de pousser un cran plus loin la recherche de l'impédance de la peau et cette dernière a pour sa part abouti à la production de HRTF calculées bien plus proches de la vérité terrain que ne le permet l'état de l'art. Ces différentes approches nécessitent maintenant de passer au crible de l'évaluation subjective et il nous faut discuter de leur généralisation à d'autres individus.

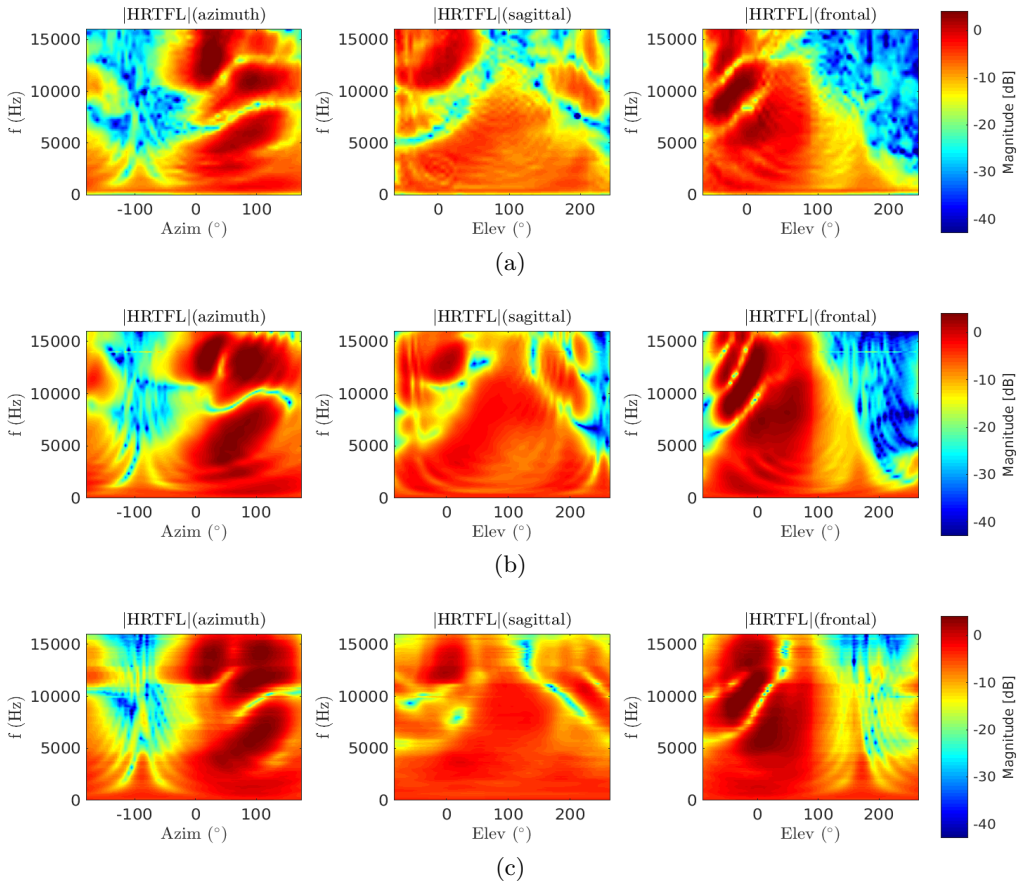


FIGURE 4.36: *En (a), le rappel de la vérité terrain. En (b) – resp. (c) –, les coupes de la HRTF gauche du sujet 006 avant – resp. après – optimisation avec torse absorbant.*

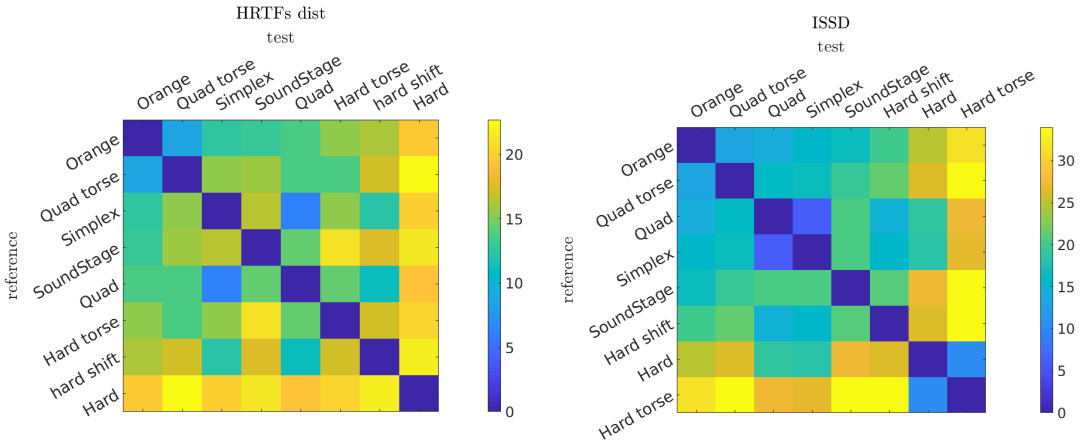
4.3.4 Évaluation et généralisation

En premier lieu, intéressons-nous aux distances entre HRTF et à ce qu'elles nous apprennent. Concrètement, deux métriques sont à notre disposition : l'ISSD d'une part et notre métrique subjective décrite annexe B d'autre part. Grâce à elles les huit jeux de HRTF suivants ont été comparés :

— *Orange* : HRTF acoustique de référence mesurée dans les laboratoires d'Orange Labs.
— *Hard* : HRTF calculée conformément à l'état de l'art, d'impédance infinie sur l'ensemble du maillage.
— *Hard shift* : HRTF hard mise à l'échelle en fréquence pour s'approcher le plus possible de la référence.
— *Simplex* : HRTF calculée avec des conditions d'impédance variables. L'optimisation utilise l'algorithme du simplex.
— *Quad* : HRTF calculée avec des conditions d'impédance variables. L'optimisation utilise un quadrillage de l'espace des coefficients de réflexion.
— *Hard torse* : HRTF hard avec des conditions d'absorbance spécifique pour le torse.
— *Quad torse* : HRTF quad avec des conditions d'absorbance spécifique pour le torse.

— *SoundStage* : HRTF acoustique mesurée grâce au Soundstage (cf. annexe A)

En calculant les distances séparant chaque couple de HRTF, nous obtenons les matrices présentées figure 4.37.



(a) (b)

FIGURE 4.37: *Matrices des distances entre HRTF. En (a), en utilisant notre métrique subjective, en (b), en utilisant l'ISSD.*

Par extraction de la première ligne de chacune d'entre elles – cf. figure 4.38 –, nous obtenons le classement des différents jeux, de la plus proche de la référence à la plus éloignée. On peut alors faire les observations suivantes :

— Quelle que soit la métrique employée, la HRTF Quad torse est la plus proche de la référence.
— Les HRTF hard sont les plus éloignées.
— La HRTF SoundStage, qui offre un second point de vue acoustique, la HRTF Quad et la HRTF Simplex sont toutes à distances équivalentes de la HRTF Orange.
— La HRTF Hard est deux fois plus éloignée de la HRTF Orange que ne l'est la HRTF Quad torse.

Afin d'affiner un peu plus notre compréhension de cet espace de HRTF, l'algorithme *isomap* a été employé pour y repositionner les jeux à notre disposition. Pour rappel, cet algorithme de réduction de dimension permet de trouver les positions optimales, dans un espace de dimension d, de points d'un espace de dimension $N > d$ à partir de la donnée des distances point à point. La variance résiduelle permet de quantifier l'information perdue lors de l'opération. En l'occurrence, en choisissant $d = 3$, nous obtenons les représentations des figures 4.39 et 4.40. Les variance résiduelles, sont quant à elles à peine de $2,79\,\%$ lorsque la distance choisie est l'ISSD et de $8,67\,\%$ lorsqu'il s'agit de la distance subjective. Dans ces conditions, on peut donc raisonnablement estimer que les représentations 3D obtenues sont pertinentes.

Et ces représentations nous montrent, plus clairement sans doute que les matrices de distances, que la HRTF Quad torse est bien la plus proche de l'objectif et que les HRTF hard sont les plus éloignées. Elles mettent également en évidence la proximité des HRTF

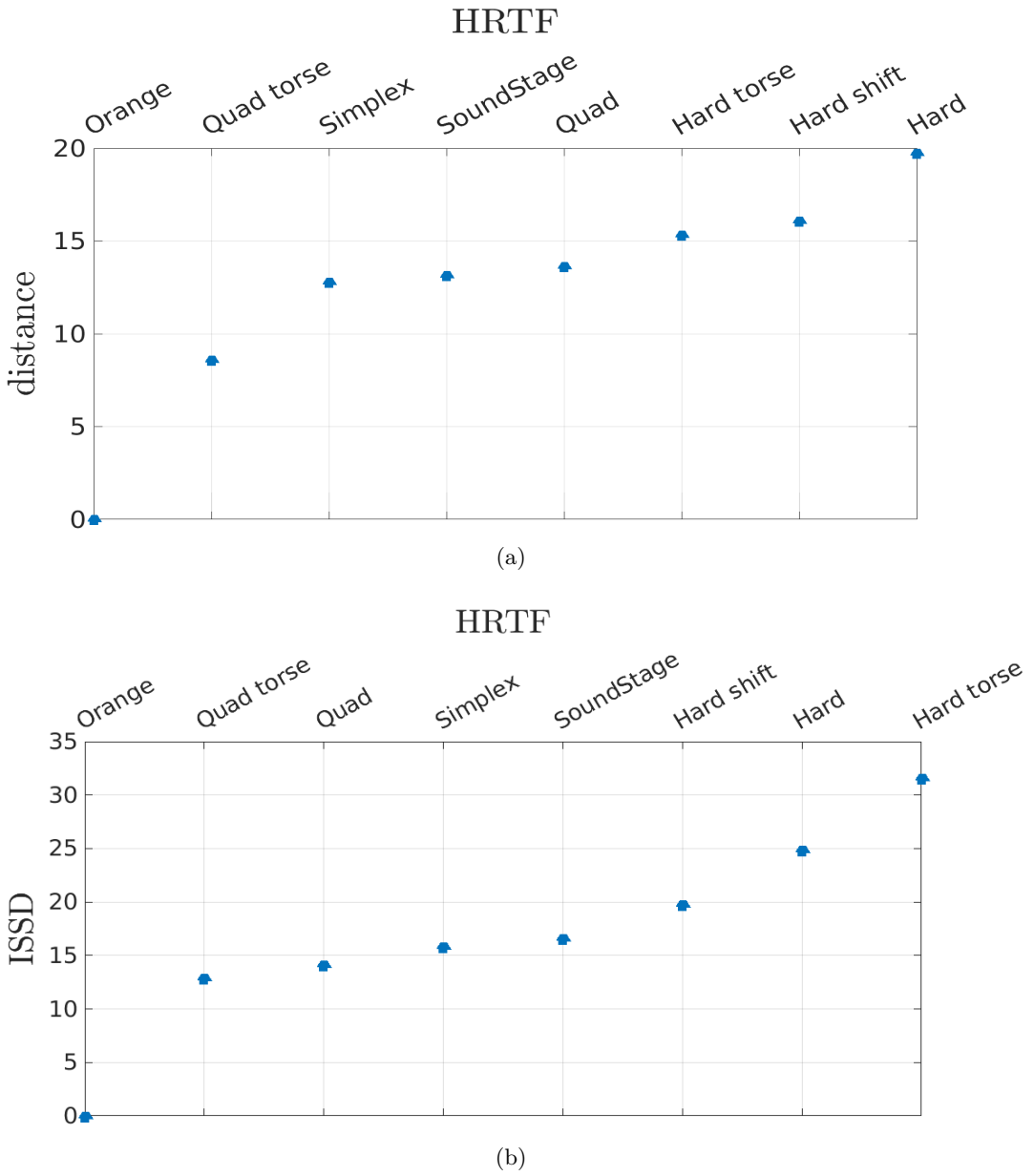


FIGURE 4.38: *Distances à la HRTF Orange. En (a), en utilisant notre métrique subjective. En (b), en utilisant l'ISSD.*

Quad et Simplex ainsi que l'isolement relatif de la HRTF SoundStage. Elles sont donc toutes trois aussi proches de l'objectif mais sans être équivalentes pour autant.

En définitive, nous avons pu exhiber un procédé de calcul de HRTF améliorant grandement la qualité du résultat, celui-ci se situant à mi-chemin entre l'objectif et une HRTF acoustique de moindre qualité. Ce procédé repose sur la connaissance de la morphologie du sujet et de l'impédance de la peau. Celle-ci n'étant pas d'emblée disponible sur l'ensemble du spectre, il a été nécessaire de la déterminer et cela s'est fait par optimisation sur simulations. Nous avons pu établir à cette occasion que l'impédance n'avait que peu d'impact sur la partie basse du spectre, jusqu'à 6 kHz. Au-delà, elle a permis de s'approcher considérablement de l'objectif.

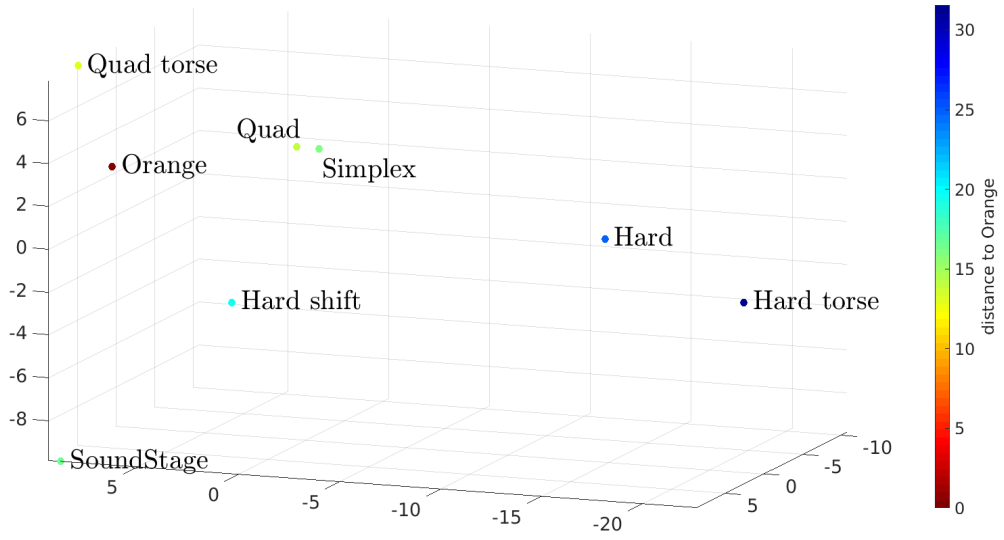


FIGURE 4.39: *Positionnement des HRTF par* isomap *dans un espace 3D en utilisant l'ISSD.*

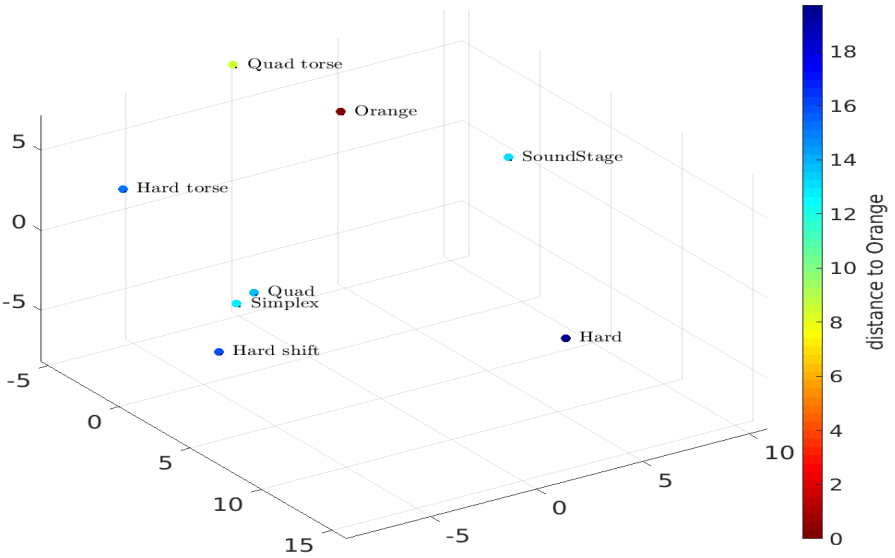


FIGURE 4.40: *Positionnement des HRTF par* isomap *dans un espace 3D en utilisant notre métrique subjective.*

Cette recherche d'impédance nous a néanmoins mis dans l'obligation d'utiliser la HRTF cible dans le processus d'optimisation, et donc de connaître par avance le résultat recherché. Or, dans le cas général, la HRTF acoustique du sujet est supposée inconnue et une seule simulation doit suffire à proposer une HRTF numérique de qualité équivalente. Les questions qui se posent donc désormais sont celles de l'évaluation subjective réelle et

de la généralisation du procédé.

HRTF	Erreur locale (°)	Erreur de quadrant (%)
Orange	31,31	5,88
Simplex	33,33	12,94
Quad	31,34	17,05
SoundStage	36,96	18,23
Quad torse	34,07	20,59
Hard shift	36,18	29,41
Hard	38,58	32,35
Hard torse	40,62	34,71
Kemar	43,43	38,23
Random	65,29	52,94

TABLE 4.3: Résultats subjectifs du sujet 006 au test de localisation en élévation avec différentes HRTF. La ligne *Random* correspond à la simulation d'un tirage aléatoire des réponses.

Pour cela, les HRTF produites jusque là ont fait l'objet des tests de localisation décrits en annexe B. Plus précisément, le sujet ayant servi aux recherches précédentes a également passé lesdits tests. Cinq séances ont été réalisées au total. Leurs résultats moyennés sont présentés table 4.3 et figure 4.41

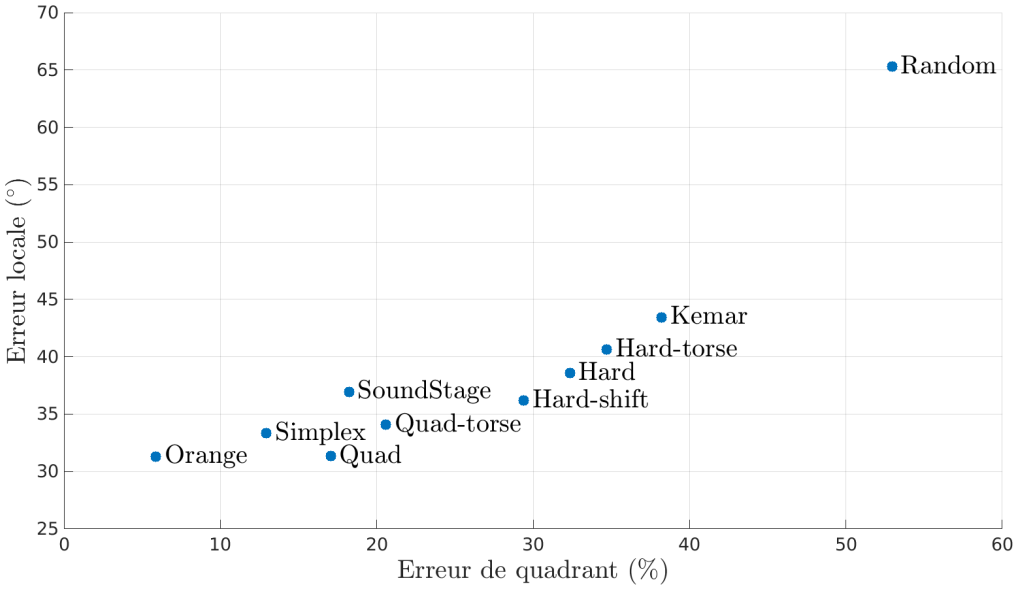


FIGURE 4.41: *Positionnement 2D des résultats subjectifs.*

Premier constat, le tirage aléatoire est très isolé des autres points. L'utilisation d'une HRTF a donc un effet certain, qu'elle soit individualisée ou non. Ensuite, la HRTF Kemar – qui représente l'absence d'individualisation – offre les plus mauvais résultats alors que la HRTF Orange – individualisée acoustiquement – offre les meilleurs.

Entre les deux se retrouvent les autres HRTF, que l'on peut scinder en deux groupes avec en premier lieu celui des HRTF *hard*. Assez proches les unes des autres, elles montrent un fort taux d'erreurs de quadrant et une erreur moyenne locale conséquente. De plus, leurs résultats sont assez proches de ceux de Kemar. On peut donc parler d'une *mauvaise* individualisation, tant leurs différences par rapport à une HRTF générique sont ténues.

De l'autre côté se trouvent les HRTF optimisées par recherche d'impédance et la HRTF SoundStage. La présence de cette dernière au sein du groupe est en soi un motif de satisfaction : nous retrouvons un niveau de performances au moins comparable à celui d'une HRTF acoustique. Ensuite, nous observons que les HRTF Simplex et Quad forment toujours le même duo, ce qui est également rassurant car conforme aux observations précédentes. Toutefois, à l'inverse de celles-ci, ce duo offre de meilleures performances que la HRTF Quad torse alors que cette dernière apparaissait jusque-là plus aboutie. La HRTF Quad, bien que moins proche d'Orange en moyenne selon les mesures d'ISSD, voit ses écarts à la référence mieux répartis sur l'ensemble du spectre. Cependant, la HRTF Simplex ne semble pas présenter le même profil, affaiblissant l'hypothèse émise. Une autre explication consiste alors à rappeler que nous avons jusqu'à présent toujours considéré la HRTF Orange comme une vérité terrain indiscutable alors qu'elle comporte elle aussi d'inévitables défauts et imprécisions. On peut donc se retrouver dans une situation où, bien que la HRTF Quad torse soit plus proche d'Orange que ne le sont les HRTF Simplex et Quad, ces dernières – et Orange – se trouvent être plus proches de la véritable HRTF du sujet 006 que ne l'est la HRTF Quad torse.

Quoi qu'il en soit une constante demeure : les HRTF à impédance optimisée ont un meilleur rendu subjectif que les HRTF en condition *hard*. Néanmoins, les expériences précédentes ne valent que pour un seul sujet et la généralisation n'a pour l'heure pas été couronnée de succès. Plus précisément, pour les quatre autres sujets dont les HRTF Orange sont disponibles, les HRTF Quad, Quad torse, Simplex, hard, hard torse et hard shift ont été calculées. Et lorsque cela a été nécessaire l'impédance optimisée issue du sujet 006 a été utilisée. Plusieurs séries de tests subjectifs de localisation ont alors été passées mais aucun résultat statistiquement probant n'a pu en être extrait. Quel que soit le paramétrage choisi les performances demeurent mauvaises et très proches de celles obtenues avec la HRTF Kemar, c'est-à-dire sans individualisation. Il est à ce moment tentant d'incriminer le procédé d'optimisation, trop spécifique à un sujet. Cependant, il est aussi important de noter que les HRTF acoustiques des autres sujets n'ont pas, elles non plus, donné de meilleurs résultats. Plusieurs hypothèses peuvent être formulées pour l'expliquer :

1. **Nous sommes en présence de mauvais localisateurs.**

   Si cette première hypothèse est exacte alors aucune amélioration n'est envisageable. Quoi que l'on fasse, les résultats seront limités par les facultés naturelles des sujets et aucune HRTF n'y pourra rien. Ceci étant, les tests de localisation en conditions réelles menés dans le SoundStage pour ces sujets ne montrent pas de problème particulier. Le SoundStage n'étant malheureusement pas anéchoïque et les HP n'étant pas exempts de coloration, il n'est pas possible de dire si les indices de localisations perçus sont uniquement issus des HRTF ou si d'autres éléments, de nature différente, ont été mis

à contribution lors de ces tests.

2. **Les HRTF acoustiques utilisées sont déficientes.**

   Cette deuxième hypothèse peut se traduire de deux façons. Soit les HRTF – ou plutôt les DTF – ne sont pas suffisantes pour spatialiser convenablement un son, soit un problème technique est venu perturber nos HRTF acoustiques lors de l'acquisition. Malheureusement, il est très difficile de trancher car, dans le premier cas, le fait que la HRTF du sujet 006 apporte les résultats escomptés fait figure de contre-exemple et, dans le second cas, il faut expliquer pourquoi la HRTF du sujet 006 n'a pas subi le même sort que ses consoeurs alors que toutes ont été acquises le même jour, dans les mêmes conditions et par les mêmes personnes.

3. **Le test de localisation ne convient pas à tout le monde.**

   Pour tester la troisième hypothèse, d'autres éléments tels que la réverbération ou le *head-tracking* ont été ajoutés mais sans grand succès.

En définitive, ces recherches nées de l'observation des différences entre HRTF acoustiques et simulées nous ont permis de proposer un coupable potentiel : l'impédance des matériaux. Nous avons prouvé que, correctement paramétrée, elle permettait la production de HRTF calculées individualisées de qualité équivalente à celle des HRTF acoustiques. Nous ne sommes cependant pas parvenu à passer l'étape de la généralisation à tout individu. Plusieurs hypothèses, nécessitant de plus amples recherches, ont néanmoins été émises pour surmonter cet obstacle. En chemin nous avons exhibé l'origine potentielle d'interférences indésirables visibles en simulation, à savoir l'absorbance acoustique réelle des épaules et du torse. De nombreuses questions restent là encore en suspens et appellent à de nouvelles recherches. Enfin, nos expériences ont mis en exergue la difficulté à obtenir des HRTF personnalisées au rendu subjectif satisfaisant, qu'elles soient issues du calcul ou de la mesure.

Chapitre 5

D'un monde à l'autre : liens entre morphologie et HRTF

L'esprit humain fut fait pour comprendre, comme l'oeil fut fait pour voir les couleurs et l'oreille pour entendre les sons.

- Johannes Kepler -

Ce dernier chapitre du manuscrit est dédié à la génération de bases de données, à leur analyse par ACP et au couplage des espaces mathématiques ainsi créés.

Nous commenceront par la création de trois bases de données de maillages et de HRTF aux caractéristiques propres. Les étapes de leur mise en place seront détaillées et seront suivies d'une analyse critique du résultat.

Nous aborderons ensuite les phases de décomposition par ACP et de couplage de ces bases. Plusieurs méthodes de couplage seront envisagées, dont deux seront évaluées ensuite.

Dans une dernière partie, une validation subjective des deux couplages sera menée à l'aide d'un simulateur de test de localisation dans le plan médian et permettra de quantifier les performances de chaque méthode. Ces analyses nous amènerons notamment à nous confronter aux méthodes de personnalisation par sélection de HRTF. Elle se terminera par une discussion des résultats obtenus, des forces et des faiblesses des différentes approches ainsi que de l'impact de la taille de la base.

5.1 Bases de données

Les travaux préalables de modélisation morphologique et de simulation numérique ayant été exposés, il est maintenant possible de s'attacher à la constitution des bases de données de maillages 3D et de HRTF correspondantes. Dans les faits, trois bases ont été créées : l'une dite *synthétique*, l'autre *aléatoire* et la dernière *mixte*. Pour ce qui est des maillages, ceux-ci sont tantôt issus du modèle synthétique, tantôt du modèle mixte. Concernant les HRTF, celles-ci ont été calculées soit en condition totalement rigide – aussi dite *hard* –, soit en incorporant l'absorbance du torse.

Le choix d'une condition *hard* se justifie car il s'agit là de l'état de l'art et qu'il facilite les comparaisons avec les travaux d'autres équipes de recherche. Le choix du torse absorbant nous semble intéressant pour l'amélioration, au moins visuelle, qu'il apporte aux HRTF. Mais qu'il s'agisse de l'un ou l'autre de ces choix, cela ne doit pas modifier fondamentalement les enseignements à tirer des expériences ci-après. La constitution de ces trois bases devra entre autres choses permettre d'en discuter. L'impédance de peau trouvée par optimisation n'ayant en revanche pas passé l'étape de la généralisation à d'autres individus, elle n'a pas été retenue pour ces expérimentations à grande échelle.

Nous proposons de les présenter en suivant systématiquement le même schéma *constitution / analyse*. La première partie aura pour objectif d'exposer les caractéristiques techniques souhaitées à l'origine tandis que la seconde effectuera un retour critique sur les données effectivement obtenues.

5.1.1 Base synthétique

Ainsi qu'on a pu le voir à la section 3.2.1, le modèle synthétique constitue un puissant outil de génération de morphologies humaines. Son réalisme, son déterminisme et l'automatisation des mesures rendent en effet possible la création d'une base de maillages répondant à un cahier des charges précis en termes de paramètres anthropométriques. Et alors que le modèle statistique d'oreille – cf. section 3.1.2 – tire sa force de son ancrage

dans la population réelle, au prix d'un certain effort d'acquisition des données et d'une absence de contrôle sur les déformations qui le composent, le modèle synthétique perd en variabilité intrinsèque ce qu'il gagne en contrôle morphologique.

5.1.1.1 Constitution

Maillages Le décor ainsi planté, intéressons-nous plus avant au jeu de maillages à produire. Comme indiqué, l'idée maîtresse est de pouvoir comparer l'effet sur les HRTF de combinaisons de déformations. Pour ce faire, le nombre de possibilités explosant rapidement, un choix restreint de paramètres - les plus influents [65] - et de valeurs est impératif. Sept paramètres d'oreilles ont ainsi été sélectionnés. Nommément, la largeur de conque d_3, la hauteur de la fosse d_4, la hauteur de conque $d_1 + d_2$, l'angle de rotation de l'oreille selon l'axe interaural θ_1, l'angle de rotation selon l'axe vertical θ_2 et les paramètres R et P. Pour ce qui est des plages de valeurs explorées, les statistiques de CIPIC ont été utilisées – lorsque cela avait un sens – en prenant en compte valeurs moyennes et écarts-types. Les paramètres R et P étant plus « exotiques », de telles statistiques sont inexistantes et les limites jugées acceptables l'ont été sur l'expérience personnelle des membres de l'équipe du laboratoire. Une fois les plages de valeurs déterminées, ne manquent plus que leurs nombres. La largeur de conque étant notoirement déterminante, quatre valeurs lui ont été attribuées. Pour d_4, $d_1 + d_2$, θ_1 et θ_2, trois valeurs ont été choisies. Pour R et P, deux seulement, nous amenant au total de 1 296 formes distinctes d'oreilles. La table 5.1 en fait le récapitulatif.

ID technique	Nb valeurs	Paramètre
d_3	4	cavum concha width
$d_1 + d_2$	3	cavum concha + cymba height
d_4	3	fossa height
θ_1	3	pinna rotation angle around axis Y
θ_2	3	pinna rotation angle around axis Z
P	2	parabole
R	2	crus of helix

TABLE 5.1: *Correspondances entre identifiants techniques de la base synthétique et paramètres morphologiques. S'il existe, l'équivalent CIPIC est indiqué.*

Dans chacun des cas, ces valeurs sont équiréparties dans la plage jugée acceptable. Les autres caractéristiques du maillages – distance interaurale, largeur d'épaules, emplacement de l'oreille sur la tête, etc. – ont été choisies de façon à s'approcher au mieux des valeurs moyennes des statistiques de CIPIC. Sont présentés figure 5.1 des exemples d'oreilles résultantes.

Simulations Les HRTF de cette base ont été calculées par pas de 100 Hz sur la plage [100, 16 000] Hz. Pour chaque maillage, seule la HRTF gauche a été générée. Une zone de 3 mm^2 située au fond du canal auditif a servi de microphone. Les techniques de maillage adaptatif et de dépendance en fréquence du maillage – cf. section 4.2 – ont été utilisées lors des simulations. La grille d'évaluation, de rayon 2 m, est icosaédrique et comporte

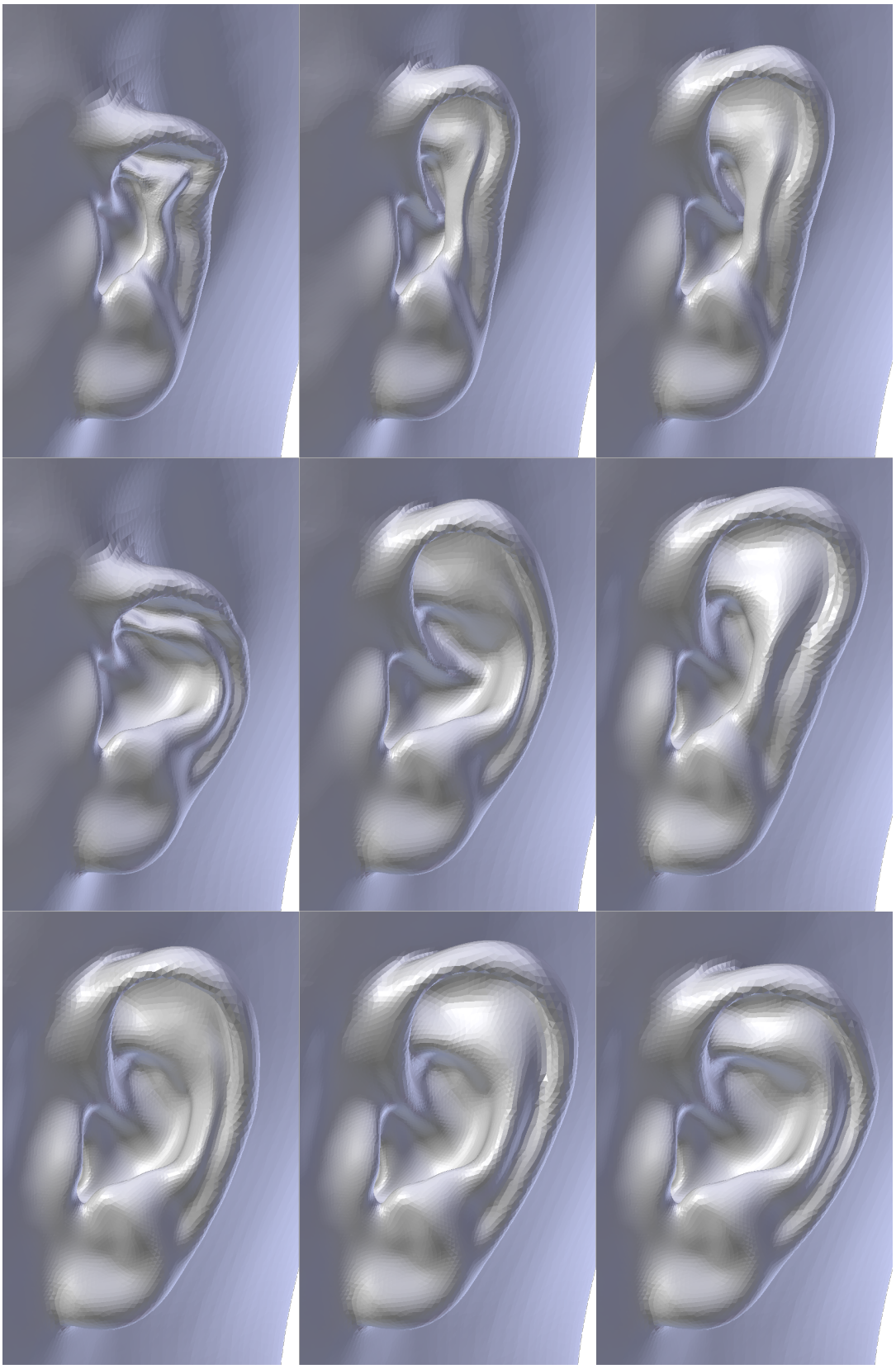


FIGURE 5.1: *Exemples d'oreilles de la base synthétique de maillages 3D.*

2 562 points. Enfin, toutes les simulations de la base synthétique utilisent une condition d'impédance *hard*.

Ceci clôt le détail des lignes directrices de cette première base. Dans les faits, sa réalisation a dû affronter les inévitables aléas du monde réel et il est nécessaire d'en tempérer quelque peu les résultats.

5.1.1.2 Analyse

Tout d'abord, si l'on a bien sélectionné 1 296 jeux de paramètres de forme, 42 d'entre eux ont abouti à des maillages inutilisables. Des intersections de faces étaient en effet présentes pour ceux-là, rendant impossible tout calcul de HRTF. C'est donc une base à 1 254 entrées que nous manipulons dans les faits.

Ensuite, le mode de production des maillages en lui-même a été source d'écarts plus ou moins grands entre les valeurs de consigne des paramètres et les valeurs réelles de sortie. La figure 5.2 présente les consignes des maillages valides. Comme on peut le voir, quatre valeurs ont été assignées au paramètre $d3$, trois à $d1 + d2$ et ainsi de suite.

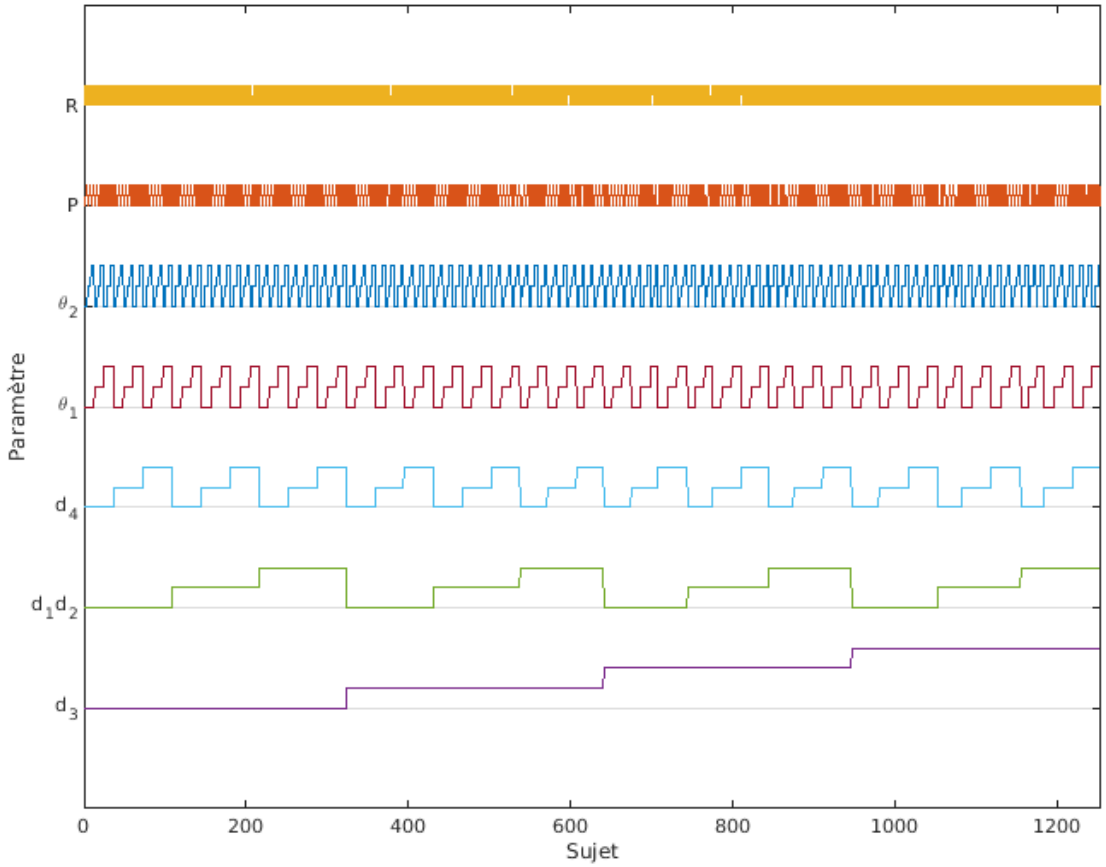


FIGURE 5.2: *Changements des valeurs des paramètres de forme.*

Ces consignes sont utilisées par le générateur de maillages, qui les lit de manière séquentielle et va chercher à optimiser le résultat paramètre après paramètre. Ci-après – cf. figure 5.3 –, les résultats des mesures pour chacun d'entre eux.

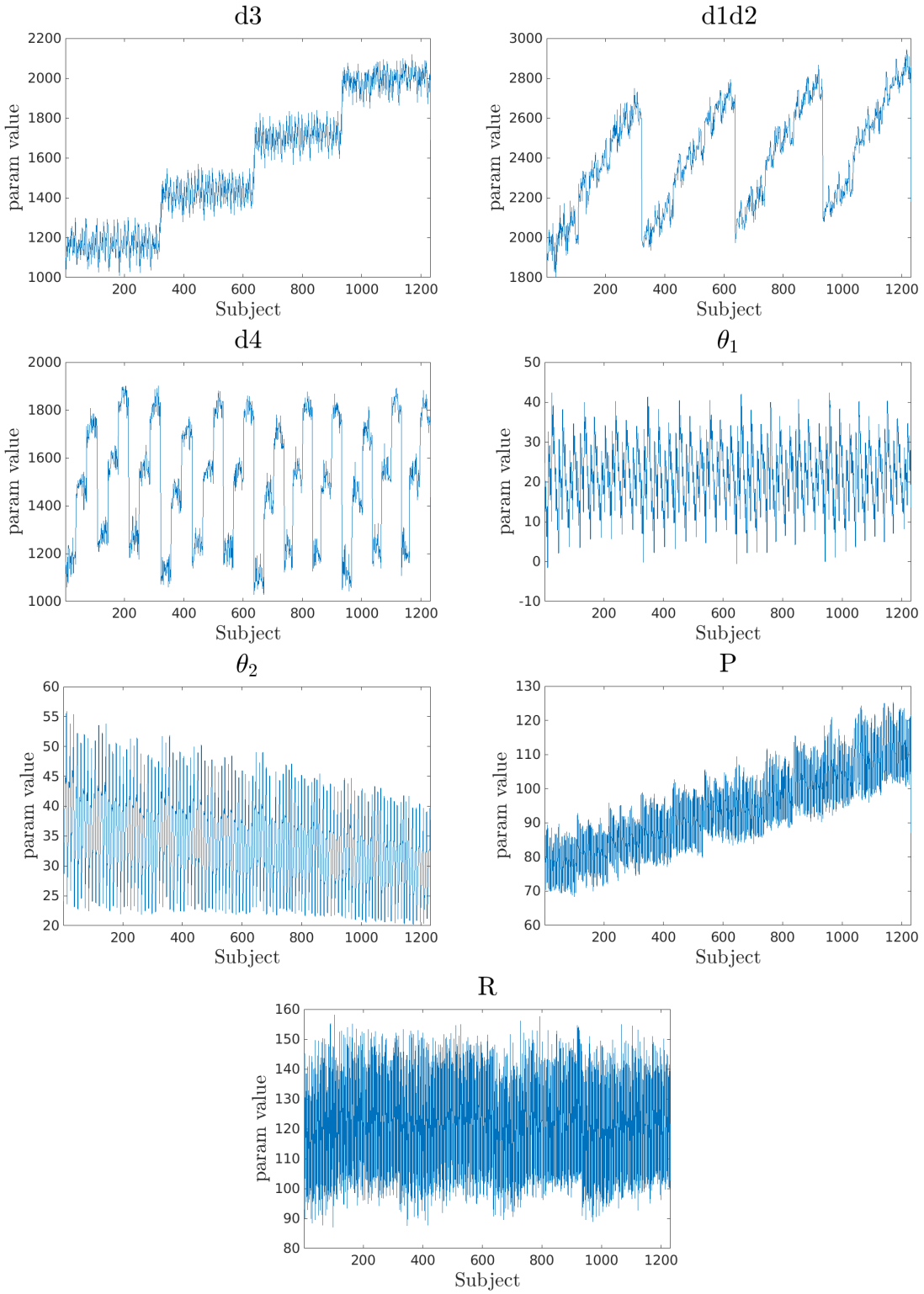
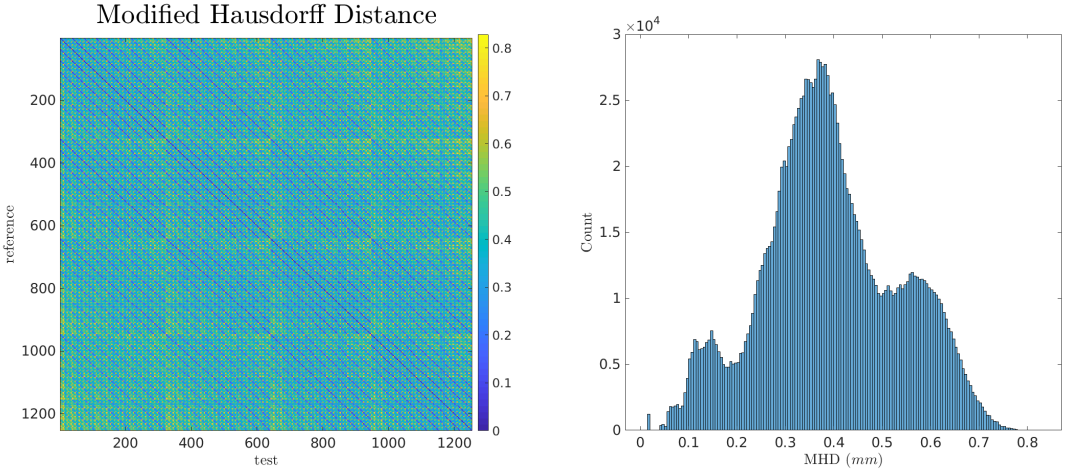


FIGURE 5.3: *Valeurs effectives des paramètres de forme, à comparer aux valeurs de consignes de la figure 5.2.*

De manière certaine, quelques grains de sable sont parfois venus gripper les rouages de notre générateur. En particulier, les paramètres $d1d2$ et P sont sujets à caution. Ils présentent manifestement des dérives et / ou des corrélations avec d'autres paramètres. Ceci peut s'expliquer par le caractère séquentiel du générateur. En effet, à l'opposée d'une optimisation globale dans laquelle toutes les distances et tous les angles de consigne seraient obtenus en même temps, l'optimisation séquentielle va les traiter les uns à la suite des autres. Or les *blendshapes* que nous manipulons pour créer nos maillages ne sont pas indépendantes les unes des autres. Par conséquent il peut arriver que l'optimisation du n-ième paramètre altère celle de l'un – ou plus – de ses prédécesseurs. Cela explique également le caractère légèrement bruité des mesures. Il s'agit là d'une contrainte d'ordre technique qu'il n'a pas été possible de surmonter, sauf à modifier radicalement le modèle 3D en lui-même.

Toutefois, ces problèmes ont davantage remis en cause la répartition des valeurs que leurs plages d'études. Et bien qu'il soit nécessaire d'en tenir compte pour toute analyse subsidiaire, cela n'est pas de nature à invalider la base en elle-même.

Pour terminer ce tour d'horizon de la partie morphologique, nous avons mesuré la distance de Hausdorff modifiée (MHD) séparant chaque couple de maillages. La MHD reposant sur des calculs de distance point à surface, elle est plus adaptée à la mesure de différences de formes que ne le sont les distances point à point. Le prix à payer est son coût calculatoire, qui rend notamment impraticable son utilisation dans beaucoup d'applications temps-réel. La figure 5.4 montre la matrice, symétrique, des MHD de la base synthétique. Comme on peut le constater, celle-ci a un fort aspect en damier, conséquence directe de la façon dont a été pensée la base, couplé à la présence de nombreuses diagonales secondaires à faibles valeurs. L'histogramme des MHD présente une répartition complexe de moyenne $0,39$ mm et d'écart-type $0,14$ mm.



