



UNIVERSITÉ DE LA ROCHELLE

ÉCOLE DOCTORALE Euclide 618

**LABORATOIRE : Laboratoire Informatique, Image et
Interaction (L3i)**

THÈSE présentée par :

Éric Méaux

soutenue le : **07 Janvier 2022**

pour obtenir le grade de : **Docteur de l'université de La Rochelle**

Discipline : **Informatique et Applications**

**Approche perceptive pour la spatialisation / localisation
sonore 3D**

JURY :

Catherine SEMAL
Gianpaolo EVANGELISTA

Olivier WARUSFEL
Sylvain MARCHAND
Myriam DESAINTE-CATHERINE
Richard KRONLAND-MARTINET

Professeure, Université de Bordeaux. Présidente du jury
Professeur, University of Music and Performing Arts, Rapporteur

Directeur de Recherche, IRCAM-STMS, Rapporteur
Professeur, Université de La Rochelle, Directeur de thèse
Professeure, Bordeaux INP
Directeur de Recherche, CNRS PRISM

Table des matières

I	Notions d'audition humaine et de spatialisation sonore	14
1	Quelques notions d'audition et de localisation humaine	16
1.1	L'oreille et l'acquisition du son par l'homme	16
1.2	La localisation du son par l'homme	17
1.2.1	Système de coordonnées et conditions	18
1.2.2	La localisation de l'azimut utilisée dans cette thèse	18
1.2.3	La localisation de la distance utilisée dans cette thèse	19
1.2.4	La localisation de l'élévation utilisée dans cette thèse	20
1.3	Fonction de transfert relative à la tête (HRTF)	21
1.3.1	Théorie de base sur les HRTFs	21
1.3.2	Les bases de données d'HRTFs et de BRIRs	22
1.3.3	Les HRTFs de la base CIPIC et leurs utilisations dans STAR	24
1.4	Indices acoustiques interauraux	26
1.4.1	Généralités	26
1.4.2	Différences interaurales de niveau (ILDs)	28
1.4.3	Différences interaurales de temps (ITDs)	30
2	Introduction à la spatialisation sonore, méthodes existantes et utilisations	33
2.1	Méthode de spatialisation au casque - le son binaural	33
2.2	Méthode de spatialisation via des réseaux de haut-parleurs	34
2.2.1	L'approche physique de spatialisation du son	34
2.2.2	Une approche mathématique de spatialisation du son	36
2.2.3	L'approche transaurale utilisée pour la méthode STAR	37
II	Localisation sonore perceptive	40
3	Méthode de localisation perceptive pour l'azimut	42
3.1	La méthode	42
3.1.1	Rappel sur les modèles d'indices acoustiques utilisés dans la méthode STAR	42
3.1.2	Le processus de localisation	43
3.2	Validation de la méthode de localisation perceptive	45
3.2.1	Étude en conditions idéales	45
3.2.2	Étude en conditions réelles	47
4	Une approche perceptive pour la localisation sonore en élévation	50
4.1	Origine des pics et nœuds	50
4.2	Étude d'un modèle existant	51
4.2.1	Le modèle développé par Iida	51
4.2.2	Application du modèle sur la base CIPIC	52
4.3	Utilisation d'une approche Shazam pour la localisation perceptive de l'élévation	58
4.3.1	Construction de la table	58

4.3.2	Utilisation de la table pour la détection de l'élévation	62
4.3.3	Cas de l'utilisation d'HRTFs théoriquement identiques	67
III	Spatialisation sonore perceptive	70
5	La spatialisation en azimuth	73
5.1	Principe général de la spatialisation de l'azimut de la méthode STAR	73
5.2	Chemins acoustiques synthétiques	75
5.2.1	Les chemins acoustiques de la méthode STAR	75
5.2.2	Discussion sur les chemins acoustiques	76
5.3	Le principe de transauralité dans la méthode STAR	78
5.4	Le déterminant	80
5.4.1	Le déterminant de la méthode STAR	80
5.4.2	Discussion sur le déterminant	81
5.5	Les coefficients	83
5.6	Les coefficients de la méthode STAR	83
5.6.1	Discussion sur les coefficients	84
6	La spatialisation en élévation	87
6.1	Les modèles utilisés pour la spatialisation en élévation	87
6.2	La méthode de spatialisation en élévation STAR	89
7	La spatialisation en distance	92
7.1	Le modèle d'absorption du son dans l'air utilisé	92
7.2	La méthode de spatialisation en distance de la méthode STAR	93
IV	Les tests de validation de la méthode STAR	95
8	Évaluation objective de l'indépendance de l'élévation et de la distance avec l'azimut	97
8.1	Indépendance entre l'azimut et l'élévation	98
8.1.1	Étude sur l'élévation en conditions anéchoïques	98
8.1.2	Étude sur l'élévation en conditions réelles	99
8.1.3	Étude sur l'élévation simulée	101
8.2	Indépendance entre l'azimut et la distance	102
8.2.1	Étude sur la distance réelle	102
8.2.2	Étude sur la distance simulée	105
9	Évaluation objective de la méthode STAR par la cartographie des ILDs	108
9.1	Cartographie des différences d'intensité interaurales (ILDs) par simulations	109
9.1.1	Protocole de simulation	109
9.1.2	Résultat des simulations	111
9.2	Cartographie des différences d'intensité interaurales (ILDs) réelles	115
9.2.1	Protocole expérimental	115
9.2.2	Résultats des expérimentations	115
10	Évaluation subjective de comparaison de méthodes	120
10.1	Protocoles des tests perceptifs de comparaison des méthodes de spatialisation	120
10.1.1	Tests préliminaires	121
10.1.2	Tests statiques	121
10.1.3	Tests dynamiques	121
10.1.4	Tests polyphoniques	122
10.2	Dispositif Expérimental	122
10.3	Résultats des tests perceptifs de comparaison des méthodes de spatialisation	123

TABLE DES MATIÈRES

10.3.1 Tests statiques	124
10.3.2 Tests dynamiques	125
10.3.3 Tests polyphoniques	127
11 Évaluation subjective pour la validation de la distance et de l'élévation, ainsi que leur étude d'indépendance	131
11.1 Protocole expérimental	131
11.2 Résultats des tests de validation de la méthode STAR pour l'élévation et la distance	132
11.3 Test d'indépendance subjectif pour l'élévation et la distance, complémentaire de l'étude objective de leurs indépendances avec l'azimut	135
 V Conclusion et perspectives	 137
 VI Annexes	 139
11.4 Liste des abréviations	140
11.5 Symboles mathématiques utilisés	140
11.6 GUI-STAR, l'IHM d'étude de la méthode STAR et de création d'exemples sonores	141
11.7 Les locaux et outils utilisés	147
11.7.1 Le dôme du SCRIME	147
11.7.2 La salle de conférence et de diffusion Hémicyclia	148
11.7.3 Salle de classe, salon et salle de concert	148
11.8 Le matériel de diffusion et d'enregistrement	148
11.8.1 Le mannequin KEMAR	148
11.8.2 Les cartes sons	149
11.8.3 Les micros	150
11.8.4 Les haut-parleurs	150
11.8.5 Les dispositifs créés pour les prises de son à l'aide du mannequin KEMAR	150
11.8.6 Les logiciels utilisés	152

Table des figures

1.1	<i>Schéma général de l'oreille humaine (Wikimédia).</i>	17
1.2	<i>Système de coordonnées utilisées dans le cadre de la thèse (Marchand (2020)).</i>	18
1.3	<i>Exemple d'une HRIR, provenant ici du sujet $s = 4$ de la base CIPIC, pour une direction $\theta = 0^\circ$ et une élévation $\phi = 25^\circ$.</i>	21
1.4	<i>Exemple d'une HRTF, provenant ici du sujet $s = 4$ de la base CIPIC, pour une direction $\theta = 0^\circ$ et une élévation $\phi = 25^\circ$.</i>	22
1.5	<i>HRTF d'un sujet unique - ici Kemar large pinnae, pour une élévation $\phi = 0^\circ$.</i>	24
1.6	<i>HRTF de la moyenne de la base de données CIPIC, pour une élévation $\phi = 0^\circ$.</i>	25
1.7	<i>Énergie des chemins acoustiques pour KEMAR large pinnae.</i>	26
1.8	<i>Énergie des chemins acoustiques pour la moyenne de la base CIPIC.</i>	26
1.9	<i>Schéma du trajet d'une onde sonore entre sa source et les oreilles de l'auditeur (Marchand (2020)).</i>	27
1.10	<i>ILDs réelles calculées à partir de la moyenne des HRTFs de la base CIPIC.</i>	28
1.11	<i>Coefficient α obtenu par la méthode des moindres carrés par correspondance du modèle aux données.</i>	29
1.12	<i>ILDs synthétiques.</i>	29
1.13	<i>ILDs moyennes en fonction de l'azimut, enregistrées en utilisant différentes méthodes de diffusions (STAR, VBAP et HOA) pour faire tourner la source, ainsi que des mesures réelles (les KEMAR).</i>	30
1.14	<i>ITDs réelles.</i>	30
1.15	<i>Coefficient β obtenu par la méthode des moindres carrés par correspondance du modèle aux données.</i>	31
1.16	<i>ITDs synthétiques.</i>	31
2.1	<i>Exemple de micro conçu pour la prise de son binaural, ici de la marque 3Dio. https://3dio-sound.com/</i>	34
2.2	<i>Schéma simplifié d'une synthèse binaurale à l'aide de BRTF/HRTF. Les BRTF/HRTF gauche et droit sont appliqués au signal mono d'entrée pour être ensuite envoyés sur les champs gauche/droit du casque de l'auditeur.</i>	34
2.3	<i>Schéma des harmonies sphériques pour les 3 premiers ordres (ordre 0 sur la première ligne, 1 sur la seconde ligne, etc.). Les parties blanches correspondent aux valeurs négatives, et celles en noir aux valeurs positives des harmoniques. (Zotter and Frank (2019).)</i>	35
2.4	<i>Exemple de système mis en place pour une diffusion avec la méthode WFS. Cette photo présente une partie de la couronne de haut-parleurs et montre bien le grand nombre de haut-parleurs nécessaires. https://easternbloc.ca</i>	36
2.5	<i>Schéma provenant de l'article de Pulkki (1997) et présentant un cas de spatialisation en 2D d'une source sonore virtuelle positionnée entre deux haut-parleurs.</i>	37
2.6	<i>Principe transaural : 4 chemins acoustiques réels (H_{LL}, H_{RL}, H_{LR}, et H_{RR}) provenant des haut-parleurs gauches et droits (LS_L et LS_R) sont utilisés afin de reproduire les 2 chemins acoustiques virtuels (H_L et H_R) provenant de la source virtuelle S.</i>	38

TABLE DES FIGURES

3.1	Coefficients α et β , obtenus par correspondance au modèle via la méthode des moindres carrés, et contenant l'ensemble des informations relatives aux têtes et oreilles de chaque sujet de la base, et ceci pour toutes les fréquences et tous les azimuts.	43
3.2	Processus de localisation perceptif.	44
3.3	Histogramme de l'azimut dans des conditions idéales (en utilisant la base CIPIC). L'énergie est très concentrée autour de la position de la source sonore (ici $\theta = 0^\circ$).	45
3.4	Précision de localisation de l'azimut en conditions idéales (anéchoïques), ceci pour cinq azimuts, les 45 sujets de la base CIPIC ainsi que 25 élévations.	46
3.5	Précision de localisation de l'azimut en conditions idéales (anéchoïques), pour le mannequin KEMAR larges oreilles provenant de la base CIPIC ainsi que 25 élévations.	47
3.6	Erreur d'estimation pour la localisation de l'azimut dans différentes configurations de pièces (BRIRs). Les résultats sont du même ordre que l'idéal dans l'intervalle $-40^\circ + 40^\circ$	48
3.7	Histogramme de répartition de l'énergie de la localisation de l'azimut dans des conditions réelles (première configuration de la base BRIR). L'exemple montre le résultat de la localisation pour une source réelle présente à 48° , l'énergie est dissipée et la source réverbérée (environ 25°) possède plus d'énergie que la source réelle.	48
3.8	Erreur d'estimation pour la localisation de l'azimut dans différentes configurations de pièces (BRIRs). Il est à noter que la pièce 1 génère de nombreux outliers, qui ne sont pas visibles sur ce tracé.	49
4.1	Schéma du pavillon de l'oreille externe (docplayer.fr).	51
4.2	Tracé du modèle d'après Iida et al. (2007) utilisé pour connaître la position des nœuds en fonction de l'élévation (N1 en bleu et N2 en rouge); la troisième composante permettant la localisation verticale étant le premier pic P1, constant aux alentours de 3500 Hz.	52
4.3	Ensemble des HRTFs de la base CIPIC, avec P1 (ligne noire), N1 (ligne bleu) et N2 (ligne rouge) représentés.	52
4.4	Visualisation des HRTFs traitées pour 6 élévations, pour le sujet 4, en azimut $\theta = 0^\circ$	54
4.5	Évolution des positions de N1 et N2 détectés en fonction de l'élévation. On observe une erreur pour l'élévation 60°	55
4.6	Nœuds détectés sur l'ensemble des élévations et pour tous les sujets. Pour chaque élévation, il y a ainsi 45 points tracés pour N1 et pour N2.	56
4.7	Répartition des nœuds détectés au plus proche de ceux attendus par le modèle.	57
4.8	Répartition du nombre de nœuds (premier graphique) et pics (second graphique) en fonction de l'élévation.	59
4.9	La table Shazam compacte pour les nœuds de la base CIPIC. Le pas idéal est de 86 Hz.	60
4.10	La table Shazam compacte pour les pics de la base CIPIC. Le pas idéal est de 87 Hz.	61
4.11	Histogrammes de table Shazam de collisions pour les nœuds.	62
4.12	Histogrammes pour un sujet appris (ici sujet 4, élévation 90°).	63
4.13	Histogrammes des conglomerats de couples NP pour un sujet appris (ici sujet 4, élévation 90°).	64
4.14	Synthèse des détections sur l'ensemble de la base pour des sujets appris, en considérant le conglomerat N1P N2P NP.	65
4.15	Histogrammes des conglomerats de couples NP pour un sujet non appris (ici sujet 4, élévation 90°).	66
4.16	Synthèse des détections sur l'ensemble de la base pour des sujets non appris, en considérant le conglomerat N1P N2P NP.	67
4.17	Utilisation de la méthode Shazam sur les HRTFs du SCRIME, enregistrées pour $\theta = 0^\circ$ et $\phi = 25^\circ$	68
4.18	Processus complet de spatialisation de la méthode STAR.	72
5.1	Configuration octophonique, représentant huit haut-parleurs. Dans cet exemple, deux sources S1 et S2, et quatre haut-parleurs actifs : S1 utilisant les haut-parleurs LS2 et LS3, et S2 utilisant les haut-parleurs LS5 et LS6.	74
5.2	Principe général de spatialisation de la méthode STAR.	74

5.3	Coefficients α et β , obtenus par correspondance au modèle via la méthode des moindres carrés, et contenant l'ensemble des informations relatives aux têtes et oreilles de chaque sujet de la base, et ceci pour toutes les fréquences et tous les azimuts.	75
5.4	Second nœud (N_2) en fonction de l'azimut θ , et pour une élévation $\phi = 0^\circ$. La symétrie des HRTFs est visible ici.	76
5.5	Erreur quadratique entre les HRTFs réelles du mannequin KEMAR Large Pinnae et les chemins acoustiques synthétiques de la méthode STAR, pour une élévation $\phi = 0^\circ$. L'erreur, hors valeurs extrêmes, est du même ordre que les erreurs entre HRTFs individualisées. . . .	77
5.6	Erreur quadratique entre la moyenne des HRTFs de la base CIPIC et les chemins acoustiques synthétiques de la méthode STAR, pour une élévation $\phi = 0^\circ$. L'erreur, hors valeurs extrêmes, est du même ordre que les erreurs entre HRTFs individualisées.	78
5.7	Le processus de spatialisation STAR pour l'azimut.	79
5.8	Norme du déterminant en fonction de l'écartement des haut-parleurs et de la fréquence. En rouge, les valeurs n'assurant pas la stabilité du système (< 0.01). L'espacement minimal des haut-parleurs est donc de 2°	80
5.9	Norme du déterminant pour l'azimut de référence $\theta = 0^\circ$	81
5.10	Déterminant initial avec une erreur sur les deux bandes jouées par les haut-parleurs de part et d'autre de l'axe transaural. Ici, les haut-parleurs sont placés au niveau de chaque changement de bande, avec l'axe transaural passant par 90° et -90° , provoquant une symétrie entre les deux haut-parleurs les entourant et ainsi un déterminant nul.	81
5.11	Déterminant une fois la méthode corrigée.	83
5.12	Amplitude des coefficients K_L (graphique du haut) et K_R (graphique du bas) en fonction de la fréquence et de l'azimut θ de la source sonore virtuelle.	84
5.13	Coefficients VBAP (en pointillé noir) et STAR (ligne verte). La comparaison est donnée pour $\theta = 0^\circ$, cas où les coefficients gauches et droits sont identiques.	85
5.14	Coefficients initiaux. (Gauche pour le graphique du haut et droit pour le graphique du bas).	85
5.15	Coefficients avec la méthode adaptée. (Gauches pour le graphique du haut et droits pour le graphique du bas)	86
6.1	Largeur (Hz) et profondeur (dB) des nœuds et pics moyens provenant de la base CIPIC, en fonction de l'élévation.	88
6.2	Courbe polynomiale représentant l'évolution de la largeur et de la hauteur des pics et des nœuds en fonction de l'élévation. Les données utilisées pour l'apprentissage proviennent de la base de données CIPIC.	89
6.3	Le processus de spatialisation STAR pour l'élévation.	90
6.4	Coefficients de mise en forme des HRTFs calculées en utilisant les paramètres provenant des modèles apprises sur la base CIPIC, et du modèle de Iida, pour une élévation de 22.5 degrés.	90
7.1	Modèle d'atténuation du son par l'air en fonction de la fréquence et de la distance (entre 0 et 100 m). Chaque courbe représente une distance, la moins élevée représentant $d = 0m$ (confondue avec l'axe des abscisses) et la plus élevées $d = 100m$	93
7.2	Le processus de spatialisation STAR pour la distance.	94
8.1	Les trois paramètres de spatialisation de STAR et leurs méthodes d'évaluation d'indépendance.	98
8.2	Localisation de l'azimut pour 5 positions, 43 sujets de la base CIPIC et 25 élévations.	98
8.3	Inclinaison du mannequin pour simuler l'élévation (ici 40°).	99
8.4	Erreur d'estimation de l'azimut pour différentes élévations en utilisant les BRIRs du SCRIME. L'erreur est cohérente avec la précision humaine et ne dépend pas de l'élévation.	100
8.5	Erreur d'estimation de l'azimut pour différentes élévations en utilisant les BRIRs du SCRIME. Les résultats suivent l'idéal dans la plage $\pm 40^\circ$, et ceci quelle que soit l'élévation.	100
8.6	Erreur d'estimation de l'azimut pour différentes élévations simulées. L'erreur est cohérente avec la précision humaine et ne dépend pas de l'élévation.	101
8.7	Erreur d'estimation de l'azimut pour différentes élévations simulées. Les résultats suivent l'idéal dans la plage $\pm 40^\circ$, et ceci quelle que soit l'élévation.	102

TABLE DES FIGURES

8.8	Exemple des tests d'enregistrement dans des conditions extérieures (ici pour un azimut $\theta = 80^\circ$ et une distance $d = 2\text{m}$).	103
8.9	Exemple d'enregistrement effectué pour les tests d'indépendance de la distance dans le studio du SCRIME. Ici pour une distance $d = 2\text{m}$ et un azimut $\theta = 0^\circ$	103
8.10	Estimation de l'azimut pour différentes distances. Les résultats suivent assez bien l'idéal. On voit aussi que plus la distance est importante et plus le décrochage se trouve dans des azimuts faibles.	104
8.11	Estimation de l'azimut pour différentes distances. L'erreur obtenue concorde avec la précision de localisation de l'humain. De nombreux outliers sont présents lorsque l'on est dans des distances plus importantes. Cependant, la moyenne et l'écart type ne dépendent pas de la distance.	105
8.12	Estimation de l'azimut pour différentes distances simulées. Les résultats suivent bien l'idéal dans une large plage d'azimuts $\pm 60^\circ$. La distance n'influe pas sur la détection de l'azimut, les paramètres sont donc indépendants.	106
8.13	Estimation de l'azimut pour différentes distances simulées. L'erreur concorde avec la précision de localisation du son par l'homme et ne dépend pas de la distance. Aucun outlier n'est présent contrairement aux résultats pour des distances réelles. Les paramètres sont donc indépendants.	106
9.1	Chemins acoustiques pour la configuration octophonique.	110
9.2	Maillage simplifié utilisé pour la simulation de cartographies des ILDs.	111
9.3	Cartographie des ILDs utilisant la méthode de simulation pour une configuration octophonique régulière, une source placée à $\theta = 30^\circ$, et un maillage de 100×100 . — Méthode de spatialisation HOA.	112
9.4	Cartographie des ILDs utilisant la méthode de simulation pour une configuration octophonique régulière, une source placée à $\theta = 30^\circ$, et un maillage de 100×100 . — Méthode de spatialisation STAR.	112
9.5	Cartographie des ILDs utilisant la méthode de simulation pour une configuration octophonique régulière, une source placée à $\theta = 30^\circ$, et un maillage de 100×100 . — Méthode de spatialisation VBAP.	113
9.6	Évolution des cartographies des ILDs pour les trois méthodes de simulation, une configuration octophonique régulière, un maillage de 100×100 , et un azimut θ allant de 10° à 160° avec un pas de 10°	114
9.7	Mise en place expérimentale pour la prise de mesures des ILDs au sein du dôme du SCRIME.	115
9.8	Simulation d'un public par la présence de 25 mannequins KEMAR, la position centrale est représentée en gras.	115
9.9	Cartographie des ILDs utilisant la méthode expérimentale pour une configuration octophonique régulière, une source placée à $\theta = 30^\circ$, et un maillage de 5×5 . — Méthode de spatialisation STAR.	116
9.10	Cartographie des ILDs utilisant la méthode expérimentale pour une configuration octophonique régulière, une source placée à $\theta = 30^\circ$, et un maillage de 5×5 . — Méthode de spatialisation HOA.	117
9.11	Cartographie des ILDs utilisant la méthode expérimentale pour une configuration octophonique régulière, une source placée à $\theta = 30^\circ$, et un maillage de 5×5 . — Méthode de spatialisation VBAP.	118
10.1	Exemple du cercle de référence permettant aux auditeurs de positionner les différents extraits joués.	121
10.2	Exemple de l'échelle de notation permettant d'attribuer un score aux différents extraits joués.	121
10.3	Choix parmi les 8 trajectoires possibles, la réponse correcte étant le cercle (dans le sens de rotation direct).	122
10.4	Photographie du dispositif expérimental mis en place dans la salle de classe.	122
10.5	Dispositif expérimental (utilisé en 2019), disposant de 8 haut-parleurs actifs, (LS 1 à LS 8), 4 inactifs utilisés comme pièges (B), et en leur centre 9 places assises (S1 à S9).	123

TABLE DES FIGURES

10.6	Résultats des tests statiques. Position perçue par les auditeurs pour les différentes méthodes, et pour un azimut $\theta = 45^\circ$.	124
10.7	Résultats des tests statiques pour la qualité estimée.	125
10.8	Résultats des tests dynamiques pour les différentes méthodes. Le score de la trajectoire des sources perçues est ici représenté en pour cent - Tests réalisés en 2019.	126
10.9	Résultats des tests dynamiques pour les différentes méthodes. Le score trajectoire des sources perçues est ici représenté en pour cent - Tests réalisés en 2015.	126
10.10	Test dynamique, évaluation de la qualité des méthodes.	127
10.11	Tests polyphoniques, qualité du son ressentie.	128
10.12	Tests polyphoniques : immersion ressentie.	129
10.13	Tests polyphoniques : intelligibilité.	130
11.1	Tests de validation de la méthode STAR pour l'élévation et la distance (les deux premières boîtes) et d'indépendance entre ces paramètres (les deux secondes boîtes).	133
11.2	Résultats de la validation de la méthode pour les paramètres distance (plage 0 - 9 m) et élévation (plage 0 - 80 °) et leur indépendance. Plus la distance/élévation entre les deux sources sont importantes et plus le pourcentage de bonnes réponses est grand (ligne verte). Lorsque le mauvais paramètre est évalué (ligne pointillée bleue), les résultats sont proches de l'aléatoire (ligne rouge).	134
11.3	Montage expérimental pour la vérification de la spatialisation de l'élévation avec les méthodes VBAP et STAR couplées lors de l'utilisation de configurations avec plusieurs couronnes de haut-parleurs. (Ici chaque couronne est représentée par un seul haut-parleur).	135
11.4	Exemple de visualisation globale de l'interface GUI-STAR. Ici la visualisation du déterminant, avec les informations liées à la figure, le choix de l'échelle, ainsi que le filtre à appliquer.	142
11.5	Exemple de visualisation globale de l'interface GUI-STAR. Ici la comparaison des coefficients des méthodes STAR et VBAP en fonction de l'azimut modifiable via le slider sur le côté droit.	142
11.6	Exemple de visualisation globale de l'interface GUI-STAR. Ici une visualisation des HRTFs avec le choix de l'échelle, ainsi que de n'importe quel sujet de la base CIPIC, y compris la moyenne.	143
11.7	Choix de la configuration de haut-parleurs dans GUI-STAR.	143
11.8	Exportation de sons spatialisés avec GUI-STAR.	144
11.9	Photographie panoramique du dôme du SCRIME et couronnes de haut-parleurs.	147
11.10	Couronnes de haut-parleurs.	147
11.11	La salle de conférence et de diffusion du SCRIME Hémicyclia.	148
11.12	Le mannequin KEMAR du SCRIME.	149
11.13	Carte son Scarlett 2i2, utilisée pour l'ensemble des tests ne nécessitant pas plus de 2 haut-parleurs, ainsi que pour tous les enregistrements sur le mannequin KEMAR.	149
11.14	Micro utilisé pour les enregistrements ne devant pas être perturbés par une morphologie humaine.	150
11.15	Enceinte Genelec 8030, utilisée pour les diffusions et tests sonores hors du studio du SCRIME.	150
11.16	Premier système de prises de mesure par le mannequin KEMAR, constitué d'une structure en bois, avec un axe de rotation et des roues pour le déplacement.	151
11.17	Second système de prises de mesure, plus efficace, constitué d'un support de chaise de bureau et un support TV inclinable solide.	152
11.18	Logos des logiciels Matlab TM et GNU Octave.	152

Introduction

INTRODUCTION

Tous les jours, et à chaque seconde, nous évoluons dans un monde en plusieurs dimensions. Notre cerveau nous permet une intégration des informations reçues pour une représentation de notre environnement en 3 dimensions. Plusieurs repères sont possibles, mais les plus facilement accessibles sont la distance, la hauteur et l'azimut. Ayant subi des milliers (voire millions) d'années d'évolution dans ce milieu 3D, nous nous sommes adaptés à cet environnement. Évidemment, nous avons plus de mal à appréhender l'élévation que d'autres animaux pouvant voler. Il n'en reste pas moins que nous sommes particulièrement bien dotés pour construire, à partir de chacune de nos modalités sensorielles, nos perceptions et émotions. Il est donc tout naturel que des études menées par de nombreux chercheurs se soient intéressées à ces facultés, et particulièrement à nos différents sens. Pour ce qui est du repère dans l'espace, on utilise évidemment la vue, légèrement le toucher et également beaucoup l'ouïe.

Nous allons nous intéresser à cette dernière, primordiale pour déterminer notre position dans l'espace. En effet, quel que soit le stimulus sonore que l'on capte, nous le localisons. En tant que prédateur, cela permet de localiser une proie, et en tant que proie de localiser le danger, et, plus généralement de nos jours, de repérer une personne qui nous parle, un oiseau qui gazouille sur une branche, ou l'arrivée d'une voiture dans notre dos. Ce repérage de sources sonores se fait naturellement et de manière presque inconsciente (même si, en se concentrant, la précision de cette localisation peut être accrue); et un challenge aujourd'hui est de comprendre ces différents procédés avec précision.

Quel peut bien être le but de comprendre ces différents procédés de localisation? Bien évidemment la curiosité humaine et la soif de savoir, mais également la possibilité de soigner certaines maladies, du moins à l'origine. Aujourd'hui, nous évoluons de plus en plus vers la digitalisation, et il est nécessaire de comprendre ces différents phénomènes. C'est là que cette thèse vient se placer dans un premier temps, avec la localisation perceptive, essayant de recréer algorithmiquement ce que le cerveau effectue de manière instinctive à long-ueur de temps. Il existe donc certaines applications utilisant ce type de localisation développé ici, comme par exemple dans notre démarche, qui consiste à localiser les différentes enceintes d'une configuration et ainsi obtenir une calibration automatique du système de spatialisation.

Et nous voilà maintenant au point majeur de cette thèse : la spatialisation sonore. Là aussi, dans l'ère du numérique, les applications pour une spatialisation sonore ne manquent pas. Que ce soit les amateurs de vidéos ou de musiques, le fait d'avoir accès à un son 3D permettant une immersion totale est un point crucial; une autre utilisation nécessitant de plus en plus l'immersion et ayant donc recours à des spatialisateurs intégrés dans les simulateurs, que l'on peut étendre au domaine des jeux vidéos (les sons faisant maintenant partie intégrante des stratégies de game design); ou encore une utilisation dans les concerts où les artistes souhaitent de plus en plus contrôler l'aspect spatial et immersif de leur public (donner l'impression que l'on se trouve au cœur de l'orchestre par exemple). Toutes ces raisons expliquent l'intérêt de développer des méthodes de spatialisation.

Il est essentiel de comprendre le fonctionnement de l'humain pour la localisation, mais pas pour toutes les méthodes de spatialisation. Ainsi, certaines ont plutôt une approche physique pour recréer l'onde acoustique telle qu'elle devrait être, d'autres vont effectuer des interpolations mathématiques, et d'autres encore vont uniquement retransmettre des sources sonores enregistrées de manière spatiale. C'est ici que vient la notion de spatialisation sonore perceptive de la méthode STAR, Synthetic Transaural Audio Rendering, et l'intérêt de la thèse permettant d'établir un lien direct et étroit entre la localisation et la spatialisation. Les travaux effectués dans cette thèse sur la localisation perceptive se basent sur la thèse de Harald Viste, ainsi que sur certains prolongements développés dans la thèse de Joan Mouba; en reprenant le travail de Joan Mouba, cette thèse présente, après résolution d'imperfections, une implémentation 3D complète de la spatialisation confortée par la validation de la méthode.