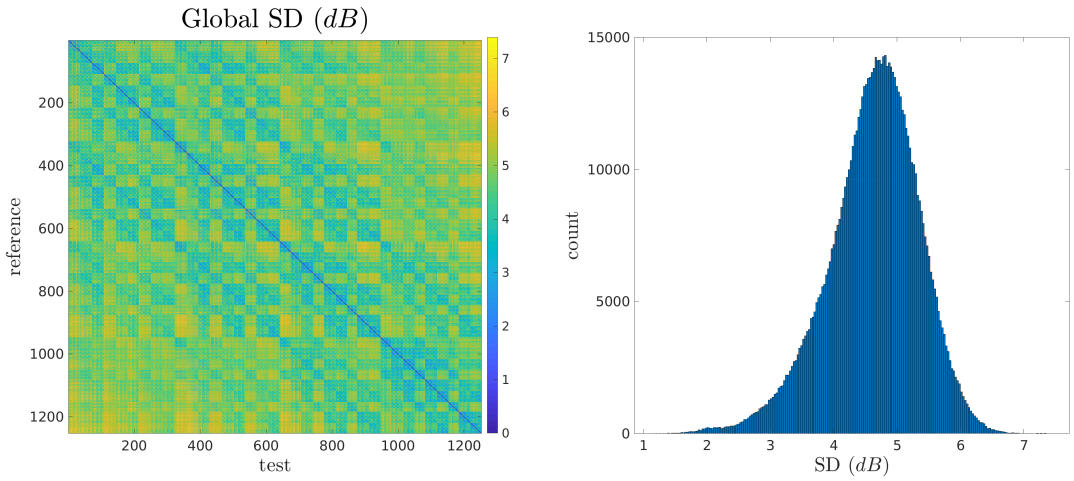
(a) (b)

FIGURE 5.4: *En (a), la matrice des MHD de la base synthétique. En (b), l'histogramme des valeurs prises par les couples de simulations distinctes.*

Ces caractéristiques sont le résultat des changements de valeurs des paramètres de formes. Les 16 plus gros carrés, par exemple, traduisent les 16 combinaisons possibles liées

à d_3, des valeurs les plus faibles en haut à gauche aux valeurs les plus fortes en bas à droite. Leurs diagonales correspondent aux configurations de paramètres où seul d_3 est modifié.

Reste désormais à évaluer la qualité des HRTF. À cette fin, la matrice des distorsions spectrales de tous les couples de HRTF a été compilée (cf. figure 5.5). Par nature celle-ci est symétrique, réelle, positive et à diagonale nulle. Elle donne également une vue d'ensemble de la base, permettant notamment de détecter les échecs de convergence qui apparaissent alors comme autant de lignes à très fortes valeurs.



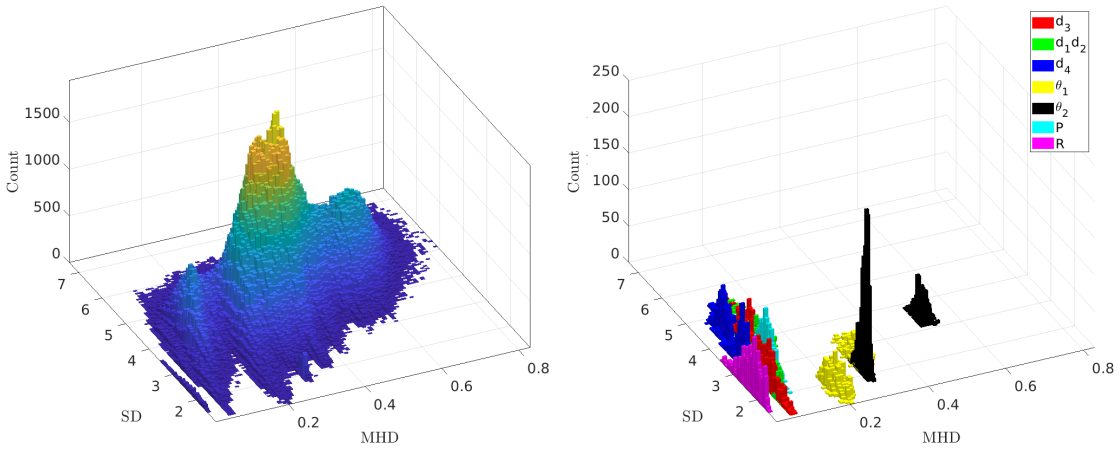
(a) (b)

FIGURE 5.5: *En (a), la matrice des distorsions spectrales de la base synthétique. En (b), l'histogramme des valeurs prises par les couples de simulations distinctes.*

Le tracé de l'histogramme nous informe quant à lui sur la répartition du nuage de points formé par les HRTF. Mis à part les 1 254 coefficients nuls de la diagonale (non pris en compte), les autres présentent une répartition gaussienne de moyenne $4,63$ dB et d'écart-type $0,72$ dB. À titre de comparaison, la moyenne et l'écart-type des distorsions spectrales des HRTF acoustiques, droite et gauche, des 10 sujets de la base Symare valent respectivement $5,28$ dB et $0,56$ dB. Les ordres de grandeur sont donc respectés et la plus faible variabilité morphologique de nos maillages synthétiques vis-à-vis de sujets réels se retrouve bien dans les moyennes des distorsions spectrales correspondantes.

À ce stade, nous avons donc observé le contenu de la base du point de vue morphologique puis du point de vue auditif. Dans les deux cas, la quantification sous-jacente des paramètres de création des maillages apparaît de manière notable. La régularité et l'aspect quadrillé des matrices de distorsion spectrale et de MHD en sont des signes indéniables. Toutefois, il faut également convenir que les motifs qu'elles présentent sont très distincts et cette différence s'oppose à l'idée qu'une proximité morphologique induise nécessairement une proximité de HRTF. Néanmoins, les modifications morphologiques sont la seule réelle variable de notre base et sont donc tenues d'expliquer toutes les observations. Dans le cas présent, nous pouvons avancer que les paramètres sélectionnés n'ont pas la même influence

sur la base selon que l'on considère son côté morphologique ou son côté auditif.



(a) (b)

FIGURE 5.6: *En (a), l'histogramme 2D des couples (SD, MHD) de la base synthétique. En (b), ce même histogramme restreint aux couples de maillages ne variant que par la valeur d'un paramètre de consigne. À chaque paramètre est associée une couleur propre.*

Pour mieux le visualiser, nous avons construit l'histogramme 2D donnant la répartition des couples d'erreurs (SD, MHD) – cf. figure 5.6. Celui-ci se révèle riche d'enseignements. Tout d'abord, l'absence de corrélation claire entre la distorsion spectrale et la distance géométrique. Cette dernière se révèle donc être un piètre prédicteur pour qui souhaite faire une sélection de HRTF fondée sur un seul critère, même s'il demeure une légère tendance faisant croître la distorsion spectrale avec la MHD. Ensuite, nous voyons clairement apparaître deux groupes. L'un, doté de plusieurs pics, représentant l'essentiel du contingent. L'autre, bien plus discret mais caractérisé par des distances de Hausdorff très faibles. Afin de mieux appréhender l'origine de cette répartition, l'histogramme a été restreint aux couples de maillages ne variant que par la valeur d'un paramètre de consigne. On voit apparaître pour chacun d'eux une ou plusieurs zones de prédominance assez franches. En particulier, on constate que le second groupe est entièrement constitué des couples de maillages dont seule diffère la valeur de R[1]. Et si ce paramètre était à l'origine classé parmi ceux de faible influence, eu égard à la répartition des distorsions spectrales associées, il faut sans doute désormais réévaluer ce jugement à l'aune du rapport entre les variations des distorsions spectrales et les variations de MHD. Il permet en effet de perturber notablement une HRTF au prix d'une déformation modique du maillage. Avec ce nouveau paradigme, les angles θ_1 et θ_2 perdent alors encore en importance car l'obtention de distorsions spectrales similaires se fait au prix de déformations bien plus grandes.

5.1.2 Base aléatoire

L'approche adoptée pour la base synthétique, qui repose sur le contrôle maximal de la morphologie, vient avec avantages et inconvénients. L'avantage majeur, nous l'avons vu, est

1. qui, rappelons-le, mesure une déformation spécifique de la conque.

de pouvoir finement étudier les effets de tel ou tel paramètre. La contrepartie est que cette base devient rapidement démesurée et qu'il nous faut donc poser des limites strictes au nombre de configurations possibles, c'est-à-dire au nombre de paramètres modifiés. L'écueil à craindre est alors de devenir sans le savoir trop restrictif et de perdre de vue des aspects plus généraux des HRTF. Nous avons pour cette raison créé une deuxième base de données, incorporant une part d'aléatoire dans les déformations morphologiques.

5.1.2.1 Constitution

Maillages Partant à nouveau du modèle synthétique, 500 maillages distincts ont été générés. Cette fois-ci, 22 *blendshapes* ont été mises à contribution et leurs valeurs tirées aléatoirement selon des lois gaussiennes. Il s'est donc tout aussi bien agi de celles liées à l'oreille que de celles liées à la tête ou au torse. La seule contrainte posée aux maillages résultants a été de présenter des dimensions – largeur d'épaules, hauteur de tête, décollement d'oreille, etc. – comprises dans des limites acceptables et de ne pas être rejetés par l'outil de simulation. La figure 5.7 en présente quelques exemples.

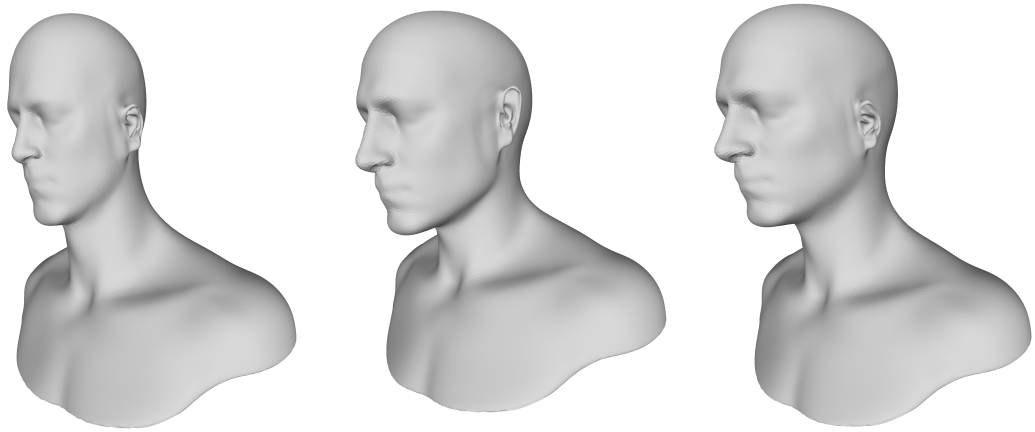


FIGURE 5.7: *Exemples de maillages de la base aléatoire.*

Simulations Les caractéristiques de simulation de la base aléatoire sont les mêmes que celles de la base synthétique, exception faite des conditions d'impédance. Nous profitons en effet de cette base pour observer à plus grande échelle l'action du torse absorbant sur les HRTF.

5.1.2.2 Analyse

Comme lors de la création la base synthétique, certains problèmes sont apparus au moment de donner corps à la base aléatoire. En effet, en raison des imprévisibles mais inévitables intersections entre *blendshapes*, un certain nombre de maillages durent être rejetés. En outre, du côté des calculs de HRTF, il est à déplorer la présence de plusieurs échecs de convergence. Au total, 43 d'entre eux n'ont pas entièrement abouti, réduisant le

nombre d'entrées de la base à 457. L'autre conséquence est que la répartition gaussienne des valeurs des paramètres n'est plus assurée.

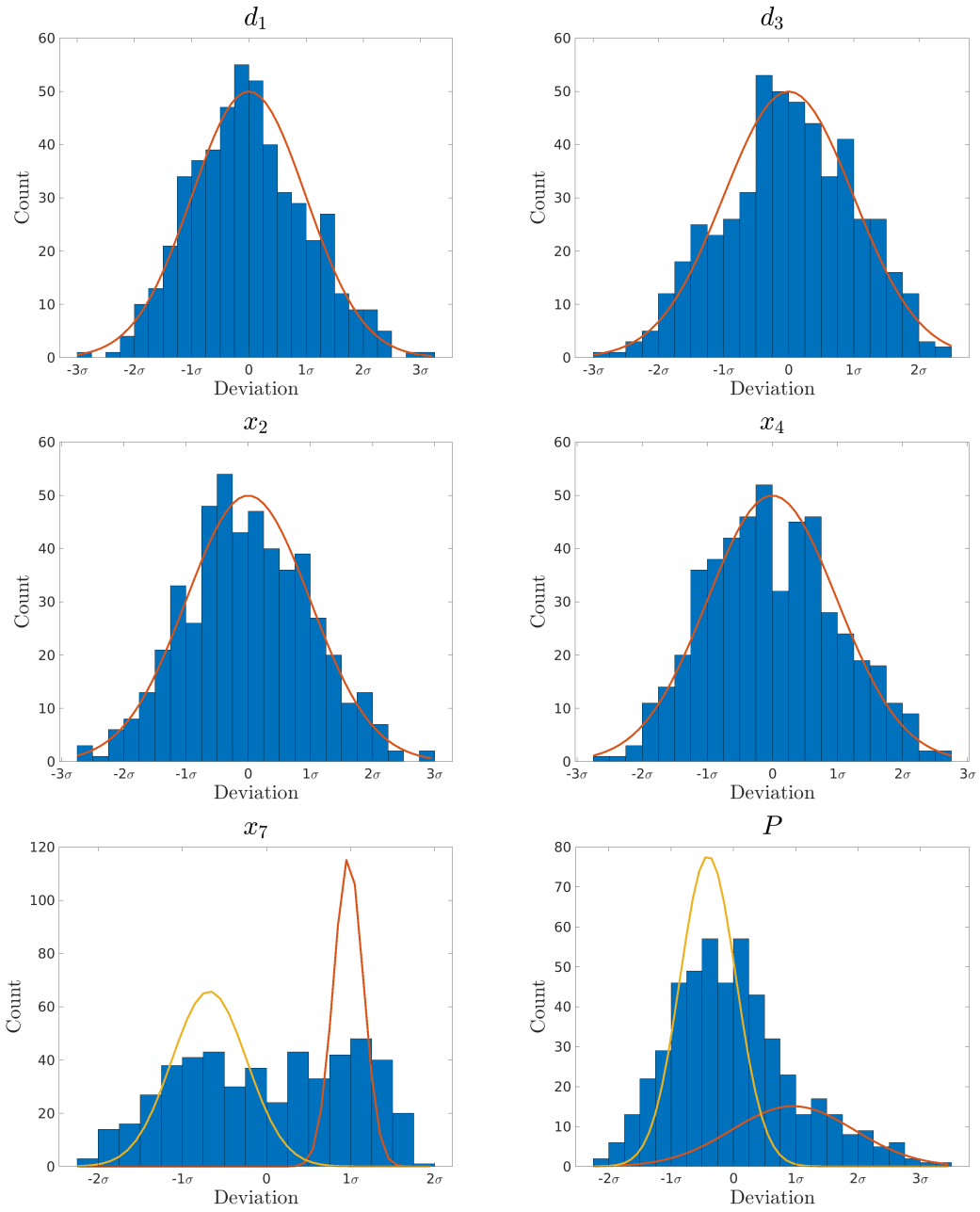


FIGURE 5.8: *Répartitions de quelques paramètres de la base aléatoire. Pour les paramètres d_1, d_3, x_2 et x_4, qui passent le test de normalité, la gaussienne correspondante est également affichée. Pour les autres, une décomposition en mixture de gaussiennes à deux composantes est surimposée. En guise d'exemples, les paramètres x_7 et P.*

Après passage du test de Kolmogorov-Smirnov avec une tolérance de 5 %, il s'avère que 9 d'entre eux ne suivent pas une loi normale, comme l'illustre la figure 5.8. Deux options s'offrent alors à nous : régulariser à posteriori les données ou tenir compte de leur biais dans la phase d'analyse. La première d'entre elles implique d'exclure les jeux de données surnuméraires jusqu'à obtenir une répartition satisfaisante. Un moyen d'y parvenir consiste

par exemple à utiliser des mixtures de gaussiennes à plusieurs paramètres pour décrire nos statistiques. Malheureusement, cette technique implique la suppression d'un trop grand nombre d'entrées de la base. Choix a par conséquent été fait de garder tous les maillages et de tenir compte de leurs biais après coup.

En parallèle, il est également nécessaire de vérifier l'indépendance des paramètres. Pour cela, la matrice de leurs corrélations a été calculée. Les coefficients de valeur absolue inférieure à $0,6$ ou de p-value supérieure à $5\,\%$ ont été écartés car non significatifs. Le résultat, visible figure 5.9, fait apparaître quelques liens résiduels notables.

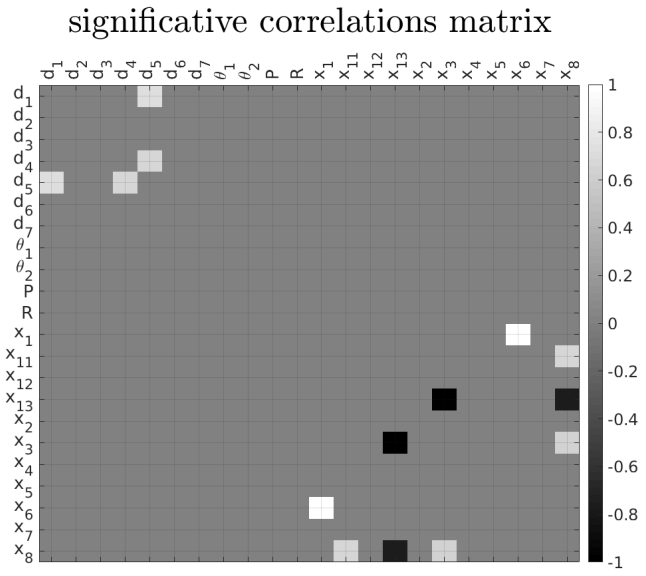
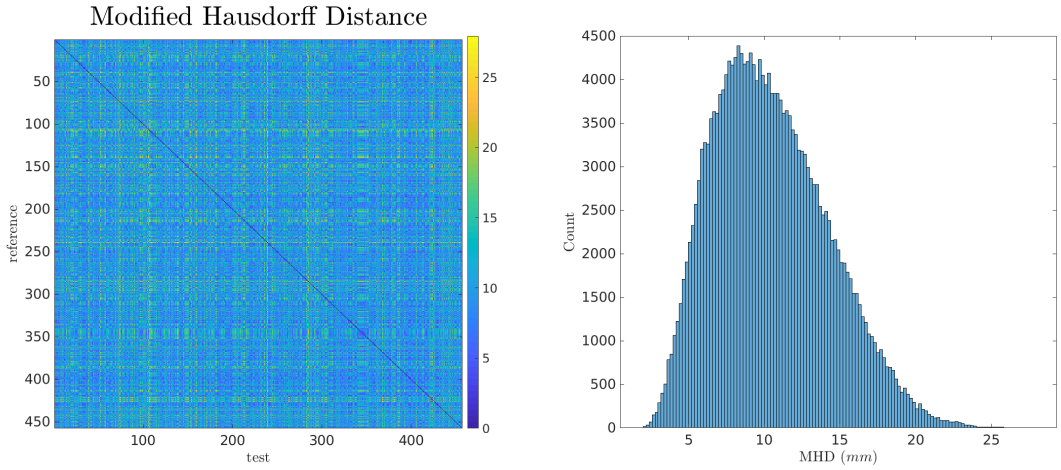


FIGURE 5.9: *Corrélations significatives entre paramètres de construction de la base aléatoire.*

On observe en effet que le paramètre d_5, qui représente la hauteur de l'oreille, est positivement corrélé à d_1 et d_4. Cela n'est pas de nature à étonner car ces derniers mesurent des sous-parties de la hauteur de l'oreille.

À l'autre extrémité de la matrice, on trouve les corrélations entre paramètres de tête et de torse. Elles montrent tout d'abord un lien entre x_1, la largeur de tête, et x_6, la largeur de cou. Après vérification, cela est dû au modèle lui-même car la *blendshape* déformant la largeur de tête déforme aussi le cou. Des raisons analogues expliquent les corrélation positives entre x_8, la profondeur du cou, et x_3, la profondeur de tête ainsi qu'entre x_8 et x_{11}, la profondeur du buste. Enfin, le paramètre x_{13}, qui représente le décalage vers l'avant de la tête par rapport au torse trouve lui aussi l'origine de ses corrélations – avec x_3 et x_8 – dans le modèle. En effet, il n'existe pas de *blendshape* pour décaler la tête par rapport au torse. Par conséquent, la valeur de x_{13} devrait rester constante. Les variations de la profondeur de la tête modifient cependant aussi légèrement qu'inévitablement la position de son centre, ce qui se répercute ensuite sur les valeurs de x_{13}. En dépouillant celles-ci on s'aperçoit d'ailleurs qu'elles varient seulement entre 2,4 mm et 5,4 mm, ce qui renforce d'autant cette explication.

En définitive, aucune corrélation anormale n'est donc présente dans les morphologies composant la base aléatoire, qui est sur ce point conforme aux attentes.



(a) (b)

FIGURE 5.10: *En (a), la matrice des MHD de la base aléatoire. En (b), l'histogramme des valeurs associées aux couples de simulations distinctes.*

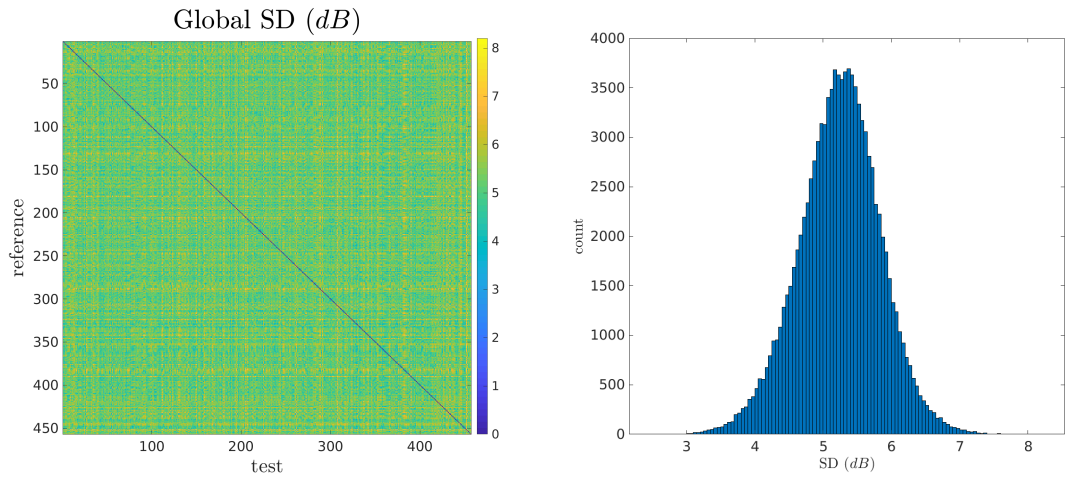
Concernant les simulations, la matrice des distorsions spectrales de tous les couples de HRTF a là aussi été compilée (cf. figure 5.11). Ici, plus de structure en damier ni de sous-diagonales apparentes. On observe en revanche la présence assez visible de lignes horizontales et verticales. Les valeurs des distorsions spectrales qu'elles contiennent ne sont pas celles de simulations en échec et ne doivent pas inquiéter. Il s'agit plus simplement des distorsions spectrales des HRTF les plus atypiques.

L'histogramme des valeurs présente de nouveau une allure gaussienne. Sa moyenne et son écart-type valent respectivement $5,24\,\text{dB}$ et $0,60\,\text{dB}$. Les variations des paramètres de tête et de torse a donc permis d'obtenir les quelques portions de dB qui nous séparaient de Symare.

Comme précédemment, nous avons aussi tracé l'histogramme 2D des couples (SD, MHD) (cf. figure 5.12). Cette fois, nous sommes en présence d'un seul groupe, réuni autour d'un seul pic, sans corrélation aucune. Il est également à noter que les distances de Hausdorff mesurées sont bien plus grandes que pour la base synthétique. Cela découle naturellement du fait que l'on a ici autorisé la tête et le torse à se déformer. En outre, on peut aussi observer que la norme SD est quant à elle restée dans la même gamme de valeurs.

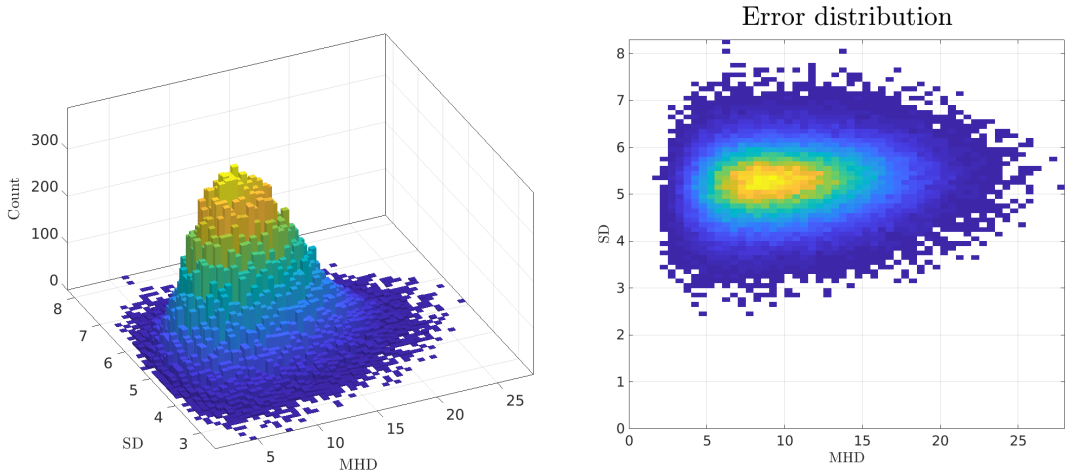
5.1.3 Base mixte

Les deux bases précédentes ont pour avantage de posséder un grand nombre d'entrées. La base aléatoire en compte 457 tandis que la synthétique en comprend 1 254 . Elles ont également pour elles de reposer sur un modèle paramétrique facilement modifiable. Le revers de la médaille est que malgré les efforts consentis pour proposer la plus grande variabilité de formes possible, seule une comparaison directe à des morphologies issues du



(a) (b)

FIGURE 5.11: *En (a), la matrice des distorsions spectrales de la base aléatoire. En (b), l'histogramme des valeurs associées aux couples de simulations distinctes.*



(a) (b)

FIGURE 5.12: *En (a), l'histogramme 2D des couples (SD, MHD) de la base aléatoire. En (b), son projeté orthogonal selon Z.*

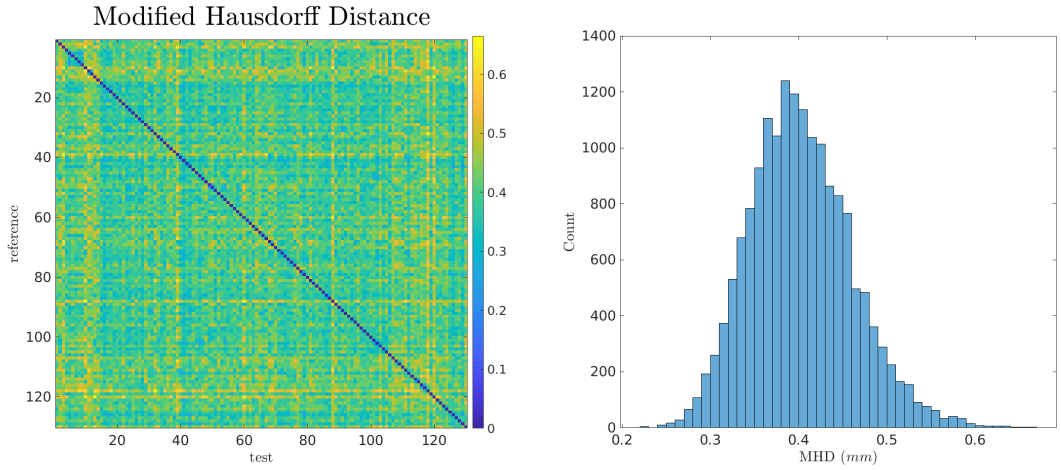
monde réel nous permet de statuer sur leurs représentativités véritables. À cette fin, nous avons donc également constitué une troisième et dernière base, dite *mixte* car issue du modèle mixte vu section 3.2.1. Elle sert de pont entre la profusion numérique des données générées jusque là et la richesse morphologique du monde réel.

5.1.3.1 Constitution

FIGURE 5.13: *Exemples de maillages de la base mixte.*

Maillages Sans surprise, les maillages utilisés pour la base mixte sont les 130 ayant permis la création du modèle mixte. Sa variabilité morphologique est donc toute entière concentrée sur le pavillon d'oreille. Par suite, les seuls paramètres pertinents sont ceux mesurant l'oreille. De plus, aucune autre contrainte n'a ici lieu d'être car les différentes formes sont issues des scans. Une précision doit toutefois être apportée car ces scans subissent une normalisation en taille et en orientation avant d'être incorporés au modèle déformable. Quelques exemples en sont présentés figure 5.13. On s'attend donc à observer moins de variations de certains paramètres, tels d_5 ou θ_1.

Simulations Les caractéristiques de simulation de la base mixte reprennent trait pour trait celles de la base synthétique.



(a) (b)

FIGURE 5.14: *En (a), la matrice des MHD de la base mixte. En (b), l'histogramme des valeurs associées aux couples de simulations distinctes.*

5.1.3.2 Analyse

Suivant le même formalisme que précédemment, il nous faut qualifier les répartitions des valeurs des paramètres morphologiques. Cette fois-ci, aucune corrélation n'apparaît, pas même celles entrevues dans la base aléatoire et liant d_5 à d_4 et d_1. Cependant, puisque les corrélations des autres bases sont liées aux *blendshapes* du modèle 3D utilisé et que celles-ci ont été remplacées par le modèle d'oreilles réelles, il n'est rien d'étonnant à ce que nous ne les retrouvions pas ici. C'est là aussi un effet lié à la taille plus réduite de la base, qui augmente les valeurs générales des p-values.

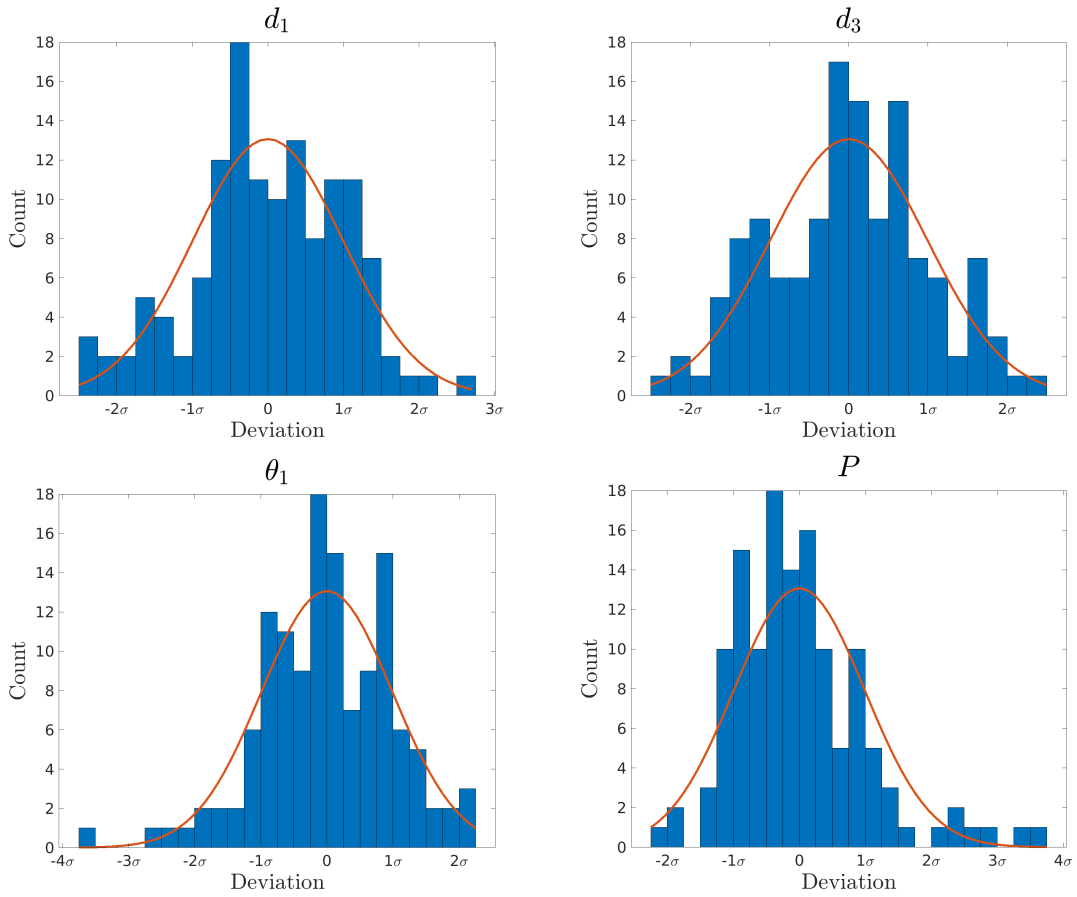
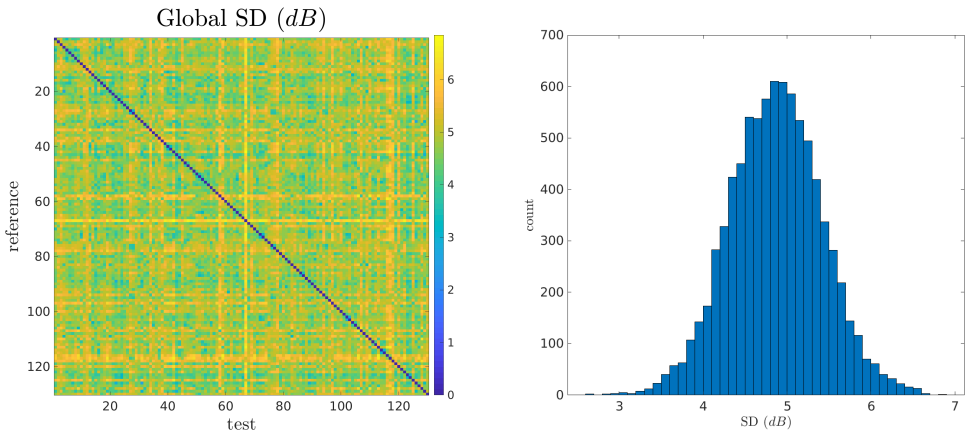


FIGURE 5.15: *Répartitions de quelques paramètres de la base mixte.*

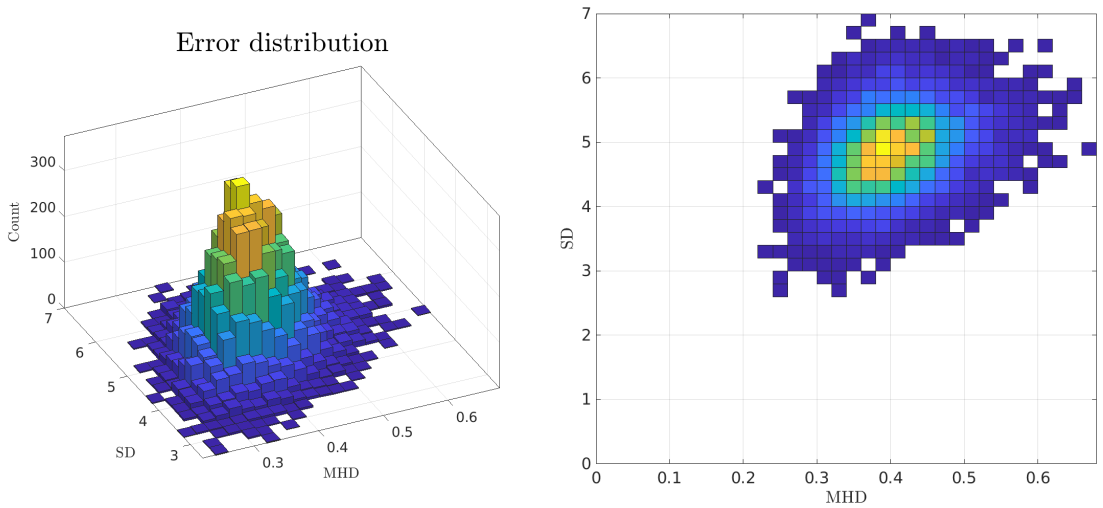
Du côté des répartitions, toutes sont gaussiennes, comme l'illustre la figure 5.15. Rétrospectivement, il s'agissait donc bien d'une caractéristique à simuler dans la base aléatoire.

Comme pour les autres bases, la matrice des distorsions spectrales de tous les couples de HRTF a été compilée (cf. figure 5.16). À nouveau, l'histogramme des valeurs présente une allure gaussienne. Sa moyenne et son écart-type valent respectivement $4,87\,\mathrm{dB}$ et $0,55\,\mathrm{dB}$. Nous sommes donc à mi-chemin entre la variabilité de la base aléatoire, dans laquelle toutes les parties de corps peuvent changer, et celle de la base synthétique, très contrainte par nature.



(a) (b)

FIGURE 5.16: *En (a), la matrice des distorsions spectrales de la base mixte. En (b), l'histogramme des valeurs associées aux couples de simulations distinctes.*



(a) (b)

FIGURE 5.17: *En (a), l'histogramme 2D des couples (SD, MHD) de la base mixte. En (b), son projeté orthogonal selon Z.*

Tout comme pour les deux autres bases, nous avons tracé l'histogramme 2D des couples (SD, MHD) (cf. figure 5.17). Cette fois, nous retrouvons une certaine corrélation positive entre les deux erreurs, que nous avions perdue chez la base aléatoire. Le fait que nous nous soyons restreint à des déformations du pavillon d'oreille n'en est à l'évidence pas étranger.

5.2 Décomposition et couplage

Maintenant que nos bases de données ont été constituées, les étapes qui suivent vont s'atteler à :

1. en extraire des espaces mathématiques dans lesquels naviguer et

2. fournir un moyen simple de passer d'un maillage à une HRTF.

5.2.1 Décomposition par ACP

5.2.1.1 Mise en place

Le premier de ces objectifs va être rempli par l'ACP. Ce choix est motivé par sa simplicité de mise en œuvre, ses nombreuses utilisations dans la littérature et son large éventail d'applications.

Pour chaque base, deux ACP sont réalisées, l'une sur les maillages et l'autre sur les HRTF. À cette fin, toutes les entrées sont vectorialisées et l'ensemble est compilé sous forme matricielle. Chaque ligne représente donc une entrée particulière. Lors de cette série d'opérations, deux points demeurent primordiaux :

1. Les entrées d'une même collection doivent être en correspondance, c'est-à-dire que leurs coordonnées doivent se rapporter aux mêmes réalités physiques. Si le n-ième triplet de coordonnées d'un maillage se rapporte à l'extrémité du nez, alors ce doit être aussi le cas pour les autres maillages. Si la i-ième coordonnée d'un vecteur HRTF est sa magnitude à 1 500 Hz dans la direction (0°, 0°), alors ce doit être aussi le cas pour les autres HRTF. En respectant cette condition, l'origine du repère que nous fournit l'ACP est de même nature que les données d'entrées. Très concrètement, on obtiendra une oreille moyenne à l'issue d'une ACP sur des formes d'oreilles et une HRTF moyenne pour une ACP sur des HRTF. À l'inverse, une ACP sur des données dont la mise en correspondance n'a pas été faite placera son origine sur un point sans signification ni structure particulière.
2. Les entrées morphologiques et les entrées auditives doivent se rapporter aux mêmes sujets. Idéalement, la i-ième ligne de la matrice des maillages et la i-ième ligne de la matrice des HRTF donnent le maillage 3D et la HRTF du sujet i.

La mise en correspondance morphologique a déjà fait l'objet d'une étude approfondie, notamment au chapitre 3. Elle est assurée par le fait que tous les maillages sont dérivés d'un même modèle déformable. On peut dès lors effectuer l'ACP et récupérer la forme moyenne et les modes de déformations. Des clichés en sont disponibles à l'annexe C.

Pour les HRTF, une mise en correspondance spécifique n'est pas nécessaire car l'utilisation de la même grille d'évaluation, du même échantillonage fréquentiel pour tous les calculs et d'un alignement des maillages avant simulation permet d'assurer la bonne cohérence des données de sortie. Des coupes azimutales et sagittales en sont disponibles à l'annexe C.

5.2.1.2 Compression des données

Une autre caractéristique de l'ACP qu'il convient, pour la suite, d'analyser est sa capacité effective à compresser l'information dans les premières composantes. Cela permet, au besoin, de mieux circonscrire le domaine d'étude. Pour cela, le cumul des quantités de variance expliquée est un indicateur classique. Les figures 5.18 à 5.20 montrent, pour chaque base, l'évolution des pourcentages de variance expliquée en fonction du nombre de premiers vecteurs propres retenus. De plus, nous y avons ajouté une mesure de la qualité de reconstruction, moyennée sur la base entière, en fonction de la proportion de premiers vecteurs propres retenus. Pour cela, l'erreur ϵ entre la donnée originelle et la donnée reconstruite avec les K premiers vecteurs propres est mesurée puis rapportée à sa valeur maximale. La qualité de reconstruction est alors donnée par $1 - \epsilon$. Nulle lorsqu'aucun vecteur propre n'est utilisé, elle vaut 1 lorsque tous le sont.

Plusieurs éléments notables apparaissent ici. Tout d'abord, la notion de « proportion de variance expliquée » est quelque peu trompeuse. En effet, elle ne coïncide pas avec la qualité de reconstruction des données mais lui est systématiquement supérieure. L'écart est d'ailleurs très important lorsque l'on compare l'évolution de la variabilité sur les maillages de la base mixte à l'évolution de la distance de Hausdorff (cf. figure 5.20). Assez inattendu, ce décalage nous semble s'expliquer par le fait que l'ACP va mesurer l'écart entre les données par une autre métrique – la moyenne des carrés des différences – que celle employée pour mesurer l'erreur de reconstruction – norme SD / MHD. En effet, la distance de Hausdorff demande à calculer une distance point à surface et notre implémentation de la norme SD pondère l'importance relative de chaque direction par le diagramme de Voronoï de la grille d'évaluation et considère des spectres de HRTF en dB. Dans les deux cas, il s'agit d'opérations très étrangères à l'ACP. La conséquence est que la troncature d'une base de vecteurs propres en fonction d'un certain pourcentage de variance résiduelle ne fixe pas le taux d'erreur résiduelle au même niveau. Cela peut s'avérer important lorsque l'on cherche à diminuer la dimensionalité des données avec une contrainte posée sur l'erreur de reconstruction.

Autre élément flagrant, la linéarité des maillages des bases synthétique et aléatoire (cf. figures 5.18 et 5.19). En effet, les 7 premiers vecteurs propres (sur 1 253) de la base synthétique et les 22 premiers (sur 456) de la base aléatoire suffisent à expliquer 99 % de la variance. Ces très faibles proportions résultent bien évidemment du caractère linéaire du processus de génération des maillages, c'est-à-dire l'utilisation de *blendshapes*. Cette caractéristique sera un atout certain en cas de réduction de dimensionalité. La base mixte, en revanche, ne présente pas du tout ce profil et plus de 60 % des vecteurs propres sont nécessaires à la récupération de 99 % de la variance.

En parallèle, nous constatons également que la variance des HRTF ne suit pas le même schéma. Certes elle atteint plus rapidement les 99 % sur les bases synthétique et aléatoire mais elle le fait bien plus lentement que celle des données morphologiques. Nous interprétons cette différence comme l'expression du caractère non linéaire du problème.

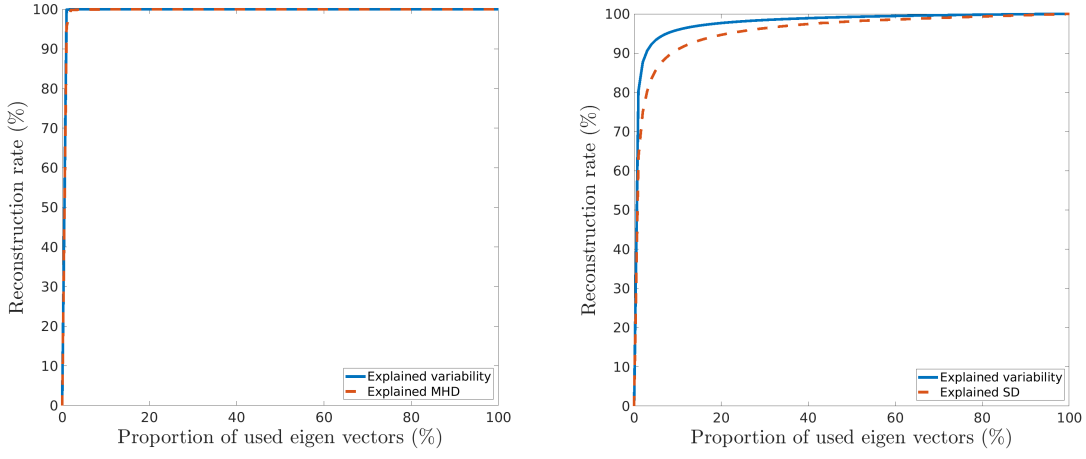
De cette étape de décomposition, nous retiendrons les éléments suivants :

— La mesure de la quantité d'information retenue par un ensemble de vecteurs propres est dépendante de la métrique utilisée et cela peut se révéler trompeur à l'usage. Dans notre cas, il ne s'agit que d'un constat sans conséquence car nous ne cherchons pas à réduire les dimensions de nos espaces de travail mais il est bon de le garder tout de même en mémoire car les besoins du jour ne sont pas forcément ceux du lendemain.

— Sur le plan de la décomposition des HRTF, les trois bases se comportent de façon assez similaire avec, bien sûr, une compression de l'information en rapport avec la variabilité sous-jacente des données. Elle est donc plus forte sur la base synthétique que sur la base aléatoire et plus forte sur la base aléatoire que sur la base mixte, comme l'on pouvait s'y attendre.

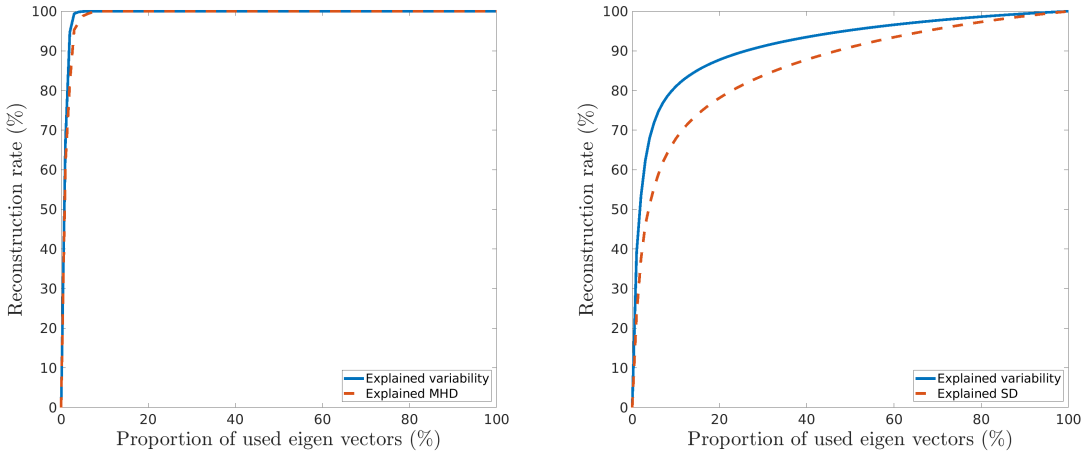
— En revanche, la différence entre la base mixte et les deux autres est très marquée sur le plan morphologique. En soi, cela justifie pleinement la constitution de la base mixte, qui affiche ainsi des caractéristiques inédites.

— De plus, si l'on adopte le point de vue des deux autres bases, la très forte compression de l'information dans les tout premiers vecteurs propres montre que l'usage de l'ACP y est particulièrement adapté. En poussant d'ailleurs le raisonnement un cran plus loin, cette particularité des bases dont les morphologies sont générées linéairement est un atout pour étudier plus facilement les liens entre la partie morphologique et la partie auditive. En effet, dans une telle situation, notre contrôle de la première partie est total et l'ensemble des efforts peut alors être porté sur les HRTF.



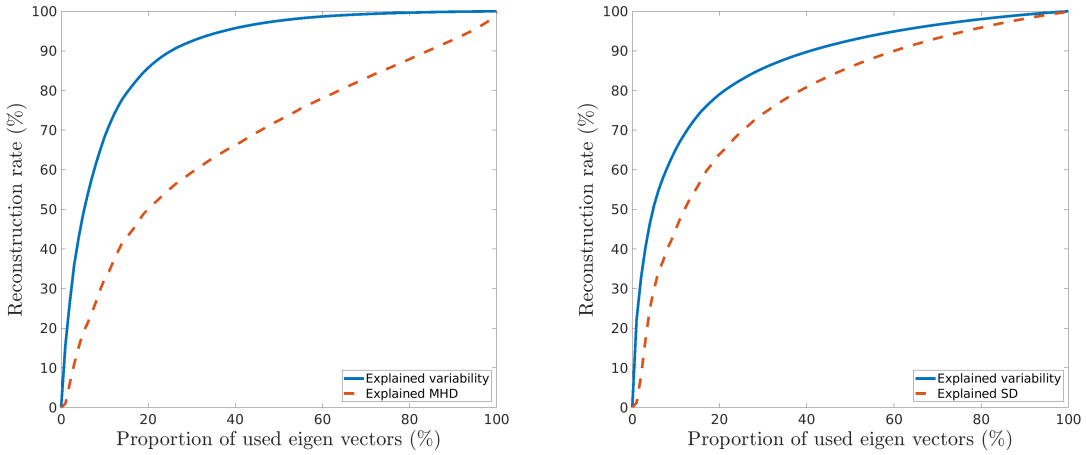
(a) (b)

FIGURE 5.18: *Base synthétique – Cumul de variance expliquée et qualité de reconstruction en fonction de la proportion de vecteurs propres retenus. En (a), après ACP sur les maillages 3D. En (b), après ACP sur les HRTF.*



(a) (b)

FIGURE 5.19: *Base aléatoire – Cumul de variance expliquée et qualité de reconstruction en fonction de la proportion de vecteurs propres retenus. En (a), après ACP sur les maillages 3D. En (b), après ACP sur les HRTF.*



(a) (b)

FIGURE 5.20: *Base mixte – Cumul de variance expliquée et qualité de reconstruction en fonction de la proportion de vecteurs propres retenus. En (a), après ACP sur les maillages 3D. En (b), après ACP sur les HRTF.*

5.2.2 Couplage

Suite à la décomposition des oreilles et des HRTF, il reste à coupler les deux espaces obtenus, c'est-à-dire à déterminer la relation de liaison qui, à chaque point de l'espace des oreilles, associera un point de l'espace des HRTF. Pour ce faire, plusieurs voies ont été explorées, chacune ayant ses avantages et inconvénients.

5.2.2.1 Couplage linéaire

Une première approche, très directe, consiste à chercher une solution linéaire au problème. À cette fin, il est utile de remarquer que les exemples d'apprentissage constituent des points de correspondance entre les deux espaces et l'on peut les exprimer sous la forme d'une relation matricielle du type $A.x_i = b_i$, où x_i est le vecteur de coordonnées de la i-ième oreille de l'ensemble d'apprentissage, b_i le vecteur de coordonnées de la HRTF correspondante et A la matrice de liaison. Les espaces ayant les mêmes dimensions, A est carrée et ses coefficients sont donnés par la résolution du système d'équations obtenu après écriture de l'ensemble des relations matricielles évoquées.

On dispose alors bien d'un moyen de passer d'une morphologie à une HRTF – et inversement – par la simple connaissance de coordonnées dans un espace ACP. Par construction, ce moyen est de plus assuré de donner une solution exacte pour chaque sujet de la base. En outre, ce pont est en pratique très simple à mettre en place et gagne en validité au fur et à mesure que de nouveaux sujets sont incorporés à la base.

Cela étant, ce procédé ne permet que la synthèse de HRTF et non de DTF alors même que la base d'apprentissage ne contient que des DTF. En effet, les vecteurs propres issus de l'ACP ne sont pas, dans le cas général, égalisés en champ diffus. Par conséquent, sauf cas particuliers, les combinaisons linéaires de vecteurs propres ne le sont pas non plus. Pour y remédier, nous ajoutons donc une étape post-synthèse d'égalisation. En outre, l'analyse de la compression par ACP nous a fait pressentir le caractère non linéaire du problème. Une solution complètement linéaire part donc avec un handicap certain.

5.2.2.2 Réseaux de neurones

Une autre approche, celle des réseaux de neurones a aussi été envisagée. L'idée est toujours de prédire les coefficients de reconstruction des HRTF à partir de ceux des morphologies mais cette fois-ci la linéarité n'est plus de mise. Pour cela, de nombreux types de réseaux neuronaux ont été envisagés et entraînés. Le nombre de couches, de neurones, d'entrées / sorties et les fonctions d'activation ont tous fait l'objet de modifications et de tests. Cependant, aucun de ces essais n'a pu montrer de véritables performances de prédiction. Au mieux certains ont réussi à fonctionner sur leur base d'entraînement, mais aucun n'a su passer l'étape de généralisation.

Ces échecs répétés peuvent se justifier par une trop grande complexité du problème vis-à-vis de la taille de la base, tant il est vrai que cette famille de méthodes est gourmande en données d'apprentissage. Pourtant, la littérature rapporte un certain nombre d'expériences utilisant les réseaux de neurones à des fins de synthèse de HRTF, et avec des bases plus petites. Bien sûr, les problèmes ne sont jamais tout à fait les mêmes mais il nous semble que nous avons avant tout manqué d'intuition dans la définition de l'architecture du réseau. Toutefois, le problème ainsi que nous l'avons posé n'assurait pas non plus que les HRTF synthétisées par le réseau neuronal soient égalisées en champ diffus. Il ne s'agissait donc pas d'une solution miracle.

5.2.2.3 Couplage barycentrique

Ces carences théoriques nous incitent alors à chercher une solution intermédiaire entre le couplage linéaire, peut-être trop simpliste, et le réseau de neurones, trop capricieux. Cette approche s'inspire des travaux développés au chapitre 3 et relatifs à la mise en correspondance des points des scans d'oreille 3D à partir de leurs équivalents 2D. À l'époque, rappelons-le, un ensemble de marqueurs des oreilles 2D avait servi de base à une triangulation, elle-même ensuite transférée à l'espace 3D. Par la suite, un jeu de recherches barycentriques effectué triangle par triangle nous permettait de mettre en correspondance de nouveaux points.

Ici, le point de départ est une nouvelle fois le transfert d'une triangulation : celle effectuée sur les points de l'espace ACP des oreilles vers l'espace ACP des HRTF. Ensuite, nous prenons en compte le fait que l'ACP concentre la variabilité de la base dans les premiers vecteurs propres pour réduire la taille de l'espace de départ à seulement K composantes. La triangulation de dimension K résulte donc en la création d'un ensemble de simplex à $K+1$ sommets pavant l'enveloppe convexe de notre base. À chaque point P compris dans l'enveloppe peut être associé un unique simplex, dans lequel P dispose de coordonnées barycentriques propres. Pour chacun de ceux hors de l'enveloppe, choix est fait de leur associer le simplex dont le barycentre est le plus proche. En reportant ces coordonnées dans l'espace d'arrivée, on obtient alors une HRTF liée à P. Cette HRTF est assurée d'être la bonne pour les exemples d'apprentissage, c'est-à-dire les sujets dont les morphologies ont servi à la triangulation. De plus, étant obtenue par combinaison linéaire de DTF, elle est en réalité elle aussi naturellement égalisée en champ diffus, ce qui confère à cette approche un avantage indéniable sur les précédentes.

Malgré tout, plusieurs critiques doivent tout de go lui être opposées. En premier lieu, les combinaisons linéaires des spectres de HRTF ont tendance à résorber la saillance des indices spectraux. Plus K sera grand et plus l'importance relative de chacun des sommets aura de probabilités d'être faible. Difficile alors de faire apparaître de façon très marquée lesdits indices. Deuxièmement, plus K sera grand et plus on aura intérêt à ce que la base soit grande elle aussi pour espérer avoir un pavage assez régulier de l'espace. Or le cardinal de la base est fixe. De plus, les analyses de nos trois bases ont montré que leurs variabilités morphologiques étaient inversement proportionnelles à leurs tailles... Enfin, l'utilisation de coordonnées barycentriques se prête en général assez peu à l'extrapolation de données et l'on peut s'attendre à observer des résultats sensiblement différents selon que l'on aura évalué un point intérieur à l'enveloppe convexe ou non.

La présentation des méthodes de couplage étant faite, il reste à effectuer leur évaluation réelle. C'est l'objet de la section qui suit.

5.3 Évaluation subjective

Comme nous l'avons déjà dit, le caractère subjectif est au cœur du problème des HRTF. Tout procédé de personnalisation doit donc se pencher sur leur évaluation finale et proposer

une phase de tests.

Dans le cas présent, nos données sont en grande partie synthétiques, sans réel humain derrière, ce qui limite de fait les possibilités en terme de tests subjectifs. Ainsi, les concepts qualitatifs, tels la préférence globale, l'émotion ou la familiarité d'un son, se voient d'entrée exclus. Parler de *head-tracking*, de temps de latence, d'indices visuels ou d'égalisation du casque n'a plus vraiment de sens non plus.

Mais pour ce qui est du quantitatif, en revanche, le simulateur de test développé par Baumgartner nous offre des possibilités sans égales. Grâce à lui, il est en effet possible d'étudier de façon exhaustive toutes les combinaisons (sujet, HRTF) imaginables. De plus, les caractéristiques spécifiques de chaque auditeur, comme une perte d'audition partielle, sa qualité de bon ou mauvais localisateur ou sa fatigue du moment peuvent être gommées, fiabilisant d'autant les comparaisons à suivre.

Ici, nous nous proposons donc d'évaluer les couplages présentés ci-avant, cœur du procédé de personnalisation, à l'aune de résultats de tests *automatisés* de localisation. Précisons à toute fin utile que nous ne parlons donc pas des tests de localisation effectués section 4.3, qui visaient à quantifier les effets de l'impédance sur les HRTF, mais de batteries de nouveaux tests de localisation ne nécessitant plus la présence d'un auditeur en chair et en os. Point commun important toutefois : les sorties du simulateur de Baumgartner nous le permettant, nous les utilisons pour calculer à nouveau notre métrique subjective décrite en annexe B. Cette métrique, exprimée en degrés, nous permet de quantifier l'erreur de localisation additionnelle que cause l'utilisation d'une autre HRTF que celle du sujet à l'étude.

5.3.1 Procédure de test

Afin de tester en masse les résultats potentiels des sujets des différentes bases de données à notre disposition, des batteries de tests automatisés ont donc été lancées. Pour chaque couple d'entrées A / B, les résultats [2] du sujet A avec les HRTF de B ont été compilés, tout comme les résultats de B avec les HRTF de A.

La stratégie de passage des tests diffère selon la base. Dans le cas des bases aléatoire et mixte, des analyses de type *leave-one-out* ont été menées. À tour de rôle, chaque entrée a tenu le rôle de sujet test tandis que l'ensemble des autres a servi de base d'apprentissage. Pour ce qui est de la base synthétique, son effectif étant notablement plus grand, il a été scindé de manière aléatoire en deux sous-blocs. 90 % des sujets – 1 128 entrées – ont alors été utilisés pour la mise en place de la chaîne de personnalisation (décomposition par ACP et couplage) tandis que les 10 % restants – 126 entrées – ont joué le rôle de testeur. En outre, la taille de cette base le permettant, d'autres répartitions (apprentissage *vs* test) ont également été utilisées. De cette façon, nous pouvons suivre l'évolution des performances en fonction du nombre d'entrées en base et en tirer des conclusions quant à la taille optimale de celle-ci.

2. Par *résultat*, nous nous référons ici aux performances de localisation dans le plan médian obtenues grâce au simulateur de test.

Pour chaque sujet test, trois méthodes de personnalisation ont été expérimentées : le couplage linéaire, le couplage barycentrique et la sélection d'une HRTF sur critère morphologique proposée par Zotkin *et al.* [171] Pour rappel, cette dernière utilise les mesures morphologiques de l'oreille du sujet dont on cherche à personnaliser les HRTF et les compare aux mesures des sujets de la base pour en sélectionner une. la HRTF choisie est celle minimisant la quantité :

$$E_j = \sum_i e_{i,j}^2, \text{ avec } e_{i,j} = \frac{\widehat{d}_i - d_i^j}{\mathrm{Var}(d_i)} \tag{5.1}$$

Ici, $\widehat{d}_i$ représente la valeur du i-ième paramètre morphologique du sujet test et d_i^j celle de ce même paramètre prise sur le j-ème sujet de la base. $\mathrm{Var}(d_i)$ représente quant à lui la variance au sein de la base des valeurs du i-ième paramètre. Dans l'article original, les auteurs excluaient certains paramètres car ils ne pouvaient être correctement mesurés à partir des photos alors utilisées. Nos données d'entrée étant tridimensionnelles, cette restriction n'a plus lieu d'être et nous pouvons choisir un nombre plus large de paramètres. Plus particulièrement, nous gardons d_1, d_1d_2, d_2, d_3, d_4, d_5, d_6, d_7, θ_1, θ_2, R et P.

5.3.2 Résultats

5.3.2.1 Base synthétique

Ci-dessous, les résultats obtenus, méthode par méthode, avec la base synthétique. En abscisse figure l'indice du sujet tandis que les mesures d'erreur sont reportées en ordonnée. La zone grise représente les niveaux d'erreur accessibles par sélection de HRTF. La ligne bleue en marque la limite inférieure, c'est-à-dire le niveau d'erreur le plus bas qu'une méthode de personnalisation par sélection de HRTF pourra jamais offrir, tandis que la ligne rouge en marque la limite supérieure, c'est-à-dire la pire sélection de HRTF imaginable. Toute méthode de ce type, et celle de Zotkin *et al.* ne fait pas exception, se situe donc nécessairement entre ces bornes. En pointillés est également affichée l'espérance mathématique associée aux statistiques de sélection. En d'autres termes, il s'agit du niveau d'erreur auquel s'attendre en faisant choix d'une sélection au hasard d'une HRTF de la base.

La figure 5.21 montre ainsi les performances d'une sélection par paramètres morphologiques. On constate que celles-ci sont en grande majorité meilleures que la sélection au hasard mais que l'on n'atteint l'optimum qu'à deux reprises.

Figure 5.22, nous avons les résultats issus du couplage linéaire. Et de manière indiscutable, ils sont excellents ! Deux sujets sont, certes, complètement manqués mais les autres sont tous très bons et 112 d'entre eux, soit 88,9 %, offrent même des performances meilleures que la meilleure des sélections possibles !

Viennent ensuite, figure 5.23, les résultats de la personnalisation par couplage barycentrique. À mi-chemin entre les deux salves de résultats précédentes, ils présentent un certain nombre de HRTF assez moyennement réussies, dont l'erreur associée frôle les 10°,

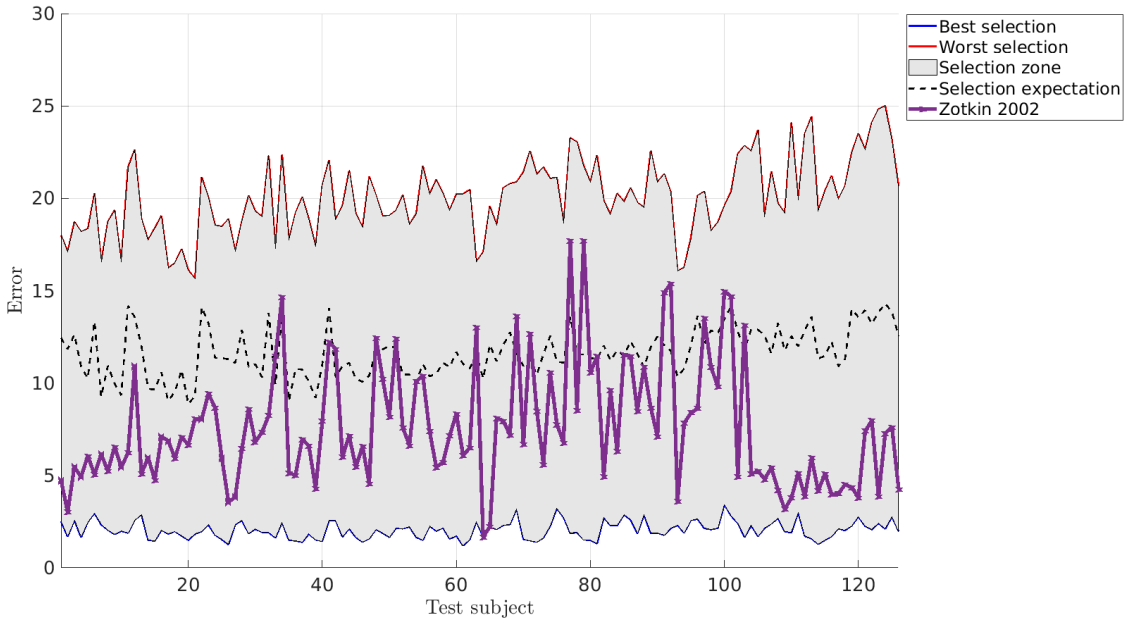


FIGURE 5.21: *Base synthétique – Performances de localisation obtenues par utilisation de la procédure de sélection de HRTF proposée par Zotkin* et al.

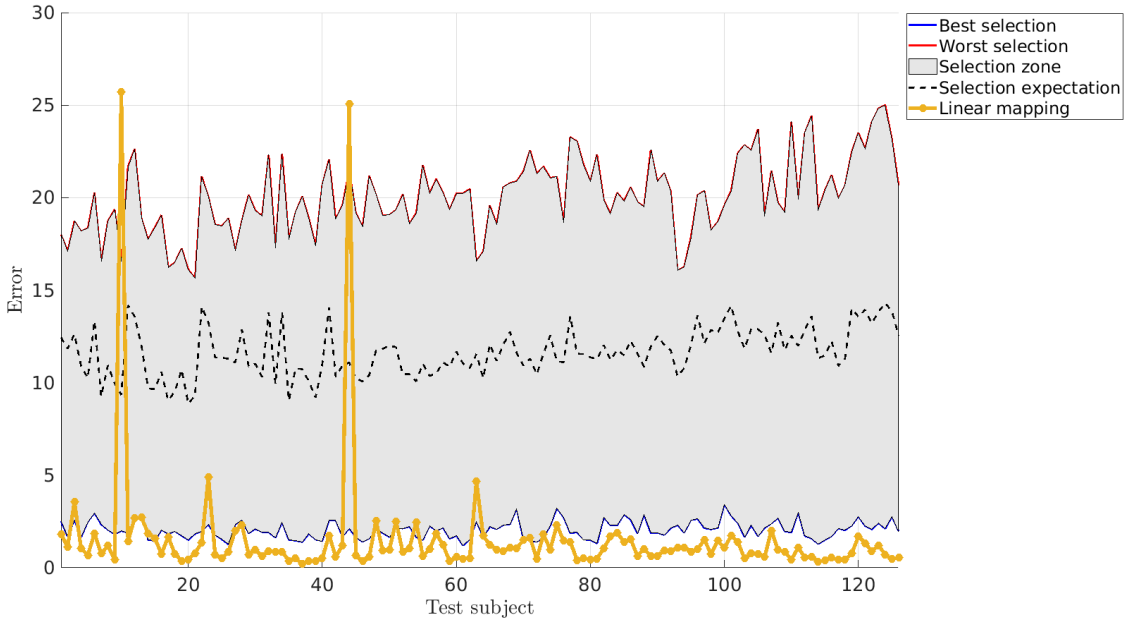


FIGURE 5.22: *Base synthétique – Performances de localisation obtenues par utilisation du couplage linéaire.*

mais parviennent également à faire mieux que la meilleure des sélections possibles dans 44,5 % des cas.

La comparaison des trois statistiques apparaît figure 5.24. On y retrouve clairement les observations faites jusqu'ici, à savoir, de très bons résultats par couplage linéaire (en dépit des deux cas particuliers), de bons résultats par couplage barycentrique et des performances honorables mais sans plus avec la personnalisation par sélection.

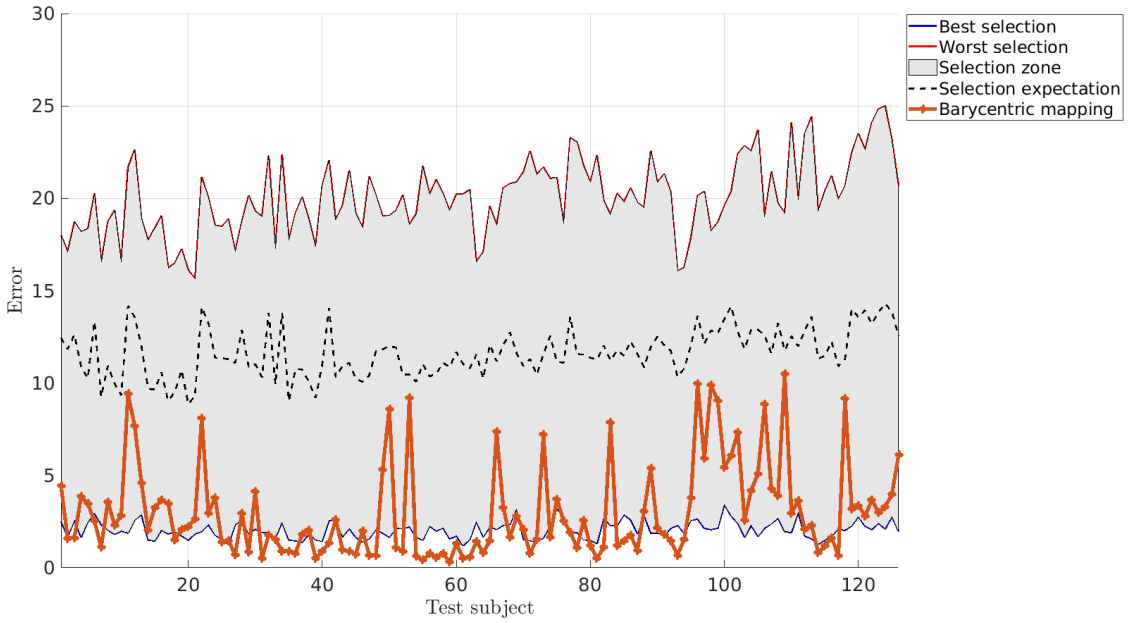


FIGURE 5.23: *Base synthétique – Performances de localisation obtenues par utilisation du couplage barycentrique.*

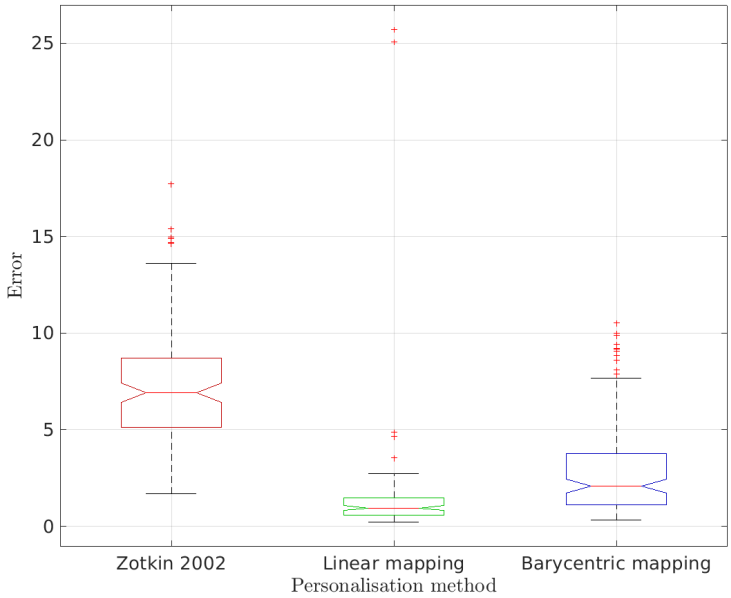


FIGURE 5.24: *Base synthétique – Comparatif des performances des différentes méthodes.*

Enfin, la figure 5.25 montre l'évolution des performances de chacune des méthodes en fonction de la proportion de sujets utilisés pour la phase d'apprentissage. Celle-ci passe de 10 % (126 sujets) à 90 % (1 128 sujets) par pas de 10 %. Pour assurer un maximum de clarté, seuls trois indicateurs ont été relevés : la médiane et les 1er et 3^{e} quartiles.

Sans surprise, les médianes sont plus basses et les quartiles plus resserrés lorsque 90 % de la base a servi à l'apprentissage. À l'inverse, les médianes sont à leur maximum et les

quartiles sont plus éloignés l'un de l'autre lorsque cette proportion tombe à 10 %.

On note également que les méthodes linéaire et barycentrique présentent des médianes qui suivent des trajectoires analogues et globalement décroissantes alors que la médiane de la méthode par sélection atteint un plateau de performance à partir de 50 %. En d'autres termes, au-delà de ce seuil, la méthode par sélection ne dégage plus de bénéfice de l'ajout de sujets en base, contrairement aux deux autres approches.

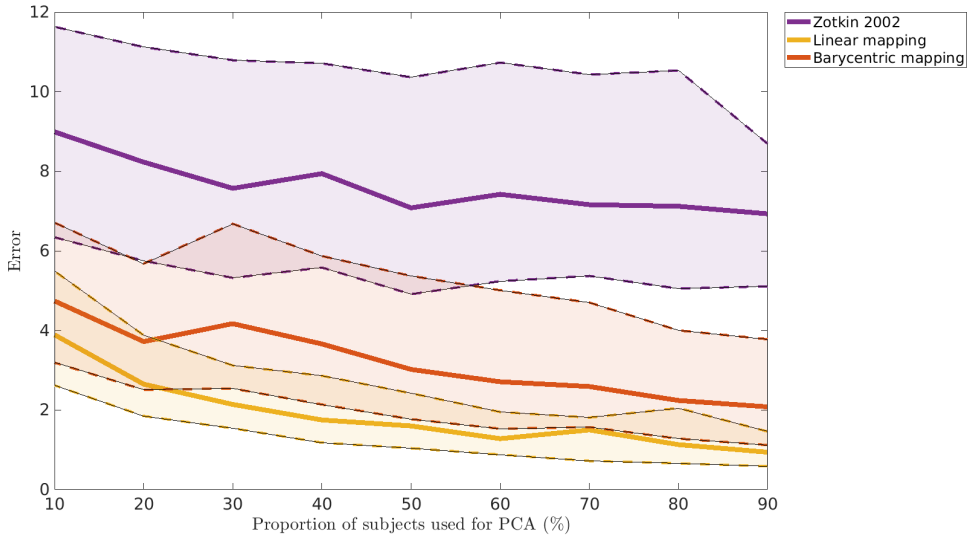


FIGURE 5.25: *Évolution des statistiques d'erreur subjective en fonction de la proportion de sujets utilisés en phase d'apprentissage (i.e. pour réaliser les ACP et le couplage). Les trois méthodes de personnalisation sont représentées (une par couleur). En trait plein, la médiane. En pointillés, les 1er et 3^{e} quartiles.*

5.3.2.2 Base aléatoire

Suivant le même type de procédure, mais en travaillant cette fois-ci avec la base aléatoire, nous pouvons faire les observations suivantes :

— En premier lieu, la sélection par paramètres morphologiques – cf. figure 5.26 – donne très souvent de meilleurs résultats que le hasard – 93,9 % du temps – et atteint même l'optimum dans 27 cas sur 457, soit 5,9 % du temps.

— Ensuite, nous constatons que le couplage linéaire – cf. figure 5.27 –, lui, est systématiquement meilleur qu'une sélection aléatoire et donne de meilleurs résultats que la sélection optimale dans 31,5 % des cas.

— Pour sa part, le couplage barycentrique – cf. figure 5.28 – déçoit quelque peu, étant en effet moins performant que le hasard dans 12,3 % des cas. Toutefois, il se révèle tout de même meilleur que la sélection optimale 7 % du temps.

— En définitive, l'analyse comparée des trois approches – cf. figure 5.29 – positionne le couplage linéaire en première place, suivi de la sélection paramétrique, au coude à coude avec le couplage barycentrique.

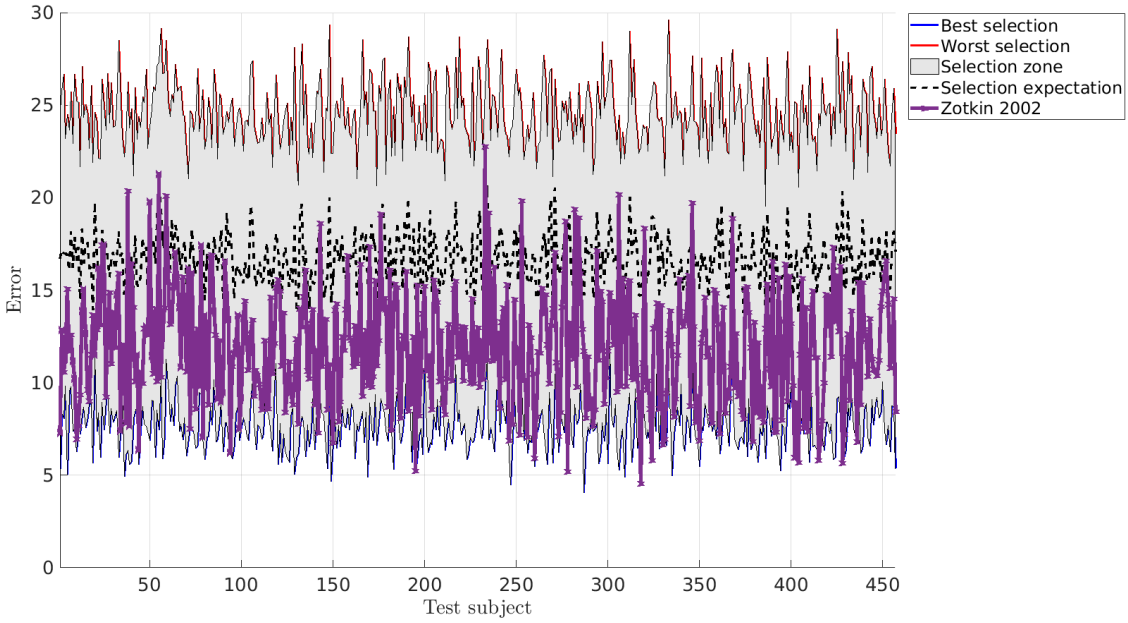


FIGURE 5.26: *Base aléatoire – Performances de localisation obtenues par utilisation de la procédure de sélection de HRTF proposée par Zotkin* et al.

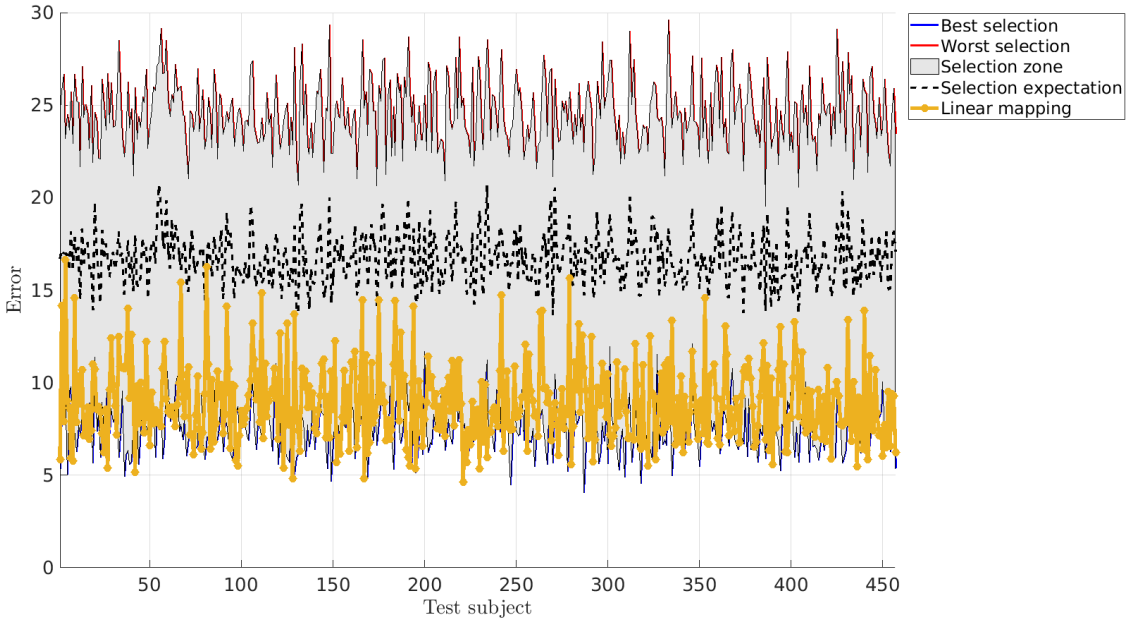


FIGURE 5.27: *Base aléatoire – Performances de localisation obtenues par utilisation du couplage linéaire.*

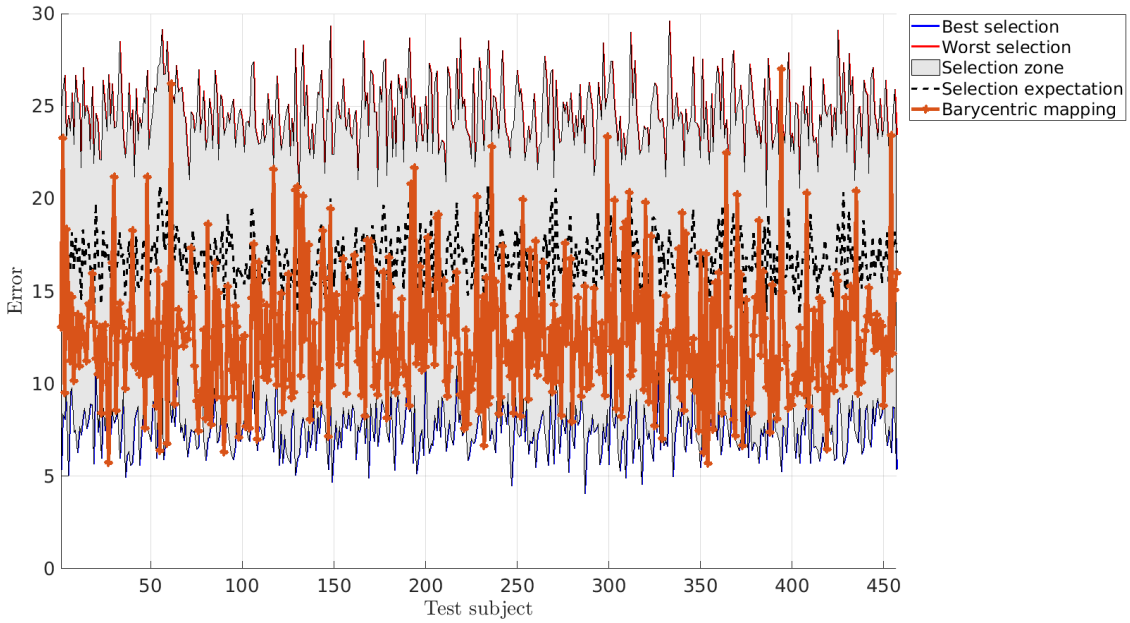


FIGURE 5.28: *Base aléatoire – Performances de localisation obtenues par utilisation du couplage barycentrique.*

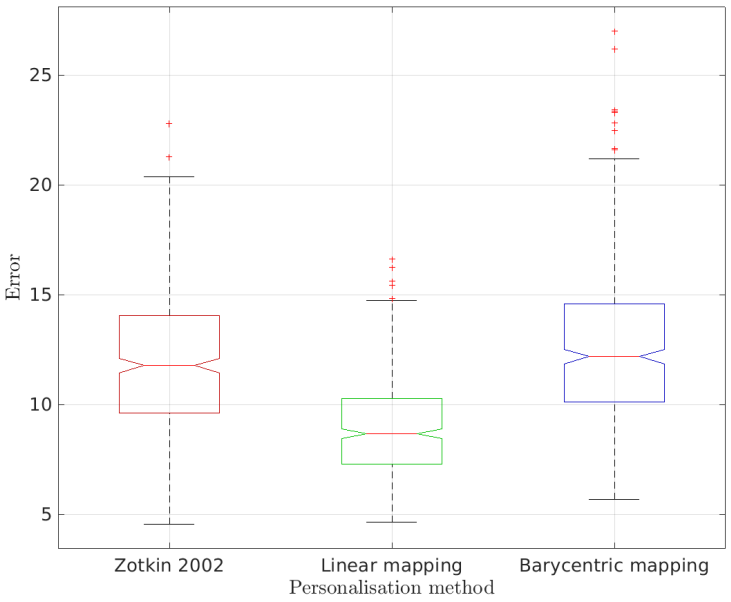


FIGURE 5.29: *Base aléatoire – Comparatif des performances des différentes méthodes.*

Enfin, la figure 5.30 montre l'évolution des performances de chacune des méthodes en fonction de la proportion de sujets utilisés pour la phase d'apprentissage. Celle-ci passe de

10 % (46 sujets) à 90 % (412 sujets) par pas de 10 %. Tout comme pour la base précédente, seuls trois indicateurs ont été relevés : la médiane et les 1er et 3e quartiles.

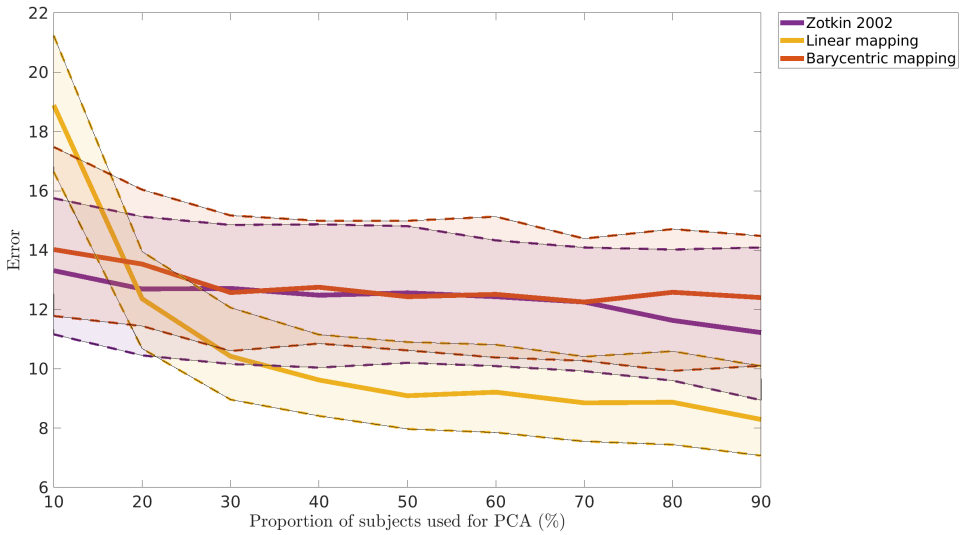


FIGURE 5.30: *Évolution des statistiques d'erreur subjective en fonction de la proportion de sujets utilisés en phase d'apprentissage (i.e. pour réaliser les ACP et le couplage). Les trois méthodes de personnalisation sont représentées (une par couleur). En trait plein, la médiane. En pointillés, les 1er et 3e quartiles.*

5.3.2.3 Base mixte

Appliquons à nouveau cette procédure, cette fois à la base mixte. Nous obtenons ce qui suit :

— Tout d'abord, les performances de la sélection par paramètres morphologiques – cf. figure 5.31 – se rapprochent dangereusement du hasard, ne faisant mieux que 60,8 % du temps. Quant à la moyenne des erreurs, elle est ici de $11,5°$ contre $12,4°$ dans le cas d'une sélection aléatoire.

— Le couplage linéaire – cf. figure 5.32 –, de son côté, opère sensiblement mieux. Il donne en effet de meilleurs résultats que le hasard dans 83,1 % des cas. Le taux de résultats meilleurs que l'optimum est cependant faible comparé aux bases précédentes, à 1,53 %.

— Enfin, le couplage barycentrique – cf. figure 5.33 – est préférable à une sélection aléatoire dans 93,6 % des cas et fait mieux que l'optimum 6,2 % du temps.

— L'analyse comparée des trois approches – cf. figure 5.34 – confirme les observations issues des deux autres bases, plaçant les méthodes de personnalisation par couplage devant la sélection sur critère morphologique.

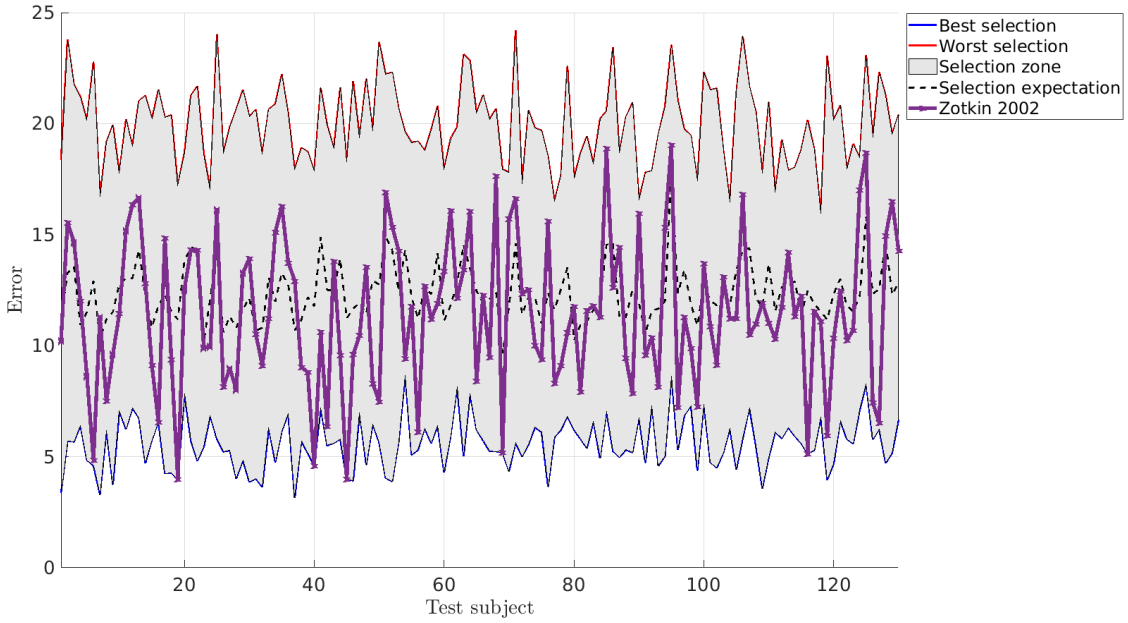


FIGURE 5.31: *Base mixte – Performances de localisation obtenues par utilisation de la procédure de sélection de HRTF proposée par Zotkin* et al.

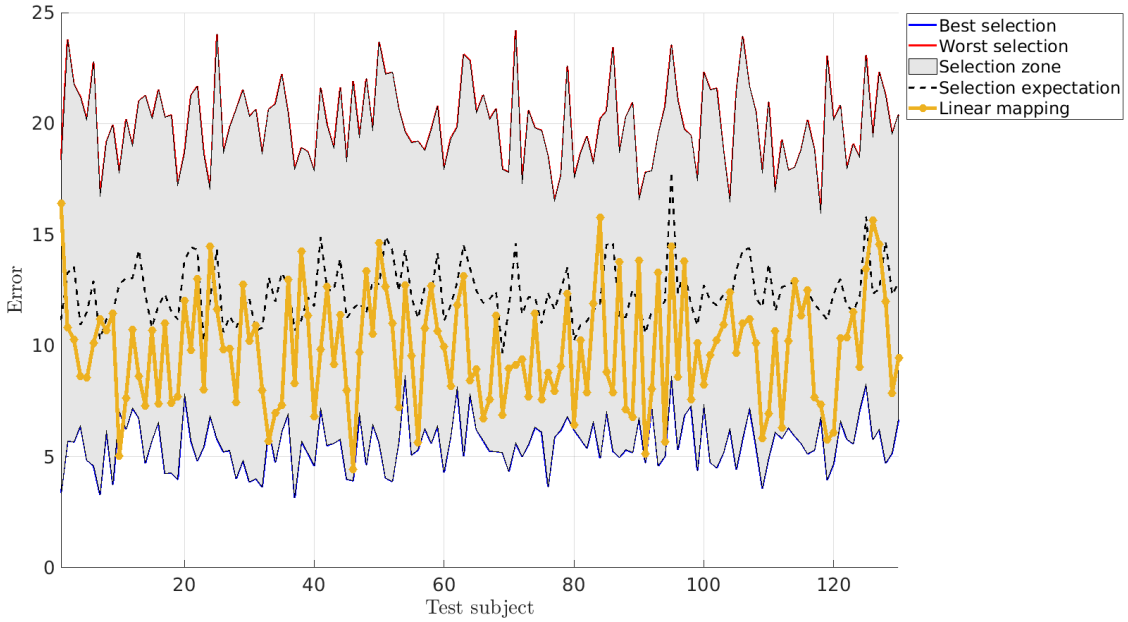


FIGURE 5.32: *Base mixte – Performances de localisation obtenues par utilisation du couplage linéaire.*

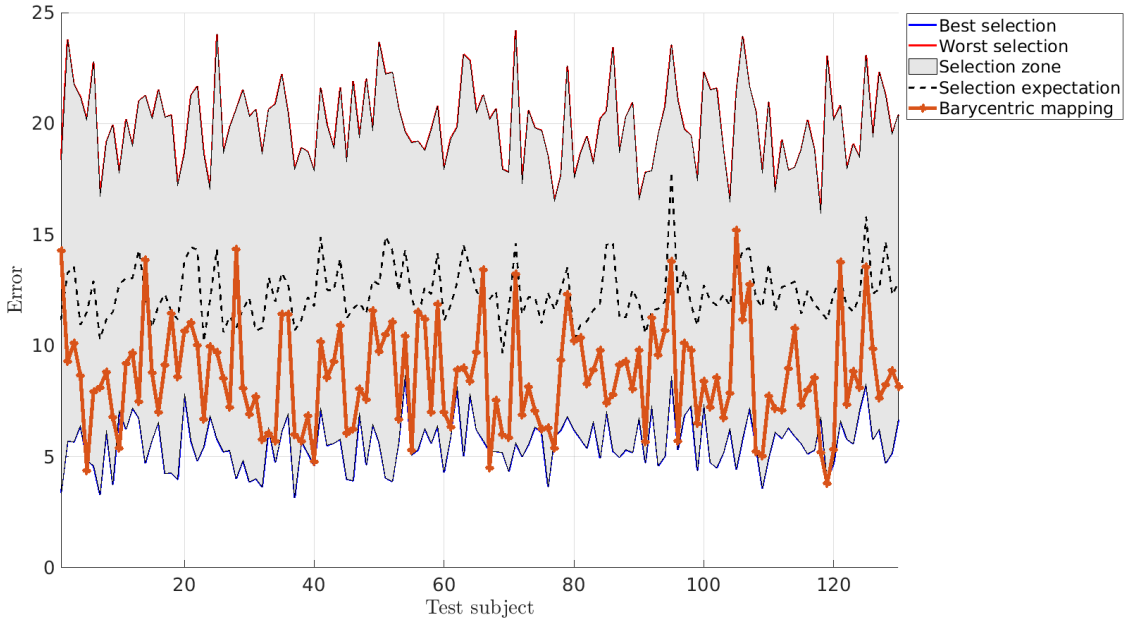


FIGURE 5.33: *Base mixte – Performances de localisation obtenues par utilisation du couplage barycentrique.*

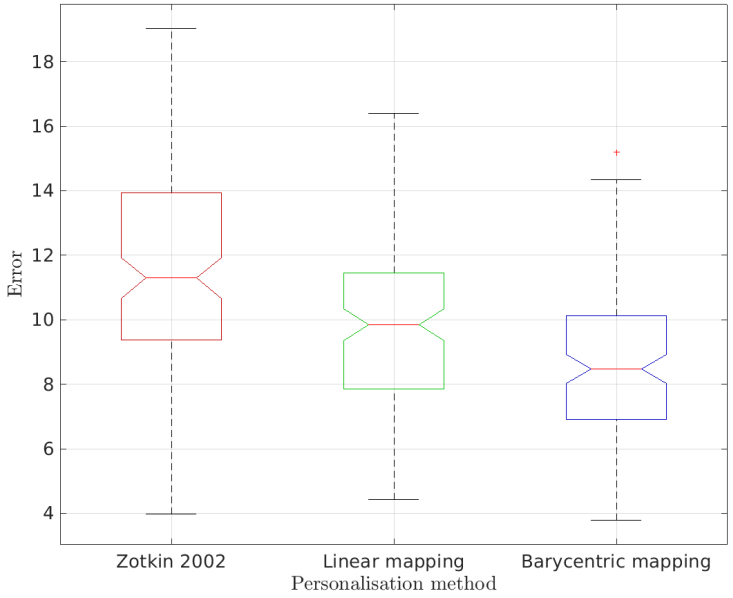


FIGURE 5.34: *Base mixte – Comparatif des performances des différentes méthodes.*

5.3.3 Discussion

Les évaluations précédentes ont eu pour but de qualifier la pertinence des méthodes de personnalisation proposées. Afin de pouvoir traiter les quantités colossales de données mises en jeu, choix a été fait de procéder à une évaluation quantitative des performances de localisation dans le plan médian qu'offre chaque méthode. Ce parti pris, critiquable par certains aspects [3], nous semble néanmoins justifié par le fait que la capacité de localisation offerte par la HRTF demeure sa caractéristique première et qu'elle est connue pour être plus complexe à obtenir en élévation qu'en azimut. En se restreignant au plan médian, d'ITD nulle, on se place donc résolument dans le cas le plus ambitieux.