La méthode STAR, expliquée tout au long de cette thèse, est une méthode qui se base sur la localisation humaine en recréant des indices perceptifs; se calquant sur le réel de la perception auditive humaine au quotidien. Ainsi, avec cette démarche initiale, la méthode prône une philosophie qui souhaite une mise en œuvre simple et pouvant prendre n'importe quel type de configuration de haut-parleurs (dans un premier temps, STAR a été conçue pour fonctionner sur une couronne dans le cas du 1D, puis a évolué pour s'appliquer

aux dômes, et a finalement été conçue pour fonctionner de manière optimale avec plusieurs couronnes de haut-parleurs ; le dispositif le plus simple de mise en œuvre). La méthode STAR doit également avoir un coût de calcul faible (permettant une spatialisation directe) et robuste quelles que soient les configurations de diffusion (représentant des conditions d'écoute réelle, avec du bruit, des réverbérations, *etc.*). C'est dans ces conditions d'utilisation souhaitées que les différents tests ont été menés. Elles correspondent également au dispositif du SCRIME (Studio de Création et de Recherche en Informatique et Musiques Expérimentales) qui a supporté logistiquement la thèse ; puisque la configuration type de STAR se base sur son dôme 3D qui possède à sa base une couronne de huit haut-parleurs (utilisés pour le 1D), expliquant l'utilisation d'une configuration octophonique dans ce travail. La philosophie de STAR nécessite également d'utiliser les indices acoustiques humains afin d'être perceptive (la spatialisation venant de la restitution de ces indices), ces indices provenant de modèles (donnant le S de la méthode pour Synthetic). Enfin, le placement des sources s'effectue de manière ponctuelle, jouées par les haut-parleurs l'encadrant (donnant le T de la méthode pour Transaural).

La première partie introduit les notions d'audition humaine ainsi que de spatialisation sonore utilisées pour ce travail de thèse, permettant d'avoir une compréhension du panel d'outils utilisés, sans proposer toutefois une étude exhaustive qui n'est pas requise ici. En effet, seules ces notions sont utiles comme support informatif pour comprendre les outils que nous utiliserons ; et reposent sur des études exhaustives réalisées par des experts dans chacun des domaines abordés dans cette partie.

La seconde partie reprend les travaux d'Harald Viste pour la localisation sonore de manière perceptive de l'azimut. Cette méthode a ici été simplifiée, puis testée dans différentes conditions, permettant de la valider et ainsi d'avoir également un outil de mesure objective pour l'azimut.

L'élévation et la distance ont été des axes de recherche de cette thèse. Ainsi, si la distance n'a pas abouti de par sa complexité, une approche exclusive et innovante est présentée pour la localisation de l'élévation. Cet axe, bien que présentant des résultats encourageants, nécessite un approfondissement.

La troisième partie constitue le cœur de cette thèse, présentant la méthode de spatialisation perceptive 3D STAR. Elle reprend les travaux de thèse de Joan Mouba pour l'azimut, pour résoudre les problèmes existants et en produisant différentes validations. L'élévation, quant à elle, est un axe neuf reposant sur les travaux de Kazuhiro Iida, qui permet de compléter la méthode STAR en restant dans une approche perceptive. Enfin, la distance, plus complexe, utilise les notions d'absorption du son par l'air et de perte de pression acoustique.

La dernière partie présente l'ensemble des tests pour la validation de la méthode STAR, effectués quasi exclusivement au cours de cette thèse. Sont ainsi exposés, d'une part les tests objectifs provenant de mesures, et d'autre part les tests subjectifs provenant d'évaluations sur des panels d'auditeurs.

REMERCIEMENTS

Je tiens dans un premier temps à remercier Sylvain Marchand, mon directeur de thèse, pour avoir encadré avec bienveillance ma thèse, en me donnant nombre de conseils qui m'ont permis de progresser et de finaliser cette thèse.

Dans un second temps, j'aimerais remercier les membres du SCRIME, et notamment ma co-directrice de thèse Myriam Desainte-Catherine, m'ayant permis d'avoir à la fois le matériel mais aussi des discussions et conseils. S'ajoutent également les deux laboratoires m'ayant accueilli, évidemment le L3i, mon laboratoire de thèse, mais aussi le LaBRI, qui m'a fourni un bureau et l'opportunité de travailler avec de nombreux collègues.

Dans un troisième temps j'aimerais remercier l'ensemble des personnes m'ayant permis d'effectuer ce travail, et notamment les personnes ayant accepté de participer aux différents tests de validation. Il est nécessaire que je remercie également l'ensemble des personnes m'ayant permis d'avoir une expérience de travail agréable ; tous mes collègues de travail et d'enseignement, les membres de l'administration, et mes élèves ; ainsi que ma famille et mes amis m'ayant fourni un cadre de vie exemplaire. Ceci avec une mention spéciale pour Raphaël Marczak, pour ses corrections et conseils au long des différentes étapes et articles ; et Laurence Méaux et Astrid Bellanger pour le temps passé à me faire un retour sur ce manuscrit. Enfin, je tiens à remercier l'ensemble des membres du jury, ainsi que les rapporteurs de cette thèse.

Première partie

Notions d'audition humaine et de spatialisation sonore

À propos de la partie – audition humaine et spatialisation sonore –

La première partie de cette thèse concerne l'audition humaine et la spatialisation sonore. Comme tout travail scientifique, il est important avant de se lancer dans une recherche - dans notre cas la spatialisation et localisation sonore perceptive - d'effectuer une étude préliminaire. Cette étude permet entre autre de connaître les travaux existants sur le même domaine, sur les domaines similaires mais aussi de regrouper les différentes informations nécessaires.

C'est pourquoi le tout premier point qu'aborde cette partie est l'audition humaine – dans le chapitre 1 –, qu'il est nécessaire de comprendre afin de pouvoir travailler sur la localisation et la spatialisation sonore. Il n'est bien évidemment pas nécessaire d'en connaître les détails comme ce que pourrait avoir besoin un médecin spécialiste ORL par exemple. Cependant, en liaison avec l'audition, il est important de mettre en avant les points clefs qui nous intéresseront dans la suite de ce travail, à savoir le fonctionnement de la localisation d'un son 3D par l'homme suivant les dimensions considérées ainsi que la précision de cette localisation – permettant un comparatif avec nos résultats.

Le second point abordé dans ce premier chapitre est intrinsèquement lié à l'audition humaine, et concerne les HRTFs/BRIRs, qui caractérisent les chemins acoustiques parcourus par les ondes sonores. Dans le cadre de notre recherche, ces chemins sont primordiaux, contenant l'ensemble des informations nécessaires à la spatialisation de notre méthode – entre autre les indices acoustiques transauraux. Ces chemins acoustiques, les bases de données les regroupant, ainsi que les indices acoustiques sont donc présentés.

Le second chapitre de cette partie concerne les méthodes de spatialisation sonore existantes. Ces méthodes nous intéressent, à la fois de par leurs fonctionnements, mais aussi à titre comparatif. Les différentes méthodes utilisées au cours de ce travail sont donc présentées.

Cette partie regroupe essentiellement des informations nécessaires à la compréhension des parties suivantes, ainsi que les informations qui seront utilisées pour le développement et l'exploitation des travaux présentés dans cette thèse. Un apport de ce travail de thèse est néanmoins présent dans le traitement des données pour leur visualisation, le modèle d'ITDs et ILDs développé pour la méthode STAR, ainsi que la création d'une base de données BRIR – SCRIME .

Chapitre 1

Quelques notions d'audition et de localisation humaine

Introduction : Si l'on veut travailler sur la localisation et la spatialisation sonore, il est primordial d'avoir certaines bases sur le fonctionnement de l'audition humaine. Ce chapitre introduit le fonctionnement du système auditif humain dans un espace 3D, donnant les bases nécessaires pour la compréhension de cette thèse.

Une fois ces bases mises en place, les outils de localisation sonore par l'homme sur lesquels repose le côté perceptif de la méthode STAR¹ sont décrits.

Enfin, ce chapitre s'attardera sur les HRTFs (Head Related Transfer Functions) et les indices acoustiques qu'ils contiennent, présentant le principal outil utilisé pour l'approche perceptive de la méthode STAR.

a. STAR (Synthetic Transaural Audio Rendering) est le nom de la méthode de spatialisation développée au cours de cette thèse. Il résume les notions fondamentales liées : l'aspect synthétique apporté par les modèles, et transaural par la considération des chemins acoustiques atteignant chacune des deux oreilles.

1.1 L'oreille et l'acquisition du son par l'homme

Chaque source sonore est modifiée entre l'émission et la réception en signaux électriques par notre cerveau, lui permettant d'en déduire des informations telles que la provenance de ces sources. Le principal outil humain pour l'acquisition d'un son est l'oreille humaine, Fig. 1.1. Anatomiquement, on distingue trois parties : l'oreille externe, l'oreille moyenne et l'oreille interne, qui permettent le processus d'acquisition du son et présentées brièvement ci-dessous.

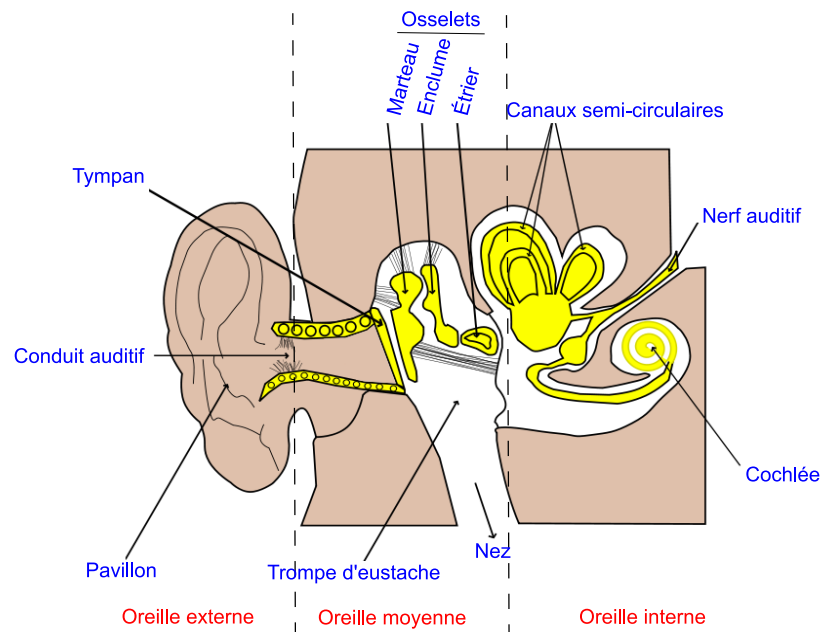


FIGURE 1.1 – Schéma général de l'oreille humaine (Wikimédia).

L'oreille externe est tout d'abord constituée du pavillon qui va guider les ondes sonores vers le conduit auditif, provoquant des modifications (formation de nœuds et de pics dans les HRTFs utilisées notamment pour la localisation en élévation). Le conduit auditif va permettre, de par son rétrécissement de diamètre, d'amplifier l'onde sonore pour le tympan.

Il est important de faire un petit aparté sur ces notions de pics et nœuds, qui seront beaucoup utilisés dans cette thèse. On appellera pic, un échantillon de données qui est sur un intervalle donné plus grand que ses deux échantillons voisins (maximum local dans le spectre), et nœuds plus petits (minimum local dans le spectre). En se rapportant à la Fig. 1.4, on observe les principaux pics à 4410, 9500, 13000, et 20000 Hz, et les principaux nœuds à 8820, 11000, 17000 Hz.

L'oreille moyenne débute au niveau du tympan, qui va vibrer lors de l'arrivée de l'onde sonore amplifiée par le conduit auditif, transmettant cette vibration au triptyque marteau/enclume/étrier. Ce dernier est fixé à la fenêtre ovale, interface entre l'oreille moyenne et interne, et joue également un rôle d'amplificateur.

L'oreille interne est constituée des canaux semi-circulaires, participant au sens de l'équilibre, et de la cochlée, organe récepteur de l'audition enroulé sur lui-même. La vibration amplifiée par la fenêtre ovale est transmise dans un liquide appelé la périlymphe contenue dans le canal entourant la base de la cochlée, puis se propage le long de la cochlée pour arriver à la membrane basilaire, sur laquelle se trouvent les cellules ciliées internes. Ce sont ces cellules, qui, en vibrant, vont entrer en contact avec une membrane (la membrane tectoriale) et libérer un neurotransmetteur, informant, via les fibres du nerf auditif, le cerveau de l'acquisition d'un son.

1.2 La localisation du son par l'homme

Le son, entre son émission et sa réception par le cerveau, est obligatoirement déformé, que ce soit causé par l'environnement dans lequel il se propage, ou par la morphologie humaine. De plus, l'être humain possède deux oreilles, se situant de part et d'autre de la tête, et ne recevant donc pas le même signal. Cette différence entre les deux oreilles est appelée différence interaurale. Les déformations du son ainsi que les différences interaurales jouent un rôle primordial pour la localisation sonore humaine.

Cette partie présente l'ensemble des processus utilisés par l'homme pour localiser une source sonore, et établir les liens avec le fonctionnement de la méthode STAR.

D'autres phénomènes interviennent également dans la localisation d'un son par l'homme, mais ne seront pas exposés ici (ces phénomènes relevant de la psychoacoustique). On peut citer par exemple :

- la vue ; le cerveau recevant des stimulus conflictuels, un biais se crée (Kohlrausch and van de Par (2005)).
- les mouvements de la tête, qu'ils soient conscients ou inconscients, augmentent grandement la précision de la localisation d'un son par l'homme (Wallach (1940), Young (1931)).

1.2.1 Système de coordonnées et conditions

Dans le cadre de la thèse, bien que d'autres systèmes de coordonnées tels que le système de coordonnées interaural-polaire-axe soient souvent utilisés pour la perception de direction de sons, il a été choisi d'utiliser majoritairement le système sphérique illustré sur la Fig. 1.2 pour sa simplicité (distance d , azimut θ , centrées sur la tête de l'auditeur, élévation ϕ), Exception faite pour la localisation du son en élévation sur la base CIPIC utilisant alors le système interaural-polaire.

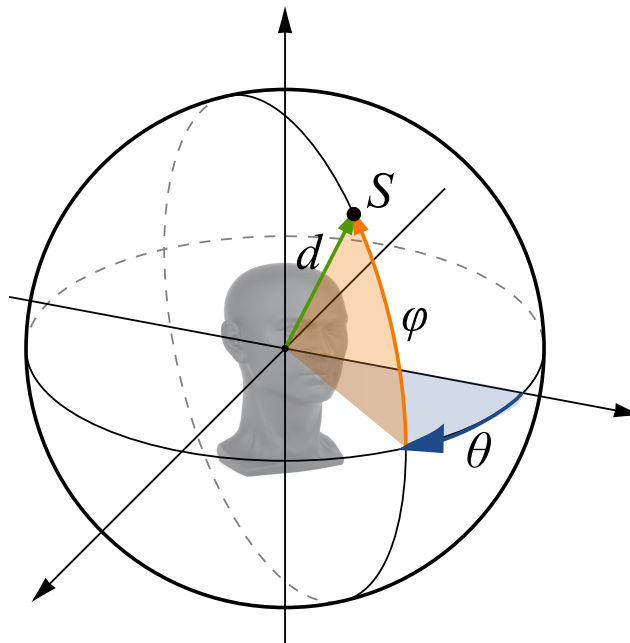


FIGURE 1.2 – Système de coordonnées utilisées dans le cadre de la thèse (Marchand (2020)).