Afin d'offrir à nos deux variantes de personnalisation par couplage quelques points de comparaison aisément accessibles, nous avons ajouté à l'étude la méthode de sélection de HRTF proposée par Zotkin *et al.* Bien que cette dernière remonte au début des années 2000, plusieurs arguments en font un choix raisonnable. En premier lieu, elle entend offrir une personnalisation à partir de données morphologiques et nécessite de disposer d'une large base de données. Par conséquent, celles que nous avons collectées lui offrent un cadre idéal d'application. Ensuite, sa relative simplicité lui confère une mise en place facile, avec un faible risque d'erreur ou de mauvaise interprétation. À titre d'exemple, la reproduction d'un réseau de neurones demanderait à connaître une foultitude de réglages tels que le nombre de couches, le nombre de neurones de chaque couche, les fonctions d'activation, la rétro-propagation, le type de pré-traitements appliqués aux données d'entrée, etc. À chaque étape des ambiguïtés peuvent apparaître et leur accumulation accroît inévitablement le risque de reproduire une architecture bien différente de l'originale. Certes, en adoptant notre approche, une part des performances peut sembler sacrifiée mais c'est au bénéfice d'une meilleure maîtrise de l'implémentation générale et donc des conclusions à tirer des résultats. Enfin, le fait de sélectionner une HRTF au sein d'un ensemble fini assure l'existence d'un optimum. Dépasser cet optimum, comme nous avons pu le faire notamment avec la base synthétique, permet alors de se comparer non plus à une méthode en particulier mais à une famille entière.

Par ailleurs, pour juger de l'efficacité des différents procédés de personnalisation, le protocole de test a été appliqué à chacune des trois bases de données en notre possession. Celles-ci diffèrent notablement les unes des autres, ce qui permet aux résultats obtenus de se compléter mutuellement. Pour rappel, et de manière très schématique, la base synthétique est la plus large et la mieux contrôlée, morphologiquement parlant, des trois. La base aléatoire, elle, est d'une taille plus modeste – mais toujours plus importante que les bases actuelles librement accessibles à travers le monde – et offre une plus grande variété de déformations, qui ne se limitent plus à l'oreille seule. La base mixte, enfin, s'appuie sur des oreilles scannées et non plus sur les déformations linéaires du modèle 3D mais « paie » cela par un effectif bien plus réduit.

Le passage des tests montre alors que la méthode par couplage linéaire est systématique-

3. *Quid* des autres critères d'évaluation tels que l'externalisation ou la coloration ? Pourquoi la qualité de la personnalisation du plan médian refléterait-elle celle de la HRTF totale ? Peut-on réduire une HRTF à un unique chiffre ?...

ment meilleure que la méthode introduite par Zotkin. Cela est tout particulièrement vrai dans le cas de la base synthétique, où les performances obtenues dépassent bien souvent les performances optimales que pourrait promettre une méthode par sélection. On notera au passage que les HRTF optimales offrent de très bonnes performances avec une erreur additionnelle aux alentours de 2°. Cela est la conséquence toute naturelle de la façon dont ont été conçus les maillages et de la proximité morphologique qu'ils affichent. Le résultat majeur de cette expérience n'est donc pas la petitesse de l'erreur de localisation en tant que telle – bien que s'approcher de l'optimum par sélection demande déjà une adresse certaine –, non, mais plutôt que nous avons fait mieux, et quasiment systématiquement, que ce que pouvait nous permettre d'espérer la meilleure HRTF présente en base !

En considérant maintenant les résultats du même couplage sur la base aléatoire, on observe à nouveau qu'il agit avec grande efficacité, amenant à de meilleures performances que n'importe quelle HRTF de la base dans près d'un tiers des cas. L'erreur, en valeur, est cette fois-ci plus importante mais cela était attendu tant les maillages sont bien plus distincts les uns des autres que dans le cas précédent et tant la taille de la base elle-même a été réduite.

Enfin, dans le cas de la base mixte, il y a encore un gain indéniable à utiliser le couplage linéaire plutôt que la sélection proposée par Zotkin mais il est bien moins marqué. Deux facteurs en particulier peuvent expliquer cette évolution et vont retenir notre attention : la taille de la base et la variabilité morphologique. En effet, il apparaît naturel que plus la base d'apprentissage est réduite et plus il est difficile d'en faire émerger de bonnes HRTF, que ce soit par sélection ou par couplage. Or l'effectif de la base aléatoire ne représente que 36,4 % de celui de la base synthétique, et celui de la base mixte n'en représente que 10,4 %. Rien d'étonnant alors à ce que les résultats se dégradent de la sorte en passant d'une base à l'autre. Néanmoins, la variabilité morphologique qu'offre chacune d'entre elles est également de nature à expliquer les observations. Rappelons en effet qu'après ACP, 7 vecteurs propres concentrent 99 % de la variance morphologique de la base synthétique, que ce chiffre passe à 22 dans le cas de la base aléatoire, et qu'il grimpe à 78 pour la base mixte.

Afin de clarifier les choses, nous avons donc joué sur l'effectif d'apprentissage des bases et réitéré les tests. Concrètement, nous avons progressivement réduit les contingents des bases synthétique et aléatoire jusqu'à ce qu'elles soient comparables à la base mixte. Alors, comme l'illustre la figure 5.25, nous pouvons observer l'impact certain qu'a la taille initial de l'effectif. Pour le couplage linéaire, la médiane des erreurs passe ainsi de $0,94°$ à $3,89°$, soit une multiplication par 4. Cela étant, cette augmentation n'est pas suffisante pour atteindre le niveau d'erreur que l'on retrouve avec la base mixte (presque 10°). Étant cette fois-ci à tailles égales, le reste de l'écart doit s'expliquer par la différence en variabilité.

Intéressons-nous maintenant au couplage barycentrique. Celui-ci aboutit à des résultats tout à fait honorables avec la base synthétique et la base mixte mais déçoit avec la base aléatoire. En soi, cela n'est pas si surprenant car cette méthode est très gourmande en ressources et nécessite de limiter la taille de l'espace de départ à quelques dimensions seulement (en l'occurrence 8). Il faut donc que dans ces quelques dimensions réside un

maximum d'information pertinente. Présenté autrement, les 8 premiers vecteurs propres décrivant l'espace morphologique doivent être les plus représentatifs possibles de la géométrie de l'oreille. Or la base aléatoire fait non seulement varier la morphologie d'oreille mais également celles de la tête et du torse, moins cruciales ici. À l'inverse, les deux autres bases ne font varier que l'oreille. Il apparaît d'ailleurs une différence flagrante dans l'évolution de l'erreur en fonction de la taille du panel d'apprentissage selon que l'on a travaillé avec la base synthétique ou l'aléatoire. Dans le premier cas, l'ajout de nouveaux sujets permet une réduction régulière de l'erreur. Dans le second, nous tombons rapidement sur un plateau. Un point toutefois doit être relevé s'agissant des performances sur la base synthétique. En effet, d'après les analyses de variabilité de cette dernière – cf. figure 5.18 – nos contraintes en puissance de calcul ne devraient pas avoir d'impact car 99 % de la variabilité des maillages est déjà contenue dans les 7 premiers vecteurs propres. Les quelque 1 246 autres se partageant donc moins de 1 %, les écarter des calculs aurait à priori dû constituer un avantage certain du couplage barycentrique sur le linéaire. Or ce dernier demeure sans conteste plus efficace. Enfin, si l'on se penche sur les résultats issus de la base mixte, où l'approche barycentrique semble la plus adaptée, les constatations précédentes nous amène à nous demander s'ils ne sont pas plutôt le produit de l'effet conjugué d'une variabilité morphologique centrée sur l'oreille et d'un nombre modeste de sujets. On vient de le voir, lorsqu'elles étaient présentes, les variations de tête et de torse ont pénalisé cette approche car elles se reflétaient en bonne place dans la liste des vecteurs propres de l'ACP morphologique. En leur absence, l'augmentation de la taille de la base a réduit les erreurs des deux couplages de manière semblable. Cependant, nous étions alors dans la situation la plus favorable à l'approche barycentrique car la quasi-totalité de la variabilité était alors utilisable – et utilisée. Or la base mixte montre une variabilité bien plus étalée et l'ajout de nouveaux sujets ne fait que renforcer cela. En d'autres termes, la conséquence attendue d'une augmentation de la taille de la base mixte est une baisse plus rapide de l'erreur du couplage linéaire que de celle du couplage barycentrique, qui ne pourra jamais exprimer son plein potentiel.

En définitive, bien qu'étant en théorie en mesure de répondre à des problématiques occultées par le couplage linéaire, le couplage barycentrique échoue à convaincre véritablement. Il est déjà quelque peu pénalisé par la limitation en ressources calculatoires, qui le bride, mais même lorsque ce n'est pas le cas, il reste moins efficace que le linéaire. Ce dernier couplage, au contraire, a dépassé nos attentes initiales et se trouve être la méthode à privilégier.

Chapitre 6

Conclusions et perspectives

L'aventure repose sur la richesse des liens qu'elle établit, des problèmes qu'elle pose, des créations qu'elle provoque.

- Antoine de Saint-Exupéry - *Pilote de guerre*

Les recherches présentées dans ce manuscrit s'inscrivent dans le cadre de la synthèse audio binaurale. Elles s'intéressent tout particulièrement à sa personnalisation à grande échelle et en étudient un procédé reposant sur l'analyse de données morphologiques. La résolution de ce problème trouve des applications industrielles directes dans de nombreux domaines, de la réalité virtuelle aux prothèses auditives en passant par la téléphonie mobile.

Le travail décrit s'est organisé en plusieurs parties, avec tout d'abord une introduction générale comprenant un rappel historique des avancées scientifiques successives dans le domaine du son, et plus spécifiquement du son 3D. Ce faisant, nous avons pu dessiner les contours des enjeux et verrous technologiques actuels.

De là, nous nous sommes penché plus avant sur les notions fondamentales en matière de son binaural, à savoir la HRTF et ses dérivés ainsi que les indices de localisation qu'il est possible d'en extraire. Nous avons également réalisé un état de l'art de la personnalisation des HRTF. Celle-ci peut se faire de manière directe, par la mesure acoustique ou la simulation numérique, ou indirecte, le plus généralement par sélection ou transformation contrôlée d'une HRTF préexistante. Un certain nombre de ces méthodes nécessitent d'avoir des HRTF à disposition, et en grandes quantités. C'est entre autres pour cette raison que de nombreuses bases de données de HRTF ont été constituées ces 20 dernières années. Nous en proposons une revue, détaillant les caractéristiques particulières de chacune d'entre elles. Nous rappelons ensuite le caractère très individuel de la perception auditive et la nécessaire validation subjective des résultats de la personnalisation binaurale. En particulier, nous faisons un tour d'horizon des protocoles le plus souvent observés dans la littérature. Enfin, nous décrivons une chaîne de personnalisation complète conçue pour répondre aux enjeux industriels exposés. Celle-ci constitue une première contribution majeure de cette thèse, donnant notamment lieu à la publication d'un brevet. Sa mise en place nécessite par ailleurs un travail de fond, à savoir l'étude des liens unissant la morphologie humaine aux HRTF. La revue des bases de données existantes nous ayant montré leur limites, une première phase de collecte de données est indispensable. De plus, l'étude des différents moyens actuels de personnalisation nous amène à nous tourner, en guise d'étape intermédiaire, vers la simulation numérique de HRTF.

La collecte de données fait l'objet du chapitre 3. Nous y détaillons les différentes contraintes à respecter pour assurer la plus grande qualité à la base en construction. Son objectif premier est de pouvoir ensuite parvenir à la création d'un modèle déformable d'oreille. Un tel modèle constitue la pierre angulaire du procédé de personnalisation et n'a encore jamais été réalisé. C'est la tâche à laquelle nous nous attelons donc et qui nécessite que l'on s'attarde, en premier lieu, sur le problème de la mise en correspondance des maillages 3D d'oreilles. Les méthodes existantes ayant été pensées pour des objets d'études sensiblement plus simples - tels les visages -, elles se trouvent inadaptées à notre cas d'étude et il nous faut en proposer une nouvelle. Celle-ci, assez intuitive et robuste, repose sur l'analyse des courbures locales et constitue une deuxième contribution majeure de ces travaux. Elle a en outre donné lieu à la publication d'un second brevet. Il est à noter que cette méthode de création de modèle déformable y est décrite dans un cadre plus large que celui des oreilles et peut ainsi s'appliquer à une multitude d'objets 3D. Une fois passée

cette étape, nous nous intéressons à la fusion de ce modèle d'oreille sur un maillage plus complet comprenant une tête et un torse. Ceci est en effet un passage obligé pour qui veut calculer des HRTF.

Les questions relatives à la production même de HRTF numériques sont abordées au sein du chapitre 4. Nous y traitons tout d'abord la question du besoin de fidélité lors de la représentation du canal auditif. Il s'agit en effet d'une zone très peu accessible, même en laboratoire, et son influence, même faible, rendrait encore plus problématique la personnalisation à grande échelle. Nous montrons que, pour autant que l'on ne s'intéresse qu'à la composante directionnelle de la HRTF, c'est-à-dire à la DTF, la forme du canal auditif n'a pas d'impact. Ce résultat, généralement implicitement admis, trouve quelques vérifications expérimentales dans les travaux d'Hammershoï & Moller, de Middlebrooks *et al.* ou encore de Weiner & Ross. L'utilisation de la simulation numérique pour traiter ce sujet permet d'évacuer certaines objections relatives aux perturbations expérimentales. La vérification du résultat mentionné par une méthode inédite constitue une autre contribution de ces travaux.

Nous passons ensuite à l'analyse de plusieurs paramétrages possibles des simulations. Nous montrons que, même pour le calcul de DTF, il est préférable de positionner la source sonore au fond du canal auditif, sans quoi certains artefacts numériques demeurent, et qu'il y a quelques avantages à la représenter par un ensemble d'éléments vibrants plutôt que par un seul. Nous détaillons également deux méthodes d'optimisation permettant d'accroître la vitesse de calcul des HRTF : le maillage adaptatif et la dépendance en fréquence. La première permet d'affiner le maillage sur les zones les plus critiques. La seconde permet d'utiliser des maillages différents selon les gammes de fréquence à l'étude. Cette dernière idée prend racine dans le code de mesh2hrtf, qui semble avoir été écrit pour accueillir à terme une fonctionnalité similaire. Toutefois, sa mise en place complète et son utilisation pratique n'ayant pas encore été rapportées, elles constituent ici une autre contribution. Au final, nous arrivons à accélérer les simulations de 10 % et à réduire considérablement les besoins en RAM des machines les réalisant. Enfin, au cours de cette thèse, nous avons observé que l'état de l'art en matière de simulation numérique peinait à produire des HRTF subjectivement convaincantes. Nous avons également observé, lorsque les HRTF acoustiques et numériques d'un même sujet étaient disponibles, que des différences notables apparaissaient. Nous avons alors émis et testé l'hypothèse que ces différences sont le résultat d'un mauvais paramétrage de l'impédance acoustique de la peau en simulation. Nous prouvons que le choix de cette impédance est de faible importance en basses fréquences mais qu'il est critique en moyennes et hautes fréquences. Toutefois, les valeurs effectives de l'impédance de la peau sur le spectre $[100, 16\,000]$ Hz n'est pas connu. Nous proposons alors deux protocoles permettant de les déterminer. Nous montrons, analyse numérique et tests de localisation à l'appui, que les modifications faites sur l'impédance permettent de produire des HRTF subjectivement pertinentes là où l'état de l'art ne fait pas mieux qu'une HRTF générique. Ceci constitue une contribution majeure de ces travaux et a donné lieu à une publication scientifique. Néanmoins, nous sommes limité par le nombre de HRTF de référence disponibles et ne pouvons passer l'étape de généralisation. En

parallèle, nous testons les effets d'une impédance différente selon les endroits du maillage. Bien que prometteuse, cette voie comporte trop d'inconnues pour l'emprunter aujourd'hui sereinement.

Une fois ces études sur la production de HRTF réalisées, nous nous attaquons à la constitution de trois bases de données de maillages et de HRTF. La première, complètement synthétique, exhibant une impédance rigide et centrée sur les formes d'oreilles, est forte de 1 254 entrées. La seconde, synthétique également mais utilisant une impédance particulière et autorisant un plus large spectre de modifications morphologiques, contient 457 entrées. Une dernière enfin, utilisant le modèle déformable d'oreilles réelles et une impédance rigide, présente 129 entrées. Une version quasi-identique de la première de ces bases, intitulée *CHEDAR*[1], a été mise en ligne sur *sofaconventions.org*[2] et a fait l'objet d'une publication scientifique. Malgré les réserves émises sur l'état de l'art de la simulation numérique, elle offre un formidable objet d'étude, tant par le type de données que par leur quantité, à toute la communauté. Cela constitue également un contribution majeure de ces travaux. La dernière partie du manuscrit est consacrée à la réalisation d'un pont rapide et fiable entre le monde morphologique et celui des HRTF. Plusieurs approches, linéaires ou non, ont été envisagées sans qu'aucune ne parvienne à résoudre tous les problèmes. La meilleure d'entre elles est pour l'heure l'approche linéaire, qui, malgré quelques handicaps théoriques, parvient à des résultats dépassant parfois les espérances. Néanmoins, elle requiert un ensemble d'apprentissage d'autant plus grand que la variabilité morphologique de la base est importante. Son application permet toutefois de tester une grande partie de la chaîne de personnalisation proposée. Par ailleurs, afin de fournir une évaluation subjective qui ne soit pas affaiblie par la question de l'impédance en simulation ou tout simplement impossible du fait de l'inexistence de sujets réels derrière les entrées des bases de données, nous proposons une utilisation à grande échelle du simulateur de tests subjectifs développé par Søndergaard & Majdak [140]. Indépendemment des résultats obtenus, cette dernière brique atteste de la faisabilité technique de la chaîne prise dans son ensemble et la dote d'une boucle de rétroaction lui permettant de s'améliorer.

Les réponses qu'elle apporte étant généralement moins nombreuses que les questions qu'elle soulève, une thèse ne se termine jamais réellement et celle-ci ne fait certainement pas exception. Un long chemin a certes été parcouru dans cette entreprise de démocratisation du son 3D mais de nombreuses questions sont apparues ainsi qu'une multitude d'améliorations possibles, comme autant de perspectives pour l'avenir. Ainsi, la fonction de liaison entre les espaces morphologique et auditif nécessite d'être améliorée afin d'en faire émerger plus naturellement et avec moins de pertes les liens sous-jacents. Les réseaux de neurones restent à cet égard un candidat sérieux sur lequel il conviendra de se pencher à nouveau. Présenté autrement, nous pensons que l'échec de notre tentative n'est aucunement rédhibitoire, que le potentiel de ce type de méthode reste intact et qu'il peut donner lieu à des recherches à part entière. Dans cette optique, les résultats obtenus grâce à l'approche linéaire concurrente forment d'ailleurs un excellent premier objectif à dépasser. Un deuxième peut alors être de

1. Computed HRTF and Ears Database for Acoustic Research
2. https://www.sofaconventions.org/data/database/chedar

conserver de meilleurs résultats que l'approche linéaire tout en réduisant la part de la base servant à l'apprentissage et un troisième étant enfin de changer de base et de se confronter à des données réelles.

Autre sujet brûlant, celui du rendu subjectif des HRTF simulées, avec en ligne de mire l'étude de l'impédance. Est-elle *la* réponse ou seulement l'une de ses composantes ? Comment traiter les cheveux ? Les vêtements ? Pourquoi l'utilisation d'un torse absorbant, bien qu'améliorant de prime abord les simulations semble-t-il plus nuire aux optimisations et aux résultats de tests subjectifs ? Certains travaux sont aujourd'hui en cours pour apporter une solution au problème de la mesure expérimentale de l'impédance dans les hautes fréquences. En attendant leurs conclusions, qui seront à n'en pas douter d'une importance capitale, il est d'ores et déjà possible d'imaginer toute une série de protocoles permettant d'avancer sur ce problème. Ainsi, pour l'expérimentateur disposant d'un système d'acquisition de HRTF acoustiques, il pourra être instructif d'étudier l'impact sur la mesure de perturbations telles que l'application de vernis sur les pavillons d'oreilles ou le remplacement du sujet par sa copie imprimée en 3D. Dans la même veine, on pourra également envisager l'acquisition de HRTF d'un mannequin type KEMAR dont plusieurs jeux d'oreilles, de formes identiques mais d'impédances bien distinctes, seraient disponibles. Quelle que soit l'option choisie, il faut s'attendre à ce que les HRTF obtenues soient notablement différentes les unes des autres, ou tout du moins suffisamment pour dégrader significativement les performances de localisation des sujets tests. Il est également à noter que dans les cas où un mannequin serait utilisé pour les mesures, le problème de l'impédance s'en trouvera grandement simplifié car celle-ci variera beaucoup moins d'une zone à une autre.

Une autre voie d'exploration est à chercher parmi les caractéristiques uniques de la base CHEDAR. En effet, grâce à ses quatre grilles d'évaluation de diamètres différents, l'étude du passage progressif du champ proche au champ lointain est à portée de main. En parallèle, elle vient aussi avec toute une série de mesures anthropométriques et laisse toute la latitude à ses utilisateurs d'en définir et réaliser de nouvelles. Cette flexibilité accrue pourra-t-elle déboucher sur un système performant de personnalisation à partir de mesures anthropométriques ? Nous ne pouvons que le souhaiter. D'ici-là, nous pouvons déjà envisager de prédire les HRTF en champ proche, très rarement disponibles, à partir de celles en champ lointain, bien plus communes. Là encore, l'entraînement d'un réseau de neurones nous semble approprié et prometteur.

Enfin, mais la liste pourrait se poursuivre encore longtemps, nous nous sommes beaucoup intéressé aux HRTF et, en comparaison, assez peu au modèle déformable. Or il est à l'évidence perfectible. Et bien que les artefacts qu'il peut produire actuellement sont corrigés après coup et ne gênent pas les traitements subsidiaires de notre cas d'utilisation, d'autres méthodes très prometteuses, s'appuyant notamment sur les *Thin Plate Splines*, sont à l'étude et pourraient aboutir à un modèle plus fiable. Un point notable d'amélioration existe aussi au niveau de la mise en correspondance, dont la première étape est aujourd'hui manuelle et n'utilise finalement qu'assez peu les caractéristiques géométriques à sa disposition. Or d'autres raffinements sont aisément implémentables. Parmi les autres raffinements envisageables, on peut penser à compléter l'information de courbure locale

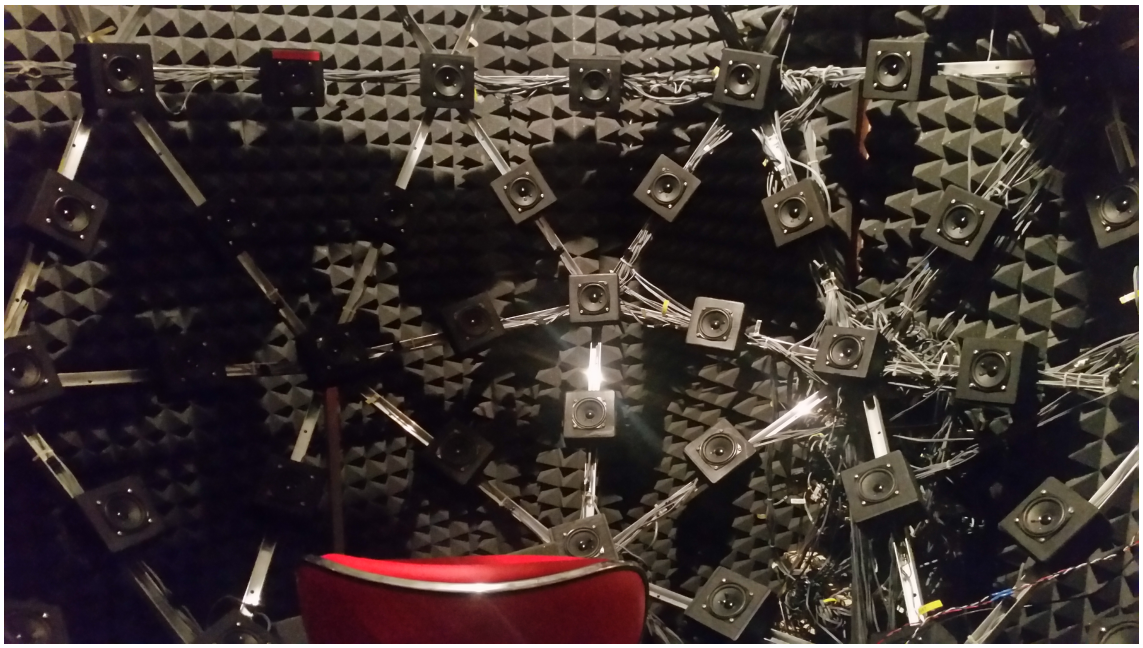
par une information de courbure moyenne autour du point, ce qui permettrait un repérage plus simple des correspondances et une stabilité générale accrue, et / ou utiliser à nouveau cette information de courbure lors de la phase de recherche barycentrique. Ce dernier point, en particulier, augmenterait encore le sens donné aux liens unissant entre eux les points en correspondance.

Annexe A

SoundStage

Bien que l'essentiel des travaux de ce manuscrit soit consacré à la production, l'analyse et l'évaluation de HRTF produites par simulations numériques, il s'est plusieurs fois avéré nécessaire de faire l'aller-retour avec l'acoustique. Malheureusement, les infrastructures permettant l'acquisition de HRTF acoustiques sont rares et difficile d'accès. Grâce à l'aimable collaboration des laboratoires d'Orange, il a été possible de réaliser la capture de HRTF de plusieurs membres de l'équipe. Ces données, utilisées comme vérité terrain lorsque cela fut possible, sont toutefois disponibles en trop faible nombre. Par ailleurs, avoir la possibilité de tester à l'envie l'audition et les facultés de localisation d'un individu en conditions réelles est chose utile et précieuse dans un domaine aussi expérimental que l'audio. Le besoin s'est donc naturellement fait sentir de construire notre propre infrastructure de capture acoustique : le *SoundStage*

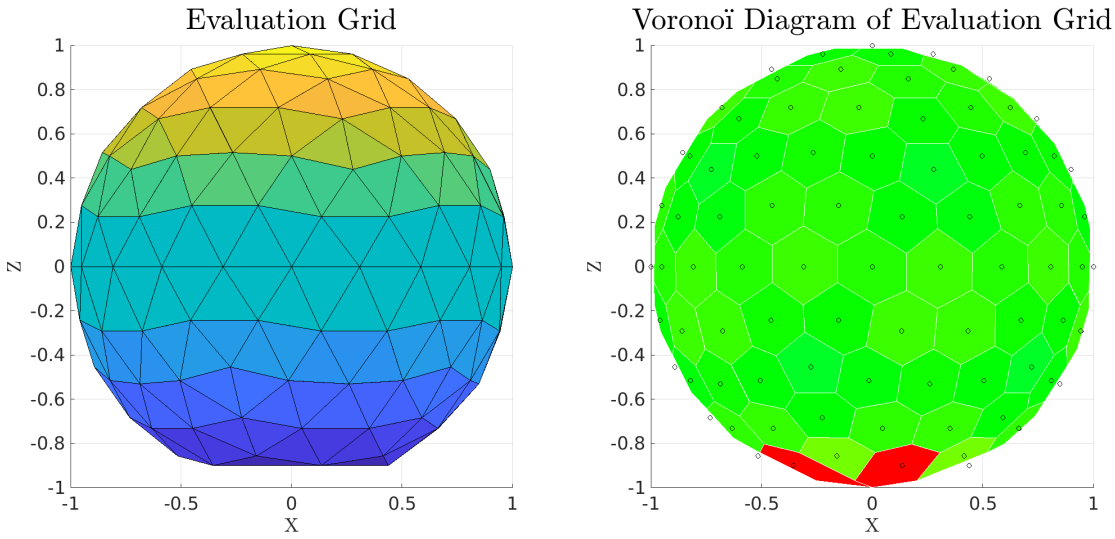
Sans chercher à concurrencer les installations professionnelles déjà existantes, il nous a permis, à faible coût, de tester / vérifier / infirmer certaines hypothèses et fait l'objet du descriptif suivant.



(a)

FIGURE A.1: *Le SoundStage construit et utilisé pour les besoins de la thèse.*

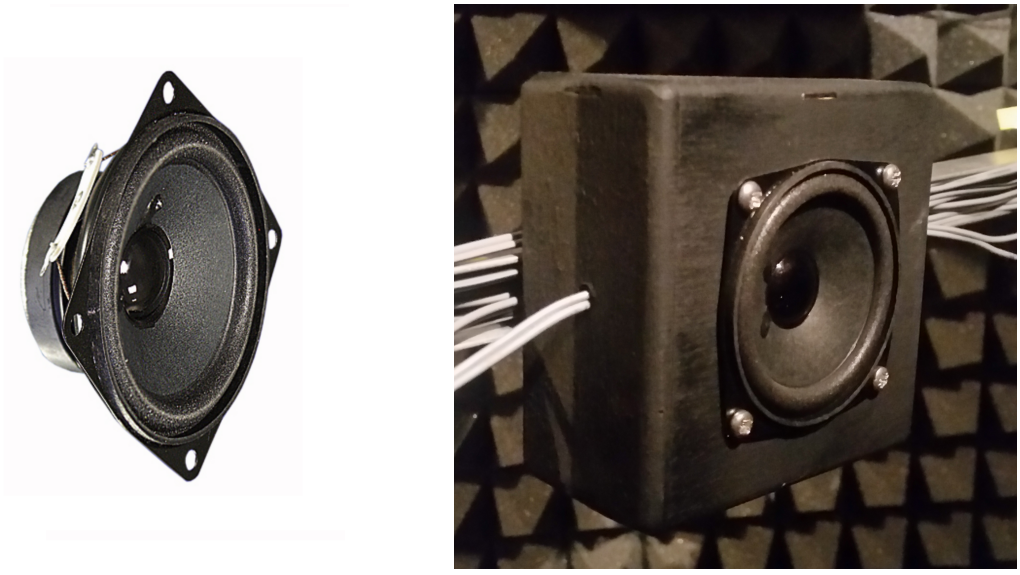
Le SoundStage consiste en une structure métallique fixe, visible figure A.1, sur laquelle sont réparties 156 haut-parleurs. Afin d'en assurer la meilleure répartition possible, une géométrie icosaédrique a été choisie. L'écartement angulaire moyen entre deux haut-parleurs voisins est de 18°. Pour permettre le passage et l'installation des sujets, quelques sommets ont été supprimés à la base de la structure. Les élévations les plus basses alors accessibles sont de $-64°$. La figure A.2 montre la grille d'évaluation obtenue et le diagramme de Voronoï associé.



(a) (b)

FIGURE A.2: *En (a), la grille d'évaluation du SoundStage. En (b), le diagramme de Voronoï associé.*

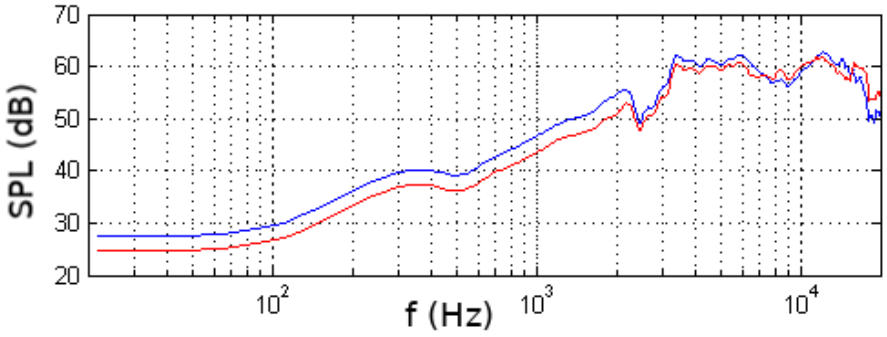
Les haut-parleurs en eux-mêmes sont de marque *Visaton*, modèle FR7. L'analyse de leur courbe de réponse en fréquence a montré que les basses fréquences sont sous-représentées et la confection d'un coffrage additionnel en bois s'est révélée indispensable - cf. figure A.3.



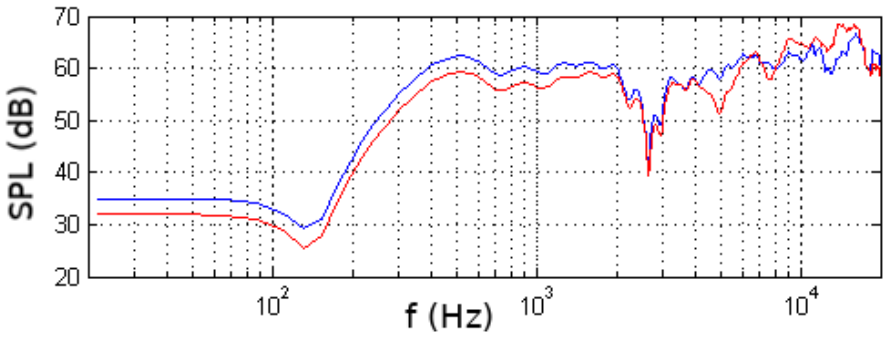
(a) (b)

FIGURE A.3: *En (a), le HP seul. En (b), le HP dans son coffrage.*

La figure A.4 montre la réponse en fréquence avant et après ajout du coffrage.



(a)



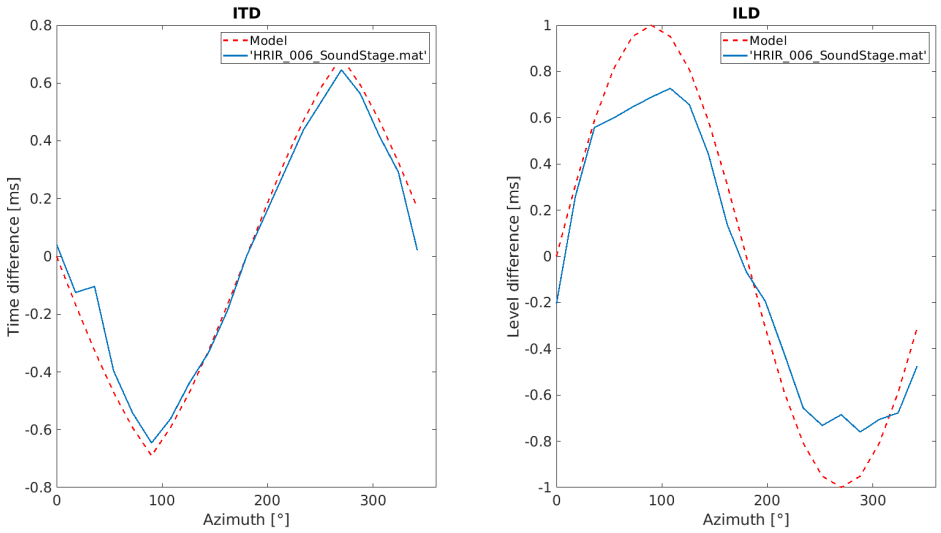
(b)

FIGURE A.4: *En (a), la réponse en fréquence du HP seul. En (b), sa réponse après ajout du coffrage. Les mesures ont été réalisées grâce à une paire de micro binauraux. Les deux canaux sont à chaque fois représentés.*

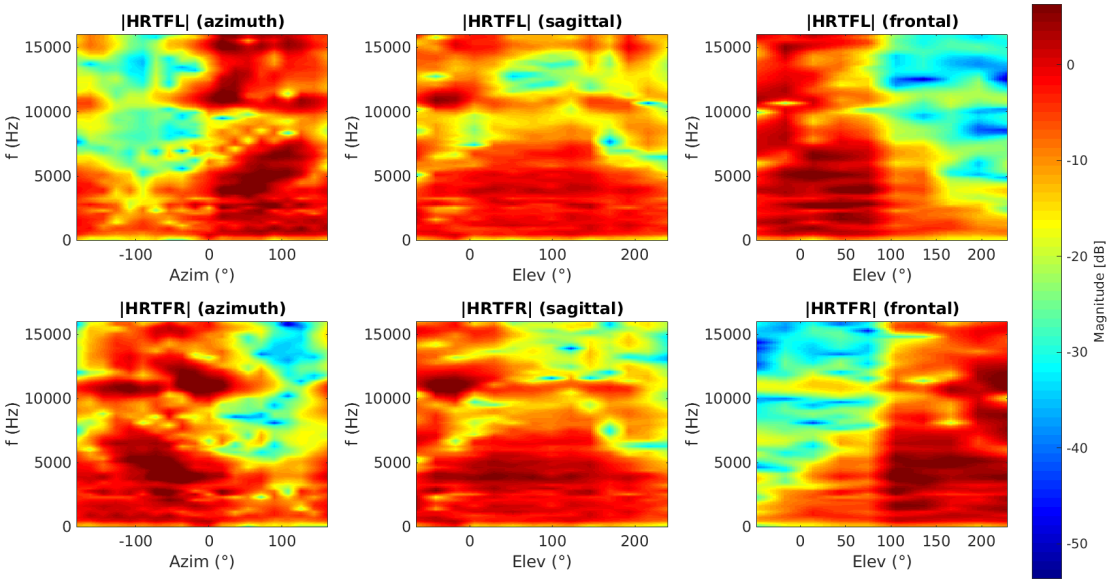
La pièce, quant à elle, est tapissée de mousse absorbante pour limiter autant que possible la présence d'échos tandis qu'une fine moquette est apposée au sol.

Lors des acquisitions de HRTF, la méthode des rampes de fréquence exponentielles a été utilisée et les réponses impulsionnelles, enregistrées à 48 kHz, ont été coupées à 208 échantillons. Comme cela a été rappelé à la section 2.2.1.3, ce type de signaux d'entrée offre une grande robustesse aux bruits extérieurs. Par ailleurs, la troncature des trames permet de supprimer en grande partie la réverbération résiduelle liée à la salle. De plus, une phase de calibration permet de tenir compte des distorsions liées aux haut-parleurs. Pour cela, une mesure à blanc avec un micro omnidirectionnel placé au centre de la sphère est réalisée. Les distorsions liées aux micros binauraux sont, elles, éliminées lors de l'égalisation en champ diffus.

Au total, une acquisition dure un peu moins de quinze minutes. Un exemple de HRTF résultante est présenté figure A.5. Elle est à comparer à sa vérité terrain - cf. figure A.6 - acquise dans les laboratoires d'Orange. On peut observer une grande similarité des ITD, ILD et du spectre pris jusqu'à environ 11 kHz, un peu moins au-delà.



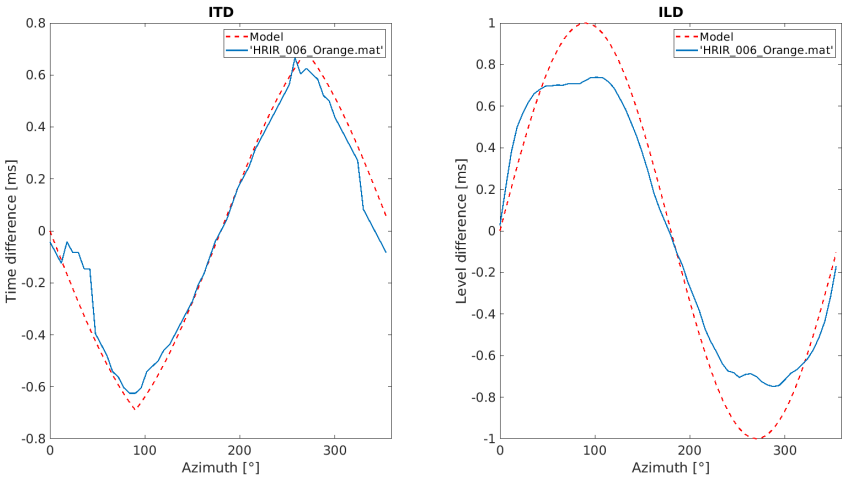
(a) Mesures d'ITD (à gauche) et d'ILD (à droite).



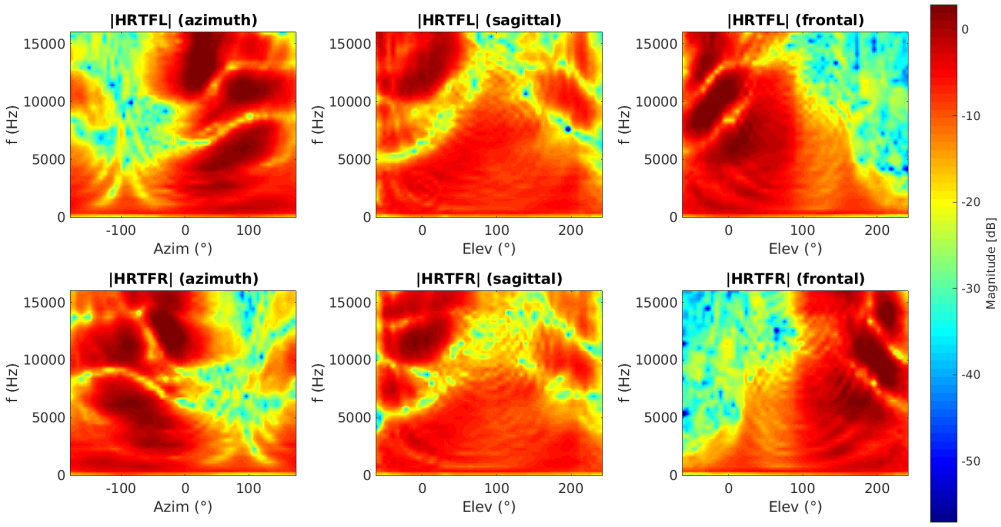
(b) Magnitude des spectres fréquentiels gauche (en haut) et droite (en bas) selon trois plans de coupe.

FIGURE A.5: *En (a), l'ITD et l'ILD extraits des mesures du SoundStage. En pointillés, un modèle type est présent et fait office de point de repère. En (b), les coupes azimutale, sagittale et frontale de la magnitude du spectre, en dB.*

Lors de tests de localisation en conditions réelles, le SoundStage a permis de vérifier que la plupart des gens partageaient un sens commun de la spatialisation sonore - ils parvenaient sans trop de difficulté à savoir quel HP était en train de jouer -, mais aussi de détecter de mauvais localisateurs. Deux en particulier ne parvenaient pas à distinguer l'avant de l'arrière. Ces mauvais localisateurs ont par la suite été exclus des campagnes de tests, leur audition étant jugée trop particulière et de nature à miner les analyses statistiques potentielles.



(a) Mesures d'ITD (à gauche) et d'ILD (à droite).



(b) Magnitude des spectres fréquentiels gauche (en haut) et droite (en bas) selon trois plans de coupe.

FIGURE A.6: *En (a), l'ITD et l'ILD extraits des mesures d'Orange Labs. En pointillés, un modèle type est présent et fait office de point de repère. En (b), les coupes azimutale, sagittale et frontale de la magnitude du spectre, en dB.*

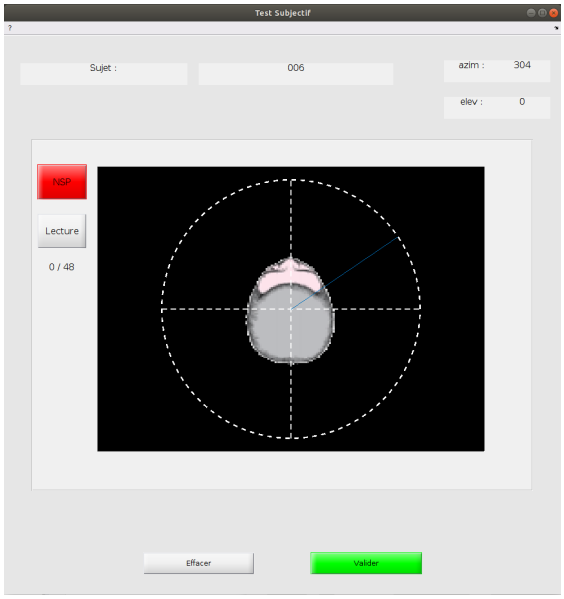
Cependant, après une première campagne portant sur une centaine de sujets, un biais manifeste est apparu dans les résultats des tests subjectifs. Le côté droit était plus facilement détecté que le gauche. L'explication la plus plausible retenue jusqu'à ce jour est que la réverbération n'est pas assez atténuée et que, ajoutée à l'asymétrie de la pièce, elle permet à de nombreux indices indésirables de localisation de subsister. Pour cette raison, la mesure des facultés de localisation en conditions réelles des individus n'a pas été plus poussée.

Annexe B

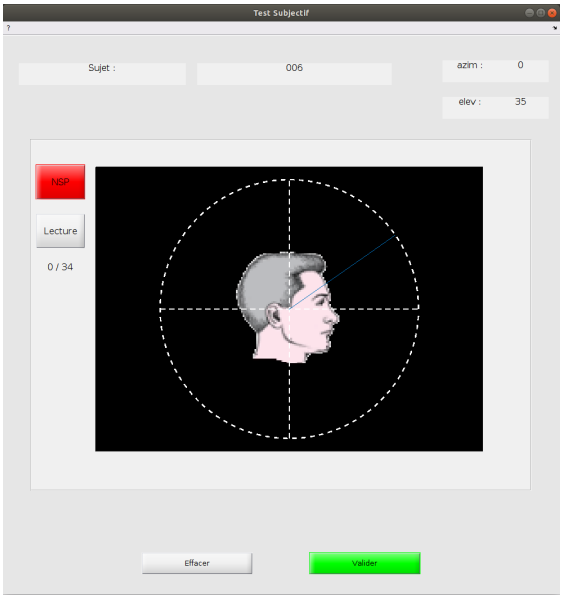
Descriptif des tests

B.1 Localisation

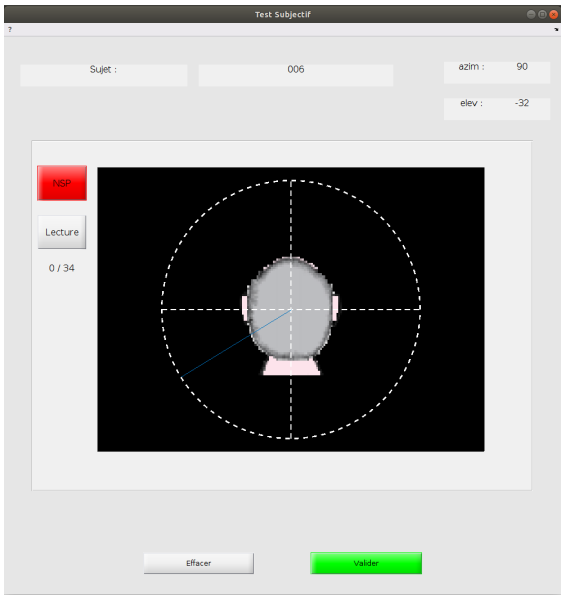
L'objectif de ce test est de quantifier la qualité du rendu d'un jeu de HRTF en terme de localisation. Pour des raisons techniques liées à l'implémentation et la maintenance du projet, le choix d'une interface graphique 2D a été fait et les HRTF des sujets sont évalués selon trois plans de coupe (azimutal, sagittal et frontal – cf. figure B.1).



(a) azimut



(b) élévation



(c) frontal

FIGURE B.1: *Interfaces graphiques du test de localisation.*

Dans chacun des cas, l'auditeur est muni d'un casque et entend un contenu binauralisé par le jeu de HRTF en cours de test. Différentes directions se succèdent et le sujet donne, via l'interface graphique, la direction d'où lui semble provenir le stimulus. Celui-ci est

sélectionné en début de test et ne change plus par la suite. Le stimulus le plus souvent utilisé est constitué d'une série de bruits blancs mais des voix, des sons de la nature ou des extraits musicaux sont également disponibles. Il n'y a pas de limite de temps et le sujet peut revenir sur ses précédentes réponses s'il le souhaite. Il peut également faire rejouer le dernier stimulus entendu autant que nécessaire. Un bouton « Ne Sais Pas » est à sa disposition s'il considère qu'aucune direction claire ne se détache.

En option, une phase d'apprentissage peut être proposée en amont du test. Dans ce cas, toutes les directions possibles sont jouées les unes à la suite des autres et une croix colorée vient matérialiser la direction de consigne sur l'interface. Les réponses peuvent être laissées libres ou contraintes aux seules directions testées.

À l'issu de la session, l'auditeur est invité à enregistrer ses réponses dans un fichier qui sera ensuite passé à l'outil d'analyse. Ce dernier permet en effet d'extraire et de visualiser un certain nombre de métriques relatives à la précision de localisation, comme l'erreur angulaire, les erreurs de cadrans ou le taux d'inversions. Les figures B.2, B.3 et B.4 en présentent un aperçu.

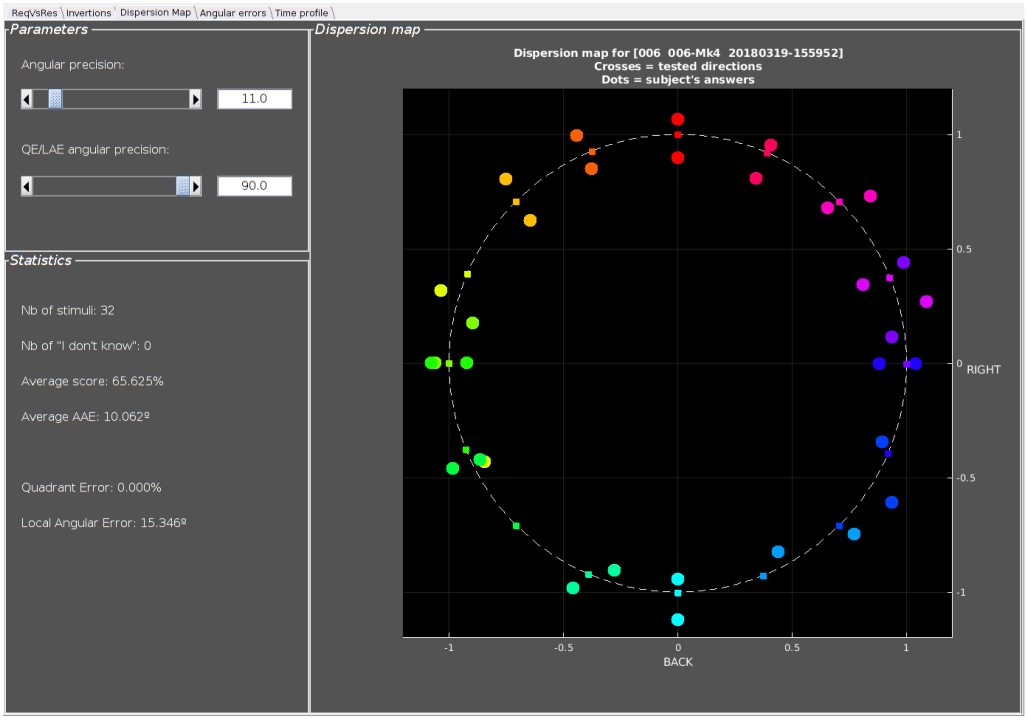


FIGURE B.2: *Carte de dispersion des réponses. Les carrés colorés sur le cercle correspondent aux directions jouées lors du test. Les points de même couleur sont les réponses données. Pour ne pas superposer les marqueurs de requête et de réponse, ces derniers voient leur distance au centre du cercle légèrement modifiée. La direction pointée, seule information véritablement importante, reste inchangée. Les éventuels « Ne Sais Pas » sont positionnés à proximité du centre. Sur le panneau gauche quelques indicateurs, comme le taux de bonnes réponses et l'erreur angulaire absolue moyenne, sont affichés.*

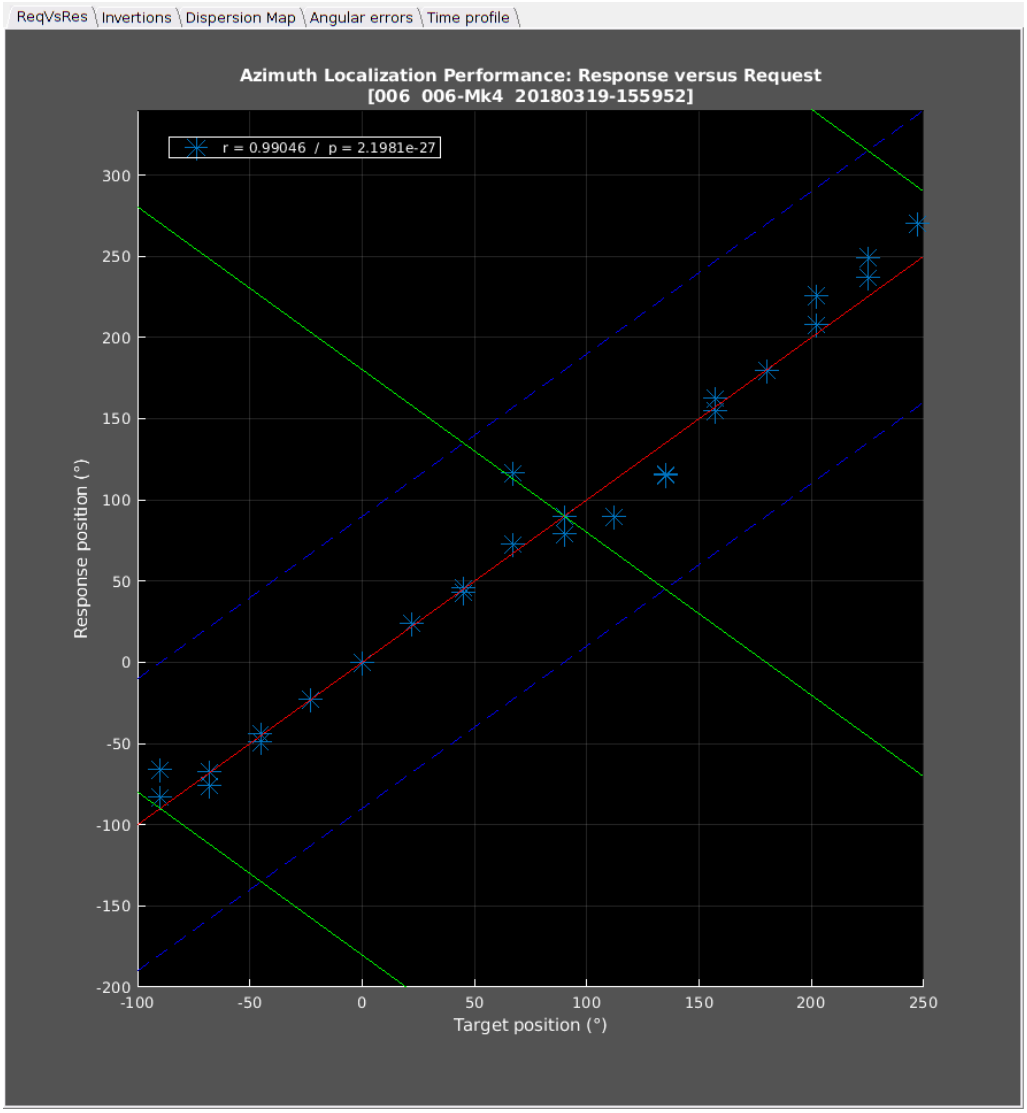


FIGURE B.3: *Graphe des réponses en fonction des requêtes. Les étoiles bleues matérialisent les réponses. Idéalement, elles sont toutes sur la ligne rouge. Une réponse sur la ligne verte est synonyme d'inversion avant-arrière. Les lignes en pointillés bleu marine indiquent un décalage de* $\pm 90°$ *par rapport à la consigne.*

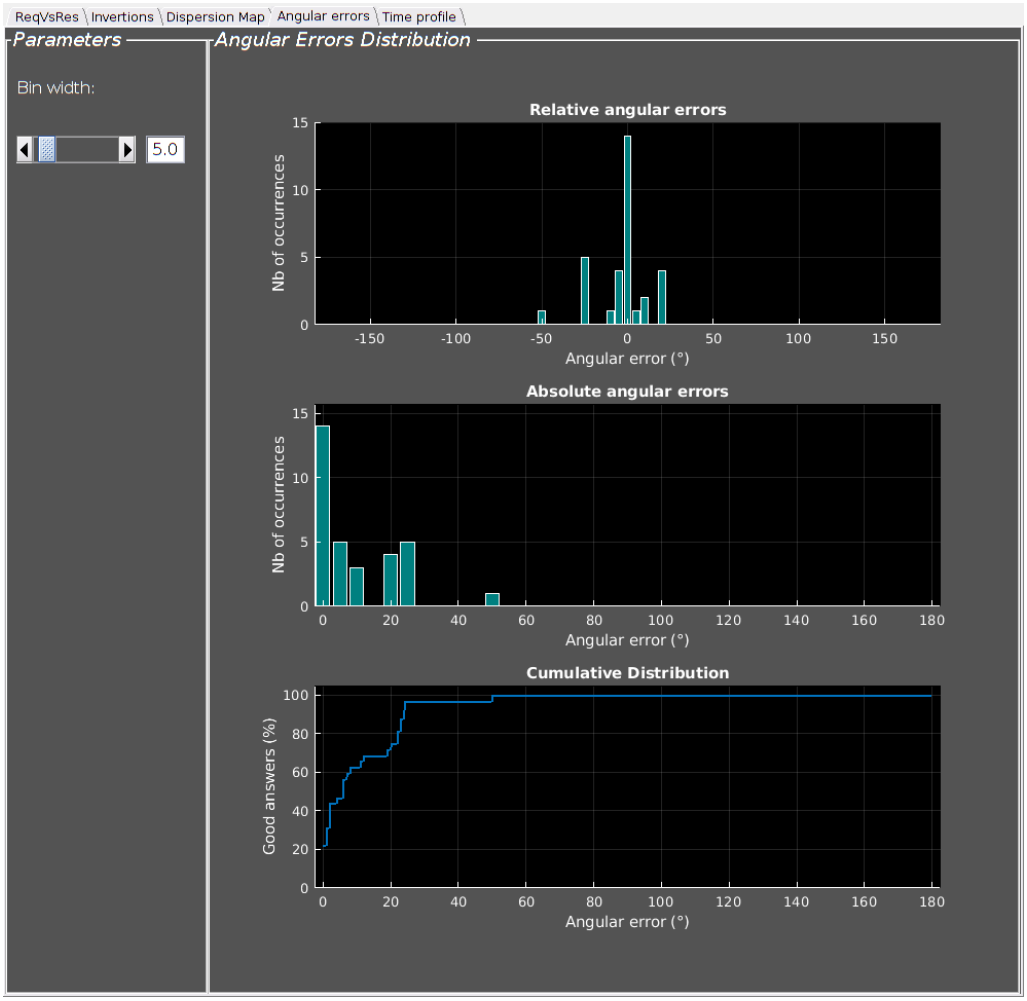


FIGURE B.4: *Histogrammes des erreurs et des erreurs en valeurs absolues et courbe du pourcentage de bonnes réponses en fonction de la marge angulaire autorisée.*

B.2 Simulateur de test

En 2014, Baumgartner *et al.* ont mis à disposition un outil de test automatique [8] dont l'objectif est la modélisation des performances auditives d'un être humain à partir de la donnée de ses HRTF. Conçu spécifiquement pour simuler et analyser les performances dans le plan sagittal, il se veut refléter l'état de l'art des connaissances en matière de psycho-acoustique. Il intègre ainsi un banc de filtres gammatones chargé de mimer les effets de la cochlée. Ce banc couvre la plage spectrale $[700, 18\,000]$ Hz. La limite inférieure de cette plage est vue par les auteurs comme la limite fréquentielle en-deçà de laquelle même le torse n'a plus d'influence sur le spectre. La limite supérieure est vue comme une limite réaliste de la perception auditive. De même, l'extraction du gradient positif du spectre figure parmi les traitements opérés. Déjà mis en évidence dans le système auditif du chat, ce gradient est supposé être également extrait par l'appareil auditif humain.

En sortie, le simulateur établit une carte de probabilités de réponse dans le plan sagittal ainsi qu'un ensemble de métriques liées à l'analyse des performances de localisation. Plus précisément, il s'agit :

— de l'*erreur moyenne locale*, soit la moyenne des écarts angulaires inférieurs à 90° entre consignes et réponses,
— du *taux d'erreur de quadrant*, soit le pourcentage de réponses éloignées de plus de 90° de la consigne,
— du *biais en élévation*, soit le biais angulaire observé dans les statistiques de réponses en élévation.

Grâce à cela, nous pouvons donc évaluer le rendu subjectif des HRTF sans se soucier de l'existence ou non de sujets réels en amont. De surcroît, nous nous affranchissons également des contraintes matérielles liées au dispositif d'écoute, de l'ordre de passage des tests, de la fatigue des sujets, etc.

Parmi les dernières spécificités de ce simulateur, notons la présence d'un peu de flou gaussien dans les réponses. S'il est le bienvenu pour accroître le réalisme du simulateur, il est en revanche gênant pour évaluer de manière robuste, c'est-à-dire reproductible, des HRTF entre elles. Pour cette raison, et sauf mention contraire, les simulations sont relancées 50 fois et la moyenne des statistiques est alors utilisée pour définir notre métrique. Plus précisément, si A et B sont deux jeux de HRTF alors nous définissons l'écart de l'un à l'autre par :

$$mes(A,B) = \frac{|eml_A(B) - eml_A(A)| + |eml_B(A) - eml_B(B)|}{2} \tag{B.1}$$

où $eml_A(B)$ est l'erreur moyenne locale obtenue par le sujet de HRTF A écoutant avec les HRTF B. Présenté autrement, $eml_A(B) - eml_A(A)$ représente l'erreur angulaire relative additionnelle commise par le sujet de HRTF A écoutant avec le jeu B. La présence des valeurs absolues est un garde-fou mathématique nous assurant que $mes(A,B) \geqslant 0$. Elles supposent implicitement qu'un auditeur obtiendra de meilleurs résultats de localisation avec ses propres HRTF. Cette mesure s'exprime en degrés (°). L'introduction de la demi-somme, quant à elle, nous permet d'avoir la propriété de symétrie : $mes(A,B) = mes(B,A)$.

Annexe C

Modèles ACP

Cette annexe compile des vues des différents modèles déformables créés par ACP à partir des bases synthétique, aléatoire et mixte vues au chapitre 5. Dans chacun des cas, la moyenne et les 10 premiers vecteurs propres sont représentés.

C.1 Modèles de la base synthétique

C.1.1 Modèle morphologique

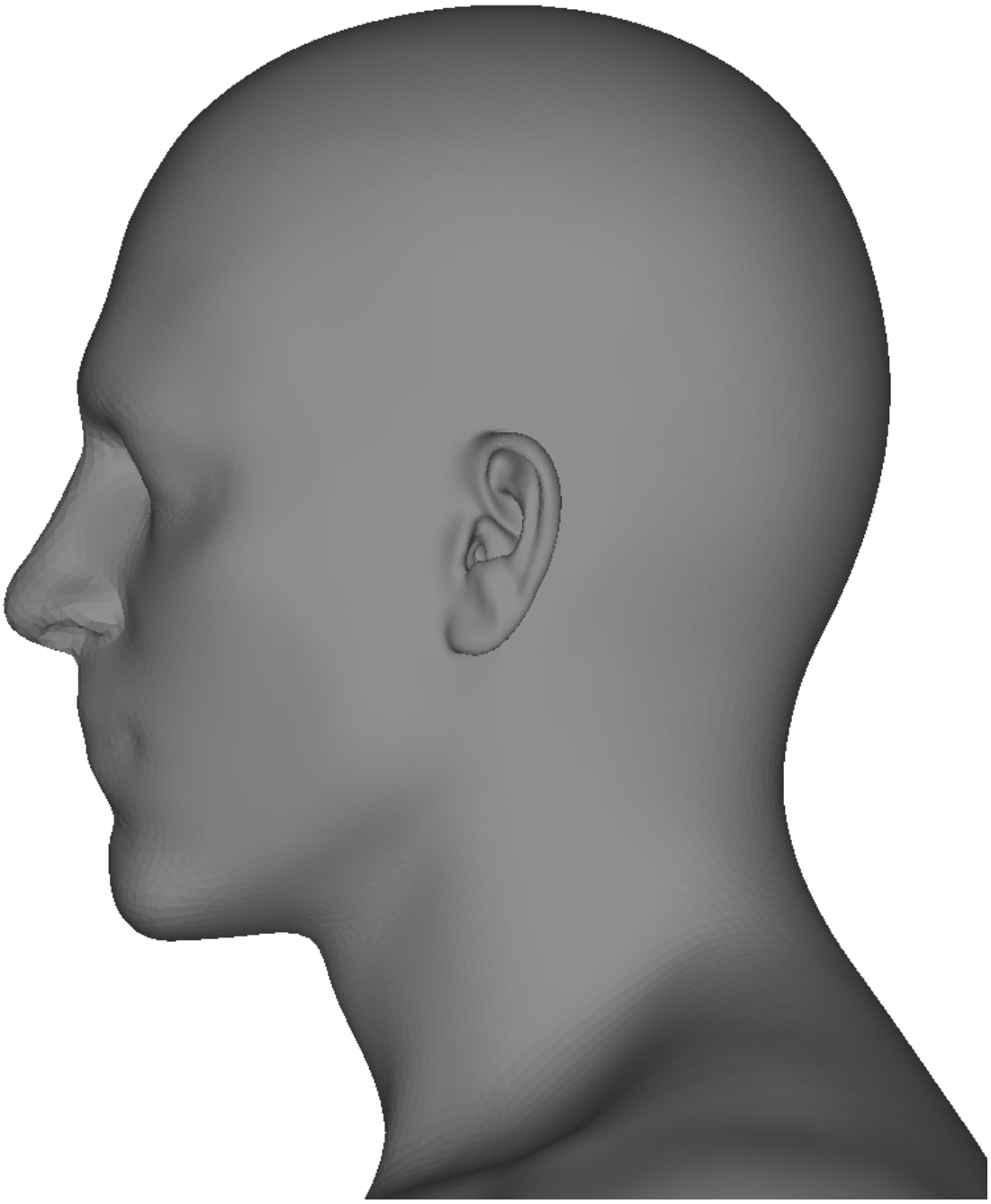


FIGURE C.1: *Forme moyenne du modèle issu de la base synthétique.*

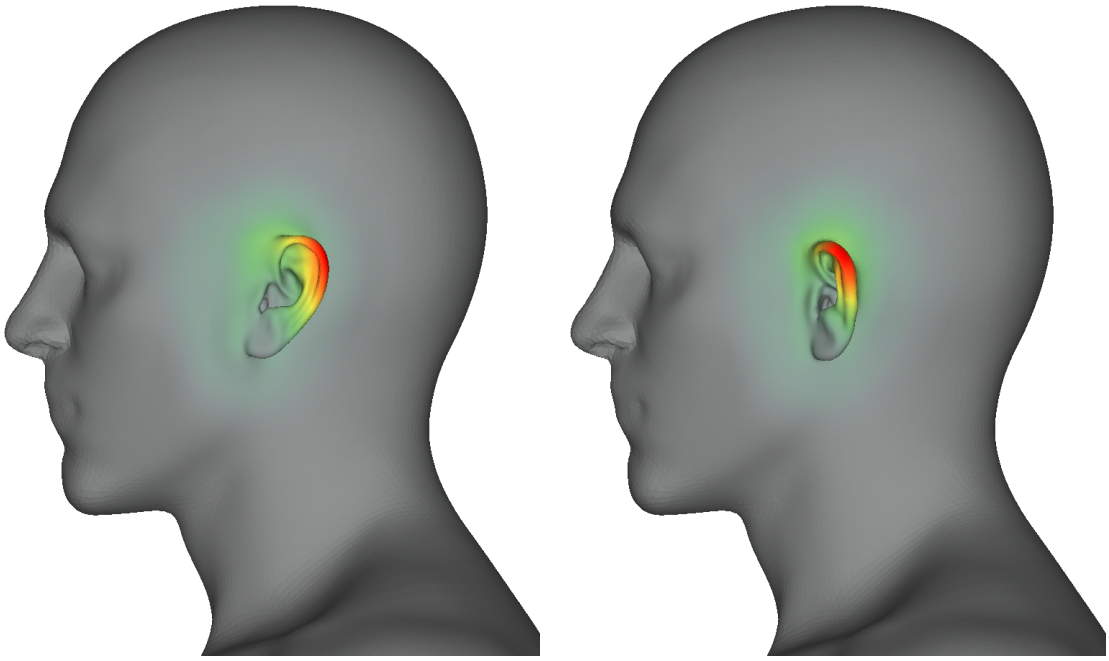


FIGURE C.2: *Effet de la* 1^e *composante - base synthétique*

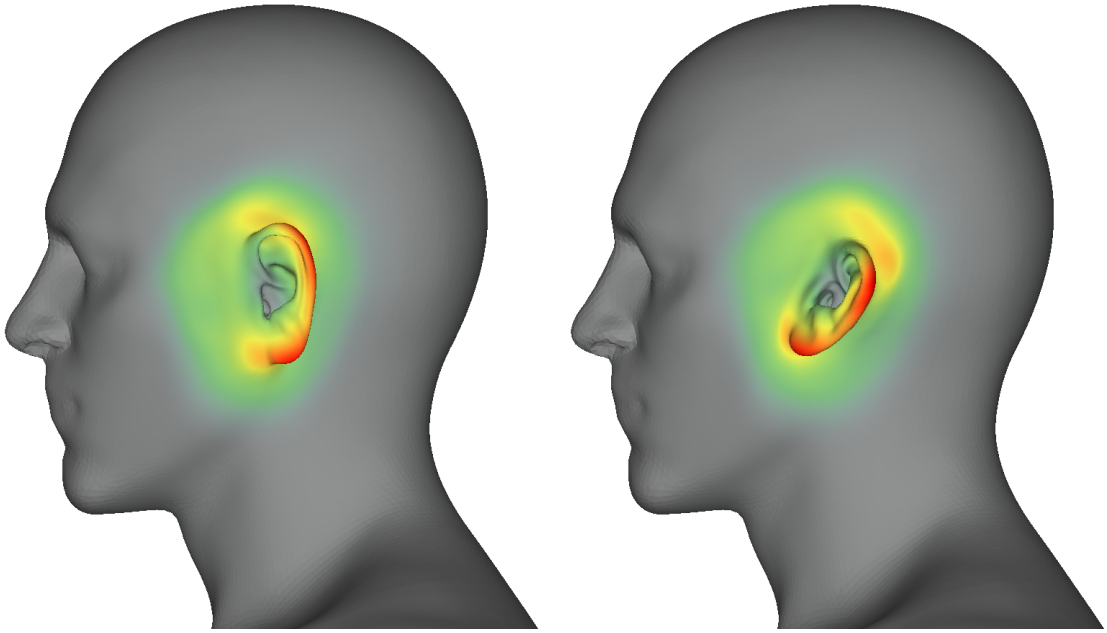


FIGURE C.3: *Effet de la* 2^e *composante - base synthétique*

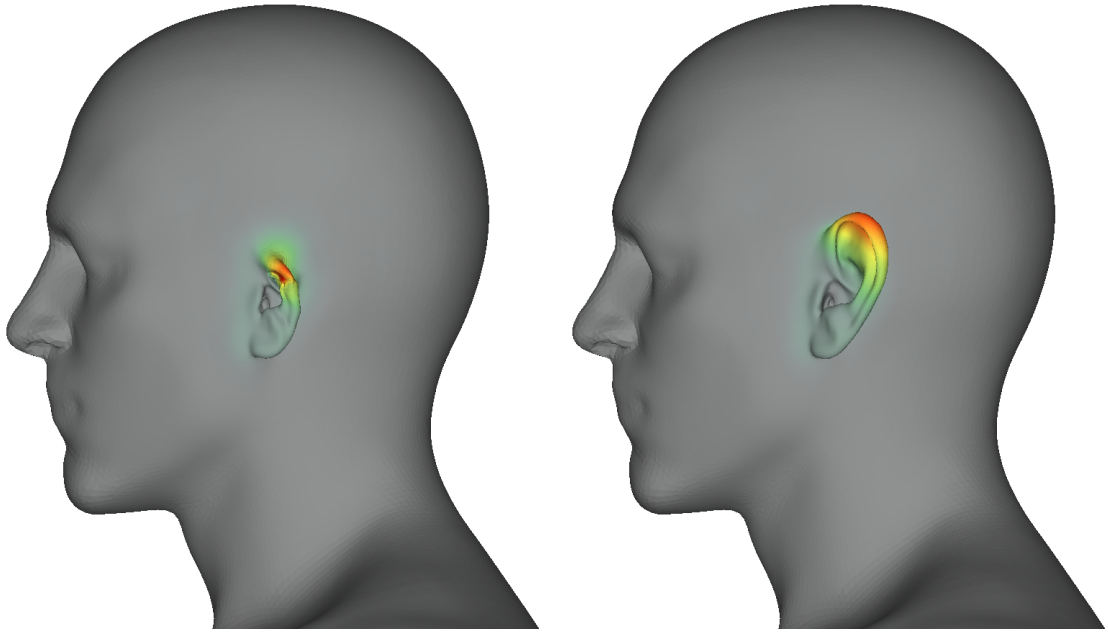


FIGURE C.4: *Effet de la* 3^e *composante - base synthétique*

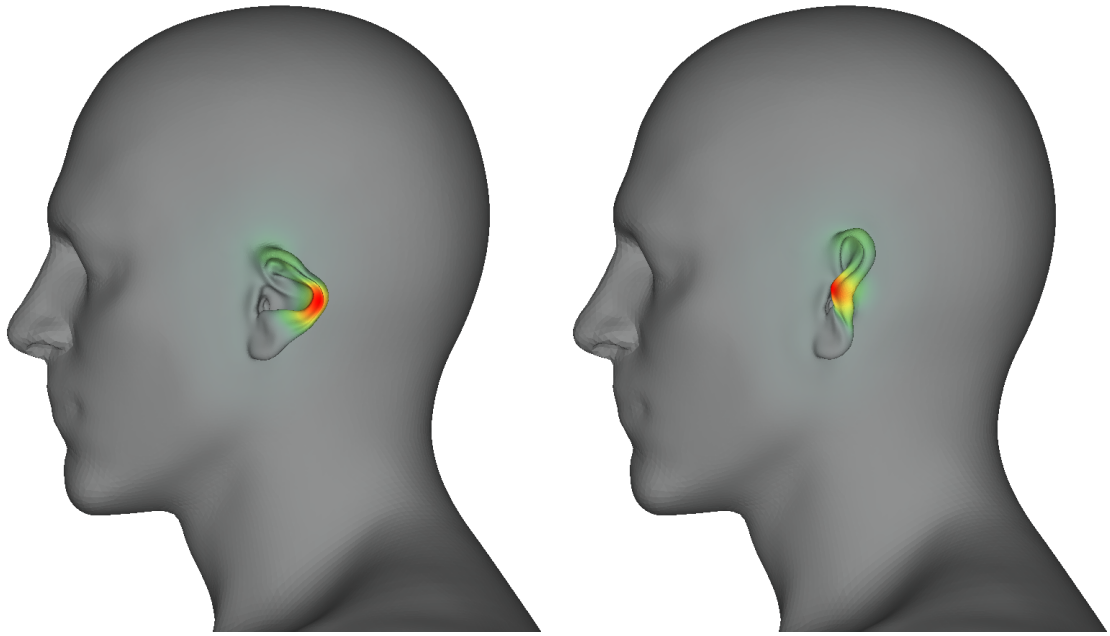


FIGURE C.5: *Effet de la* 4^e *composante - base synthétique*

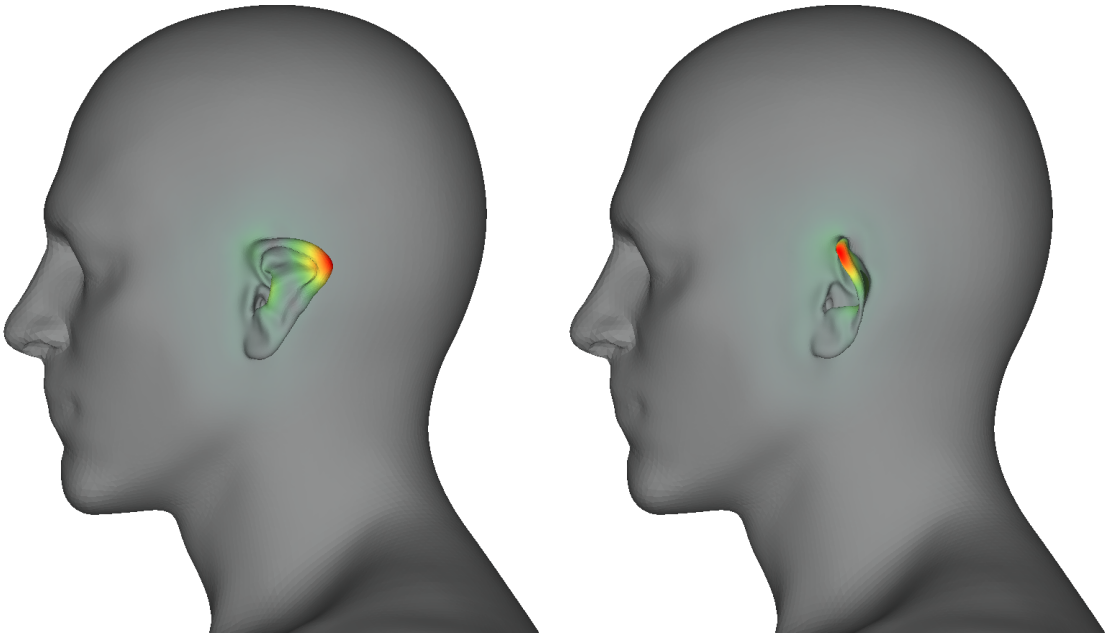


FIGURE C.6: *Effet de la* 5^e *composante - base synthétique*

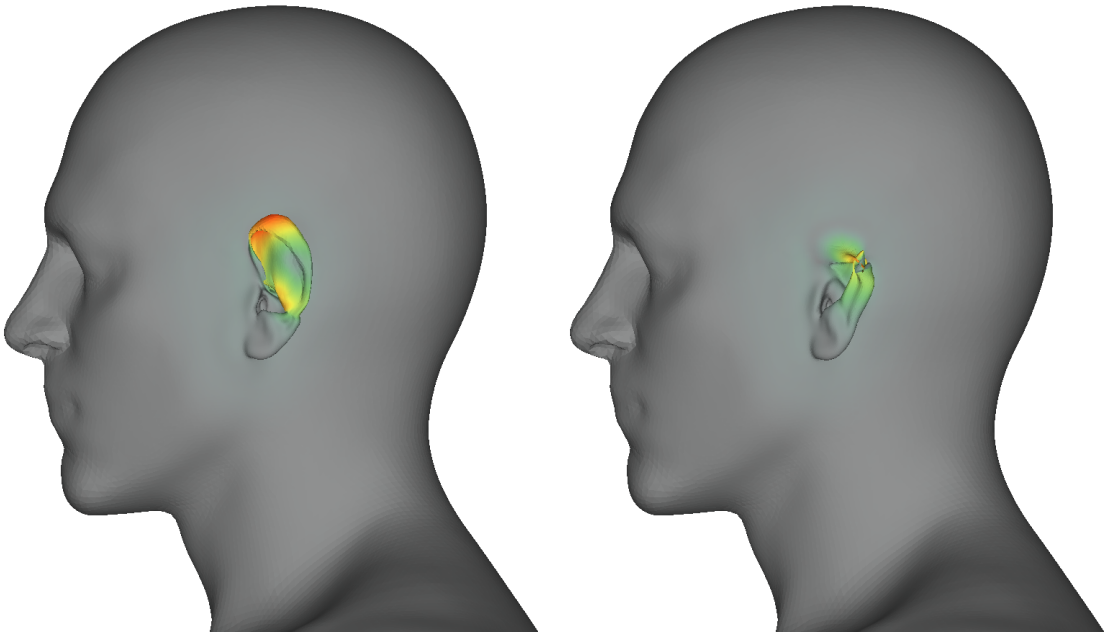


FIGURE C.7: *Effet de la* 6^e *composante - base synthétique*

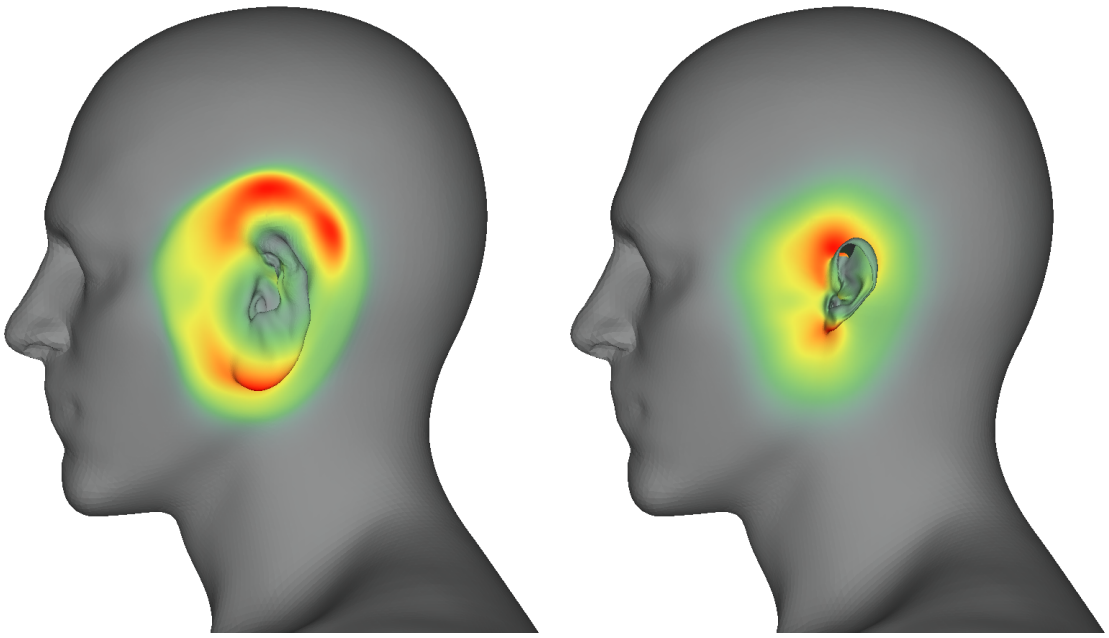


FIGURE C.8: *Effet de la 7^{e} composante - base synthétique*

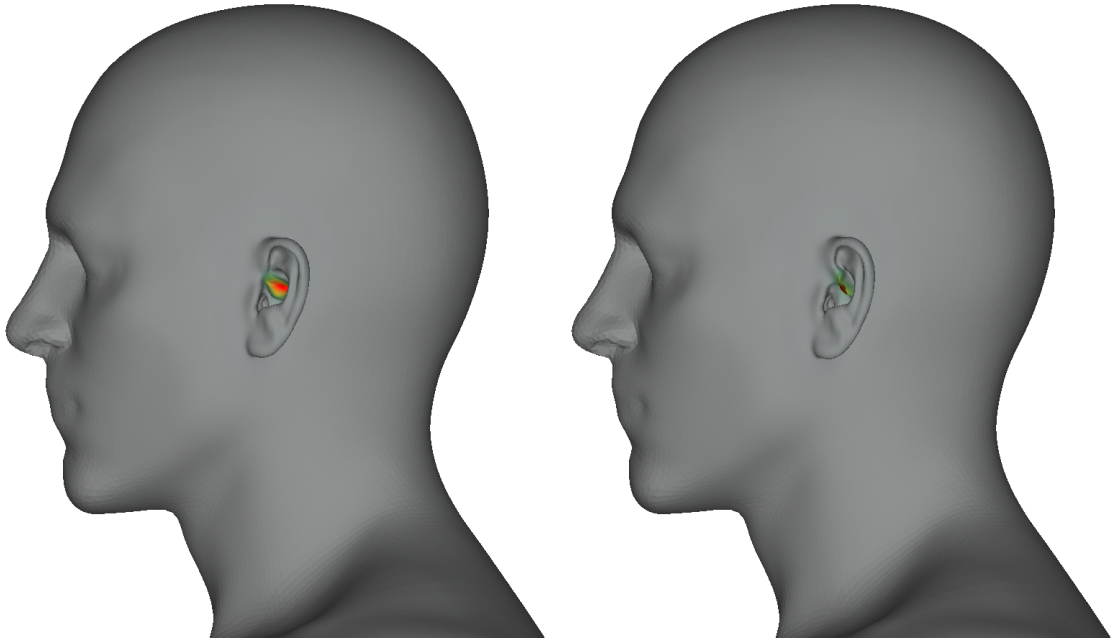


FIGURE C.9: *Effet de la 8^{e} composante - base synthétique*

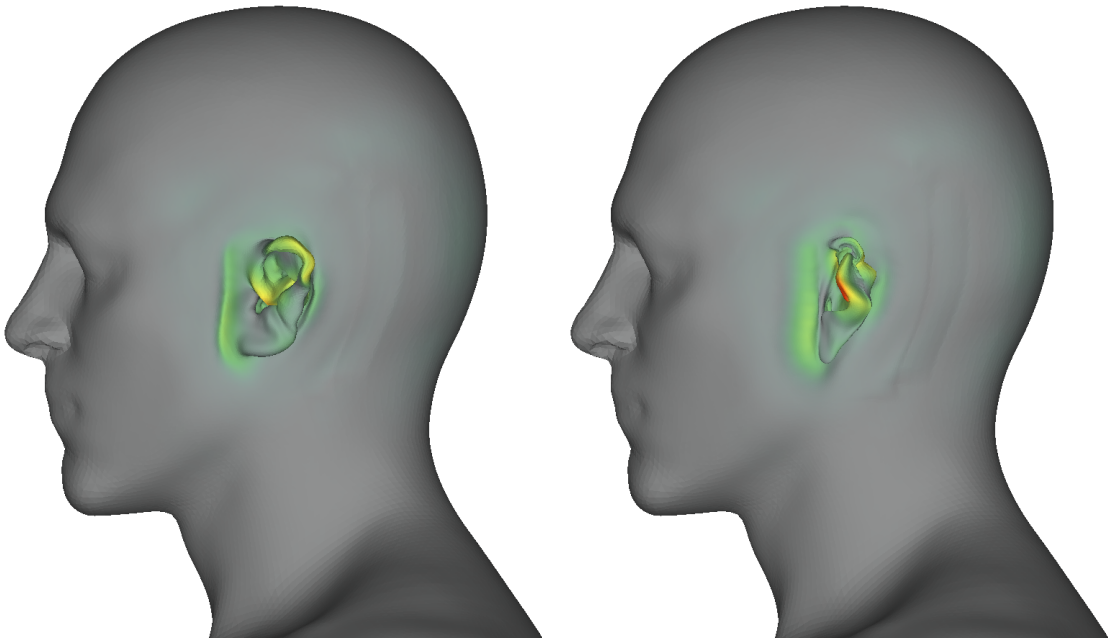


FIGURE C.10: *Effet de la* 9^e *composante - base synthétique*

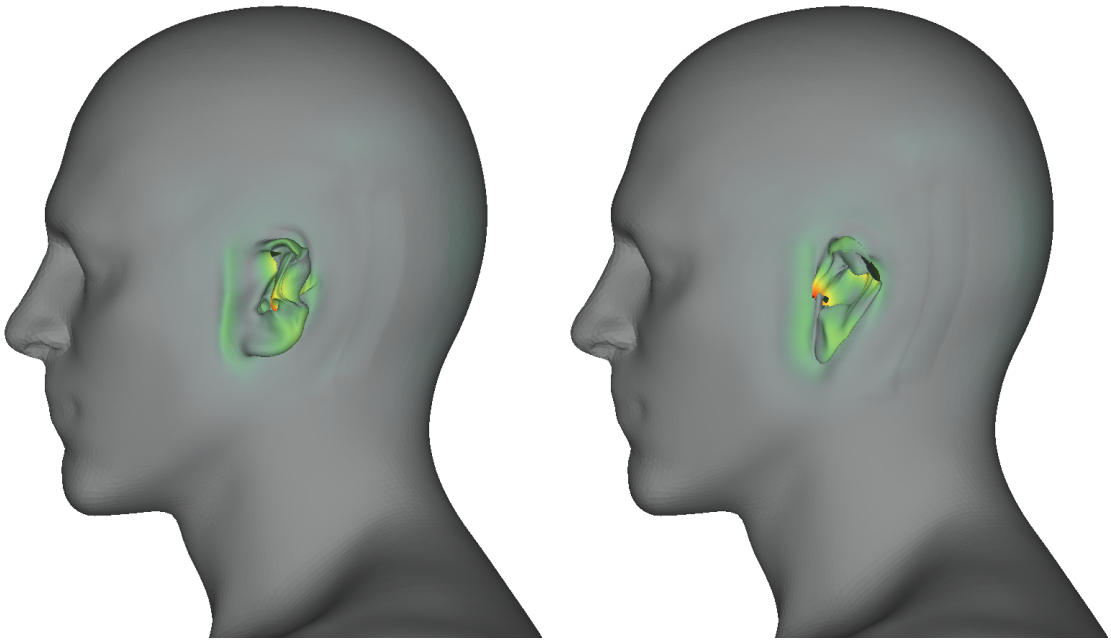


FIGURE C.11: *Effet de la* 10^e *composante - base synthétique*

C.1.2 Modèle binaural

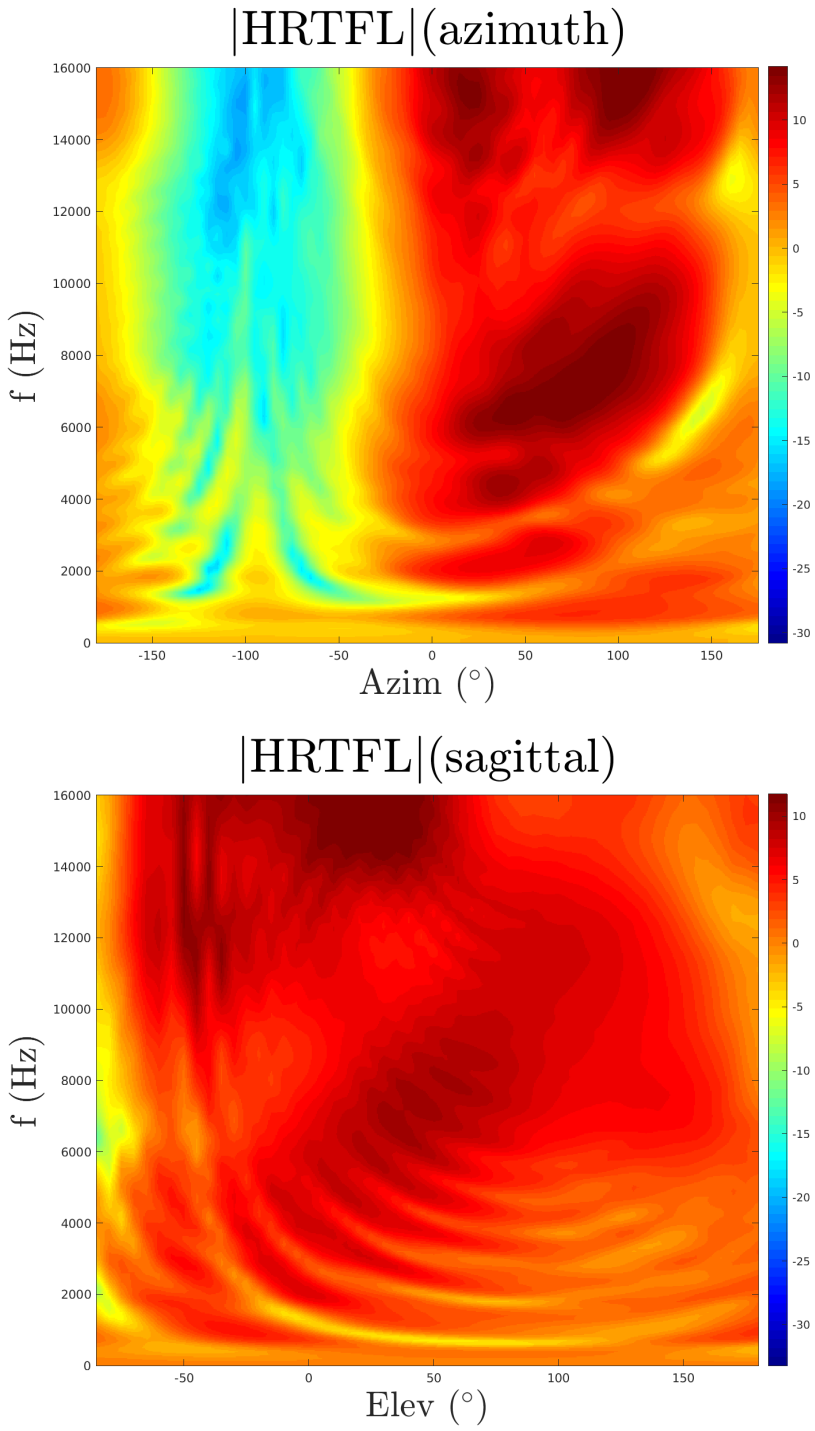


FIGURE C.12: *Coupes azimutale (en haut) et sagittale (en bas) de la HRTF moyenne du modèle issu de la base synthétique.*