1.2.2 La localisation de l'azimut utilisée dans cette thèse

Localisation en azimut, utilisation des indices binauraux

Pour cette dimension, l'utilisation des différences interaurales ¹ a été mise en avant dans les années 1900, par William Strutt (Lord Rayleigh) (1877, 1907). Les relations entre ces différences et la localisation en azimut pour un signal sonore seront, quant à elles, quantifiées dans les années 1960, rapportant des valeurs extrêmes (azimut de 90°) de 10 dB pour les ILDs (Interaural Level Differences) et de 1 ms pour les ITDs (Interaural Time Differences) Sayers (1964). Les indices acoustiques ainsi que les HRTFs permettant la localisation en azimut sont essentiels dans la méthode STAR, une partie leur est donc dédiée en Sec. 1.3.

Performance humaine de la localisation en azimut

La précision de l'être humain pour la localisation en azimut est de l'ordre de 5° en moyenne, Letowski and Letowski (2011); Risoud et al. (2018). Certaines recherches montrent également que la localisation par

1. Les indices interauraux seront abordés en détail dans une partie dédiée.

les auditeurs est meilleure sur l'avant/arrière, que sur les côtés.

1.2.3 La localisation de la distance utilisée dans cette thèse

La distance est la seconde des dimensions dans la localisation d'un son dans l'espace. On peut différencier la localisation dans l'espace péripersonnel (< 1 m) et extrapersonnel (> 1 m).

Localiser la distance d'une source sonore est un processus compliqué pour les êtres humains et repose sur une grande diversité de variables. Kolarik et al. (2016) fournissent une étude complète de la perception auditive de la distance par l'être humain. Cette partie résume les principales variables jouant un rôle dans la localisation de la distance d'une source sonore.

Niveau de Pression Acoustique SPL

C'est le niveau global du son aux oreilles. Il va augmenter lorsque la source se rapprochera et diminuer lorsqu'elle s'éloignera; ceci pouvant créer une confusion entre la distance et le niveau de diffusion (Stevens and Guirao (1962)). Autrement dit, pour une même source sonore, la pression acoustique diminue lorsque la distance augmente et inversement. La littérature considère que le SPL suit une loi carré inversement proportionnelle à la distance, ceci en conditions anéchoïques (-6 dB à chaque fois que la distance double). Cependant, pour des environnements réels, la valeur moyenne est d'environ -4 dB (Kolarik et al. (2016)) expliquée par la présence de réverbérations.

Le SPL affecte grandement l'image sonore (Gardner (1969)) mais ne se suffit pas à lui-même pour localiser la distance d'une source. En effet, bien qu'il soit un bon indice de spatialisation, il permet uniquement le jugement relatif de la distance. Une exception est celle où on a connaissance du niveau de référence de la source, cas du signal vocal contenant cette information intrinsèque, rendant ainsi possible la localisation avec uniquement le niveau sonore (Brungart and Scott (1962)).

Le SPL est donc essentiellement un indice acoustique de jugement relatif de la distance, et est donc intéressant pour la méthode STAR, permettant de spatialiser de manière relative différentes sources.

Rapport champ direct à champ réverbéré (Direct-to-Reverberant Ratio , DRR)

C'est une autre variable importante pour la localisation d'un son en distance. Lorsque l'on réceptionne un son, nous recevons à la fois l'onde directe et toutes celles réverbérées par l'environnement de propagation (le sol, les murs, le mobilier *etc.*). L'énergie du champ direct décroît avec la distance alors que les champs réverbérés sont eux plus ou moins constants. Ainsi, plus la distance séparant la source de l'auditeur est grande, plus le DRR sera faible (du fait de la faible intensité du champ direct). Réciproquement, lorsque la source sera proche, le champ direct sera important et donc le DRR fort. Cet indice est très important dans les lieux clos (Mershon and King (1975)). Cependant, le DRR reste peu performant en champ libre et est, de plus, un mauvais indice de localisation sonore relatif, nécessitant une combinaison avec le niveau sonore pour être précis (Zahorik and Wightman (2001a)).

Le DRR est donc un indice essentiellement utilisé pour les lieux clos, qui dépend grandement de la taille et du coefficient d'absorption de l'environnement, ce qui en fait un indice non polyvalent, ne correspondant pas à la philosophie de STAR. De plus, la méthode STAR est actuellement conçue pour le champ libre, cas ne prenant donc pas en compte les réverbérations. Cet indice ne sera donc pas intéressant pour la méthode STAR.

Contenu spectral et indices binauraux

Le contenu spectral et les indices binauraux affectent également la localisation de la distance.

Dans le champ proche ou péripersonnel, la diffraction de l'onde sonore par tête va dépendre de la fréquence ainsi que de la distance entre la source et l'auditeur; tout comme les indices binauraux qui sont également des indices importants de localisation pour les champs très proches (Brungart and Rabinowitz (1999)), de par une ILD dépendant de la distance. En champ proche, on peut également citer un phénomène

1.2. LA LOCALISATION DU SON PAR L'HOMME

de parallaxe, l'angle entre la source et le plan sagittal de l'oreille gauche est différent de celui de l'oreille droite ; et cette différence augmente proportionnellement au rapprochement de la source.

Dans le champ lointain (> 15 m), l'absorption acoustique de l'air est utilisée (Blauert (1997)), utilisant la propriété du son à avoir les hautes fréquences davantage absorbées que les basses fréquences.

La méthode STAR peut donc utiliser le contenu spectral pour spatialiser des sources sonores lointaines, grâce à l'absorption du son par l'air. Elle reproduit déjà les ILDs via un modèle mais qui ne permet pas actuellement une adaptation en fonction de la distance. De la même manière, le phénomène de parallaxe est très spécifique et ne correspond pas à la volonté de simplicité de STAR.

Performance humaine de la localisation en distance

L'étude de la performance des humains pour la localisation en distance est bien moins étudiée que la localisation pour l'élévation et l'azimut. La raison en est que c'est la plus compliquée, l'incertitude dépendant fortement du stimulus sonore et des conditions. De manière générale, la distance sera sur-estimée pour des sources proches et sous-estimée pour des sources lointaines (Kearney et al. (2015); Zahorik and Wightman (2001b)).

1.2.4 La localisation de l'élévation utilisée dans cette thèse

L'être humain est constitué d'une forme complexe (tête et torse), ce qui engendre des irrégularités complexes dans les ondes acoustiques incidentes arrivant aux oreilles. Les fonctions de transferts caractérisant le trajet de l'onde acoustique sont donc très différentes en fonction de la provenance de la source. Le pavillon de l'oreille joue notamment un rôle prépondérant dans les déformations sur les HRTFs pour la localisation en élévation (Butler (1969)), et une étude complète du rôle du pavillon est donnée par (Moore (2013)).

Les indices spectraux pour la localisation de l'élévation

Dans le cadre de la méthode STAR, il s'est posé la question des indices utilisés pour la localisation verticale. Des études ont été menées depuis les années 1970 afin d'identifier ces indices spectraux, et ainsi de connaître quelle partie du spectre jouait un rôle important pour cette localisation verticale. Hebrank et Wright ont montré que les composantes de fréquence au-dessus de 16 kHz et au-dessous de 3,8 kHz n'affectent pas la précision de la localisation de l'élévation (Hebrank and Wright (1974)). Il est maintenant bien connu que les pics et nœuds spectraux contribuent à la perception de l'angle vertical (Hebrank and Wright (1974); Musicant and Butler (1984)); ceux-ci étant en grande partie formés par la forme du pavillon de l'oreille. Plus récemment, Iida et al. (2007) ont proposé un modèle permettant de localiser un son en utilisant un pic et deux nœuds. Le pic principal situé vers 4000 Hz étant constant, indépendamment de l'élévation verticale de la source, et les deux principaux nœuds N1 et N2, sont montrés comme suffisants pour une bonne localisation en élévation.

La méthode STAR se base sur ce modèle afin d'obtenir des indices spectraux pour la spatialisation verticale. Le chapitre 4 s'inspire également de ces travaux afin de proposer une méthode de localisation perceptive.

Performance humaine de la localisation en élévation

La localisation en élévation utilise des indices monauraux, contrairement à l'azimut qui utilise des indices binauraux, raison pour laquelle la précision est moindre, et va dépendre grandement du stimulus. Ainsi, par exemple, pour une voix familière, la précision sera d'environ 9° ; alors que pour une voix inconnue, la précision sera d'environ 17° (Blauert (1969)); ce qui est expliqué par le fait que si l'auditeur a connaissance du signal qu'il devrait recevoir, il peut comparer le filtre du signal reçu avec celui dont il a connaissance. Wettschurek (1970) a aussi montré que pour un bruit blanc, la précision est d'environ 4° . Ce très bon résultat peut être expliqué par la présence de l'ensemble du spectre, permettant au cerveau une meilleure utilisation des indices spectraux.

1.3 Fonction de transfert relative à la tête (HRTF)

1.3.1 Théorie de base sur les HRTFs

En champ libre, les ondes sonores, lors de leurs parcours entre la source d'émission et la réception par l'auditeur, sont affectées par différents facteurs. Ceux-ci sont les lobes d'oreilles, le torse et la tête de l'auditeur.

Les HRTFs caractérisent les changements physiques dans le domaine spectral dus à ces déformations.

Les HRIRs

Les HRIRs (Head Related Impulse Responses) sont l'expression temporelle des HRTFs. La Fig. 1.3 montre une réponse impulsionnelle provenant de la base CIPIC Algazi et al. (2001)¹ (ici sujet numéro $s = 4$), avec la source sonore située face au sujet ($\theta = 0^\circ$) et légèrement au-dessus de lui ($\phi = 25^\circ$).

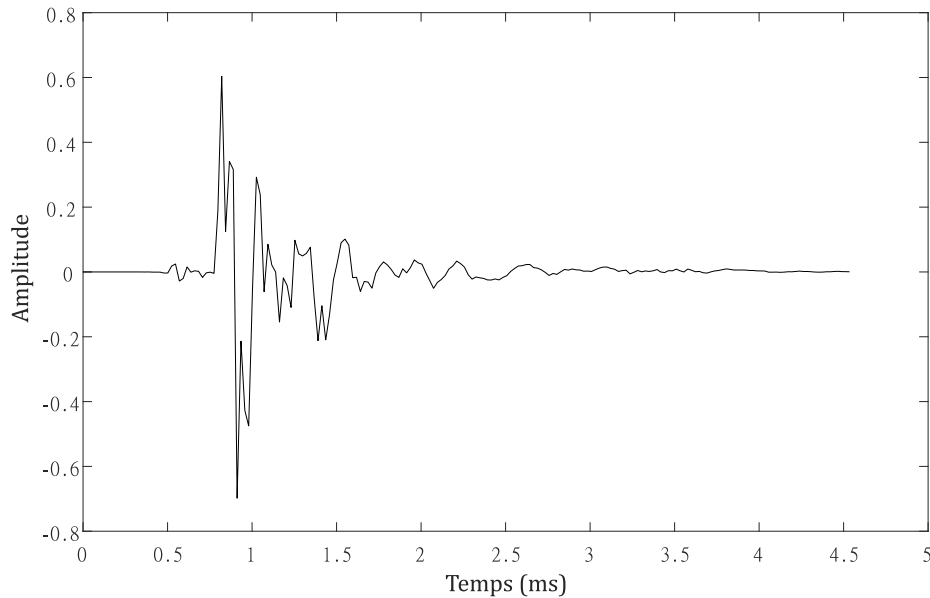


FIGURE 1.3 – Exemple d'une HRIR, provenant ici du sujet $s = 4$ de la base CIPIC, pour une direction $\theta = 0^\circ$ et une élévation $\phi = 25^\circ$.

Les HRTFs

La Fig. 1.4, représente le spectre d'une HRTF provenant de la base CIPIC, Algazi et al. (2001), (ici sujet numéro $s = 4$), avec la source sonore située face au sujet ($\theta = 0^\circ$) et légèrement au-dessus de lui ($\phi = 25^\circ$). 0 dB représentant l'amplitude d'un son sans déformation, les pics indiquant une augmentation des SPLs due à l'ensemble tête/torse/oreilles, et les nœuds une diminution.

1. La base CIPIC (Center for Image Processing and Integrated Computing) est une série d'enregistrement de réponse impulsionnelle effectuée avec de nombreux sujets et pour différents azimuts et élévations.

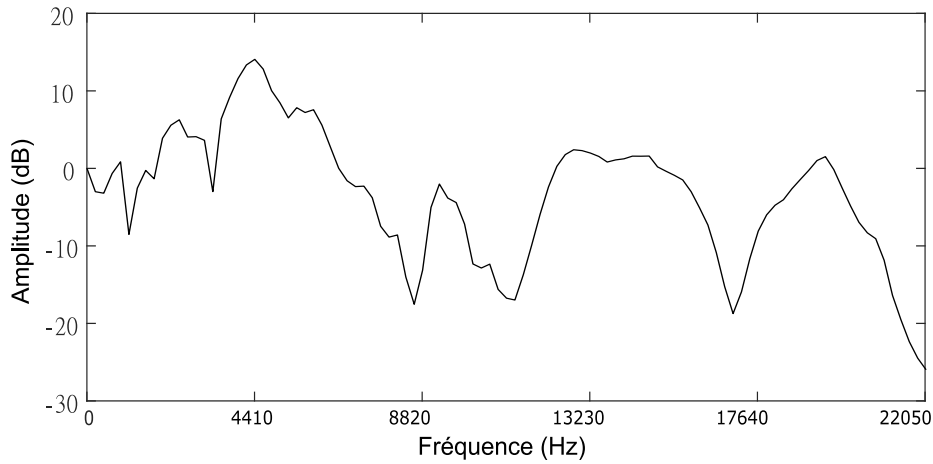


FIGURE 1.4 – Exemple d’une HRTF, provenant ici du sujet $s = 4$ de la base CIPIC, pour une direction $\theta = 0^\circ$ et une élévation $\phi = 25^\circ$.

Les HRTFs sont définies comme suivant l’Eqn. 1.1 :

$$Hrtf_{l,r}(s, \theta, \phi, d, f) = G_{l,r}(s, \theta, \phi, d, f) / F(\theta, \phi, d, f) \quad (1.1)$$

Où G est la fonction de transfert de la source sonore à l’entrée des oreilles, et F la fonction de transfert entre la source sonore et le centre de la tête de l’auditeur.

L’Eqn. 1.1 montre que les HRTFs dépendent des trois dimensions de l’espace.

À noter que les Fig. 1.3 et 1.4 sont la représentation pour l’oreille gauche ou l’oreille droite. En effet, l’azimut $\theta = 0^\circ$ a été choisi pour la symétrie gauche/droite.