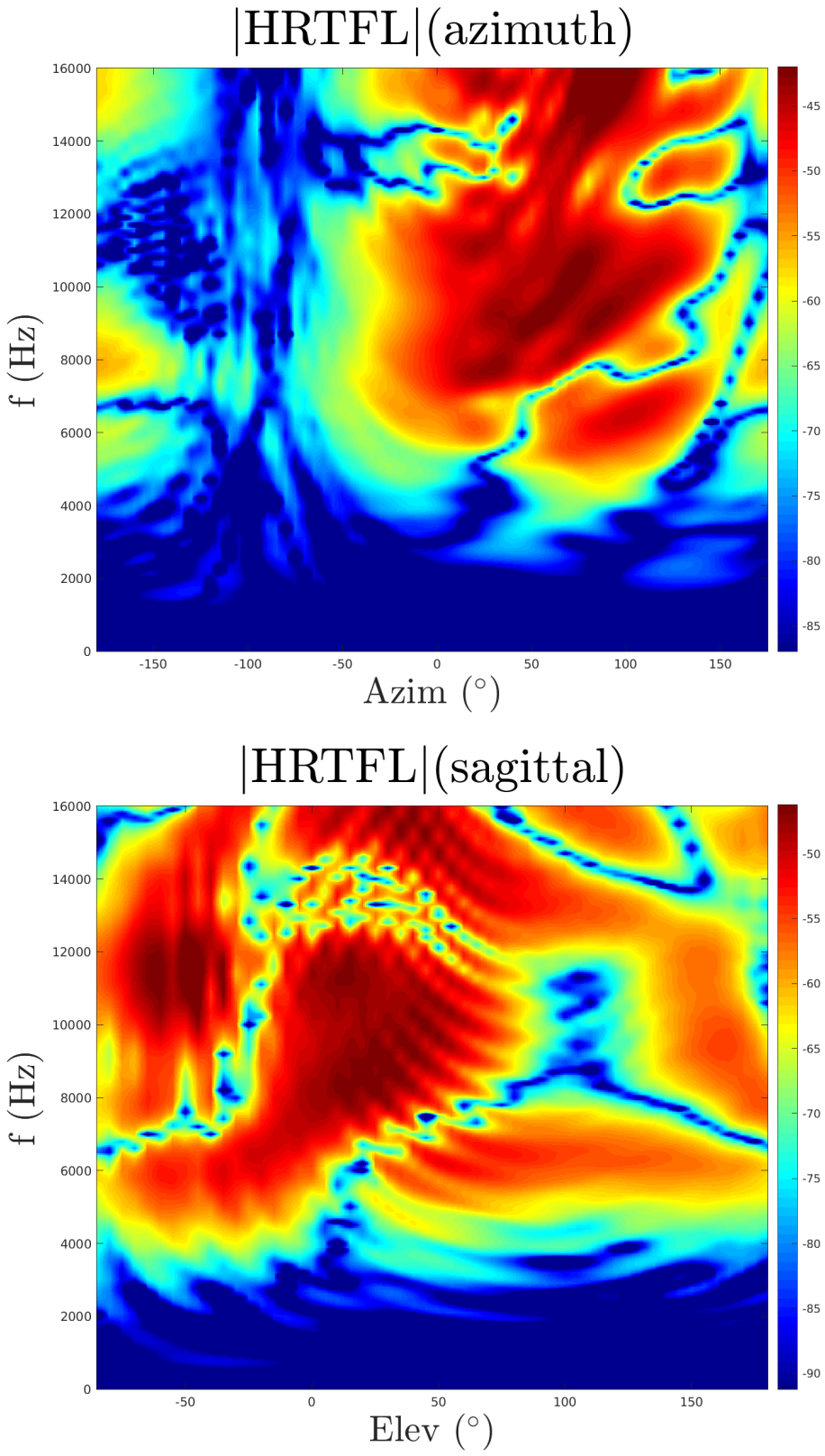


FIGURE C.13: *Effet de la* 1^e *composante - base synthétique*

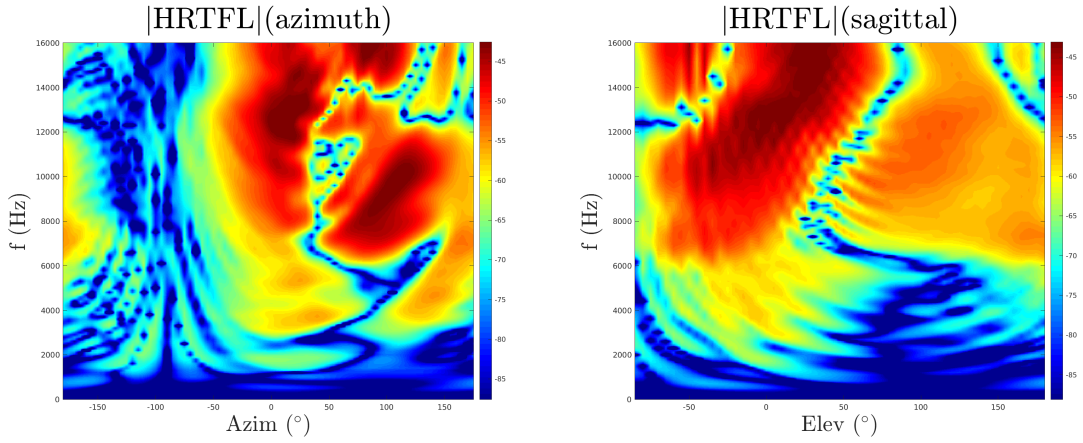


FIGURE C.14: *Effet de la 2^e composante - base synthétique*

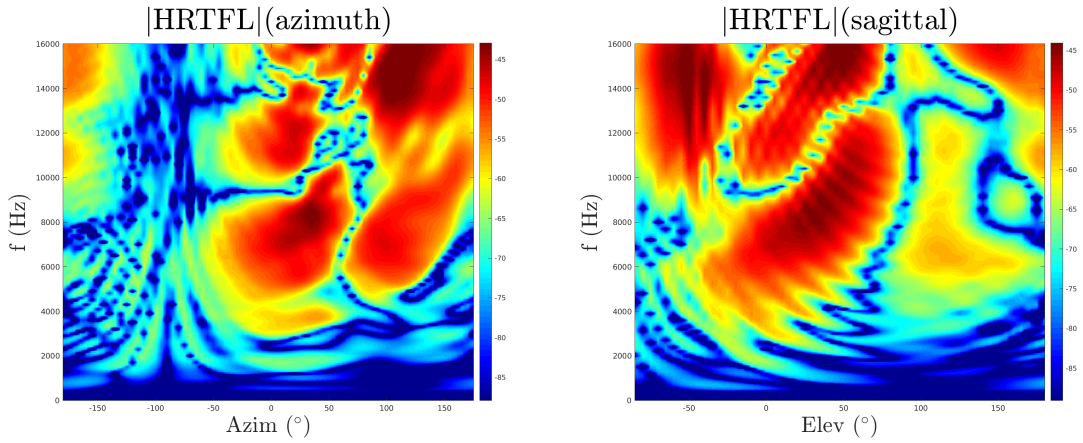


FIGURE C.15: *Effet de la 3^e composante - base synthétique*

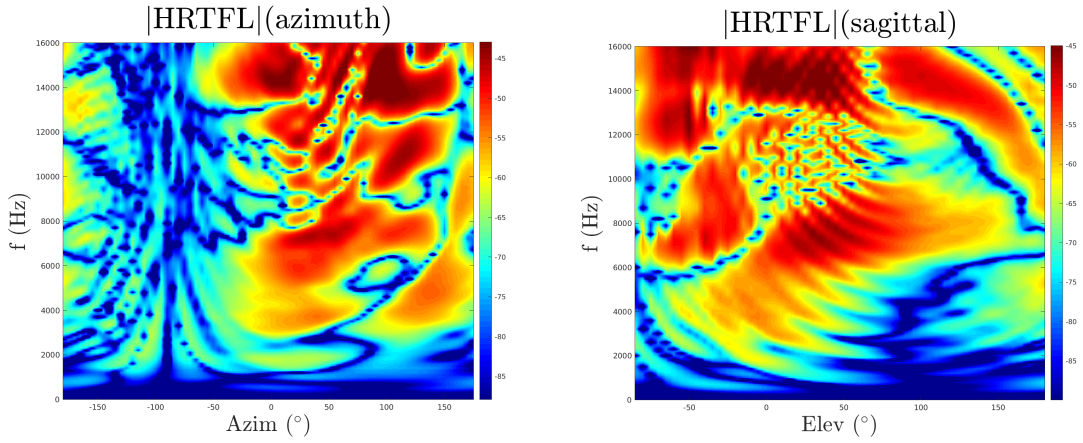


FIGURE C.16: *Effet de la 4^e composante - base synthétique*

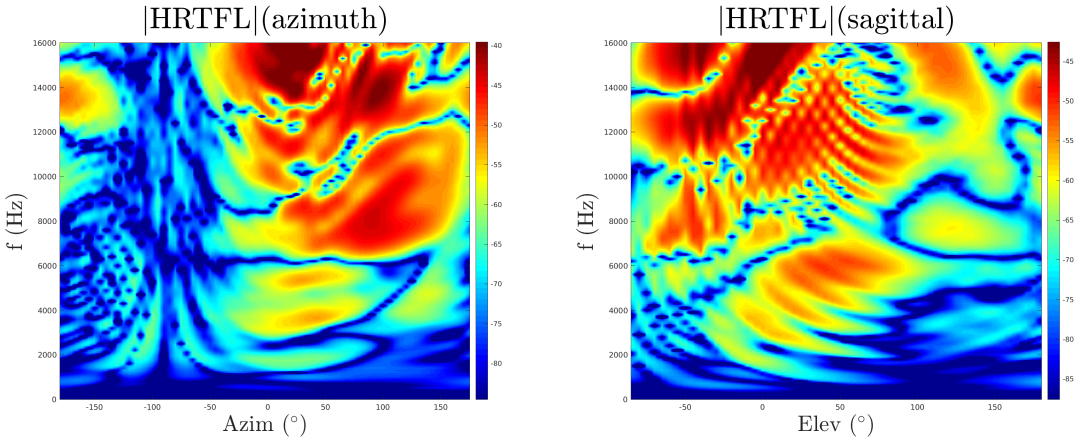


FIGURE C.17: *Effet de la 5^e composante - base synthétique*

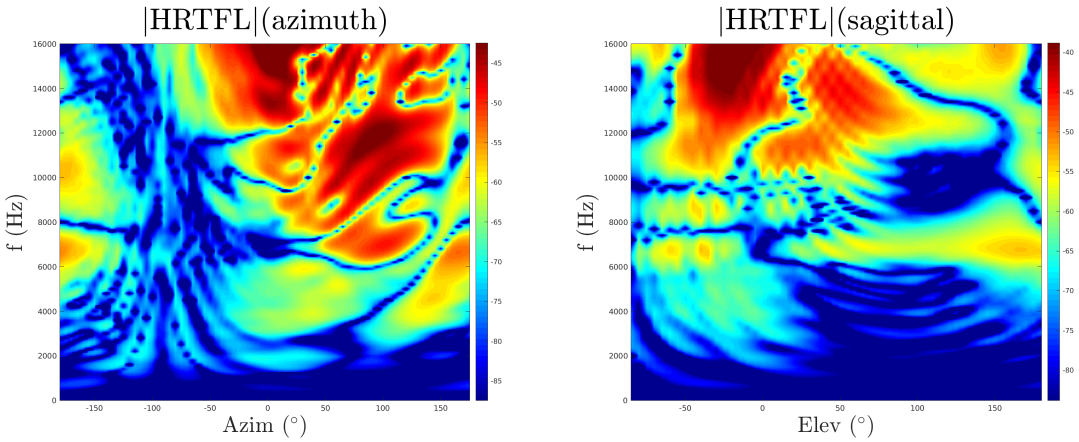


FIGURE C.18: *Effet de la 6^e composante - base synthétique*

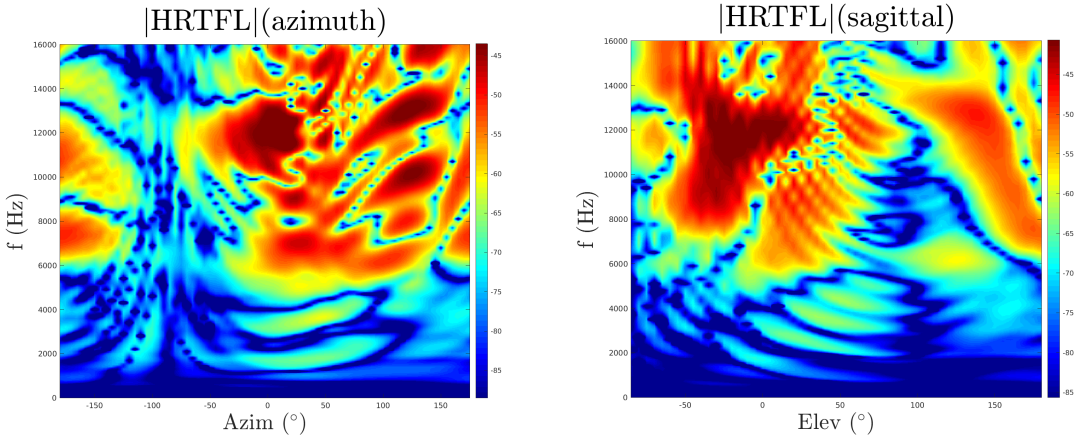


FIGURE C.19: *Effet de la 7^e composante - base synthétique*

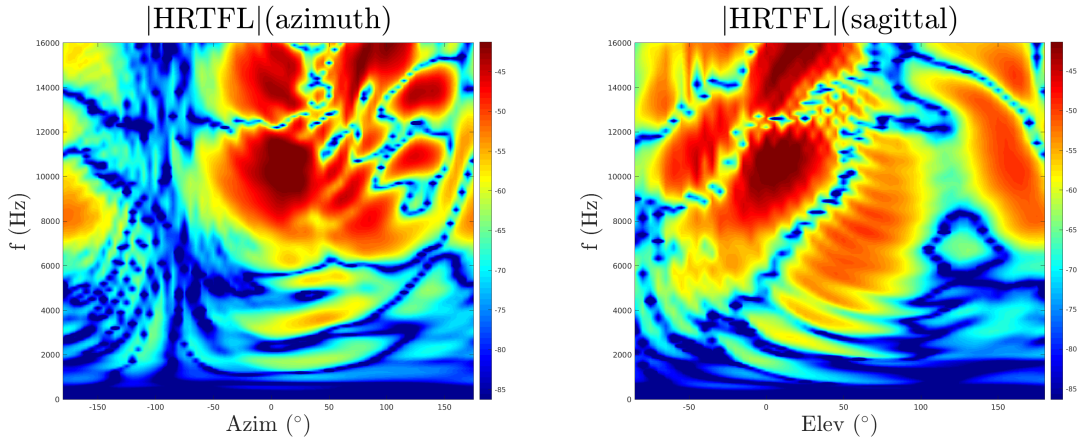


FIGURE C.20: *Effet de la 8^e composante - base synthétique*

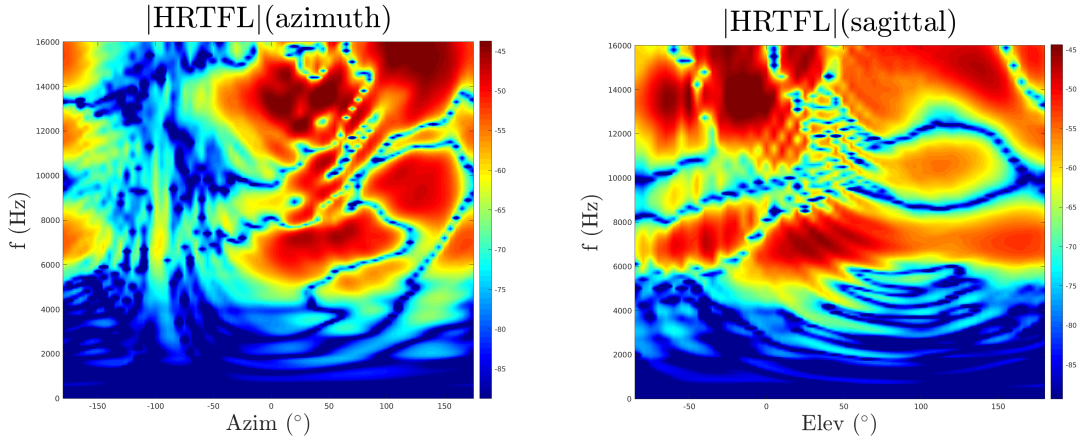


FIGURE C.21: *Effet de la 9^e composante - base synthétique*

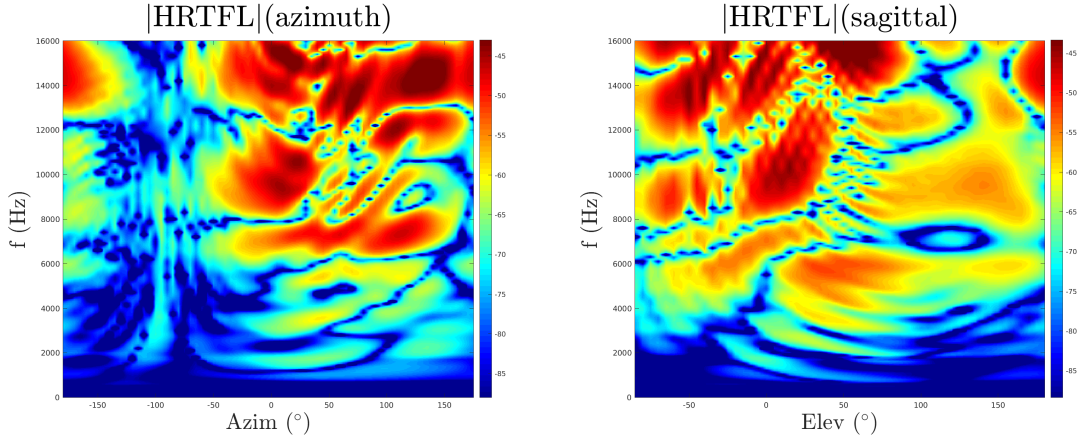


FIGURE C.22: *Effet de la 10^e composante - base synthétique*

C.2 Modèles de la base aléatoire

C.2.1 Modèle morphologique

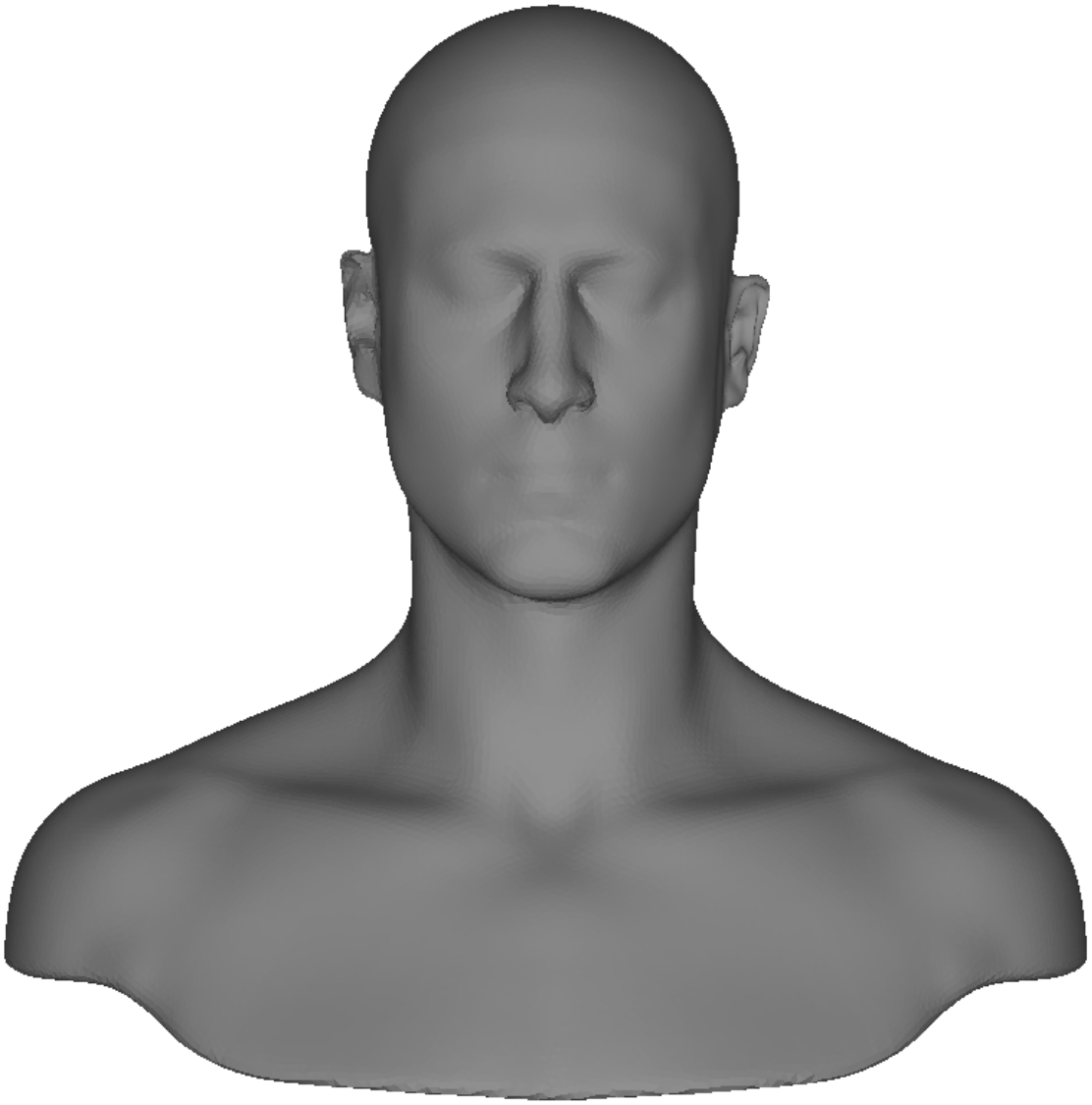


FIGURE C.23: *Forme moyenne du modèle issu de la base aléatoire.*

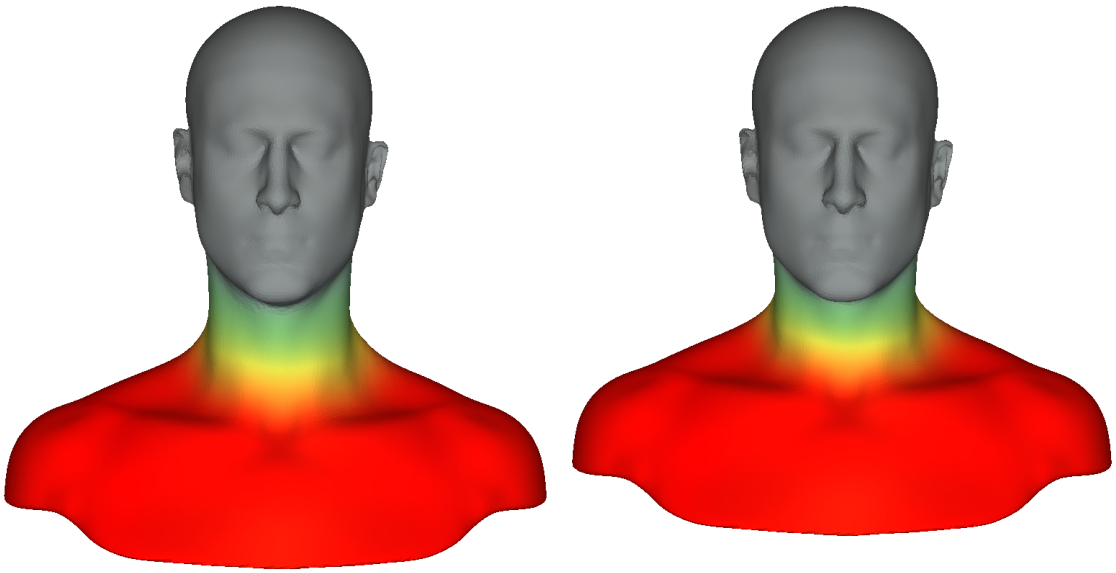


FIGURE C.24: *Effet de la* 1^e *composante - base aléatoire*

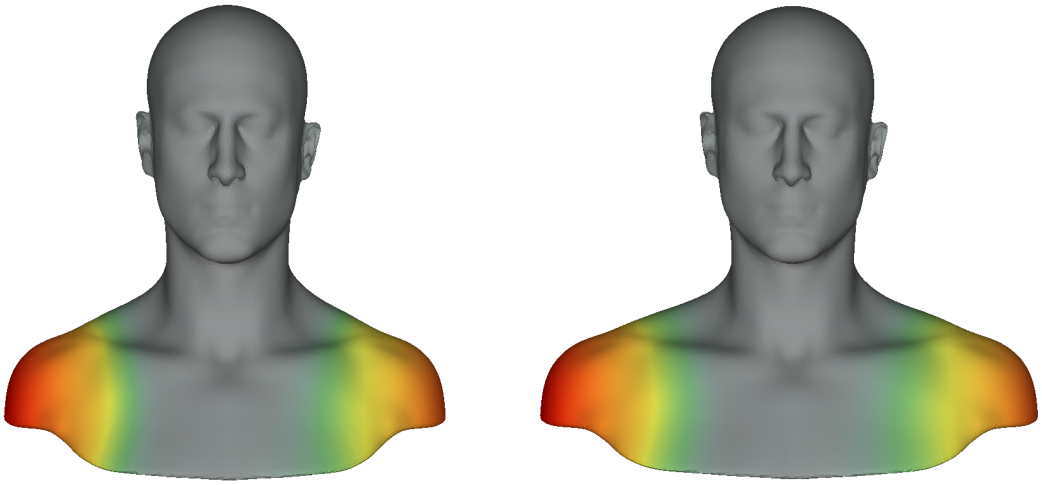


FIGURE C.25: *Effet de la* 2^e *composante - base aléatoire*

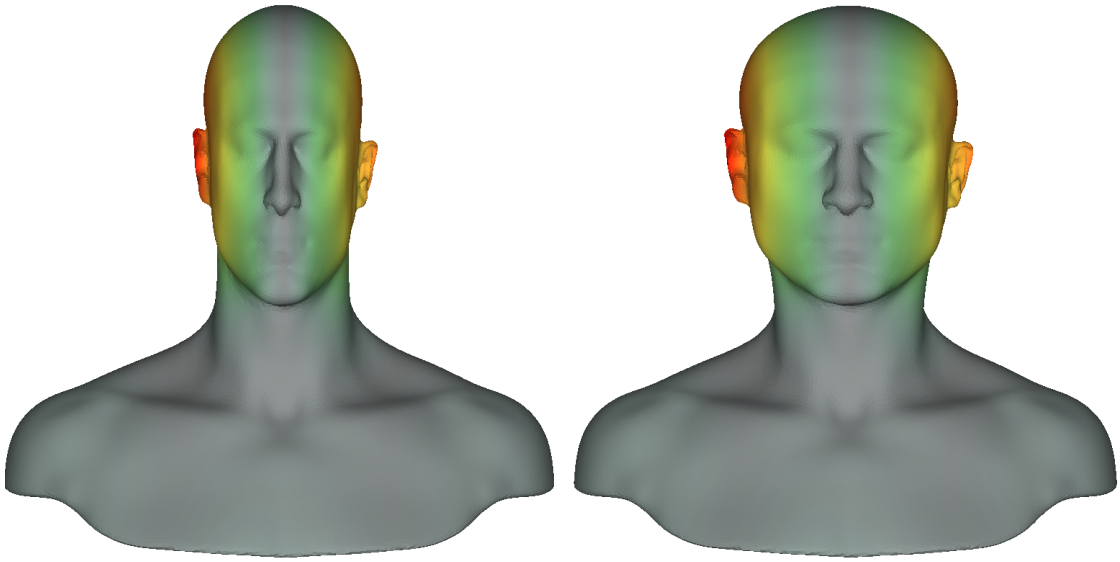


FIGURE C.26: *Effet de la* 3^e *composante - base aléatoire*

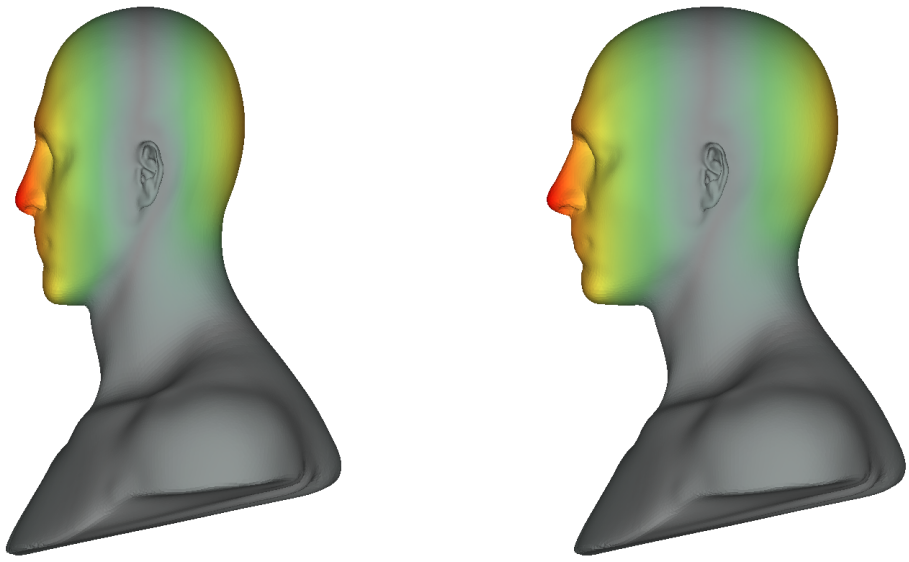


FIGURE C.27: *Effet de la* 4^e *composante - base aléatoire*

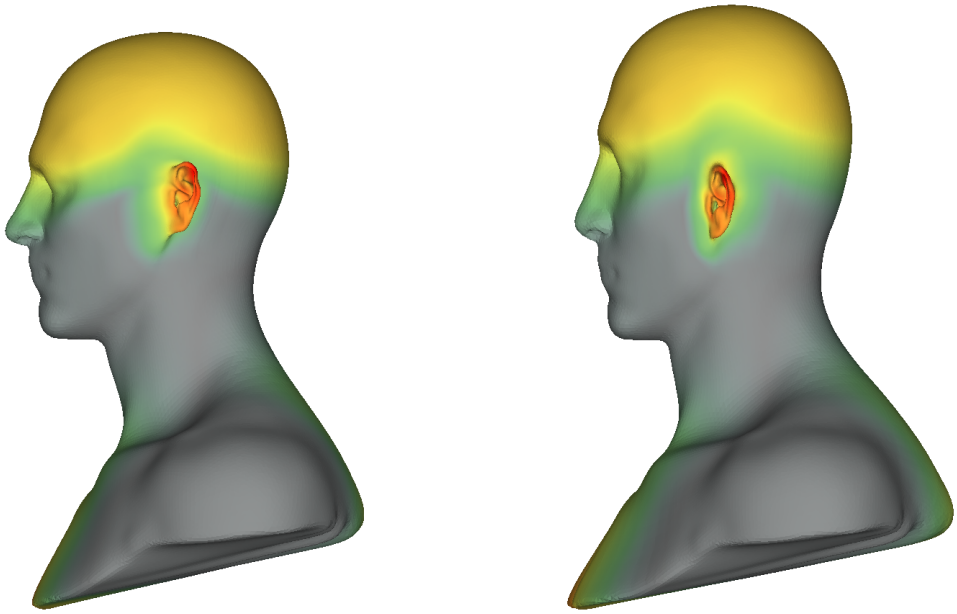


FIGURE C.28: *Effet de la* 5^e *composante - base aléatoire*

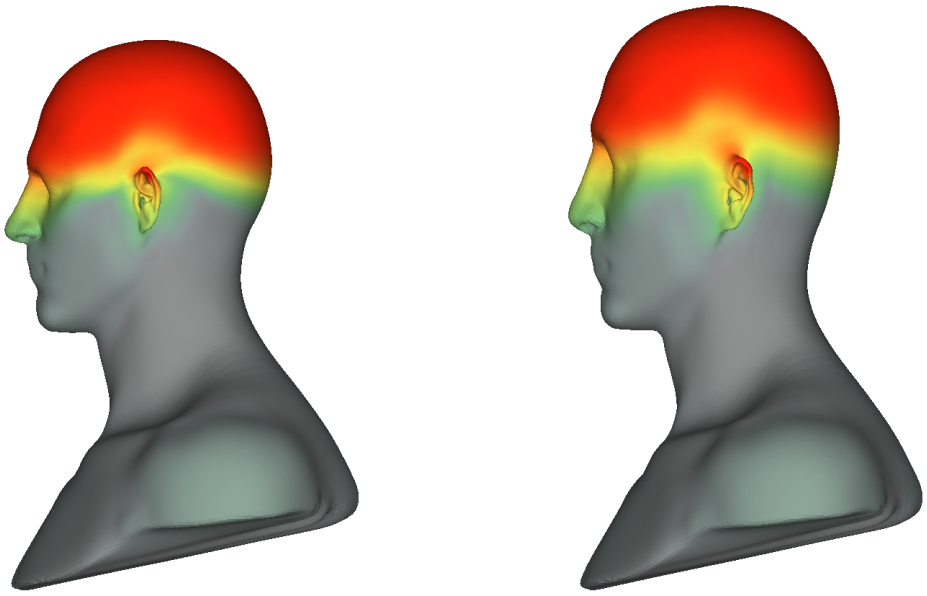


FIGURE C.29: *Effet de la* 6^e *composante - base aléatoire*

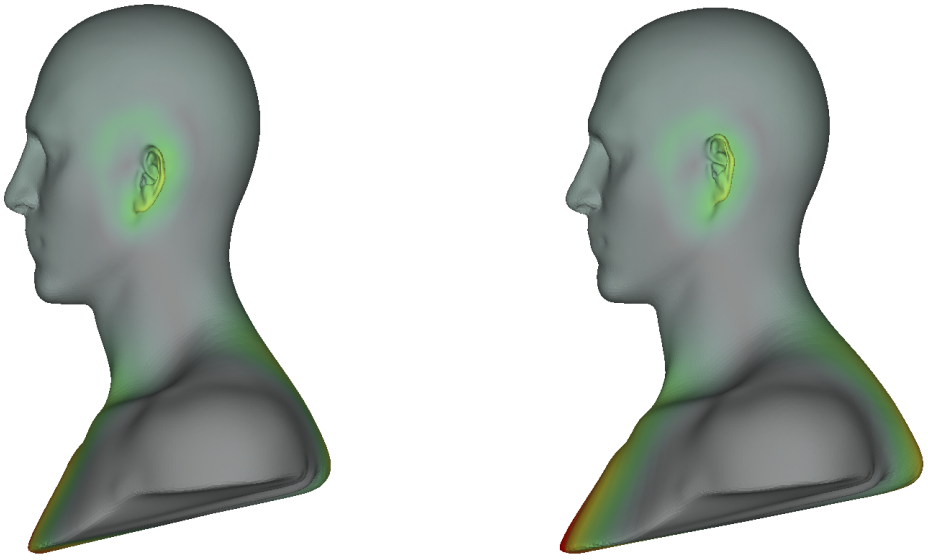


FIGURE C.30: *Effet de la* 7^e *composante - base aléatoire*

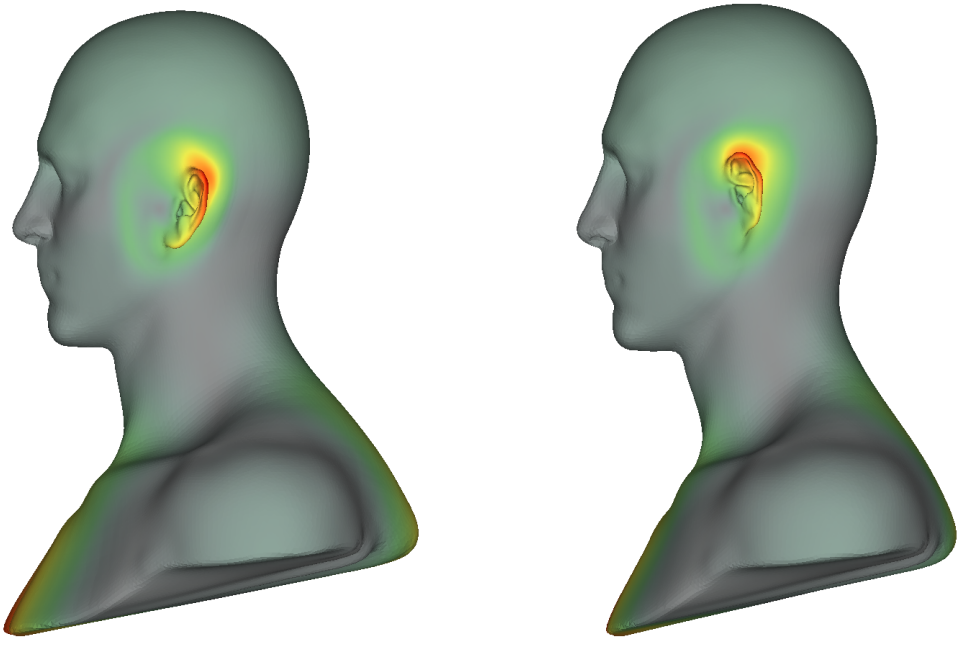


FIGURE C.31: *Effet de la* 8^e *composante - base aléatoire*

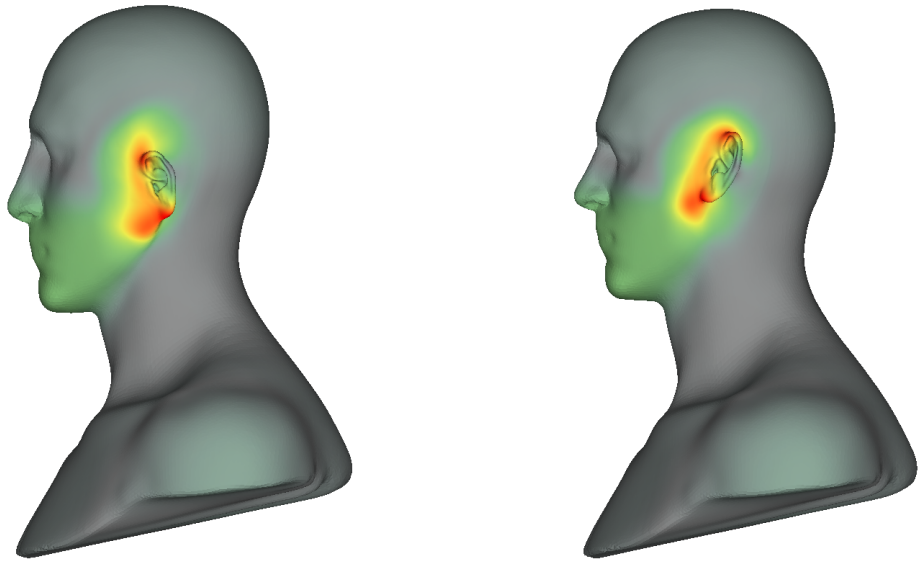


FIGURE C.32: *Effet de la 9^e composante - base aléatoire*

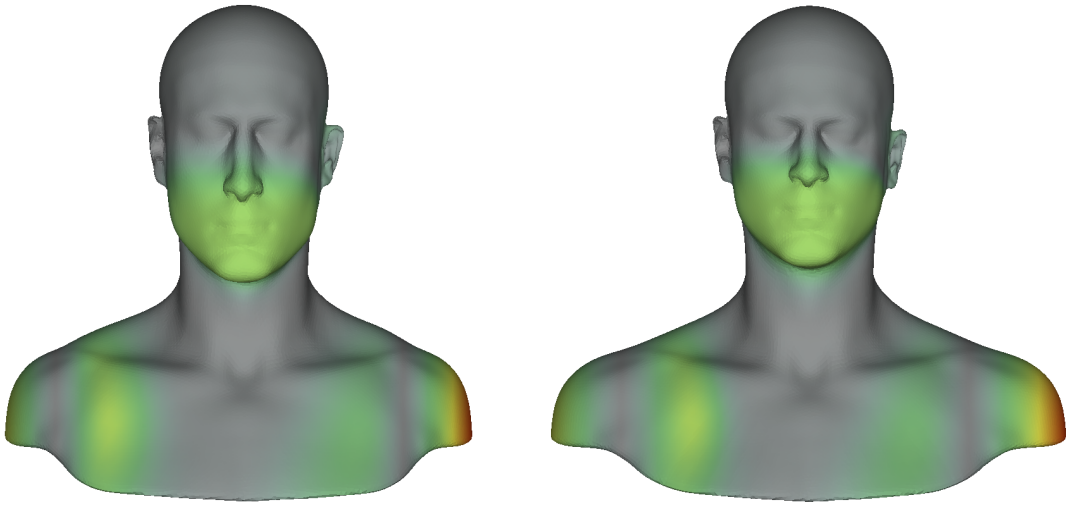


FIGURE C.33: *Effet de la 10^e composante - base aléatoire*

C.2.2 Modèle binaural

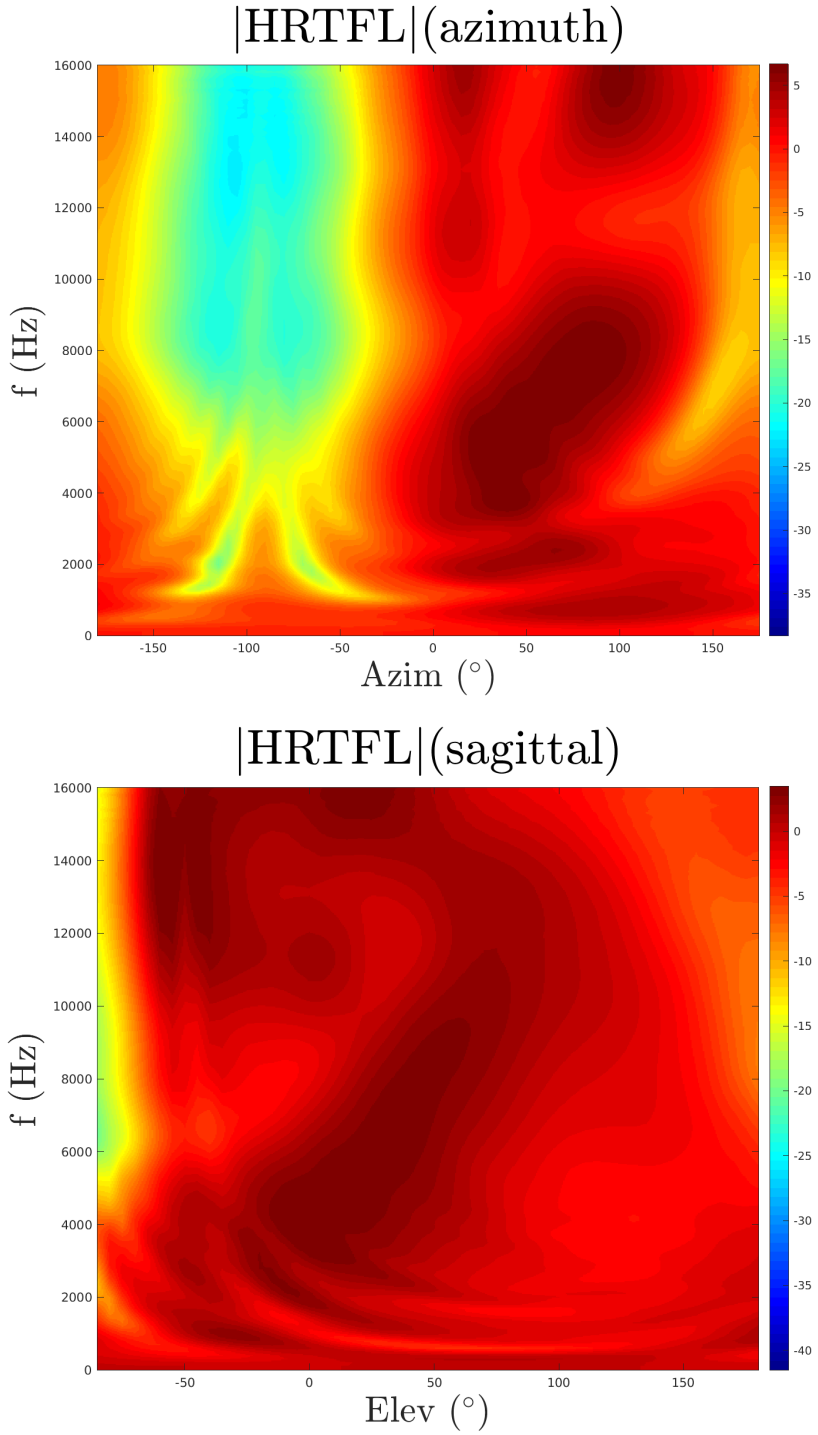


FIGURE C.34: *Coupes azimutale (en haut) et sagittale (en bas) de la HRTF moyenne du modèle issu de la base aléatoire.*