Les BRIRs

Les BRIRs (Binaural Room Impulse Responses) sont l’équivalent des HRTFs lorsque l’on ne se trouve pas en champ libre. Elles prennent donc en compte les caractéristiques de la pièce dans lesquelles elles sont acquises (réverbérations, bruits, etc.).

Les BRIRs sont donc plus représentatives de la réalité que les HRTFs, et seront utilisées dans ce travail de thèse afin de se placer en conditions réelles.

1.3.2 Les bases de données d’HRTFs et de BRIRs

Les bases de données HRTF

Dans les recherches pour la spatialisation perceptive, la méthode STAR se place dans un cadre perceptif. Aussi, l’ensemble des modèles proposés sont calculés à partir de bases de données (dans notre cas CIPIC, Algazi et al. (2001)). Il existe de nombreuses bases de données, montrant un intérêt, et pouvant être utilisées à la place de la base CIPIC pour la création des modèles. Un résumé des principales bases de données HRTF, ainsi que de leurs caractéristiques sont listées dans le Tab. 1.1, avec :

- La base HRTF du centre de traitement d’images et d’informatique intégrée (CIPIC) proposée par Algazi et al. (2001),
- La base HRTF proposée par l’Austrian Research Institute (2021),
- La base HRTF de l’institut de recherche de communications électriques proposée par Watanabe et al. (2014),
- La base HRTF de l’institut de technologie de Chiba du Spatial Hearing Laboratory (2017),
- La base HRTF de l’institut de Recherche et Coordination Acoustique/Musique (2021),
- Et la base du Massachusetts Institute of Technology Media Lab Machine Listening Group (1994)

1.3. FONCTION DE TRANSFERT RELATIVE À LA TÊTE (HRTF)

	CIPIC	ARI	RIEC	CIT	IRCAM	MIT
Sujets	45	200	105	61	50	KEMAR
Signal source	MESM ¹	MESM	Log Sweep	Swept-Sine	Log sweep	MLS
Échantillons	200	256	512	512	8192	512
F_e	44100	48000	48000	48000	44100	44100
Directions	1250	1550	865	7-148 ²	187	710
Microphones	ER-7C	KE-4-211-2	FG3329	WM64AT102	FG3329	ER-11
Enceintes	Acoustimass	10 BGS	FE83E	Panasonic	System600	ROP 7

TABLE 1.1 – Tableau comparatif des principales bases de données d’HRTFs.

Afin d’acquérir ces bases de données, les micros miniatures sont positionnés à l’entrée du canal auditif de chaque oreille des sujets, ou sont équipés dans des mannequins anthropométriques tels que le mannequin KEMAR (Knowles Electronics Mannequin for Acoustics Research). Ces mannequins sont utiles car plus faciles de mise en œuvre que de vrais sujets et donnent théoriquement des résultats identiques à un être humain de taille moyenne. De plus, ces mannequins étant normés, il est possible de comparer les jeux de données (les résultats devant être identiques si l’on suit le même protocole expérimental). Les caractéristiques détaillées du mannequin KEMAR (utilisé par la base CIPIC ainsi que pour nos enregistrements) sont détaillées par Burkhard and Sachs (1975).

Les caractéristiques de ces bases permettent en effet d’effectuer des choix judicieux en fonction du matériel disponible et du but recherché. C’est ainsi que pour la méthode STAR, le choix s’est porté sur la base CIPIC. En effet, d’anciens travaux utilisaient déjà cette base, et le matériel du SCRIME³ est similaire (même mannequin et mêmes microphones), et les fréquences d’échantillonnage des enregistrements identiques.

Les bases de données BRIR

Tout comme les HRTFs, il existe un certain nombre de bases de données BRIR, qui sont cependant moins connues et utilisées que celles des HRTFs. La principale différence entre ces bases est que les HRTFs caractérisent uniquement les transformations causées par la morphologie de l’auditeur alors que les BRIRs utilisent des systèmes de simulations anthropomorphiques pour caractériser à la fois la morphologie et les caractéristiques de la pièce. C’est aussi pour cela que les temps d’acquisitions pour les BRIRs sont plus importants que pour les HRTFs (on souhaite acquérir le signal direct ainsi que toutes les réverbérations, alors que pour les HRTFs, seul le signal direct est acquis, les réverbérations étant en théorie nulles).

Les quatre bases de données de BRIR citées dans le Tab. 1.2 ont montré un intérêt pour les expérimentations de la méthode STAR. En effet, le fait que les BRIRs soient le reflet de la réalité correspond à la philosophie de STAR. Chacune de ces bases a un intérêt, mais seuls les résultats de la base de Rostok et du SCRIME ont été utilisés.

- La base BRIR de Rostok proposée par Erbes et al. (2015), propose 7 configurations de pièces (en disposant à différents endroits des éléments absorbants). Cette base est intéressante car elle propose une gamme de pièces différentes avec de nombreux azimuts.
- La base BRIR du SCRIME ne possède qu’une configuration de pièce. Elle a été conçue dans l’optique de compléter la base BRIR de Rostok qui ne proposait pas différentes élévations et distances.
- La base Air (Aachen Impulse Response) proposée par Jeub et al. (2009) propose quatre pièces distinctes avec différentes positions dans celles-ci, et est donc intéressante pour la distance. Elle fait cependant doublon avec les deux précédentes bases, et n’est donc pas exploitée.
- Enfin la base Like-Apartement proposée par Winter et al. (2016) présente un appartement avec plusieurs positions d’enregistrement et d’émission, pouvant être assimilé à 4 pièces, ainsi qu’un enregistrement de cheminement au sein de cet espace. Il a été préféré ici d’utiliser la base de Rostok afin de

1. MESM pour Multiple Exponential Sweep Method est une méthode de sweep mise au point afin d’accélérer la mesure d’HRTF, procédé coûteux en temps Majdak et al. (2007).

2. Le nombre de directions dépend des sujets, allant de 7 à 148.

3. Le SCRIME (Studio de Création et de Recherche en Informatique et Musiques Expérimentales) est une structure partenaire au cours de la thèse, qui a fourni des locaux d’enregistrements, ainsi que du matériels et diverses ressources.

1.3. FONCTION DE TRANSFERT RELATIVE À LA TÊTE (HRTF)

comparer différentes pièces.

	Rostok	SCRIME	Air	Like-apartement
Système d'enregistrement	KEMAR	KEMAR	HMS2 Head	KEMAR
Type de signal source	Sweep Linéaire	Bruits blancs	MLS	Log sweep
Taille du signal	2^{20}	2^{18}	2^{15}	2^{19}
Fréquence d'échantillonnage	44100	44100	16000	44100
Nombre de directions	81	405+distance	1+distance	79+chemin
Microphones utilisés	G.R.A.S. 40AG	G.R.A.S. 40AG	RME Octamic II	RA0241
Nombres de pièces	7	1	4	4
Enceintes utilisées	Neumann KH 120 A	Genelec 8040	Genelec 8130	Genelec 8020

TABLE 1.2 – Tableau comparatif des principales bases de données de BRIRs.

1.3.3 Les HRTFs de la base CIPIC et leurs utilisations dans STAR

Nous avons vu dans la Sec. 1.3 ce que sont les HRTFs, mais aussi qu'elles contiennent de nombreuses informations pour la localisation sonore humaine. La méthode STAR se base dessus pour la conception des différents modèles (notamment des HRTFs provenant de la base CIPIC Sec. 1.3.2). Il est donc intéressant de visualiser et comparer ces données.

Visualisation des HRTFs de la base CIPIC

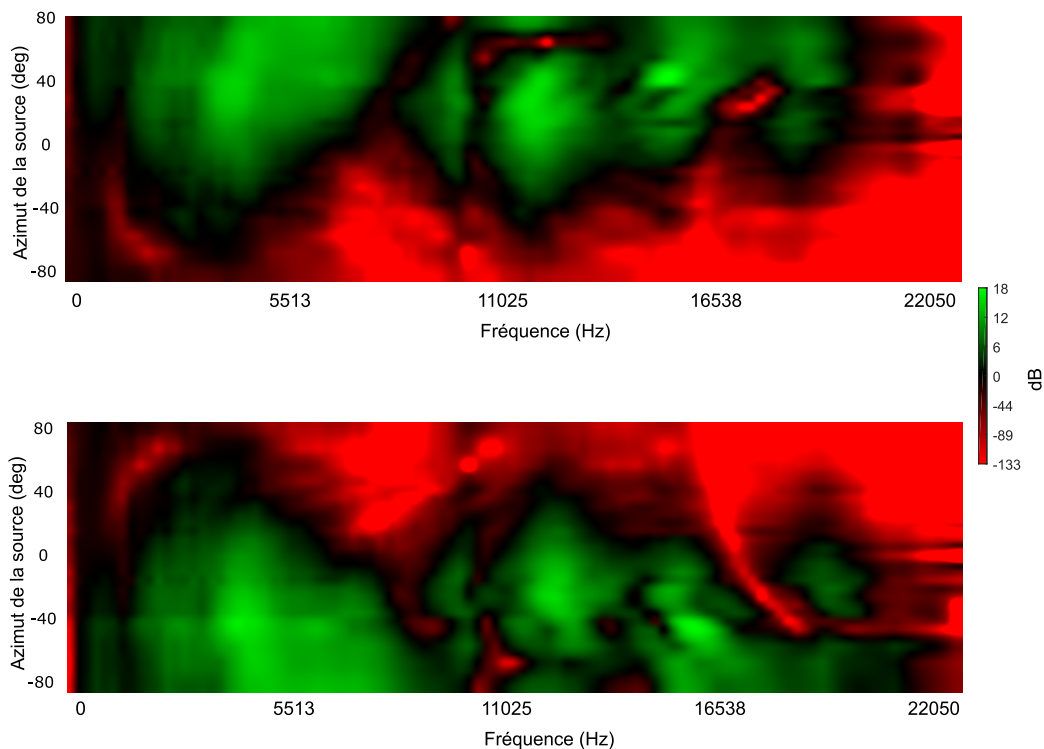


FIGURE 1.5 – HRTF d'un sujet unique - ici Kemar large pinnae, pour une élévation $\phi = 0^\circ$.

La Fig. 1.5 présente les HRTFs gauche et droite ¹ pour un sujet unique (ici il a été choisi celles provenant du mannequin KEMAR, le même que celui dont le SCRIME dispose (présenté en Sec. 11.8.1).

1. Sauf précision, la gauche est représentée dans la figure du haut, et la droite dans celle du bas.

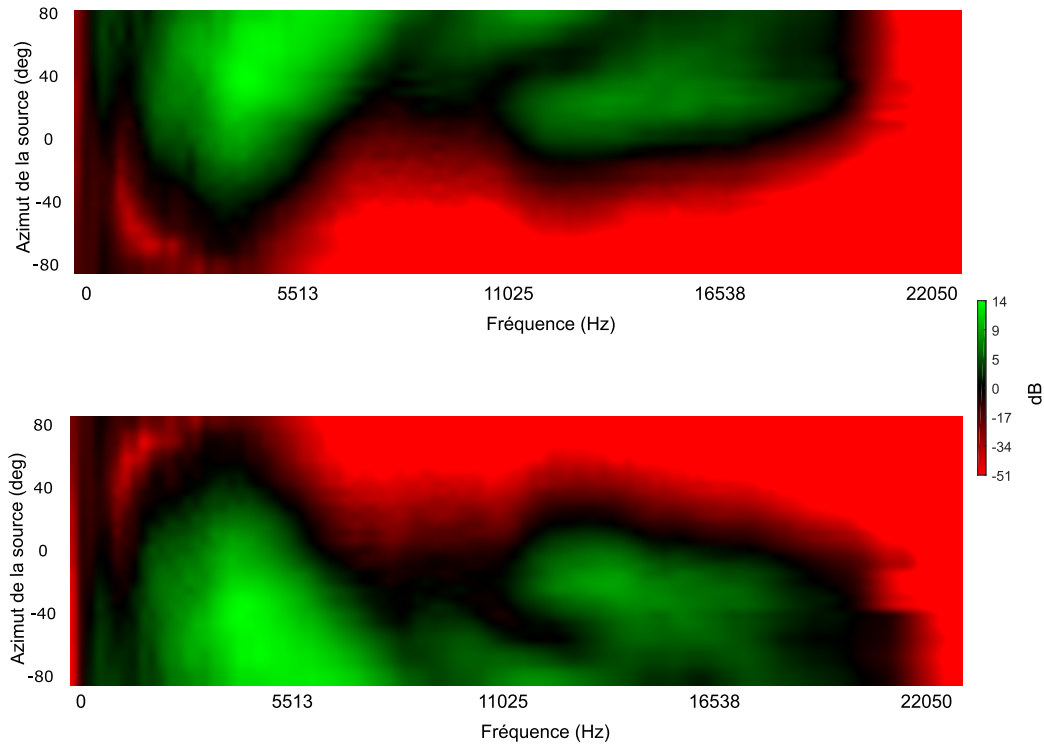


FIGURE 1.6 – *HRTF de la moyenne de la base de données CIPIC, pour une élévation $\phi = 0^\circ$.*

La Fig. 1.6 représente la moyenne de tous les HRTFs de la base CIPIC, pour tous les sujets. Ces deux figures donnent une visualisation simple de l'ensemble des HRTFs pour une même élévation, l'axe des abscisses présentant la variation fréquentielle, et l'axe des ordonnées les azimuts de -80 à 80 degrés.

C'est cette moyenne d'HRTFs provenant de la base CIPIC qui sera la plus utilisée par la suite. En effet, la démarche de STAR est d'obtenir une méthode de spatialisation globalement bonne pour un grand nombre de spectateurs, il n'est donc pas judicieux d'utiliser des HRTFs individualisées¹.

Si on compare les Fig. 1.5, et 1.6, on observe que les HRTFs d'un individu isolé sont très ressemblantes à la moyenne des HRTFs de la base. Cependant, la moyenne a un aspect moins chaotique et des valeurs extrêmes moins élevées. Deux phénomènes allant dans le sens de la méthode STAR et permettant de conforter le choix de prendre la moyenne plutôt que des HRTFs individualisées.

Visualisation de l'énergie des HRTFs

Il peut être intéressant pour la comparaison des données entre la moyenne de la base CIPIC de regarder l'énergie contenue dans les HRTFs (correspondant au produit des HRTFs gauche et droit). Les Fig. 1.7 et 1.8 montrent des aspects très similaires bien que plus chaotiques pour KEMAR, avec deux pics d'énergies ne dépendant pas de l'azimut, et présents à environ 3900 Hz et 1200 Hz. Nous noterons également qu'en moyenne l'énergie est majoritairement située aux alentours de 1.

1. Bien que ce ne soit pas la philosophie de la méthode, il est très facile de relancer STAR en utilisant des HRTFs personnalisées afin d'obtenir un meilleur rendu pour une utilisation personnelle par exemple.

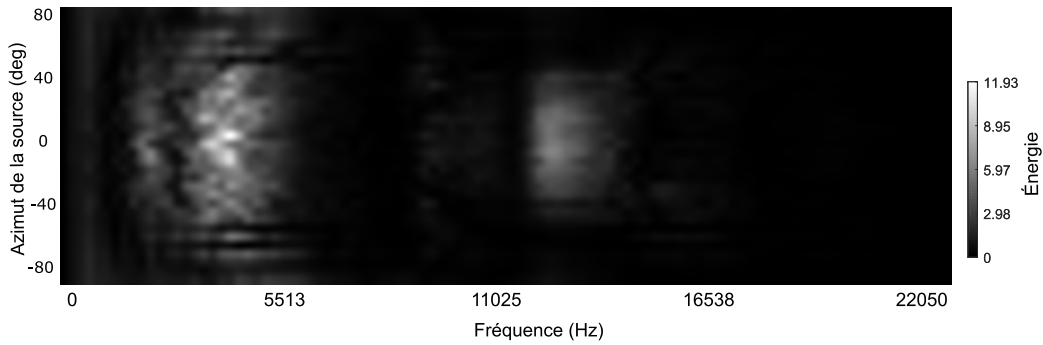


FIGURE 1.7 – Énergie des chemins acoustiques pour KEMAR large pinnae.

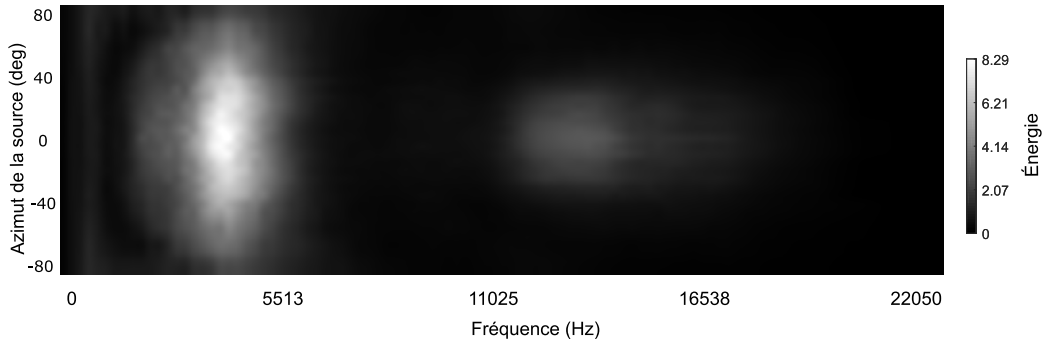


FIGURE 1.8 – Énergie des chemins acoustiques pour la moyenne de la base CIPIC.

Ces deux pics sont intéressants, comme nous le verrons dans la suite, notamment pour la spatialisation en élévation. En effet, comme présenté en Sec. 4.1, les pics d'énergies sont dus à la forme de l'oreille, et sont fixes pour chaque individu contrairement aux nœuds. Ces pics jouent un rôle important dans la localisation de l'élévation.

1.4 Indices acoustiques interauraux

Les indices acoustiques interauraux ont déjà été abordés. Ces indices jouent un rôle primordial dans la méthode STAR et sont repris en détail dans cette section. En plus de définir et présenter ces indices, cette section présentera également les modèles synthétiques d'ITDs et ILDs de la méthode STAR. Le système auditif humain utilise également les différences interaurales de manière dynamique, puisqu'en effet, une multitude de positions dans l'espace peuvent engendrer des ITDs et ILDs similaires (von Hornbostel and Wertheimer (1920)), que l'on nomme cônes de confusion. L'homme va alors s'affranchir de ces cônes à l'aide de mouvements de la tête. Ces indices dynamiques n'étant pas utiles pour la méthode STAR, ils ne seront pas approfondis dans cette section.

1.4.1 Généralités

L'onde sonore émise arrive aux oreilles avec des instants et des intensités différents, dûs aux trajets différents effectués.

En supposant une onde plane (supposée à une distance infinie), il est possible de caractériser les trajets acoustiques gauches et droits.

1.4. INDICES ACOUSTIQUES INTERAURAUX

Ainsi, on peut poser l'Eqn. 1.2, avec $\Delta_d = r.\sin(\theta)$, r étant le rayon de la tête, d la distance de la source et G et D désignant l'oreille gauche et droite, présenté sur la Fig. 1.9.

$$\begin{aligned} d_G &= d - \Delta_d \\ d_D &= d + \Delta_d \end{aligned} \quad (1.2)$$

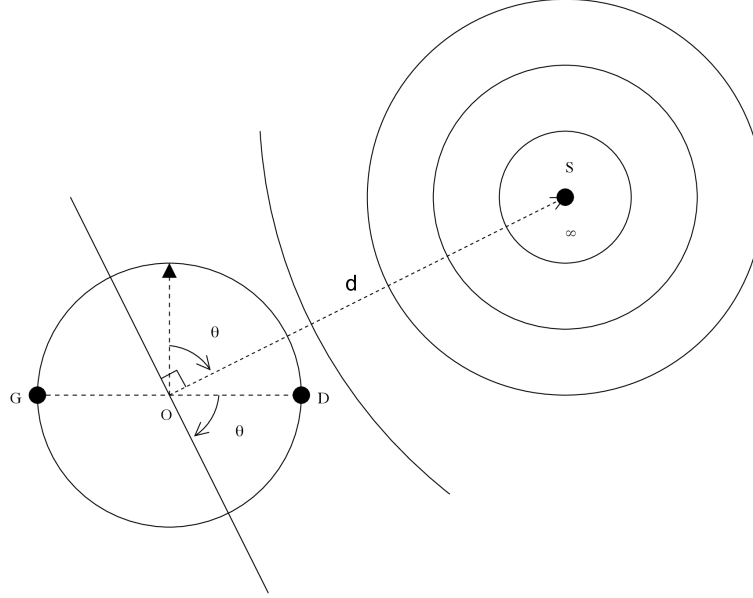


FIGURE 1.9 – Schéma du trajet d'une onde sonore entre sa source et les oreilles de l'auditeur (Marchand (2020)).

On obtient ainsi la différence de temps entre les deux oreilles sans prise en compte de la tête de l'auditeur, proposée par von Hornbostel and Wertheimer (1920) :

$$\Delta_\theta = (d_G - d_D)/c = 2\Delta_d/c = 2r/c.\sin(\theta) \quad (1.3)$$

En réalité, la tête étant un obstacle à l'onde acoustique en considérant la tête ronde, Woodworth and Schlosberg (1954) proposent l'équation suivante.

$$\Delta_T(\theta, f) = r/c.(\sin(\theta) + \theta) \quad (1.4)$$

La tête n'étant pas ronde, il est nécessaire d'appliquer un facteur correctif en fonction de la fréquence (Wightman and Kistler (1954)).

Mouba and Marchand (2006), puis Méaux and Marchand (2019) ont montré qu'il était possible de simplifier ce modèle par l'Eqn. 1.11 présentée dans la Sec. 1.4.3; Δ_θ étant l'ITD.

Enfin l'intensité sonore I étant inversement proportionnelle au carré de la distance, on obtient :

$$\begin{aligned} \Delta_L &= 10 \log_{10}(I_D) - 10 \log_{10}(I_G) \\ &= C \log_{10}(d_G/d_D) \end{aligned} \quad (1.5)$$

$$= C \log_{10}((d + \Delta_d)/(d - \Delta_d)) \quad (1.6)$$

$$= C \sum_{n=0}^{+\infty} 1/(2n+1)(\Delta_d/d)^{2n+1} \quad (1.7)$$

Pour $\Delta_D \ll d$ un développement limité à l'ordre 1 est une bonne approximation de l'ILD et est proportionnelle à l'ILD. Viste (2004) propose alors une équation simplifiée, utilisée comme base pour les ILDs synthétiques afin d'obtenir l'Eqn. 1.9 proposée dans la Sec. 1.4.2.

1.4.2 Différences interaurales de niveau (ILDs)

ILDs réelles

Les différences interaurales de niveau (ILDs) représentent la différence d'intensité entre les deux oreilles. Ces indices interauraux peuvent être obtenus à partir des HRTFs présentées dans la Sec. 1.3.

Plus précisément, l'équation pour obtenir les ILDs à partir des HRTFs est donnée dans l'Eqn. 1.8.

$$\text{ILD}_{\text{real}}(f) = 20 \log_{10}(|\text{HRTF}_L(f)/\text{HRTF}_R(f)|) \quad (1.8)$$

Si l'on trace les différentes ILDs calculées ainsi, on obtient la Fig. 1.10. Seuls les azimuts présents dans la base ont été tracés. Si l'on souhaite utiliser les ILDs réelles obtenues à partir d'HRTFs, il faudrait combler les fréquences manquantes (par une interpolation par exemple).

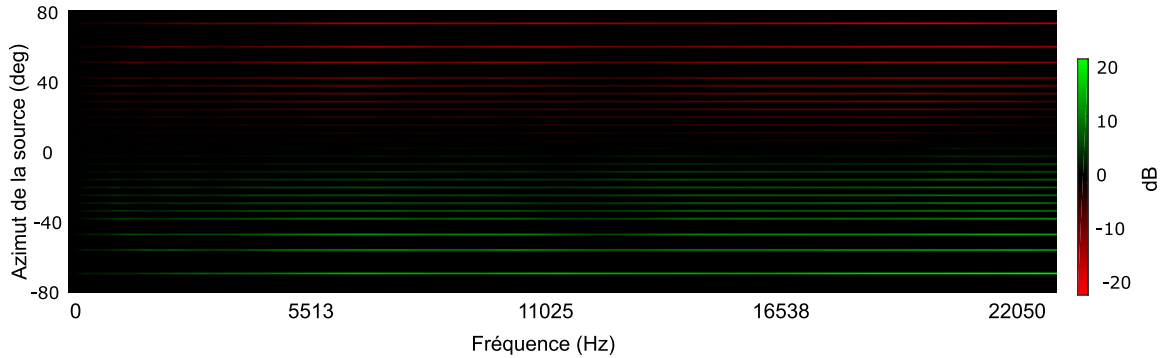


FIGURE 1.10 – ILDs réelles calculées à partir de la moyenne des HRTFs de la base CIPIC.

ILDs synthétiques de STAR

Comme déjà énoncé précédemment, la méthode STAR a une approche synthétique, basée sur des modèles. Le premier modèle est donc celui d'ILDs, inspiré par le travail de Viste (2004) et proposé par Mouba et al. (2008), donné dans l'Eqn. 1.9.

$$\text{ILD}(\theta, f) = \alpha(f) \sin(\theta) \quad (1.9)$$

À partir du modèle, et par correspondance en utilisant la méthode des moindres carrés sur les données, on obtient ainsi un coefficient α , présenté sur la Fig. 1.11. Le coefficient permet d'avoir la variabilité fréquentielle, et le modèle la variabilité temporelle. L'erreur moyenne commise, pour l'ensemble des sujets, azimuts et fréquences de la base CIPIC est de 4,29 dB.

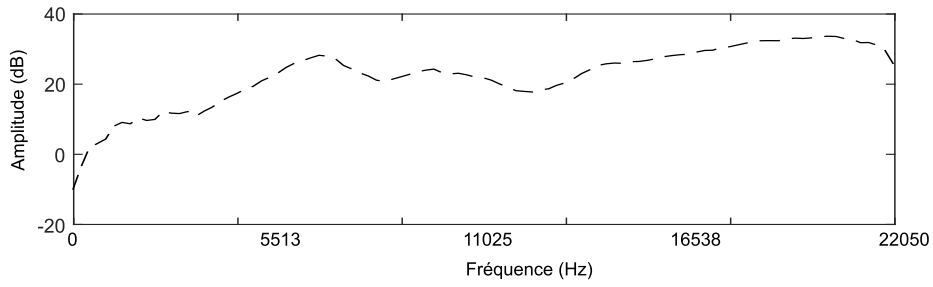


FIGURE 1.11 – Coefficient α obtenu par la méthode des moindres carrés par correspondance du modèle aux données.

À partir de ce modèle et du coefficient α , on peut ainsi tracer les ILDs en fonction de l'azimut (Fig. 1.12). L'avantage de ce modèle est qu'il est continu en fonction de l'azimut, et ne nécessite pas d'interpolation (contrairement aux ILDs réelles Fig. 1.10).

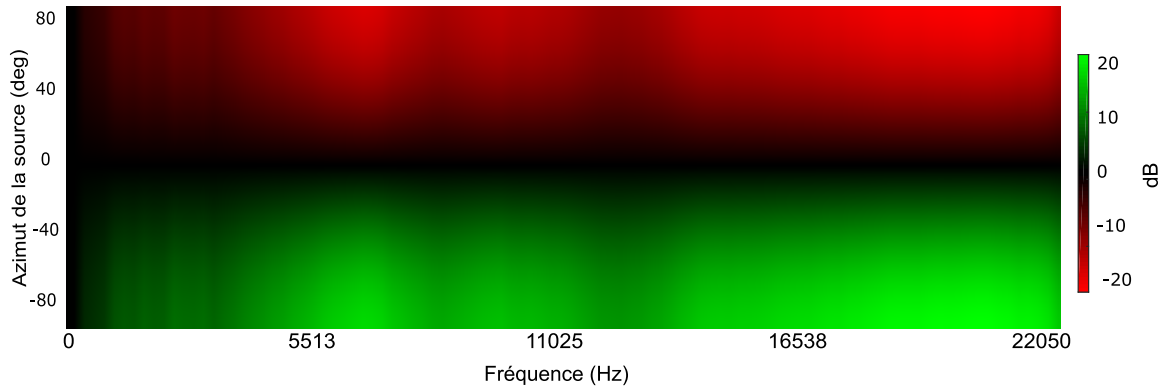


FIGURE 1.12 – ILDs synthétiques.

Pour valider que les ILDs obtenues correspondent à la réalité, un test a été mené, présenté Fig. 1.13. Une source spatialisée avec différentes méthodes (STAR, VBAP et HOA, présentées dans le chapitre 2) est jouée, balayant l'ensemble des azimuts. Ces sources sont enregistrées à l'aide d'un mannequin KEMAR, permettant ainsi de calculer les ILDs en utilisant un rendu binaural et ainsi comparer les différentes méthodes. La comparaison prend aussi en compte les ILDs provenant de simulations à l'aide des HRTFs du KEMAR de la base CIPIC (en noir) et de celles enregistrées au SCRIME (en utilisant cette fois la position réelle de la source sonore et non sa simulation). Chaque ILD obtenue correspond à la moyenne des fréquences.