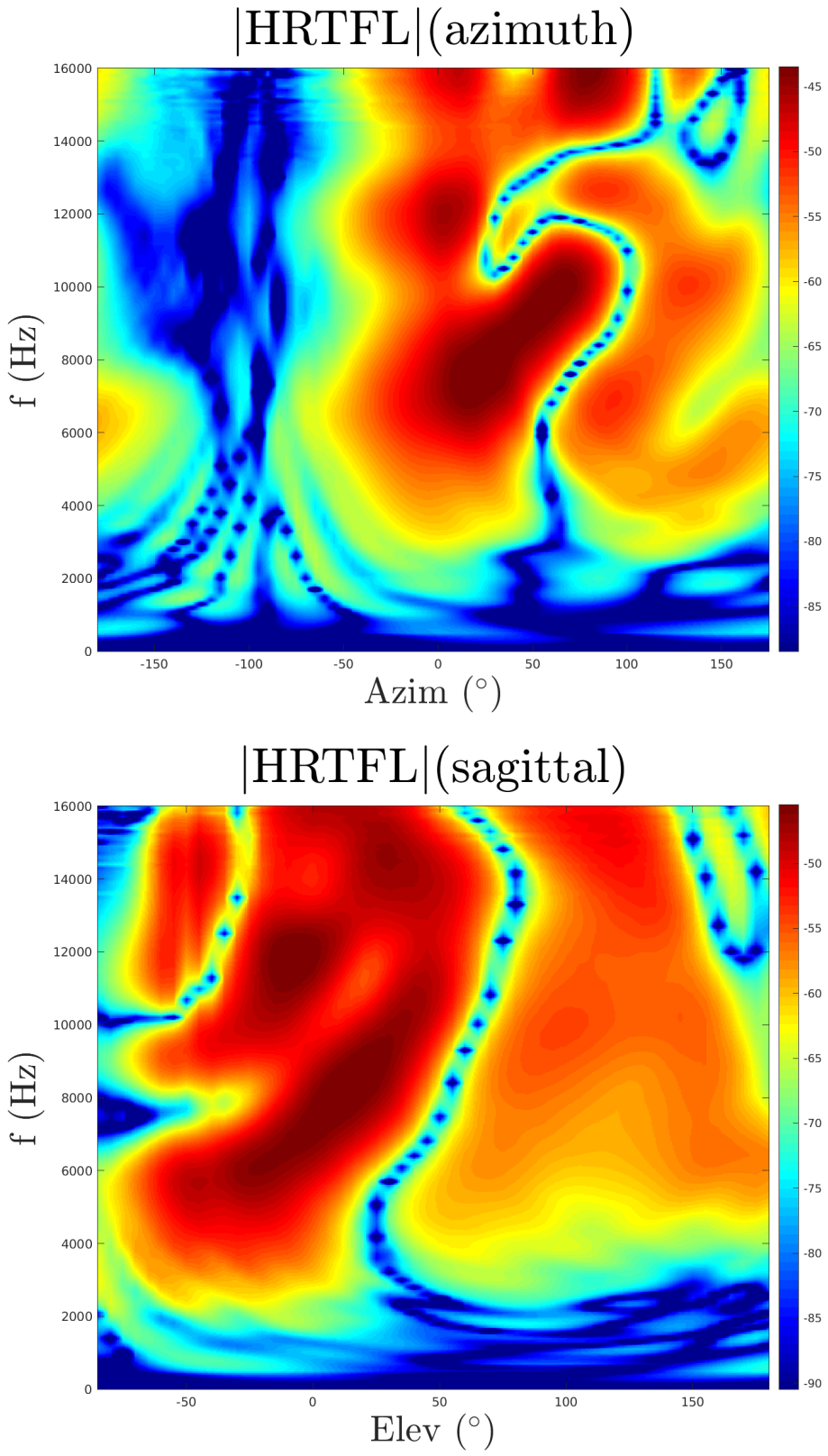


FIGURE C.35: *Effet de la 1^e composante - base aléatoire*

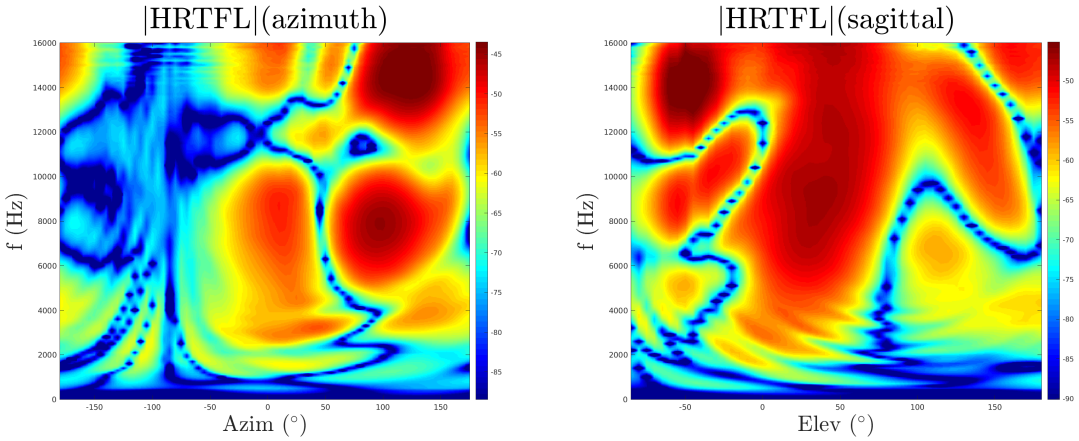


FIGURE C.36: *Effet de la 2^e composante - base aléatoire*

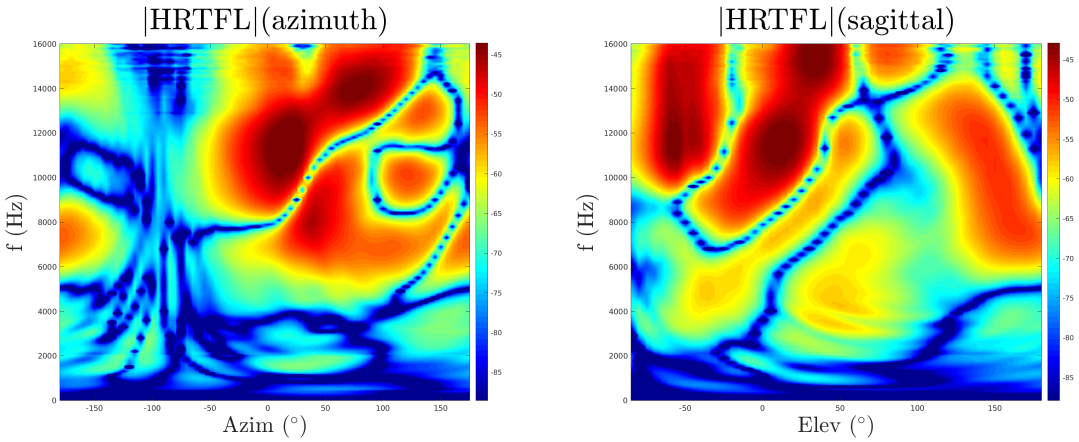


FIGURE C.37: *Effet de la 3^e composante - base aléatoire*

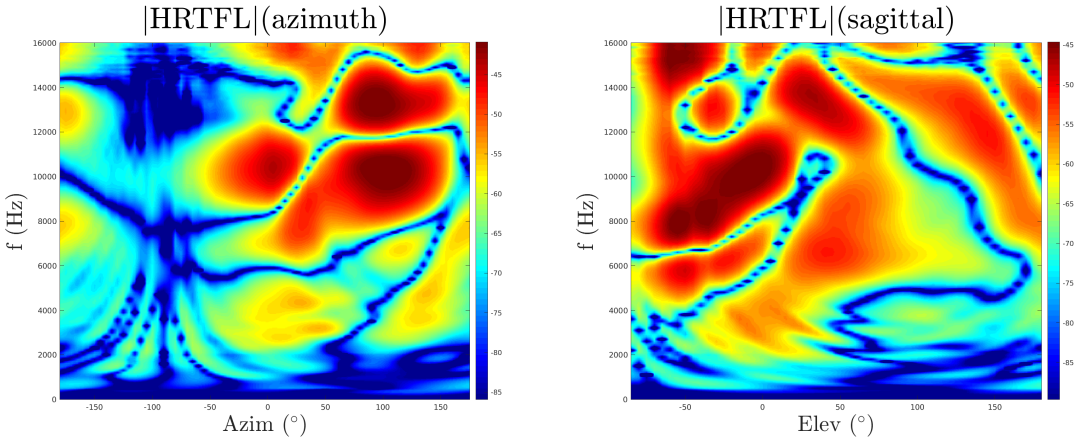


FIGURE C.38: *Effet de la 4^e composante - base aléatoire*

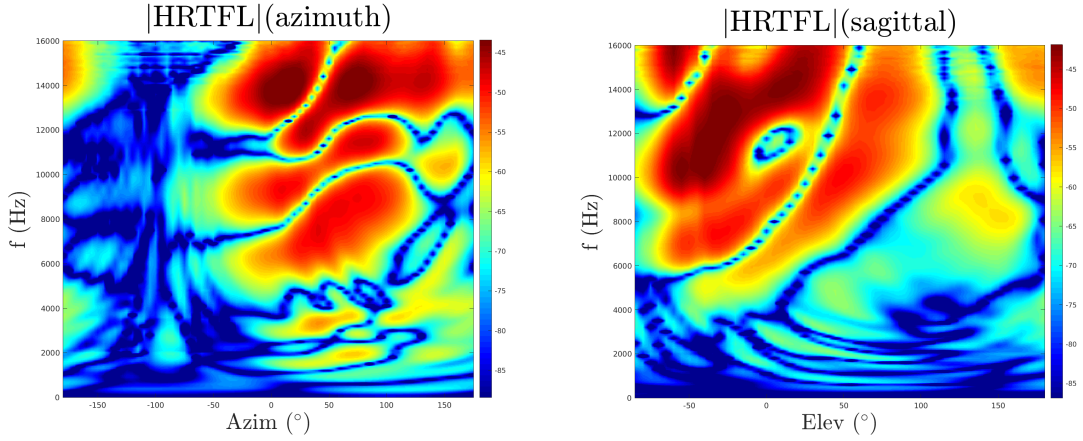


FIGURE C.39: *Effet de la 5^e composante - base aléatoire*

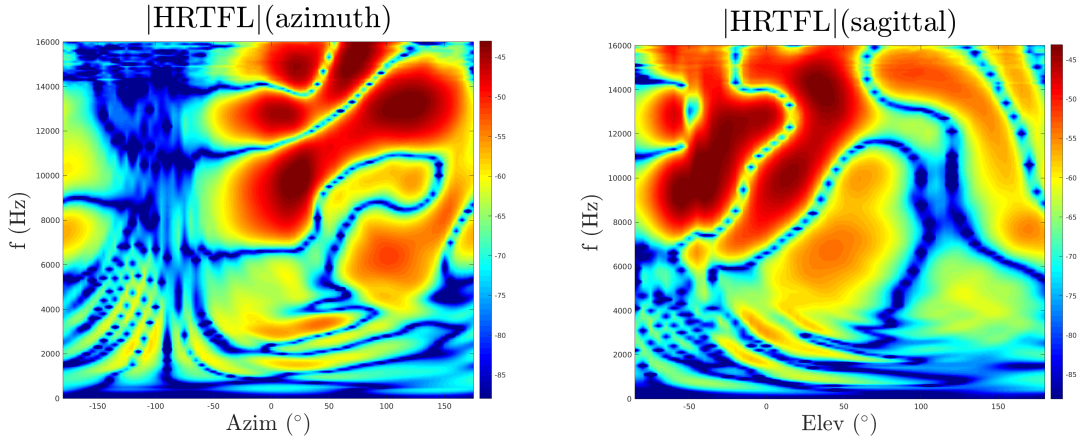


FIGURE C.40: *Effet de la 6^e composante - base aléatoire*

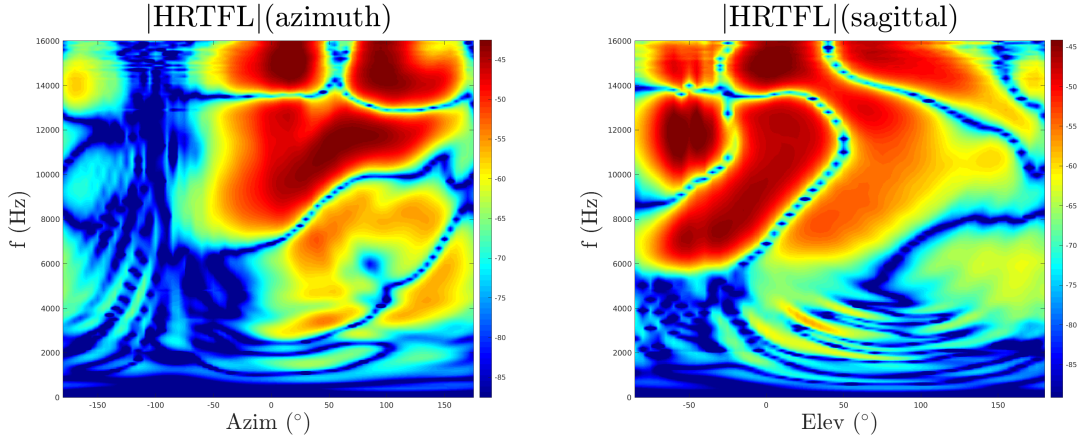


FIGURE C.41: *Effet de la 7^e composante - base aléatoire*

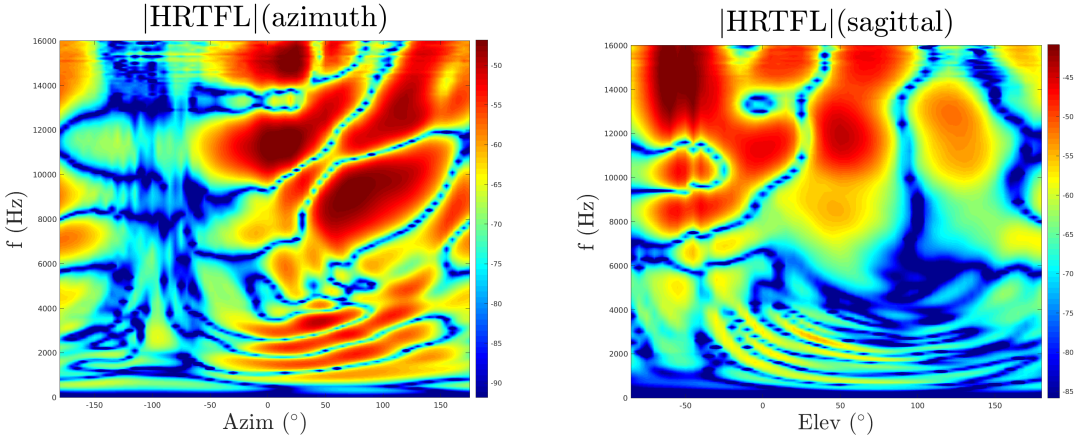


FIGURE C.42: *Effet de la* 8^e *composante - base aléatoire*

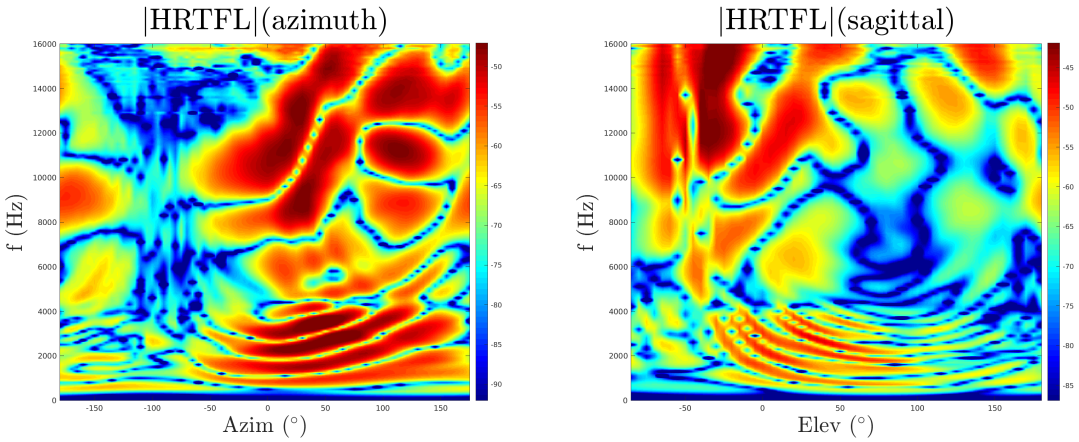


FIGURE C.43: *Effet de la* 9^e *composante - base aléatoire*

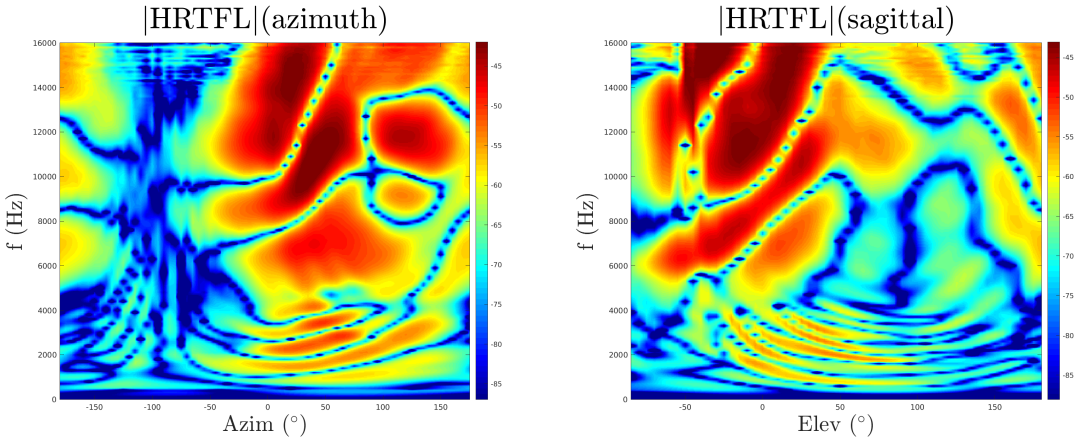


FIGURE C.44: *Effet de la* 10^e *composante - base aléatoire*

C.3 Modèles de la base mixte

C.3.1 Modèle morphologique

FIGURE C.45: *Forme moyenne du modèle issu de la base mixte.*

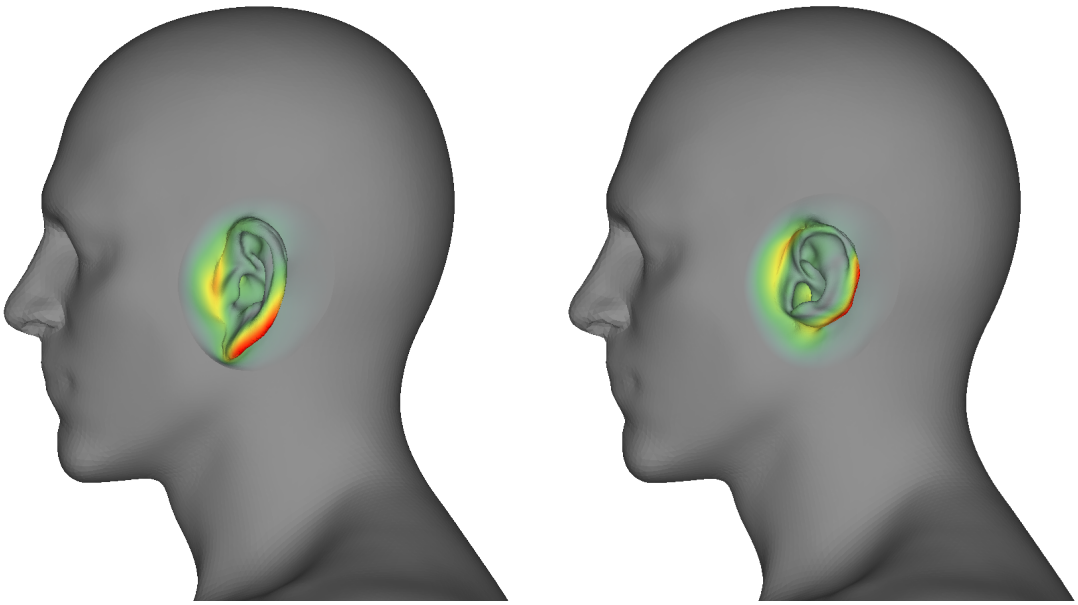


FIGURE C.46: *Effet de la* 1^e *composante - base mixte*

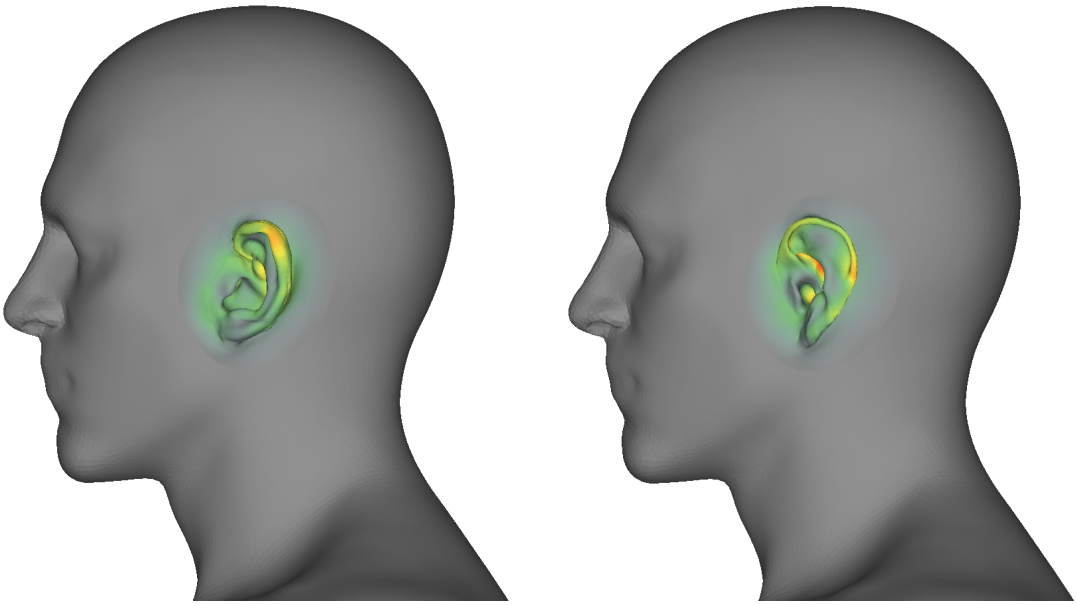


FIGURE C.47: *Effet de la* 2^e *composante - base mixte*

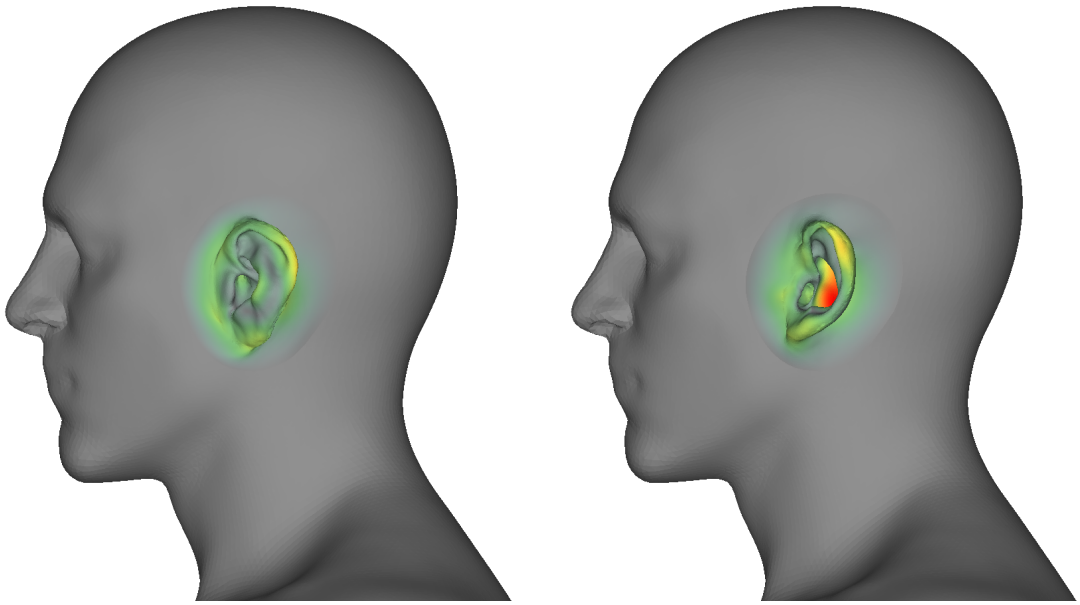


FIGURE C.48: *Effet de la* 3^e *composante - base mixte*

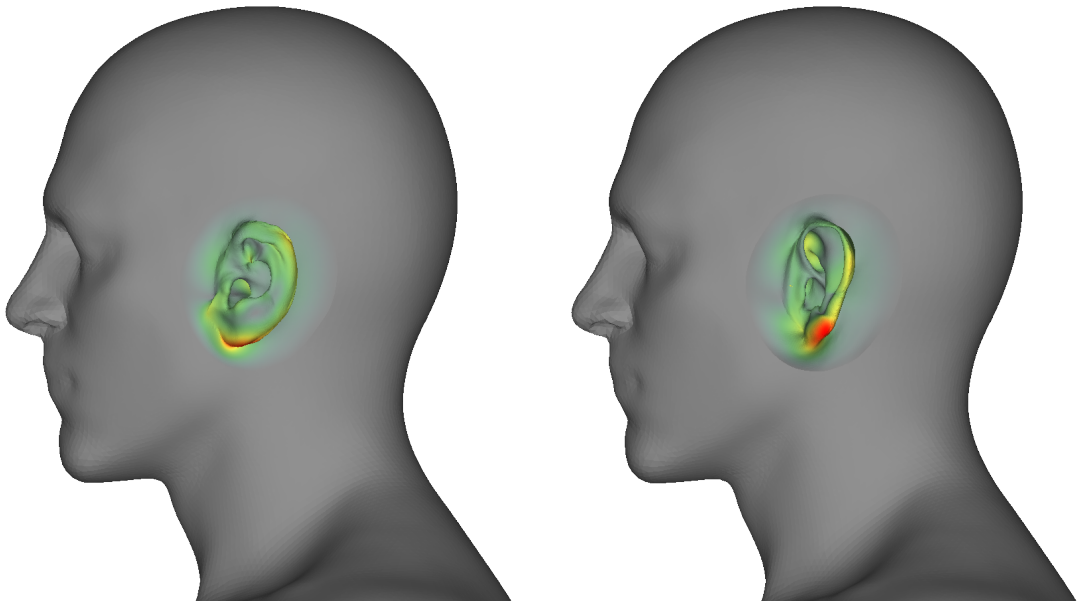


FIGURE C.49: *Effet de la* 4^e *composante - base mixte*

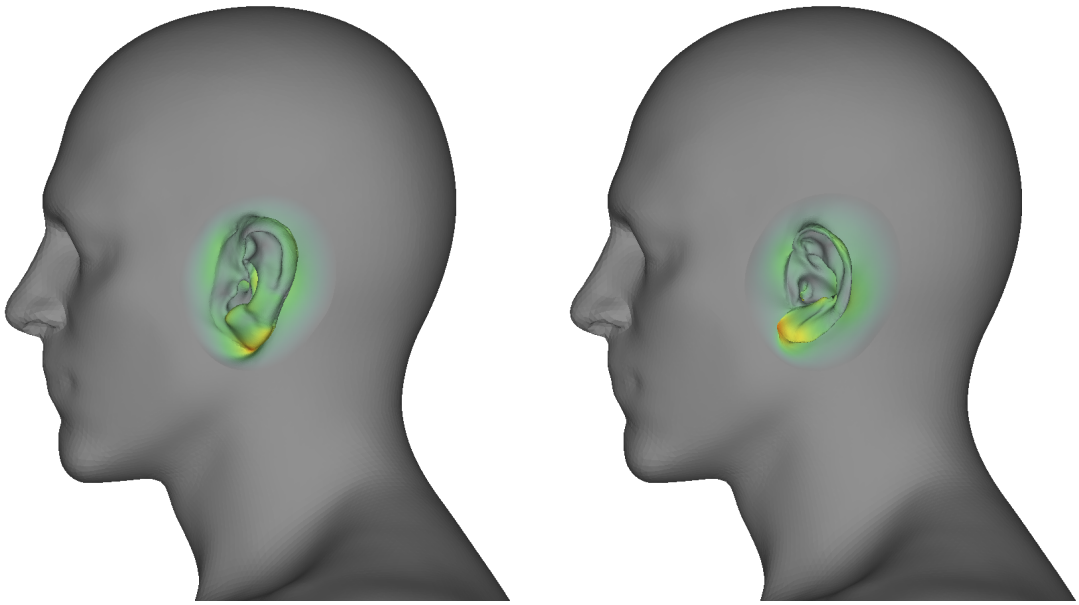


FIGURE C.50: *Effet de la* 5^e *composante - base mixte*

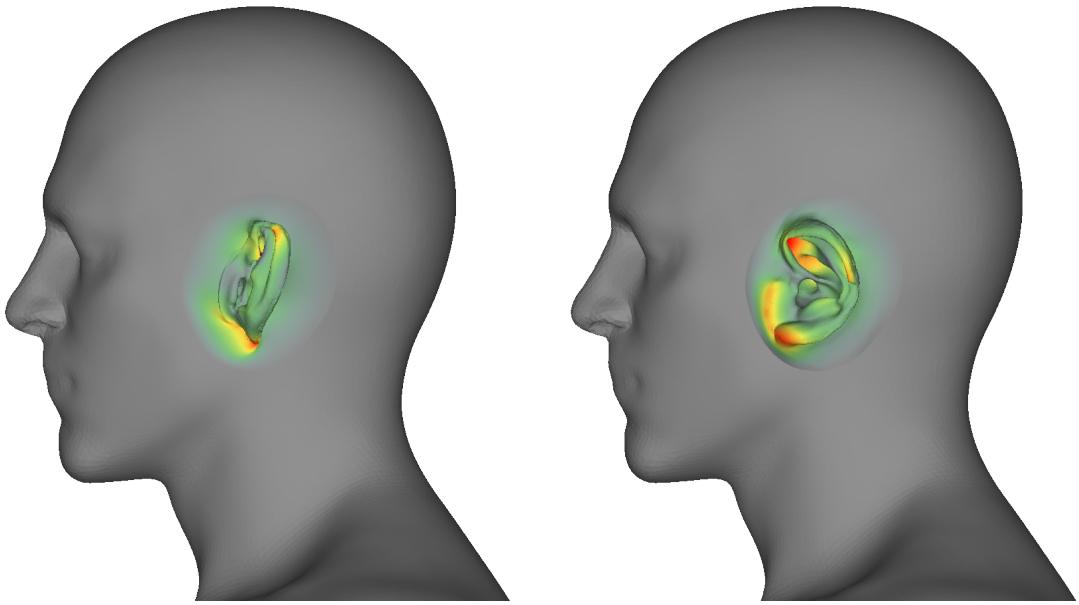


FIGURE C.51: *Effet de la* 6^e *composante - base mixte*

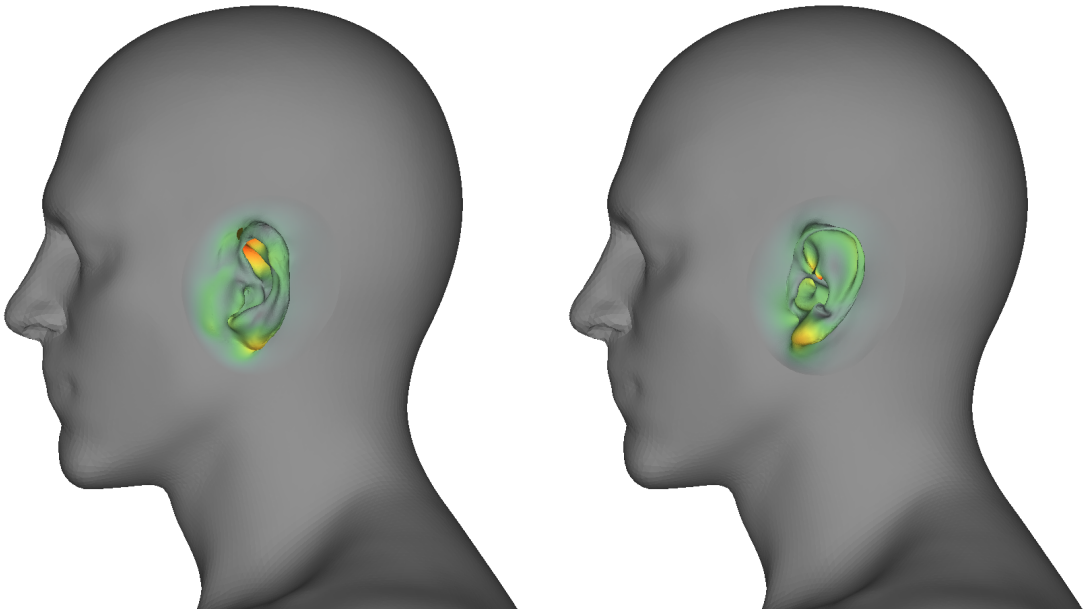


FIGURE C.52: *Effet de la 7^e composante - base mixte*

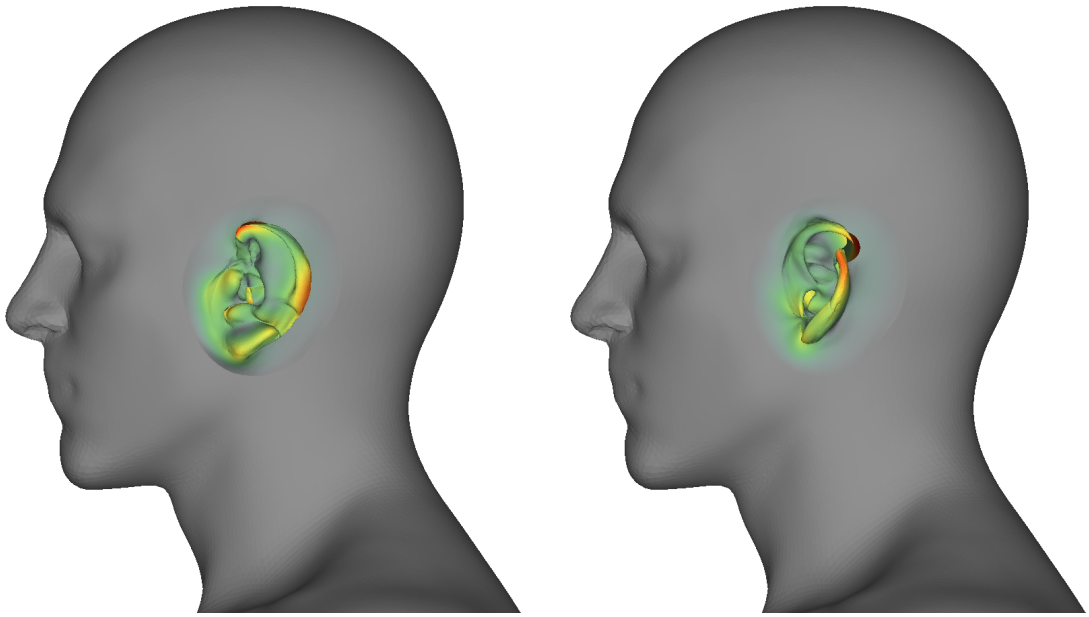


FIGURE C.53: *Effet de la 8^e composante - base mixte*

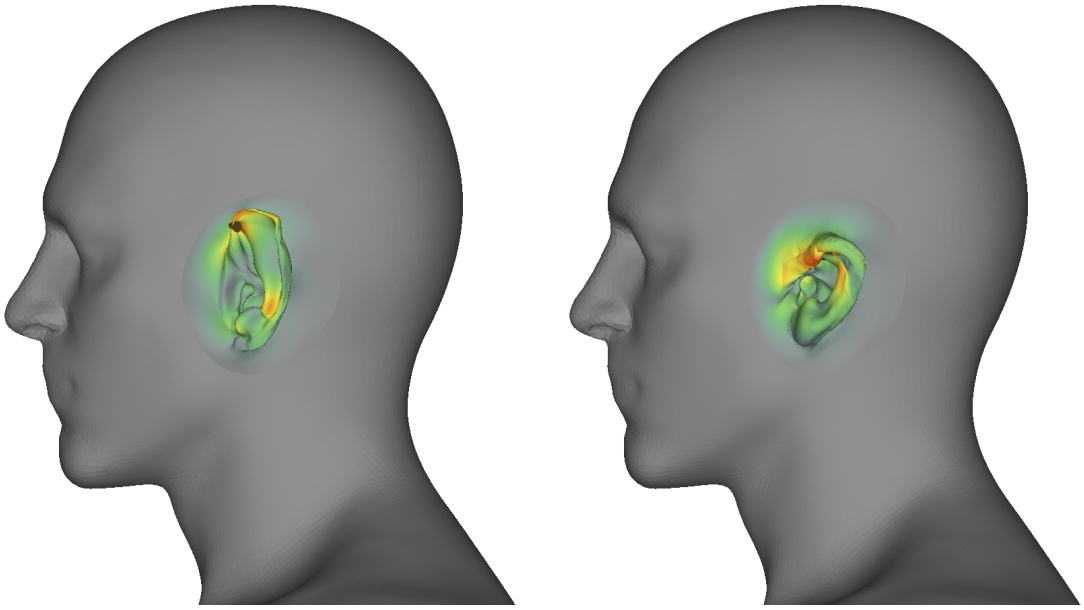


FIGURE C.54: *Effet de la* 9^e *composante - base mixte*

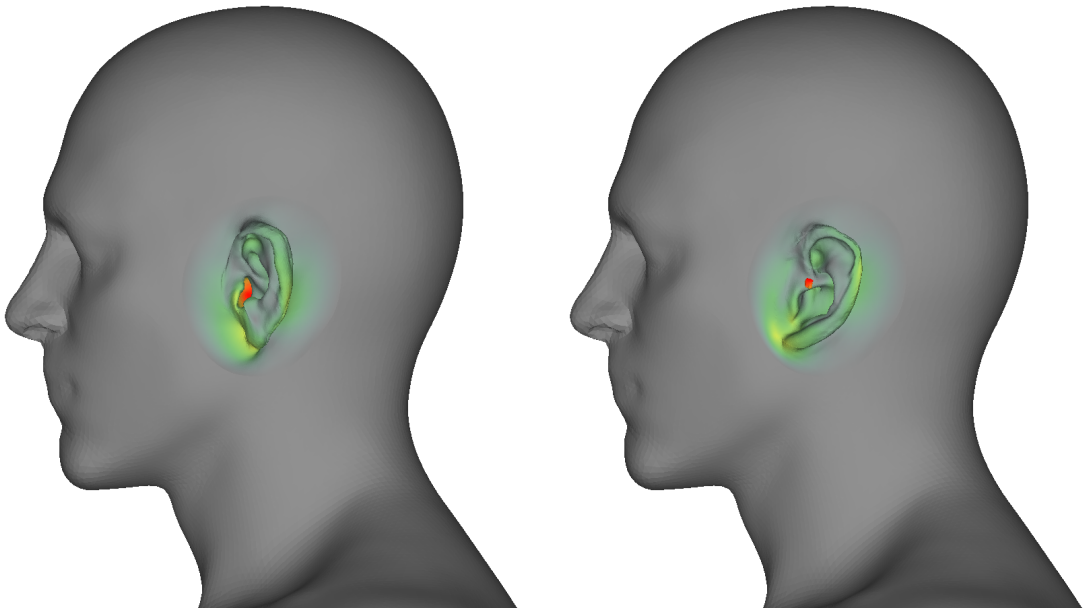


FIGURE C.55: *Effet de la* 10^e *composante - base mixte*

C.3.2 Modèle binaural

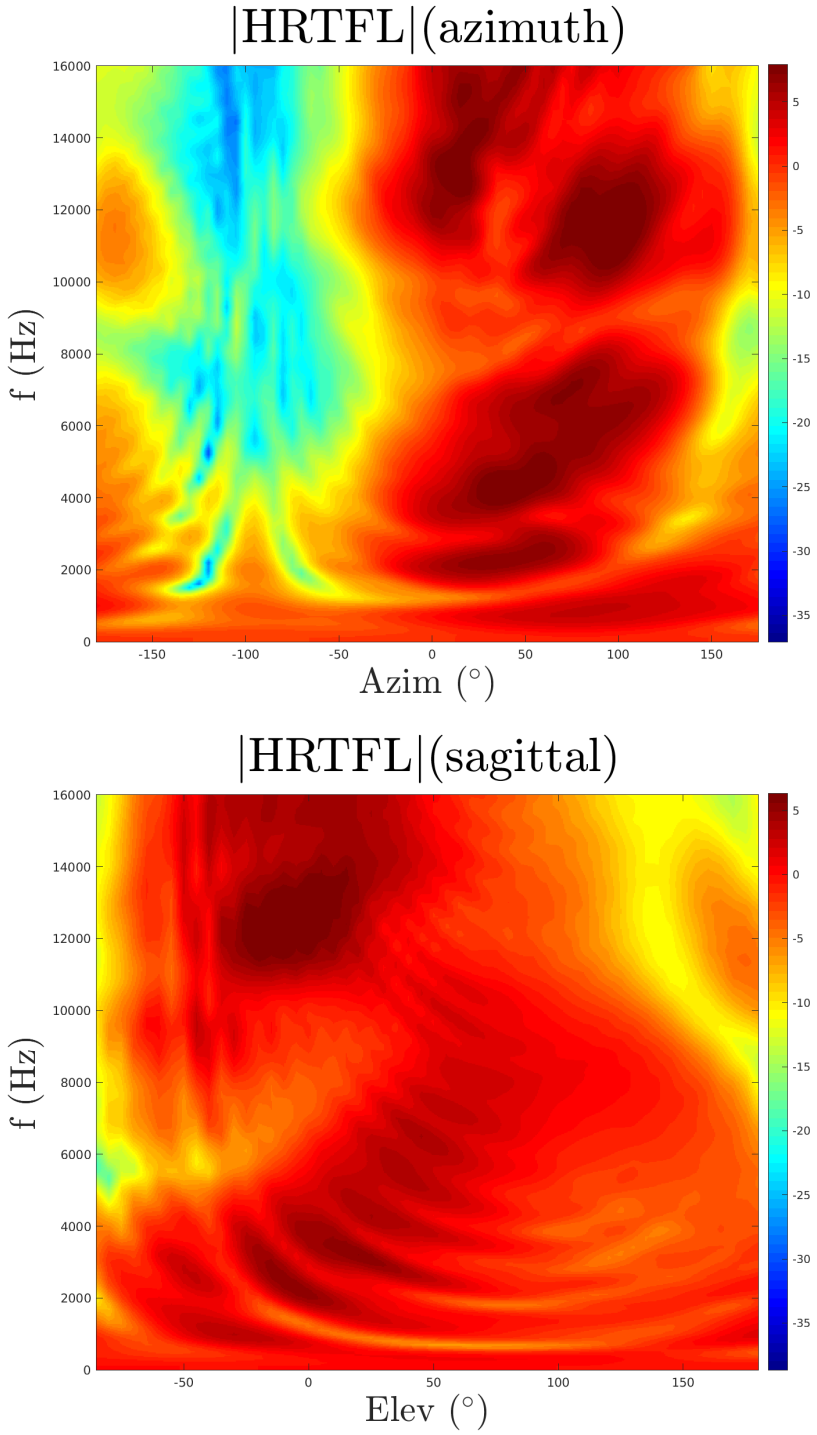


FIGURE C.56: *Coupes azimutale (en haut) et sagittale (en bas) de la HRTF moyenne du modèle issu de la base mixte.*