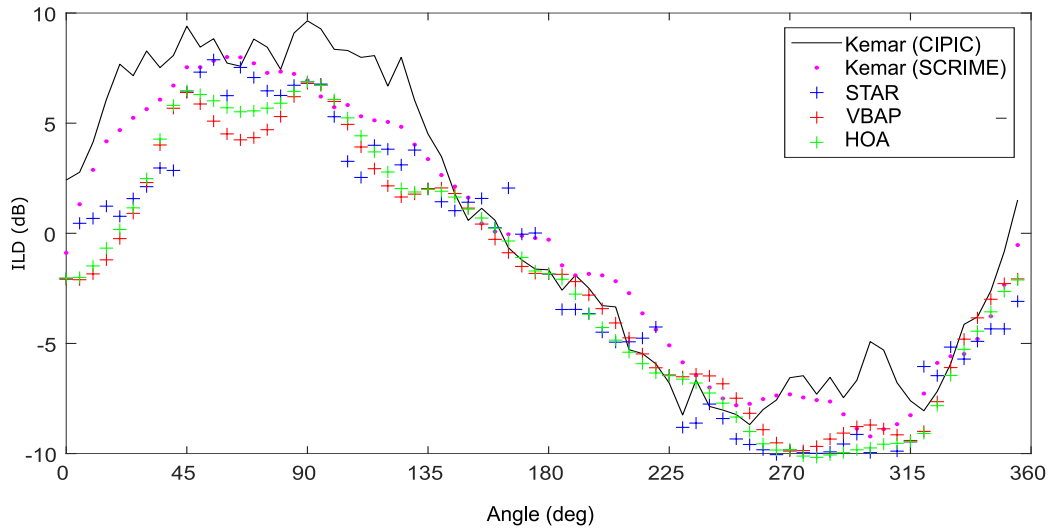


FIGURE 1.13 – *ILDs moyennes en fonction de l'azimut, enregistrées en utilisant différentes méthodes de diffusions (STAR, VBAP et HOA) pour faire tourner la source, ainsi que des mesures réelles (les KEMAR).*

La Fig. 1.13 permet ainsi d'affirmer que les ILDs obtenues grâce au modèle d'ILDs contenu dans la méthode STAR sont concordantes avec la réalité (cas des mesure KEMAR), et que toutes les méthodes sont assez proches.

1.4.3 Différences interaurales de temps (ITDs)

ITDs réelles

Les différences interaurales de temps (ITDs) représentent la différence de temps entre les deux oreilles pour la réception d'un son. Ces indices interauraux peuvent, tout comme les ILDs, être obtenus à partir des HRTFs présentées dans la Sec. 1.3.

Plus précisément, l'équation pour obtenir les ITDs à partir des HRTFs est donnée dans l'Eqn. 1.10, où c est la vitesse du son dans l'air (335 m/s), et r le rayon de la tête de l'auditeur. À noter également que les ITDs sont modulo 2π .

$$\text{ITD}_{\text{real}}(f) = \angle(\text{HRTF}_L(f)/\text{HRTF}_R(f))/(2\pi) \quad (1.10)$$

Tout comme les ILDs, en traçant les ITDs on obtient la Fig. 1.14.

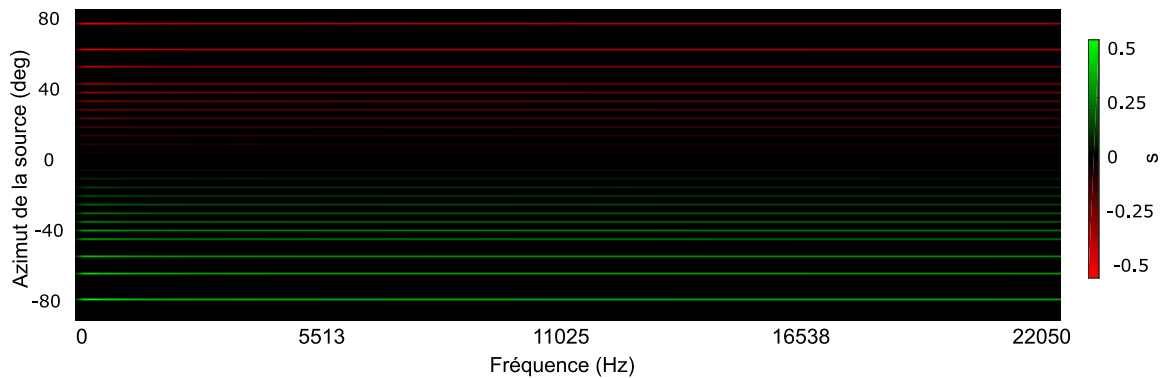


FIGURE 1.14 – *ITDs réelles.*

ITDs synthétiques de STAR

Le second modèle est celui d'ITDs, également inspiré par le travail de Viste (2004) et proposé par Mouba et al. (2008), donné dans l'Eqn. 1.11, mais simplifié lors de cette thèse, les résultats étant pratiquement identiques ¹ (initialement le modèle proposait $\beta(f)r(\sin(\theta) + \theta)/c$).

$$\text{ITD}(\theta, f) = \beta(f)r \sin(\theta)/c \quad (1.11)$$

Avec le même procédé que pour les ILDs, on obtient un coefficient β , présenté sur la Fig. 1.15.

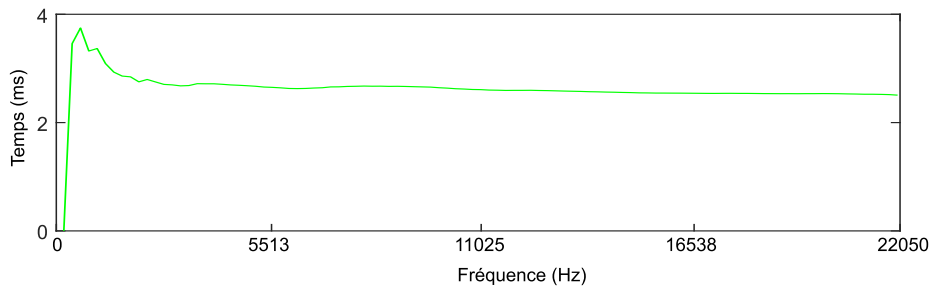


FIGURE 1.15 – Coefficient β obtenu par la méthode des moindres carrés par correspondance du modèle aux données.

On peut alors tracer les ITDs en fonction de l'azimut (Fig. 1.16). L'avantage de ce modèle est qu'il est continu en fonction sur l'azimut, ne nécessitant pas d'interpolation (contrairement aux ITDs réelles Fig. 1.14).

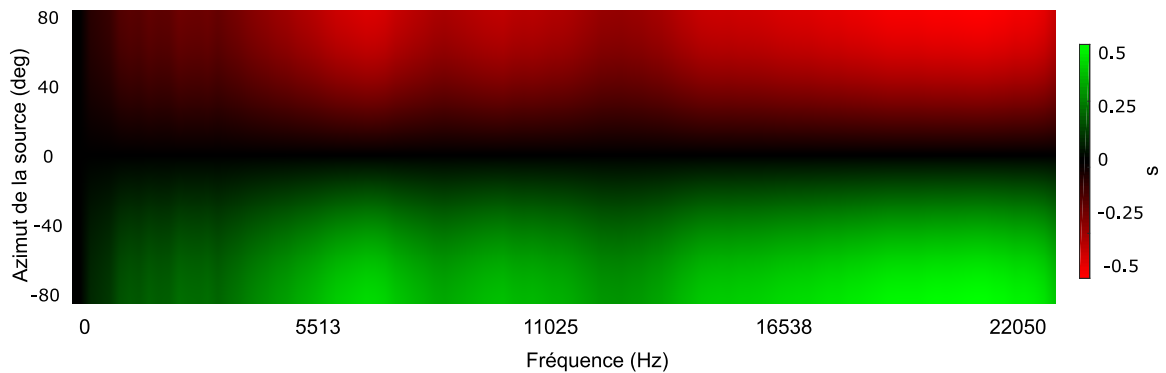


FIGURE 1.16 – ITDs synthétiques.

1. Les différents modèles d'ITDs ayant été utilisés et comparés

Conclusion : Ce chapitre a permis de poser de nombreuses bases ; à commencer par le fonctionnement de l'audition humaine, phénomène important puisque sans lui la spatialisation/localisation sonore n'aurait pas de sens pour les humains. De plus, comprendre ce fonctionnement et établir ses limites (précision des localisations selon les dimensions) vont permettre par la suite d'analyser de manière juste nos différents résultats.

Ensuite, les HRTFs et BRIRs sont largement discutées et expliquées ; celles-ci étant la 'matière première' utilisée par la méthode STAR, il est primordial de bien les avoir prises en considération.

Découlant de ces HRTFs, différents indices acoustiques peuvent être obtenus, les deux principaux, les ILDs et ITDs, sont le dernier sujet abordé dans ce chapitre.

Ainsi, après avoir vu ces bases, il est possible de se pencher sur les techniques de spatialisation et de localisation sonores ; sujets des prochains chapitres.

Ce chapitre, bien que présentant essentiellement les outils nécessaires à la compréhension de la thèse, introduit également certains travaux effectués, notamment sur les bases de données HRTF et BRTF, et les modèles d'ITDs et d'ILDs, inspirés par les travaux de Viste (2004) et Mouba and Marchand (2006), et repris et simplifiés dans cette thèse, présentés également dans les publications Méaux and Marchand (2019) et Méaux and Marchand (2021a).

Chapitre 2

Introduction à la spatialisation sonore, méthodes existantes et utilisations

– Introduction –

La perception spatiale est une part importante de la cognition humaine. C'est pourquoi, à l'heure où les technologies tendent à proposer des immersions de plus en plus réalistes, la spatialisation sonore joue un rôle très important. Elle présente en effet de vastes champs d'applications dans les domaines multimédia ; pour l'immersion dans des films, la retransmission de concerts spatialisés, des conversations plus réalistes, ou encore les jeux vidéo.

Le but de la spatialisation sonore est de reproduire un champ acoustique (voulu en théorie parfaite), ou l'illusion perceptive de cette source. Il existe cependant de nombreuses contraintes selon les utilisations qui expliquent la variété des méthodes, pouvant dépendre de leurs complexités de mises en œuvre, du coût de calcul disponible, du type de diffusion souhaité, du dispositif sonore utilisé (simple couronne de haut-parleurs, dôme 3D, etc.).

Il existe maintenant de nombreuses techniques de spatialisation, qui vont de techniques simples telles que la prise de sons spatialisés pour une ré-émission, à des systèmes plus complexes tendant à reproduire parfaitement l'onde sonore en un point de l'espace, voire en tous points. Ce chapitre a pour utilité de présenter les principales méthodes de spatialisation sonore existantes.

2.1 Méthode de spatialisation au casque - le son binaural

La notion de son binaural est ancienne. Elle a vu le jour en 1881, avec l'invention du théâtrophone, qui permettait à des auditeurs loin du théâtre (2 km) d'écouter un son distinct avec chacune de leurs oreilles. L'enregistrement réellement binaural est né dans les années 1920-30, avec l'invention de têtes artificielles "Oscar" et "Kunstkopf", développées par les laboratoires Bell et Philips. C'est en 1931 qu'Alan Blumlein présente un son binaural, qui en réalité correspond à la méthode bien connue aujourd'hui de la stéréophonie. Il faudra ensuite attendre les années 1960 afin que le son binaural devienne un concept, notamment grâce aux trois ingénieurs Georg Plenge, Ralf Kürer et Henning Wilkens, et leur invention du premier mannequin (KU80 tête factice) en 1969.

À l'origine de la méthode, le son est enregistré au niveau des oreilles, afin d'être restitué au casque (diffusant en théorie exactement ce que l'on doit entendre).

Il existe maintenant différentes ressources afin d'enregistrer directement ces sons binauraux, tels que des écouteurs comportant également un micro (écouteurs Sennheiser Ambeo par exemple) ou des micros spécialisés possédant des oreilles tels que ceux de la marque 3Dio (Fig. 2.1).



FIGURE 2.1 – Exemple de micro conçu pour la prise de son binaural, ici de la marque 3Dio. <https://3dio-sound.com/>

Les méthodes binaurales actuelles permettent maintenant de diffuser également n'importe quel son, en appliquant les caractéristiques de la personne (HRIR) ou de la pièce (BRIR) (Fig. 2.2).

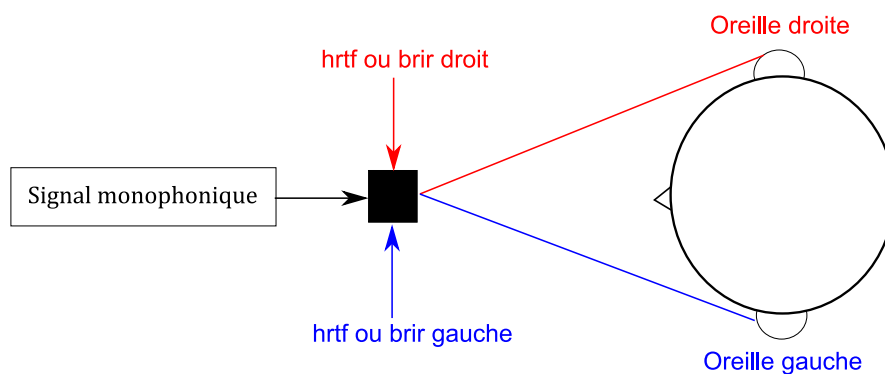


FIGURE 2.2 – Schéma simplifié d'une synthèse binaurale à l'aide de BRTF/HRTF. Les BRTF/HRTF gauche et droit sont appliqués au signal mono d'entrée pour être ensuite envoyés sur les champs gauche/droit du casque de l'auditeur.

Les méthodes binaurales sont très utilisées pour les podcasts, les fictions audio, les émissions radio, *etc.* Elles sont également très utilisées pour l'immersion audio, le rendu étant très bon car non perturbé, et simples de mise en œuvre (utilisation pour retranscrire des orchestres symphoniques, pour les jeux vidéo (cas pour l'Audio Spatializer SDK de Unity très utilisé dans ce domaine)).

Cette méthode bien que performante présente néanmoins un inconvénient principal, qui est de devoir porter un casque, pouvant ainsi nuire à l'expérience de l'auditeur, et inenvisageable dans certaines circonstances (cas des concerts par exemple). C'est pourquoi la méthode STAR, ainsi que d'autres méthodes (présentées en Sec. 2.2) se placent dans un champ libre en utilisant des réseaux de haut-parleurs.

2.2 Méthode de spatialisation via des réseaux de haut-parleurs

2.2.1 L'approche physique de spatialisation du son

High Order Ambisonic (HOA)

La technique dite ambisonique date des années 1970 principalement développée par Gerzon (1973), historiquement pour 4 canaux audio. L'ensemble des théories relatives à l'ambisonique est présenté dans le livre

2.2. MÉTHODE DE SPATIALISATION VIA DES RÉSEAUX DE HAUT-PARLEURS

de Zotter and Frank (2019). Cette méthode fut ensuite généralisée aux ordres supérieurs par Daniel (2001) (Higher Order Ambisonics ou HOA) dans son travail de thèse. Cette méthode basée sur un modèle mathématique des ondes acoustiques tente de reproduire un champ primaire (champ de la source) à l'aide de champs secondaires (les haut-parleurs) en décomposant l'espace sonore sur la base d'harmoniques sphériques, présenté dans la Fig. 2.3.

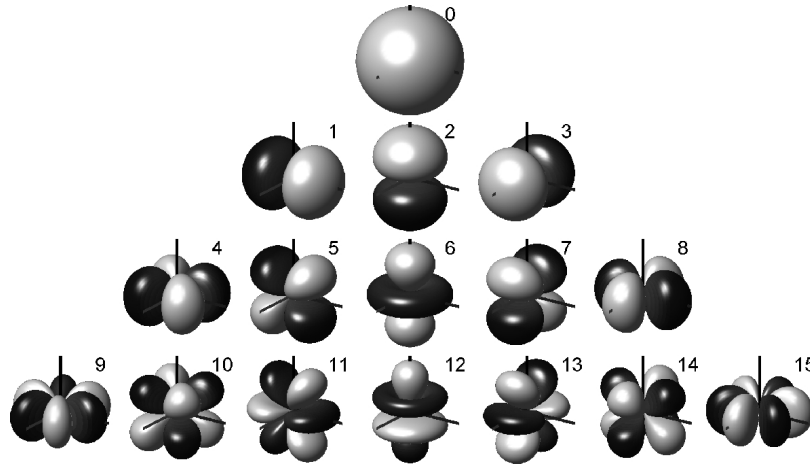


FIGURE 2.3 – Schéma des harmoniques sphériques pour les 3 premiers ordres (ordre 0 sur la première ligne, 1 sur la seconde ligne, etc.). Les parties blanches correspondent aux valeurs négatives, et celles en noir aux valeurs positives des harmoniques. (Zotter and Frank (2019).)