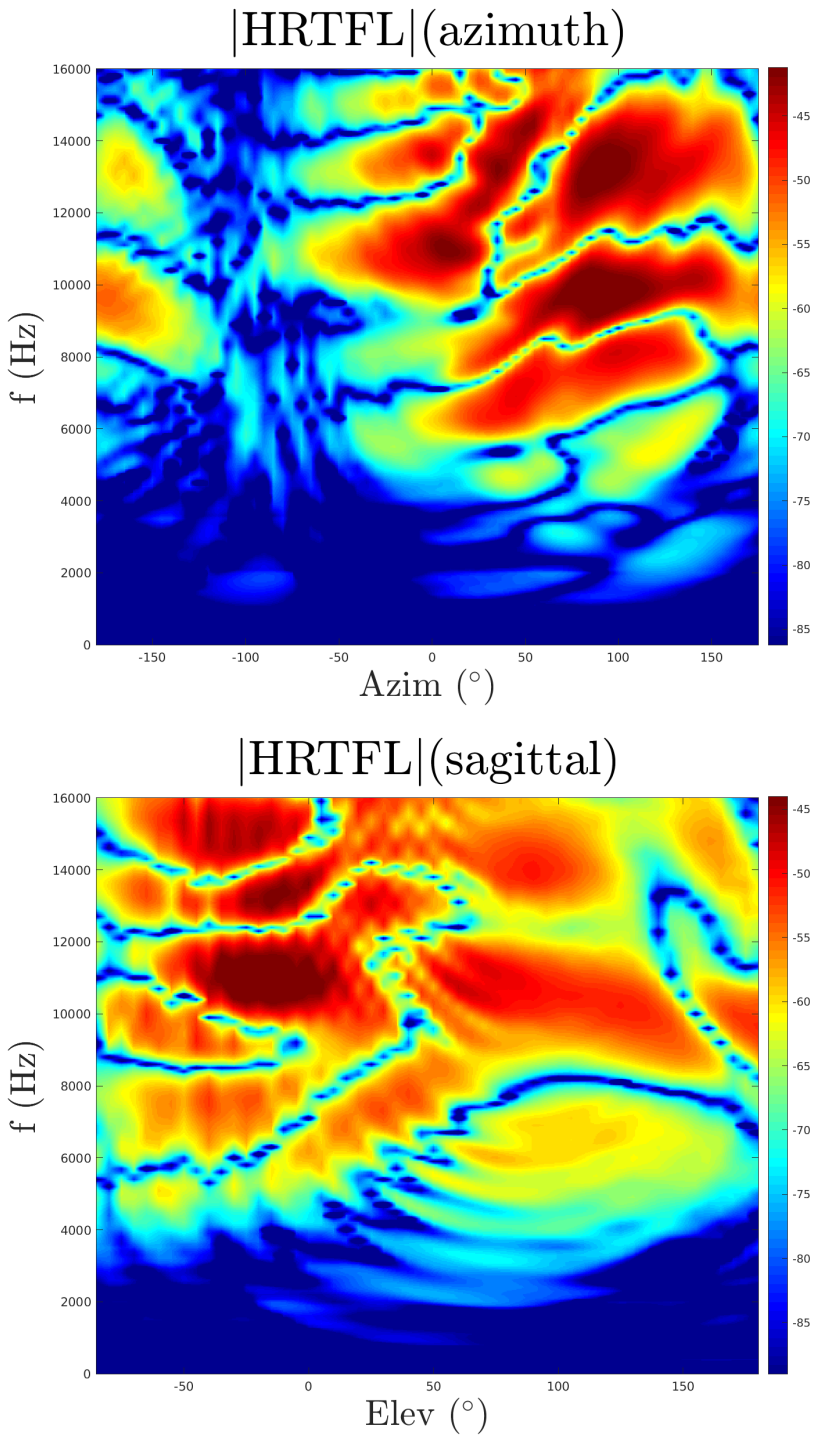


FIGURE C.57: *Effet de la* 1^e *composante - base mixte*

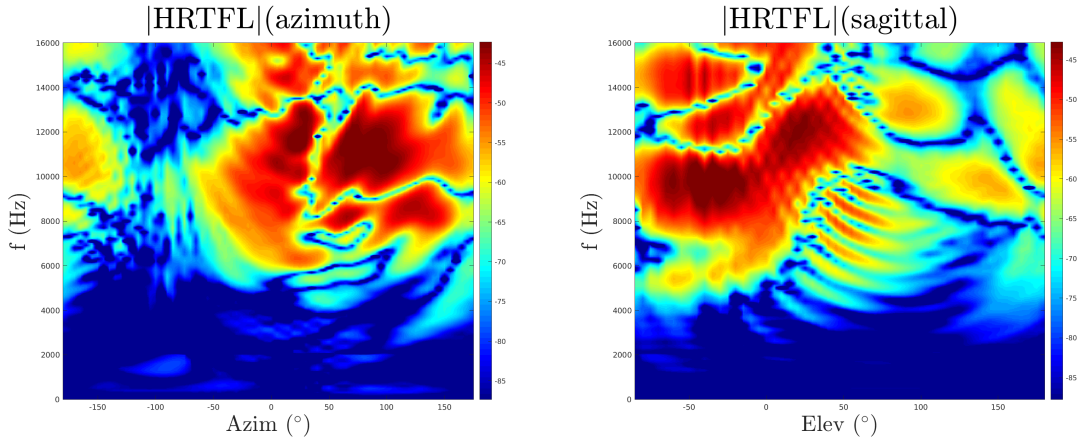


FIGURE C.58: *Effet de la 2^e composante - base mixte*

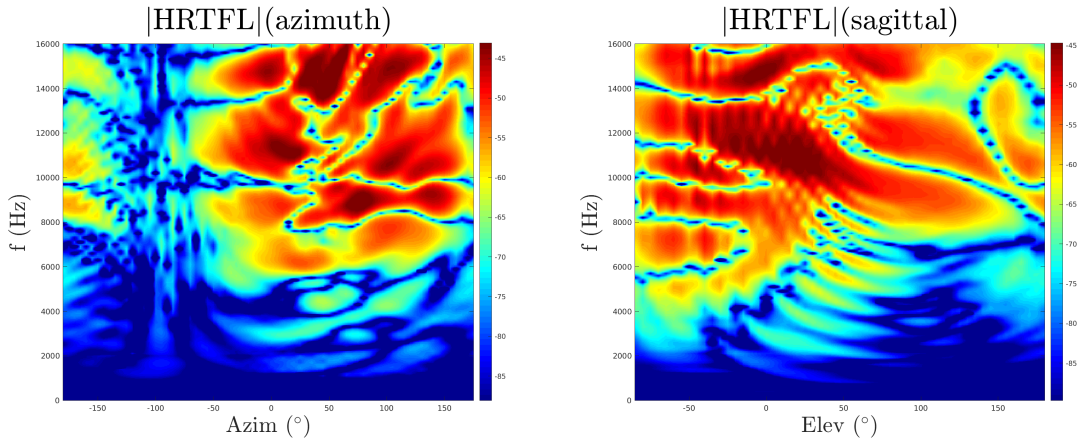


FIGURE C.59: *Effet de la 3^e composante - base mixte*

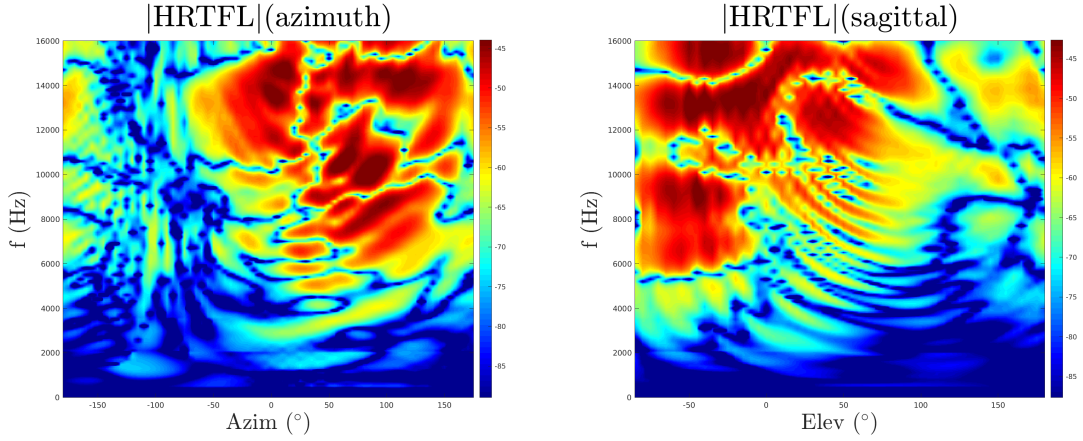


FIGURE C.60: *Effet de la 4^e composante - base mixte*

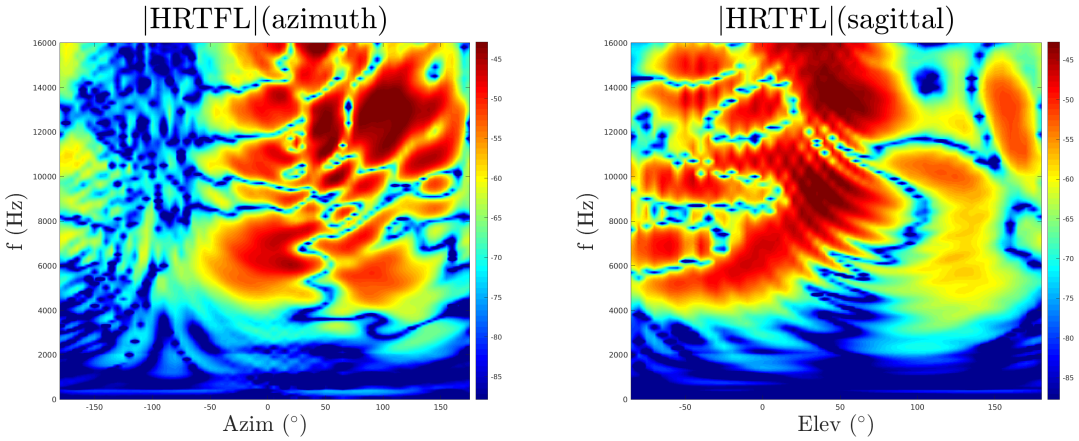


FIGURE C.61: *Effet de la* 5^e *composante - base mixte*

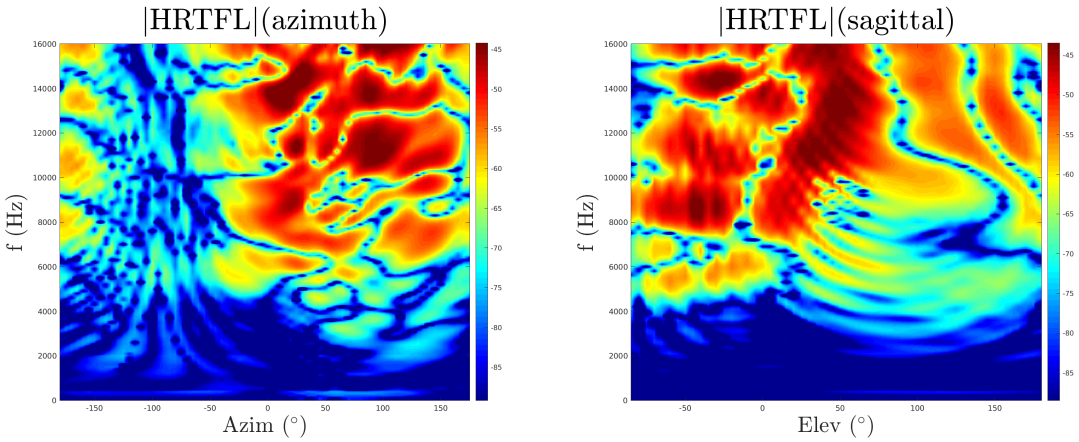


FIGURE C.62: *Effet de la* 6^e *composante - base mixte*

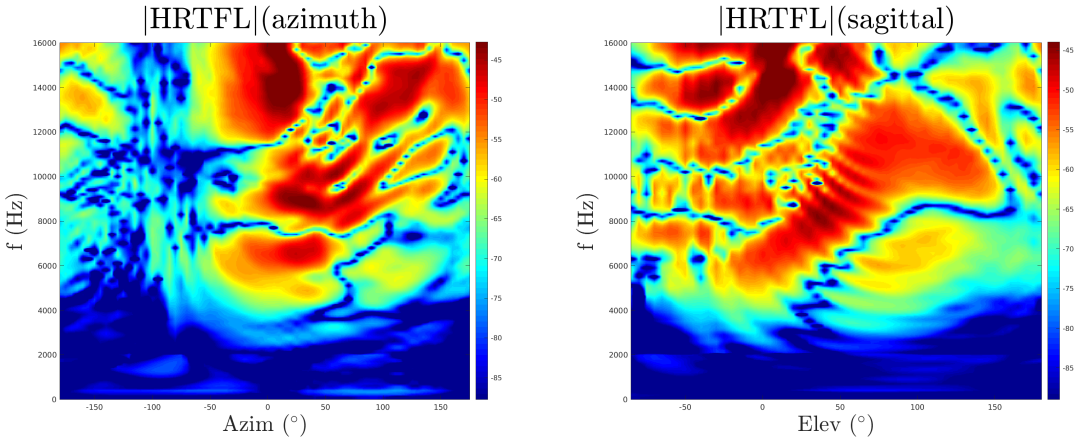


FIGURE C.63: *Effet de la* 7^e *composante - base mixte*

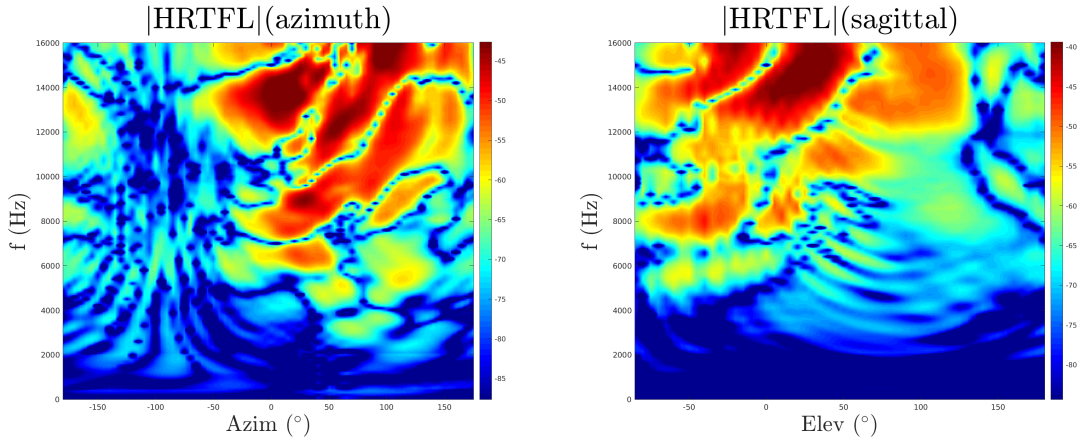


FIGURE C.64: *Effet de la 8^e composante - base mixte*

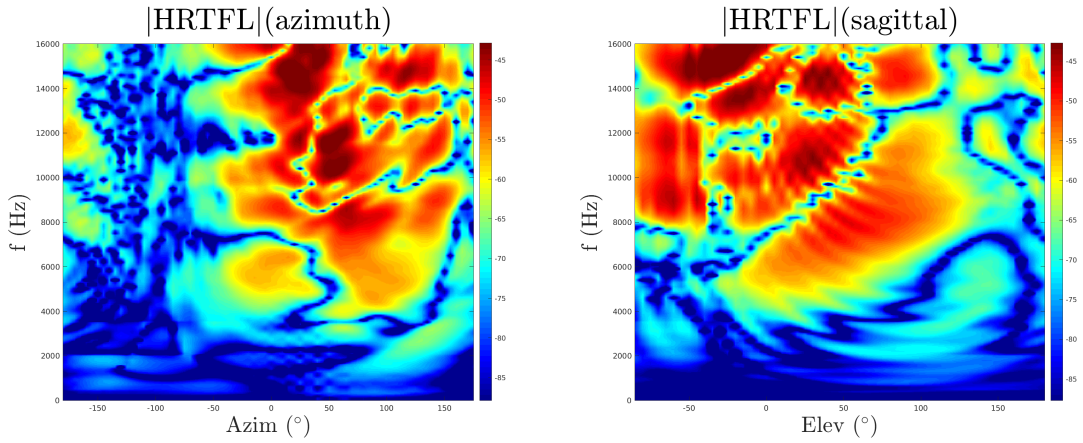


FIGURE C.65: *Effet de la 9^e composante - base mixte*

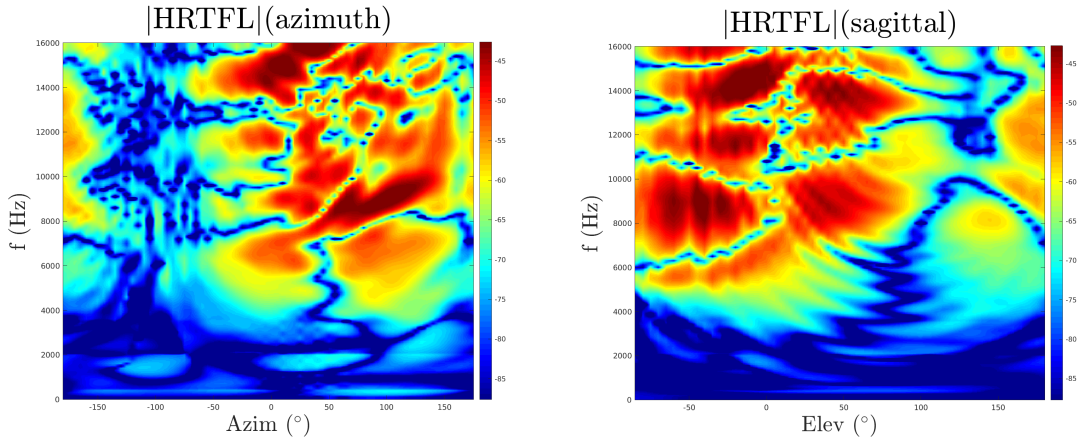


FIGURE C.66: *Effet de la 10^e composante - base mixte*

BIBLIOGRAPHIE

[1] V Ralph Algazi, Carlos Avendano et Richard O Duda, « Estimation of a spherical-head model from anthropometry », in : *Journal of the Audio Engineering Society* 49.*6* (2001), p. 472-479.

[2] V Ralph Algazi et al., « The cipic hrtf database », in : *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, IEEE, 2001, p. 99-102.

[3] Brett Allen, Brian Curless et Zoran Popović, « The space of human body shapes : reconstruction and parameterization from range scans », in : *ACM transactions on graphics (TOG)*, t. 22, ACM, 2003, p. 587-594.

[4] Nobuharu Aoshima, « Computer-generated pulse signal applied for sound measurement », in : *The Journal of the Acoustical Society of America* 69.*5* (1981), p. 1484-1488.

[5] Matthieu Aussal, « Méthodes numériques pour la spatialisation sonore, de la simulation à la synthèse binaurale », thèse de doct., Ecole Polytechnique X, 2014.

[6] The Acoustics Research Institute of the Austrian Academy of Sciences, *The ARI HRTF database*, http://www.kfs.oeaw.ac.at/hrtf.

[7] Dwight W Batteau, « The role of the pinna in human localization », in : *Proc. R. Soc. Lond. B* 168.*1011* (1967), p. 158-180.

[8] Robert Baumgartner, Piotr Majdak et Bernhard Laback, « Modeling sound-source localization in sagittal planes for human listeners », in : *The Journal of the Acoustical Society of America* 136.*2* (2014), p. 791-802.

[9] Durand R Begault, « Perceptual effects of synthetic reverberation on three-dimensional audio systems », in : *Journal of the Audio Engineering Society* 40.*11* (1992), p. 895-904.

[10] Durand R Begault et Elizabeth M Wenzel, « Headphone localization of speech », in : *Human Factors* 35.*2* (1993), p. 361-376.

[11] Durand R Begault, Elizabeth M Wenzel et Mark R Anderson, « Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source », in : *Journal of the Audio Engineering Society* 49.*10* (2001), p. 904-916.

[12] Jacob Benesty et al., « Advances in network and acoustic echo cancellation », in : Springer, 2001.

[13] Augustinus J Berkhout, Diemer de Vries et Marinus M Boone, « A new method to acquire impulse responses in concert halls », in : *The Journal of the Acoustical Society of America* 68.*1* (1980), p. 179-183.

[14] J Michael Berman et Laurie R Fincham, « The application of digital techniques to the measurement of loudspeakers », in : *Journal of the Audio Engineering Society* 25.*6* (1977), p. 370-384.

[15] Volker Blanz et Thomas Vetter, « A Morphable Model for the Synthesis of 3D Faces », in : *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '99, New York, NY, USA : ACM Press/Addison-Wesley Publishing Co., 1999, p. 187-194, isbn : 0-201-48560-5, doi : 10.1145/311535.311556, url : http://dx.doi.org/10.1145/311535.311556.

[16] Volker Blanz et Thomas Vetter, « Face recognition based on fitting a 3D morphable model », in : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 25.*9* (2003), p. 1063-1074.

[17] J von Blauert, P Laws et H J Platte, « Impulsverfahren zur Messung von AuBenohriibertragungsfunktionen », in : *Acustica* 31 (1974), p. 35.

[18] Jens Blauert, « Sound localization in the median plane », in : *Acta Acustica united with Acustica* 22.*4* (1969), p. 205-213.

[19] Jens Blauert, *Spatial hearing : the psychophysics of human sound localization*, MIT press, 1997.

[20] Ramona Bomhardt, Matias de la Fuente Klein et Janina Fels, « A high-resolution head-related transfer function and three-dimensional ear model database », in : *Proceedings of Meetings on Acoustics 172ASA*, t. 29, ASA, 2016, p. 050002.

[21] Fabian Brinkmann et al., « A cross-evaluated database of measured and simulated HRTFs including 3D head meshes, anthropometric features, and headphone impulse responses », in : *Journal of the Audio Engineering Society* 67.*9* (2019), p. 705-718.

[22] Fabian Brinkmann et al., « A high resolution head-related transfer function database including different orientations of head above the torso », in : *proceedings of the International Conference on Acoustics*, AIA-DAGA, Berlin : DEGA, mar. 2013, p. 596-599.

[23] Fabian Brinkmann et al., *The FABIAN head-related transfer function database*, rapp. tech., Technical University Berlin, 2017.

[24] Fabian Brinkmann et al., *The HUTUBS head-related transfer function (HRTF) database*, rapp. tech., Technical University Berlin, 2019.

[25] Adelbert W Bronkhorst, « Localization of real and virtual sound sources », in : *The Journal of the Acoustical Society of America* 98.*5* (1995), p. 2542-2553.

[26] Alexander M Bronstein, Michael M Bronstein et Ron Kimmel, « Audio- and Video-Based Biometric Person Authentication : 4th International Conference, AVBPA 2003 Guildford, UK, June 9–11, 2003 Proceedings », in : Berlin, Heidelberg : Springer Berlin Heidelberg, 2003, chap. Expression-Invariant 3D Face Recognition, p. 62-70, isbn : 978-3-540-44887-7, doi : 10.1007/3-540-44887-X_8, url : http://dx.doi.org/10.1007/3-540-44887-X_8.

[27] Douglas S Brungart et W Rabinowitz, « Auditory localization of nearby sources. Head-related transfer functions. », in : *The Journal of the Acoustical Society of America* 106 3 Pt 1 (1999), p. 1465-79.

[28] Douglas S Brungart et Brian D Simpson, « Effects of bandwidth on auditory localization with a noise masker », in : *The Journal of the Acoustical Society of America* 126.*6* (2009), p. 3199-3208.

[29] MD Burkhard et RM Sachs, « Anthropometric manikin for acoustic research », in : *The Journal of the Acoustical Society of America* 58.*1* (1975), p. 214-222.

[30] David S Burnett, « A three-dimensional acoustic infinite element based on a prolate spheroidal multipole expansion », in : *The Journal of the Acoustical Society of America* 96.*5* (1994), p. 2798-2816.

[31] AJ Burton et GF Miller, « The application of integral equation methods to the numerical solution of some exterior boundary-value problems », in : *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 323.*1553* (1971), p. 201-210.

[32] Sylvain Busson, « Individualisation d'Indices Acoustiques pour la Synthèse Binaurale », thèse de doct., Université de la Méditerranée - Aix-Marseille II, 2006.

[33] John D Bustard et Mark S Nixon, « 3D morphable model construction for robust ear and face recognition », in : *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, IEEE, 2010, p. 2582-2589.

[34] Robert A Butler, « Monaural and binaural localization of noise bursts vertically in median sagittal plane », in : *Journal of Auditory research* 9.*3* (1969), p. 230-235.

[35] Robert A Butler et Rosemary Flannery, « The spatial attributes of stimulus frequency and their role in monaural localization of sound in the horizontal plane », in : *Perception & psychophysics* 28.*5* (1980), p. 449-457.

[36] Robert A Butler, Richard A Humanski et Alan D Musicant, « Binaural and monaural localization of sound in two-dimensional space », in : *Perception* 19.*2* (1990), p. 241-256.

[37] Sharon Cameron et Harvey Dillon, « Development of the listening in spatialized noise-sentences test (LISN-S) », in : *Ear and hearing* 28.*2* (2007), p. 196-211.

[38] Andrea Capra et al., « Listening tests of the localization performance of Stereodipole and Ambisonic systems », in : *Audio Engineering Society Convention 123*, Audio Engineering Society, 2007.

[39] Simon Carlile, *Virtual auditory space : Generation and applications*, Springer Science & Business Media, 2013.

[40] Thibaut Carpentier et al., « Measurement of a head-related transfer function database with high spatial resolution », in : *7th Forum Acusticum (EAA)*, 2014.

[41] Jasmina Catic, Sébastien Santurette et Torsten Dau, « The role of reverberation-related binaural cues in the externalization of speech », in : *The Journal of the Acoustical Society of America* 138.*2* (2015), p. 1154-1167.

[42] AES Standards Committee et al., « AES69-2015 AES Standard for File Exchange-Spatial Acoustic Data File Format », in : *Audio Engineering Society, Inc* (2015).

[43] Arthur Conan Doyle, *The Sign of the Four*, Lippincott's Monthly Magazine, 1889.

[44] Clément Cornuau, « Étude et optimisation de la synthèse transaurale à deux canaux », in : *Mars* (2011), p. 1.

[45] Camilo Klinkert Correa, Song Li et Jürgen Peissig, « Analysis and Comparison of different Adaptive Filtering Algorithms for Fast Continuous HRTF Measurement », in : *DAGA 2017* (2017), p. 1049-1052.

[46] P von Damaske et B Wagener, « Richtungshörversuche über einen nachgebildeten Kopf (Localisation experiments via a head replica) », in : *Acustica* 21.*1* (1969), p. 30-35.

[47] Jérôme Daniel, « Représentation de champs acoustiques, application à la transmission et à la restitution de scènes sonores complexes dans un contexte multimédia », thèse de doct., Paris 6, 2000.

[48] Pascal Dietrich, Bruno Masiero et Michael Vorländer, « On the optimization of the multiple exponential sweep method », in : *Journal of the Audio Engineering Society* 61.*3* (2013), p. 113-124.

[49] Rob Drullman et Adelbert W Bronkhorst, « Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation », in : *The Journal of the Acoustical Society of America* 107.*4* (2000), p. 2224-2235.

[50] Richard O Duda, Carlos Avendano et V Ralph Algazi, « An adaptable ellipsoidal head model for the interaural time difference », in : *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, t. 2, IEEE, 1999, p. 965-968.

[51] Richard O Duda et William L Martens, « Range dependence of the response of a spherical head model », in : *The Journal of the Acoustical Society of America* 104.*5* (1998), p. 3048-3058.

[52] Nathaniel I Durlach et al., « On the externalization of auditory images », in : *Presence : Teleoperators & Virtual Environments* 1.*2* (1992), p. 251-257.

[53] Gerald Enzner, « 3D-continuous-azimuth acquisition of head-related impulse responses using multi-channel adaptive filtering », in : *Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA'09. IEEE Workshop on*, IEEE, 2009, p. 325-328.

[54] Gerald Enzner, « Analysis and optimal control of LMS-type adaptive filtering for continuous-azimuth acquisition of head related impulse responses », in : *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, IEEE, 2008, p. 393-396.

[55] Angelo Farina, « Simultaneous measurement of impulse response and distortion with a swept-sine technique », in : *Audio Engineering Society Convention 108*, Audio Engineering Society, 2000.

[56] Simone Fontana, Angelo Farina et Yves Grenier, « Binaural for popular music : a case of study », in : *Proceedings of the 13th International Conference on Auditory Display*, ICAD, Montréal, Canada, juin 2007.

[57] Takaya Fujimoto, « A study of TSP signal getting higher SN ratio at low frequency bands », in : *Proc. Autumn Meet. Acoust. Soc. Jpn., 1999* (1999).

[58] Mark B Gardner, « Some monaural and binaural facets of median plane localization », in : *The Journal of the Acoustical Society of America* 54.*6* (1973), p. 1489-1495.

[59] Mark B Gardner et Robert S Gardner, « Problem of localization in the median plane : effect of pinnae cavity occlusion », in : *The Journal of the Acoustical Society of America* 53.*2* (1973), p. 400-408.

[60] Michele Geronazzo, Simone Spagnol et Federico Avanzini, « Estimation and modeling of pinna-related transfer functions », in : *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10)*, 2010, p. 6-10.

[61] Slim Ghorbal, Xavier Bonjour et Renaud Séguier, « Computed Hrirs and Ears Database for Acoustic Research », in : *Audio Engineering Society Convention 148*, Audio Engineering Society, Vienne, Austria, mai 2020, url : https://hal.archives-ouvertes.fr/hal-02484097.

[62] Slim Ghorbal, Renaud Séguier et Xavier Bonjour, « Procédé d'élaboration d'un modèle déformable en trois dimensions d'un élément, et système associé », FR1654765 (France), 2016, url : https://hal.archives-ouvertes.fr/hal-01831361.

[63] Slim Ghorbal, Renaud Séguier et Xavier Bonjour, « Procédé et système d'évaluation d'une fonction de transfert relative à la tête adaptée à un individu », PCT/EP2016/065839 (France), 2016, url : https://hal.archives-ouvertes.fr/hal-01831370.

[64] Slim Ghorbal, Renaud Séguier et Xavier Bonjour, « Process of HRTF Individualization by 3D Statistical Ear Model », in : *Audio Engineering Society Convention 141*, Audio Engineering Society, 2016.

[65] Slim Ghorbal et al., « Pinna morphological parameters influencing HRTF sets », in : *International Conference on Digital Audio Effects*, Edinburgh, United Kingdom, 2017, url : https://hal.archives-ouvertes.fr/hal-01831314.

[66] Felipe Grijalva et al., « Anthropometric-based customization of head-related transfer functions using Isomap in the horizontal plane », in : *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, IEEE, 2014, p. 4473-4477.

[67] Gaël Guennebaud, Marcel Germann et Markus Gross, « Dynamic sampling and rendering of algebraic point set surfaces », in : *Computer Graphics Forum*, t. 27, Wiley Online Library, 2008, p. 653-662.

[68] Gaël Guennebaud et Markus Gross, « Algebraic point set surfaces », in : *ACM Transactions on Graphics (TOG)*, t. 26, ACM, 2007, p. 23.

[69] Pierre Guillon, « Individualisation des indices spectraux pour la synthèse binaurale : recherche et exploitation des similarités inter-individuelles pour l'adaptation ou la reconstruction de HRTF », thèse de doct., Université du Maine, 2009.

[70] Pierre Guillon et al., « Creating the Sydney York morphological and acoustic recordings of ears database », in : *Multimedia and Expo (ICME), 2012 IEEE International Conference on*, IEEE, 2012, p. 461-466.

[71] Navarun Gupta et al., « HRTF database at FIU DSP lab », in : *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, IEEE, 2010, p. 169-172.

[72] Dorte Hammershoï et Henrik Møller, « Sound transmission to and within the human ear canal », in : *The Journal of the Acoustical Society of America* 100.*1* (1996), p. 408-427.

[73] Rudolf Häusler, S Colburn et E Marr, « Sound localization in subjects with impaired hearing : spatial-discrimination and interaural-discrimination tests », in : *Acta Oto-Laryngologica* 96.*sup400* (1983), p. 1-62.

[74] Jack Hebrank et D Wright, « Spectral cues used in the localization of sound sources on the median plane », in : *The Journal of the Acoustical Society of America* 56.*6* (1974), p. 1829-1834.

[75] V Cutanda Henriquez et Peter Møller Juhl, « OpenBEM-An open source Boundary Element Method software in acoustics », in : *Internoise 2010* (2010), p. 13-16.

[76] Marko Hiipakka, Teemu Kinnari et Ville Pulkki, « Estimating head-related transfer functions of human subjects from pressure–velocity measurements », in : *The Journal of the Acoustical Society of America* 131.*5* (2012), p. 4051-4061.

[77] Hongmei Hu et al., « Head related transfer function personalization based on multiple regression analysis », in : *2006 International Conference on Computational Intelligence and Security*, t. 2, IEEE, 2006, p. 1829-1832.

[78] W Wahab Hugeng et Dadang Gunawan, « Improved method for individualization of head-related transfer functions on horizontal plane using reduced number of anthropometric measurements », in : *arXiv preprint arXiv :1005.5137* (2010).

[79] Jyri Huopaniemi et Matti Karjalainen, « Review of digital filter design and implementation methods for 3-D sound », in : *Audio Engineering Society Convention 102*, Audio Engineering Society, 1997.

[80] Jyri Huopaniemi et Julius O Smith III, « Spectral and time-domain preprocessing and the choice of modeling error criteria for binaural digital filters », in : *Audio Engineering Society Conference : 16th International Conference : Spatial Sound Reproduction*, Audio Engineering Society, 1999.

[81] Naoya Inoue et al., « Evaluation of HRTFs estimated using physical features », in : *Acoustical science and technology* 26.*5* (2005), p. 453-455.

[82] M O Irfanoglu, B Gokberk et L Akarun, « 3D shape-based face recognition using automatically registered facial surfaces », in : *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, t. 4, août 2004, 183-186 Vol.4, doi : 10.1109/ICPR.2004.1333734.

[83] Yukio Iwaya, « Individualization of head-related transfer functions with tournament-style listening test : Listening with other's ears », in : *Acoustical science and technology* 27.*6* (2006), p. 340-343.

[84] Craig T Jin et al., « Creating the sydney york morphological and acoustic recordings of ears database », in : *IEEE Transactions on Multimedia* 16.*1* (2014), p. 37-46.

[85] G Jin et al., « Generation of customised three dimensional sound effects for individuals », US Patent 7,542,574, US Patent 7,542,574, juin 2009, url : http://www.google.com/patents/US7542574.

[86] G Jin et al., « Generation of customized three dimensional sound effects for individuals », US Patent 7,209,564, US Patent 7,209,564, avr. 2007, url : http://www.google.com/patents/US7209564.

[87] Jean-Marc Jot, Veronique Larcher et Olivier Warusfel, « Digital signal processing issues in the context of binaural and transaural stereophony », in : *Audio Engineering Society Convention 98*, Audio Engineering Society, 1995.

[88] Yuvi Kahana, « Numerical modelling of the head-related transfer function », thèse de doct., Institute of Sound et Vibration Research, 2000.

[89] Brian Fredrick Gray Katz, « Acoustic absorption measurement of human hair and skin within the audible frequency range », in : *The Journal of the Acoustical Society of America* 108.*5* (2000), p. 2238-2242.

[90] Brian Fredrick Gray Katz, « Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation », in : *The Journal of the Acoustical Society of America* 110.*5* (2001), p. 2440-2448.

[91] Brian Fredrick Gray Katz, « Measurement and calculation of individual head-related transfer functions using a boundary element model including the measurement and effect of skin and hair impedance », thèse de doct., The Pennsylvania State University, nov. 1998.

[92] Brian Fredrick Gray Katz et Markus Noisternig, « A comparative study of interaural time delay estimation methods », in : *The Journal of the Acoustical Society of America* 135.*6* (2014), p. 3530-3540.

[93] Brian Fredrick Gray Katz et Gaëtan Parseihian, « Perceptually based head-related transfer function database optimization », in : *The Journal of the Acoustical Society of America* 131.*2* (2012), EL99-EL105.

[94] Brian Fredrick Gray Katz et D Schönstein, « Procédé de sélection de filtres hrtf perceptivement optimale dans une base de données à partir de paramètres morphologiques », PCT/FR2011/050,840, oct. 2011, url : https://www.google.com/patents/WO2011128583A1?cl=fr.

[95] Robert B King et Simon R Oldfield, « The impact of signal bandwidth on auditory localization : Implications for the design of three-dimensional audio displays », in : *Human factors* 39.*2* (1997), p. 287-295.

[96] Doris J Kistler et Frederic L Wightman, « A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction », in : *The Journal of the Acoustical Society of America* 91.*3* (1992), p. 1637-1647, doi : http://dx.doi.org/10.1121/1.402444, url : http://scitation.aip.org/content/asa/journal/jasa/91/3/10.1121/1.402444.

[97] Abhijit Kulkarni, SK Isabelle et HS Colburn, « Sensitivity of human subjects to head-related transfer-function phase spectra », in : *The Journal of the Acoustical Society of America* 105.*5* (1999), p. 2821-2840.

[98] Erno HA Langendijk et Adelbert W Bronkhorst, « Fidelity of three-dimensional-sound reproduction using a virtual auditory display », in : *The Journal of the Acoustical Society of America* 107.*1* (2000), p. 528-537.

[99] Véronique Larcher, « Techniques de spatialisation des sons pour la réalité virtuelle », thèse de doct., Paris 6, 2001.

[100] Véronique Larcher et Jean-Marc Jot, « Techniques d'interpolation de filtres audio-numériques, Applicationa la reproduction spatiale des sons sur écouteurs », in : *Proc. CFA : Congres Français d'Acoustique*, Citeseer, 1997.

[101] Bruno Lévy et al., « Least squares conformal maps for automatic texture atlas generation », in : *ACM transactions on graphics (TOG)*, t. 21, ACM, 2002, p. 362-371.

[102] Chen Li et al., « A novel 3D ear reconstruction method using a single image », in : *Intelligent Control and Automation (WCICA), 2012 10th World Congress on*, IEEE, 2012, p. 4891-4896.

[103] Ewan A Macpherson et John C Middlebrooks, « Listener weighting of cues for lateral angle : the duplex theory of sound localization revisited », in : *The Journal of the Acoustical Society of America* 111.*5* (2002), p. 2219-2236.

[104] Piotr Majdak, Peter Balazs et Bernhard Laback, « Multiple exponential sweep method for fast measurement of head-related transfer functions », in : *Journal of the Audio Engineering Society* 55.*7/8* (2007), p. 623-637.

[105] Katuhiro Maki et Shigeto Furukawa, « Reducing individual differences in the external-ear transfer functions of the Mongolian gerbil », in : *The Journal of the Acoustical Society of America* 118.*4* (2005), p. 2392-2404, doi : http://dx.doi.org/10.1121/1.2033571, url : http://scitation.aip.org/content/asa/journal/jasa/118/4/10.1121/1.2033571.

[106] William L Martens, « Rapid psychophysical calibration using bisection scaling for individualized control of source elevation in auditory display », in : *Proc. Int. Conf. on Auditory Display*, juil. 2002, p. 199-206.

[107] W McLean, « Ears in Action », in : *A motion picture made by the US Naval Ordnance* (1959).

[108] Sünke Mehrgardt et Volker Mellert, « Transformation characteristics of the external human ear », in : *The Journal of the Acoustical Society of America* 61.*6* (1977), p. 1567-1576.

[109] John C Middlebrooks, « Individual differences in external-ear transfer functions reduced by scaling in frequency », in : *The Journal of the Acoustical Society of America* 106.*3* (1999), p. 1480-1492.

[110] John C Middlebrooks, « Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency », in : *The Journal of the Acoustical Society of America* 106.*3* (1999), p. 1493-1510.

[111] John C Middlebrooks et David M Green, « Directional dependence of interaural envelope delays », in : *The Journal of the Acoustical Society of America* 87.*5* (1990), p. 2149-2162.

[112] John C Middlebrooks, James C Makous et David M Green, « Directional sensitivity of sound-pressure levels in the human ear canal », in : *The Journal of the Acoustical Society of America* 86.*1* (1989), p. 89-108.

[113] Allen William Mills, « On the minimum audible angle », in : *The Journal of the Acoustical Society of America* 30.*4* (1958), p. 237-246.

[114] Pauli Minnaar et al., « The interaural time difference in binaural synthesis », in : *Audio Engineering Society Convention 108*, Audio Engineering Society, 2000.

[115] Parham Mokhtari, Ryouichi Nishimura et Hironori Takemoto, « Toward HRTF personalization : an auditory-perceptual evaluation of simulated and measured HRTFs », in : *Auditory Display (ICAD), International Conference on*, International Community for Auditory Display, 2008.

[116] Alfonso R Molares et Manuel A Sobreira-Seoane, « Benchmarking for acoustic simulation software », in : *Journal of the Acoustical Society of America* 123.*5* (2008), p. 3515-3515.

[117] Henrik MØLLER et al., « Binaural technique : Do we need individual recordings ? », in : *Journal of the Audio Engineering Society* 44.*6* (1996), p. 451-469.

[118] Naoya MORIYA et Yutaka KANEDA, « Study of harmonic distortion on impulse response measurement with logarithmic time stretched pulse », in : *Acoustical science and technology* 26.*5* (2005), p. 462-464.

[119] Alan D MUSICANT et Robert A BUTLER, « Influence of monaural spectral cues on binaural localization », in : *The Journal of the Acoustical Society of America* 77.*1* (1985), p. 202-208.

[120] Takanori NISHINO et al., « Estimation of HRTFs on the horizontal plane using physical features », in : *Applied Acoustics* 68.*8* (2007), p. 897-908.

[121] Gang PAN, Zhaohui WU et Yunhe PAN, « Automatic 3D face verification from range data », in : *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*, t. 3, avr. 2003, 193-196 vol.3, DOI : 10.1109/ICASSP.2003.1199140.

[122] Louis T More PHD, « XXXVII. On the localization of the direction of sounds », in : *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 18.*104* (1909), p. 308-319, DOI : 10.1080/14786440808636703, eprint : https://doi.org/10.1080/14786440808636703, URL : https://doi.org/10.1080/14786440808636703.

[123] Jan PLOGSTIES et al., « Audibility of all-pass components in head-related transfer functions », in : *Audio Engineering Society Convention 108*, Audio Engineering Society, 2000.

[124] Danièle PRALONG et Simon CARLILE, « Measuring the human head-related transfer functions : A novel method for the construction and calibration of a miniature "in-ear"recording system », in : *The Journal of the Acoustical Society of America* 95.*6* (1994), p. 3435-3444.

[125] Lionel REVERET et al., « Morphable Model of Quadrupeds Skeletons for Animating 3D Animals », in : *Proceedings of the 2005 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '05, Los Angeles, California : ACM, 2005, p. 135-142, ISBN : 1-59593-198-8, DOI : 10.1145/1073368.1073386, URL : http://doi.acm.org/10.1145/1073368.1073386.

[126] Klaus AJ RIEDERER, « Repeatability analysis of head-related transfer function measurements », in : *Audio Engineering Society Convention 105*, Audio Engineering Society, 1998.

[127] Douglas D RIFE et John VANDERKOOY, « Transfer-function measurement with maximum-length sequences », in : *Journal of the Audio Engineering Society* 37.*6* (1989), p. 419-444.

[128] Suzanne K ROFFLER et Robert A BUTLER, « Factors that influence the localization of sound in the vertical plane », in : *The Journal of the Acoustical Society of America* 43.*6* (1968), p. 1255-1259.

[129] I Rudomin, A Bojorquez et H Cuevas, « Statistical generation of 3D facial animation models », in : *Shape Modeling International, 2002. Proceedings*, 2002, p. 219-226, doi : 10.1109/SMI.2002.1003549.

[130] Felipe Rugeles, Marc Emerit et Brian Fredrick Gray Katz, « Évaluation objective et subjective de différentes méthodes de lissage des HRTF », in : *Cong Français d'Acoustique (CFA)*, CFA, 2014, p. 2213-2219.

[131] Lord Rayleigh O M Pres. R S, « XII. On our perception of sound direction », in : *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 13.*74* (1907), p. 214-232, doi : 10.1080/14786440709463595, eprint : https://doi.org/10.1080/14786440709463595, url : https://doi.org/10.1080/14786440709463595.

[132] Jesper Sandvad, « Dynamic aspects of auditory virtual environments », in : *Audio Engineering Society Convention 100*, Audio Engineering Society, 1996.

[133] Jesper Sandvad et Dorte Hammershoi, « Binaural auralization, comparison of FIR and IIR filter representation of HIRs », in : *Audio Engineering Society Convention 96*, Audio Engineering Society, 1994.

[134] Patrick Satarzadeh, V Ralph Algazi et Richard O Duda, « Physical and filter pinna models based on anthropometry », in : *Audio Engineering Society Convention 122*, Audio Engineering Society, 2007.

[135] Lauri Savioja et al., « Creating interactive virtual acoustic environments », in : *Journal of the Audio Engineering Society* 47.*9* (1999), p. 675-705.

[136] Bernhard U Seeber et Hugo Fastl, « Subjective selection of non-individual head-related transfer functions », in : *Proceedings of 9th International Conference on Auditory Display*, Georgia Institute of Technology, Boston, MA, USA, juil. 2003.

[137] EAG Shaw, « Acoustic response of external ear with progressive wave source », in : *The Journal of the Acoustical Society of America* 51.*1A* (1972), p. 150-150.

[138] Shoji Shimada, Nobuo Hayashi et Shinji Hayashi, « A clustering method for sound localization transfer functions », in : *Journal of the Audio Engineering Society* 42.*7/8* (1994), p. 577-584.

[139] BG Shinn-Cunningham, Scott Santarelli et Norbert Kopco, « Distance perception of nearby sources in reverberant and anechoic listening conditions : Binaural vs. monaural cues », in : *Assoc Res Otolaryn. Meeting*, t. 23, 2000.

[140] Peter L Søndergaard et Piotr Majdak, « The auditory modeling toolbox », in : *The technology of binaural listening*, Springer, 2013, p. 33-56.

[141] Simone Spagnol, Michele Geronazzo et Federico Avanzini, « On the relation between pinna reflection patterns and head-related transfer function features », in : *IEEE transactions on audio, speech, and language processing* 21.*3* (2013), p. 508-519.

[142] Yôiti Suzuki et al., « An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses », in : *The Journal of the Acoustical Society of America* 97.*2* (1995), p. 1119-1123, DOI : 10.1121/1.412224, eprint : https://doi.org/10.1121/1.412224, URL : https://doi.org/10.1121/1.412224.

[143] Robert P Tame, Daniele Barchiese et Anssi Klapuri, « Headphone Virtualization : Improved Localization and Externalization of Non-Individualized HRTFs by Cluster Analysis », in : *Audio Engineering Society Convention 133*, Audio Engineering Society, mai 2012.

[144] R Teranishi et EAG Shaw, « External-Ear Acoustic Models with Simple Geometry », in : *The Journal of the Acoustical Society of America* 44.*1* (1968), p. 257-263.

[145] Bradley E Treeby, Jie Pan et Roshun M Paurobally, « An experimental study of the acoustic impedance characteristics of human hair », in : *The Journal of the Acoustical Society of America* 122.*4* (2007), p. 2107-2117.

[146] Bradley E Treeby, Jie Pan et Roshun M Paurobally, « The effect of hair on auditory localization cues », in : *The Journal of the Acoustical Society of America* 122.*6* (2007), p. 3586-3597.

[147] Bradley E Treeby, Roshun M Paurobally et Jie Pan, « The effect of impedance on interaural azimuth cues derived from a spherical head model », in : *The Journal of the Acoustical Society of America* 121.*4* (2007), p. 2217-2226.

[148] Julia Turku et al., « Perceptual evaluation of numerically simulated head-related transfer functions », in : *Audio Engineering Society Convention 124*, Audio Engineering Society, 2008.

[149] Jesper Udesen, Tobias Piechowiak et Fredrik Gran, « Vision affects sound externalization », in : *Audio Engineering Society Conference : 55th International Conference : Spatial Audio*, Audio Engineering Society, 2014.

[150] Hans Wallach, « The role of head movements and vestibular and visual cues in sound localization. », in : *Journal of Experimental Psychology* 27.*4* (1940), p. 339.

[151] Kanji Watanabe et al., « Dataset of head-related transfer functions measured with a circular loudspeaker array », in : *Acoustical science and technology* 35.*3* (2014), p. 159-165.

[152] Elizabeth M Wenzel, « What perception implies about implementation of interactive virtual acoustic environments », in : *Audio Engineering Society Convention 101*, Audio Engineering Society, 1996.

[153] Francis M Wiener et Douglas A Ross, « The pressure distribution in the auditory canal in a progressive sound field », in : *The Journal of the Acoustical Society of America* 18.*2* (1946), p. 401-408.

[154] Frederic L Wightman et Doris J Kistler, « Resolution of front–back ambiguity in spatial hearing by listener and source movement », in : *The Journal of the Acoustical Society of America* 105.*5* (1999), p. 2841-2853.

[155] Frederic L Wightman et Doris J Kistler, « The dominant role of low-frequency interaural time differences in sound localization », in : *The Journal of the Acoustical Society of America* 91.*3* (1992), p. 1648-1661.

[156] Robert Sessions Woodworth, « Experimental Psychology. New York : Holt, 1938 », in : *Department of Psychology Dartmouth College Hanover, New Hampshire* (1937).

[157] Robert Sessions Woodworth et Harold Schlosberg, *Experimental psychology*, Holt, Rinehart et Winston, 1962.

[158] Donald Wright, John H Hebrank et Blake Wilson, « Pinna reflections as cues for localization », in : *The Journal of the Acoustical Society of America* 56.*3* (1974), p. 957-962.

[159] Bosun Xie et Zhaojun Tian, « Improving Binaural Reproduction of 5.1 Channel Surround Sound Using Individualized HRTF Cluster in the Wavelet Domain », in : *Audio Engineering Society Conference : 55th International Conference : Spatial Audio*, Audio Engineering Society, août 2014.

[160] Bosun Xie, Xiaoli Zhong et Nana He, « Typical data and cluster analysis on head-related transfer functions from Chinese subjects », in : *Applied Acoustics* 94 (2015), p. 1-13.

[161] Song Xu, Zhizhong Li et Gavriel Salvendy, « Improved method to individualize head-related transfer function using anthropometric measurements », in : *Acoustical Science and Technology* 29.*6* (2008), p. 388-390, doi : 10.1250/ast.29.388.

[162] Satoshi Yairi, Yukio Iwaya et Yôiti Suzuki, « Individualization feature of head-related transfer functions based on subjective evaluation », in : *Proc. of International Conference on Auditory Display (ICAD2008), Paris*, 2008.

[163] Bao-Cai Yin et al., « MPEG-4 compatible 3D facial animation based on morphable model », in : *Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on*, t. 8, août 2005, 4936-4941 Vol. 8, doi : 10.1109/ICMLC.2005.1527812.

[164] Pavel Zahorik, « Limitations in using Golay codes for head-related transfer function measurement », in : *The Journal of the Acoustical Society of America* 107.*3* (2000), p. 1793-1796.

[165] Elias Zea, « Binaural in-ear monitoring of acoustic instruments in live music performance », in : *15th International Conference on Digital Audio Effects, DAFx 2012, 17 September 2012 through 21 September 2012, York*, 2012, p. 1-8.

[166] M Zhang et al., « Statistical method to identify key anthropometric parameters in HRTF individualization », in : *Hands-free Speech Communication and Microphone Arrays (HSCMA), 2011 Joint Workshop on*, IEEE, 2011, p. 213-218.

[167] Harald Ziegelwanger, Wolfgang Kreuzer et Piotr Majdak, « Mesh2hrtf : Open-source software package for the numerical calculation of head-related transfer functions », in : *22st International Congress on Sound and Vibration*, 2015.

[168] Harald Ziegelwanger, Piotr Majdak et Wolfgang Kreuzer, « Numerical calculation of listener-specific head-related transfer functions and sound localization : Microphone model and mesh discretization », in : *The Journal of the Acoustical Society of America* 138.*1* (2015), p. 208-222.

[169] Reza Zolfaghari et al., « Large deformation diffeomorphic metric mapping and fast-multipole boundary element method provide new insights for binaural acoustics », in : *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, IEEE, 2014, p. 2863-2867.

[170] Dmitry N Zotkin et al., « Fast head-related transfer function measurement via reciprocity », in : *The Journal of the Acoustical Society of America* 120.*4* (2006), p. 2202-2215.

[171] Dmitry N Zotkin et al., « HRTF personalization using anthropometric measurements », in : *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.* Oct. 2003, p. 157-160, doi : 10.1109/ASPAA.2003.1285855.

DOCTORAT BRETAGNE LOIRE / MATHSTIC

CentraleSupélec

Titre : Personnalisation de l'écoute binaurale par modèle déformable d'oreille

Mot clés : HRTF ; modèle déformable ; personnalisation ; oreille ; binaural ; son 3D

Résumé : Le terme « binaural » fait référence au champ de recherche visant à comprendre et maîtriser les mécanismes permettant à l'être humain de percevoir l'origine spatiale des sons. Cette perception émerge de notre faculté à détecter certains indices de localisation au sein de notre environnement sonore et ces indices, quant à eux, naissent de l'interaction des sons avec notre corps et en particulier nos oreilles. Pour reproduire au casque un effet de spatialisation sonore il faut donc d'une part réintroduire ces indices et d'autre part en personnaliser la génération en l'adaptant à la morphologie de l'auditeur.
Pour y parvenir nous avons imaginé un procédé original fondé sur l'utilisation d'un modèle déformable 3D d'oreille et l'étude en amont des liens entre morphologies et HRTF. Nous en démontrons ici la faisabilité en le mettant en pratique grâce à des bases de données synthétiques créées pour l'occasion. Cette génération de données nous a par ailleurs amené à proposer des optimisations au calcul numérique de HRTF et à réfléchir aux améliorations possibles pour en fiabiliser le rendu subjectif.

Title: Binaural listening personalised through morphable ear model

Keywords: HRTF ; morphable model ; personalisation ; ear ; binaural ; 3D sound

Abstract: The term "binaural" refers to the field of research dedicated to the understanding and mastering of the mechanisms allowing humans to perceive the spatial origin of sounds. This perception emerges from our ability to detect some specific indices of localisation within our sound environment and these indices, for their part, arise from the interaction of sounds with our body and, more specifically, our ears. To reproduce a sound spatialisation effect with headphones, it is therefore necessary on the one hand to reintroduce these indices and on the other hand to personalise their generation by adapting it to the morphology of the listener.
To achieve this, we have imagined an original process based on the use of a 3D morphable model of the ear and the study of the relationships between morphologies and HRTFs. Here we demonstrate its feasibility by putting it into practice thanks to synthetic databases created for the occasion. This data generation has led us to suggest optimisations for numerical computation of HRTFs and to propose possible improvements to make their subjective rendering more reliable.