L'ordre ambisonique correspond au nombre d'harmoniques sphériques utilisé pour la création du champ sonore. Plus l'ordre est important, plus en théorie la résolution est grande. L'ordre 0 correspondant au cas de la monophonie, alors que la base théorique est de dimension infinie.

En 2D (à élévation nulle), les harmoniques sphériques correspondent à une base de Fourier, et en 3D, à des fonctions de Fourier-Bessel.

La technique ambisonique utilise une théorie d'encodage et de décodage, l'enregistrement se faisant à l'aide de microphones spécifiques (Core Sound TetraMic ou SoundField par exemple) et le décodage grâce aux harmoniques sphériques qui dépendent de l'ordre utilisé. Il n'est bien sûr pas obligatoire d'enregistrer un son par une méthode ambisonique afin de la restituer en ambisonique. On obtiendra alors un son spatialisé via la méthode ambisonique.

Cette technique produit des résultats optimaux dans des conditions contrôlées ; salles non réverbérantes, configuration régulière de haut-parleurs, auditeur au centre du dispositif (Gerzon (1973))... En théorie, plus l'ordre est important, plus la qualité sera grande. Cependant, le nombre de haut-parleurs nécessaires devient alors une contrainte non négligeable. L'Eqn2.1, (o correspondant à l'ordre) donne le nombre minimal de haut-parleurs. Le nombre de haut-parleurs explose donc très rapidement lorsque l'on augmente l'ordre.

$$\begin{aligned} (2D) & : n_{hautsparleurs} \geq 2.o + 1 \\ (3D) & : n_{hautsparleurs} \geq (o + 1)^2 \end{aligned} \quad (2.1)$$

La technique de l'ambisonique, coûteuse en calculs et matériel, a grandement profité de l'évolution des outils informatiques, entraînant de nombreuses recherches, ainsi qu'un certain nombre de décodeurs. Certaines de ces recherches proposent des ajustements afin de diminuer les contraintes du nombre de haut-parleurs ainsi que leurs placements. Cette méthode reste cependant coûteuse en calcul et nécessite une installation précise et calibrée.

Dans cette thèse, la méthode HOA sera utilisée à des fins de comparaisons. En effet, son fonctionnement et ses performances ont pu être testés, par exemple dans le travail de Bertet (2009), donnant ainsi un référentiel potentiel à la méthode STAR pour des évaluations subjectives (dans le cas où il n'est pas possible d'effectuer des tests objectifs).

Wave Field Synthesis (WFS)

L'holophonie (Wave Field Synthesis) se base sur le même fonctionnement que l'holographie, et repose sur le principe que la source primaire est remplacée par n sources secondaires. Elle fut énoncée par Christian Huygens (1690) et quantifiée au 19^e siècle par Gustav Kirchhoff et Hermann von Helmholtz.

Les premiers concepts sont présentés par Snow (1955), puis Berkhout et al. (1993). Ils ont appliqué cette théorie pour l'ensemble du champ acoustique en lui donnant le nom de WFS. Les ondes sonores étant enregistrées par des microphones directifs, et, après traitement, ré-émises par des réseaux de haut-parleurs dans le but d'obtenir des fronts d'ondes sonores identiques à ceux enregistrés. Cette technique s'applique généralement via des réseaux de haut-parleurs linéaires, bien que des recherches étendent cette méthode en 3D ou l'appliquent sur des haut-parleurs répartis non uniformément (Spors et al. (2008)).

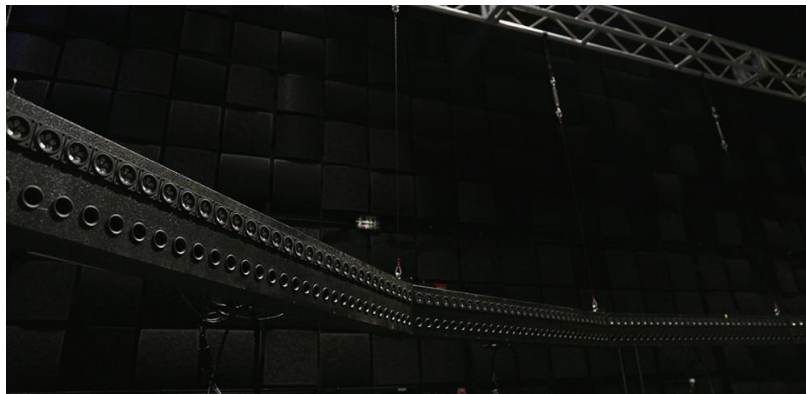


FIGURE 2.4 – Exemple de système mis en place pour une diffusion avec la méthode WFS. Cette photo présente une partie de la couronne de haut-parleurs et montre bien le grand nombre de haut-parleurs nécessaires. [https : //easternbloc.ca](https://easternbloc.ca)

Ainsi, cette technique nécessite un grand nombre de haut-parleurs (Fig.2.4) mais aussi de microphones, ainsi qu'une puissance de calcul importante, ce qui ne correspond pas à la philosophie de la méthode STAR, ni aux moyens dont nous disposons. Cette méthode ne sera donc pas utilisée dans ce travail.

2.2.2 Une approche mathématique de spatialisation du son

Vector base amplitude panning (VBAP)

VBAP, introduite dans Pulkki et al. (1996), puis proposée de manière complète dans Pulkki (1997), est une approche permettant l'utilisation d'un nombre limité de haut-parleurs autour de l'auditeur qui vont répartir l'amplitude entre les haut-parleurs entourant la source en fonction de la position de spatialisation souhaitée.

La configuration pour laquelle la méthode est désignée est une pièce peu réverbérante, équipée de haut-parleurs placés de manière équidistante.

Comme son nom l'indique, VBAP utilise des vecteurs afin de simplifier les calculs. La méthode telle que présentée par Pulkki (1997) peut se décrire pour la 2D par

$$p = g_1 l_1 + g_2 l_2 \quad (2.2)$$

où les g sont les facteurs de gain, l les vecteurs de direction entre l'auditeur et les haut-parleurs, et p le vecteur de direction entre l'auditeur et la source virtuelle, tel que décrit sur la Fig. 2.5.

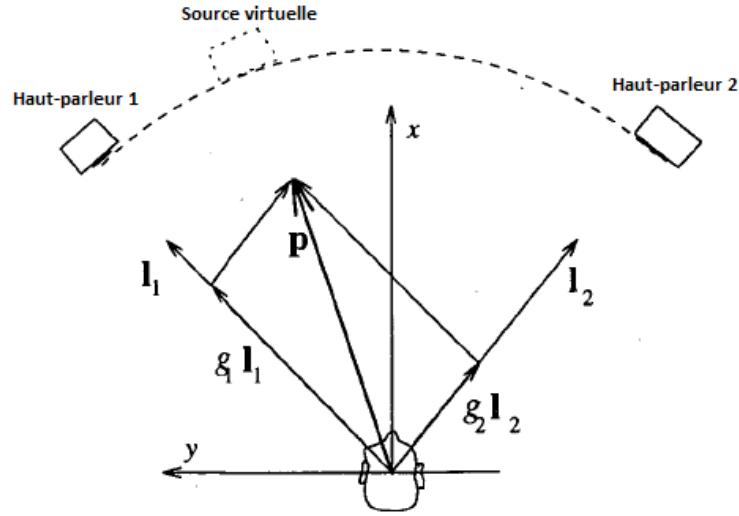


FIGURE 2.5 – Schéma provenant de l'article de Pulkki (1997) et présentant un cas de spatialisation en 2D d'une source sonore virtuelle positionnée entre deux haut-parleurs.

Sous forme vectorielle, on peut écrire $l_1 = [l_{11} l_{12}]^T$, $l_2 = [l_{21} l_{22}]^T$, $p = [p_1 p_2]^T$ et $g = [g_1 g_2]^T$. On peut ainsi écrire l'Eqn. 2.2, sous forme vectorielle (Eqn. 2.3), où $L_{12} = [l_1 l_2]^T$

$$p^T = g^T L_{12} \quad (2.3)$$

Les facteurs de gains, (Eqn. 2.3), sont l'unique inconnue de l'Eqn. 2.2. L'Eqn. 2.4 pourra être résolue si l'inverse de la matrice L_{12} existe bien.

$$g = p^T L_{12}^{-1} \quad (2.4)$$

Pour un système en trois dimensions, VBAP utilise les 3 haut-parleurs entourant la source, contrairement au système 2D. Une dimension est ainsi ajoutée à l'Eqn. 2.2, donnant $p = g_1 l_1 + g_2 l_2 + g_3 l_3$. La résolution se fait de façon similaire à la version 2D, mais en utilisant des vecteurs de dimension 3.

Si on se place dans les contextes de son 3D de la thèse (cas par exemple des couronnes de haut-parleurs), les deux haut-parleurs entourant la source seront sélectionnés avant d'appliquer la méthode VBAP (trois si on se place dans un contexte 3D).

Tout comme HOA, VBAP est une méthode dont les performances ont été testées (Pulkki (1997)). De plus, ces deux méthodes ont également été comparées comme dans les travaux de Marentakis et al. (2014), et Gandemer et al. (2018). Nous utiliserons ainsi VBAP, tout comme HOA, comme 'référence' pour les tests objectifs.

Cette méthode est également utilisée dans le cas où le dispositif dispose de plusieurs couronnes pour positionner la source sonore entre ces couronnes de haut-parleurs. Ceci permet d'améliorer la spatialisation STAR pour l'élévation implémentée pour une couronne unique. Le procédé est décrit dans le chapitre 6.

2.2.3 L'approche transaurale utilisée pour la méthode STAR

Les méthodes transaurales sont des méthodes utilisant plusieurs haut-parleurs et permettant de reproduire le signal obtenu à l'entrée des oreilles d'un auditeur dans une configuration originale, dans une seconde configuration.

La notion de transauralité a été introduite par Bauer (1961), et la première méthode proposée par Schroeder and Atal (1963) pour deux haut-parleurs placés de manière symétrique et de manière stéréophonique,

2.2. MÉTHODE DE SPATIALISATION VIA DES RÉSEAUX DE HAUT-PARLEURS

recréant le signal dans une salle de concert. D'autres travaux ont ensuite été menés comme ceux de Møller (1992).

Le principe des méthodes transaurales est exposé dans la Fig. 2.6. Le but est de recréer les chemins acoustiques H_L et H_R ¹, entre la source virtuelle et les oreilles (droite/gauche) de l'auditeur. Pour cela, les chemins acoustiques réels entre les haut-parleurs et les oreilles de l'auditeur sont utilisés (H_{LR} , H_{LL} , H_{RR} , H_{RL}). La notation H_{LR} est attribuée pour le chemin entre le haut-parleur gauche et l'oreille droite, H_{RL} pour celui entre le haut-parleur droit et l'oreille gauche (et ainsi de suite).

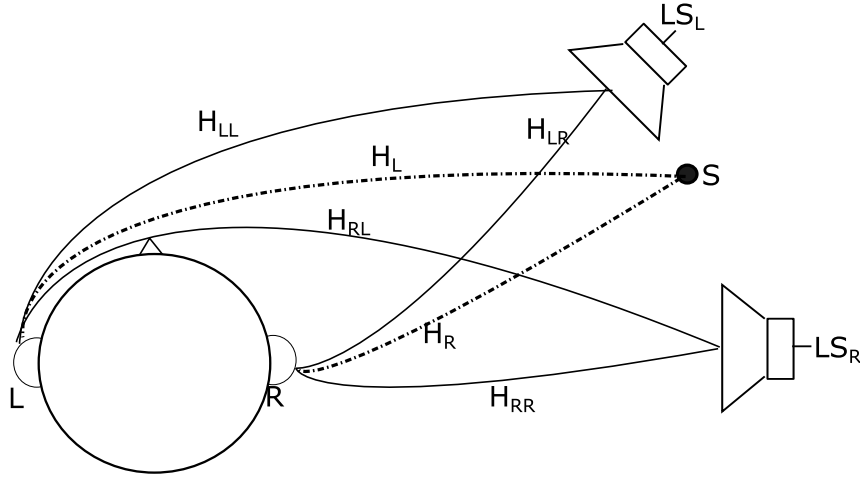


FIGURE 2.6 – Principe transaural : 4 chemins acoustiques réels (H_{LL} , H_{RL} , H_{LR} , et H_{RR}) provenant des haut-parleurs gauches et droits (LS_L et LS_R) sont utilisés afin de reproduire les 2 chemins acoustiques virtuels (H_L et H_R) provenant de la source virtuelle S .

Ainsi, pour une source sonore s (ou S si l'on se place dans le domaine spectral), les sons reçus aux oreilles gauche et droite doivent être respectivement $H_L \cdot S$ et $H_R \cdot S$. Étant donné que la source s est virtuelle, il est nécessaire d'utiliser les deux haut-parleurs entourant celle-ci, devant alors vérifier le système donné dans l'Eqn. 2.5.

$$\begin{aligned} H_L \cdot S &= K_L \cdot H_{LL} \cdot S + K_R \cdot H_{RL} \cdot S \\ H_R \cdot S &= K_L \cdot H_{LR} \cdot S + K_R \cdot H_{RR} \cdot S \end{aligned} \quad (2.5)$$

où K_L et K_R sont les coefficients devant être appliqués aux haut-parleurs gauche et droit respectivement. Ces coefficients étant obtenus par la résolution du système de deux équations à deux inconnues ; les chemins acoustiques étant connus pour l'ensemble des positions (haut-parleurs et source virtuelle).

La Fig. 2.6 montre uniquement les trajets directs. Si l'on souhaite recréer exactement le champ sonore dans sa configuration originale, il est possible d'inclure des réflexions. Cependant ceci peut engendrer certains problèmes, la durée de la réponse impulsionnelle pour un son avec réverbération pouvant être importante.

Des méthodes transaurales utilisant de nombreux haut-parleurs ont également été proposées, ayant notamment l'avantage d'élargir le sweet spot Bauck (2001); Gálvez and Fazi (2015).

Ces méthodes, notamment dans le cas de l'utilisation de nombreux haut-parleurs, nécessitent toutefois de bons étalonnages et mesures, ainsi qu'une bonne calibration et installation de la pièce.

La méthode STAR utilise la notion de transaural (synthetic Transaural audio rendering). Cependant, contrairement aux techniques tendant à recréer le son arrivant dans la condition initiale, le transaural est appliqué sur des chemins acoustiques synthétiques provenant d'un modèle d'indices acoustiques. Cette différence a comme principal avantage de simplifier grandement la mise en œuvre tout en donnant de bons résultats (comme discutés dans la Part. IV), ce qui est conforme à la philosophie et aux objectifs de la méthode STAR.

1. L pour Left (gauche) et R pour Right (droite); LS(LoudSpeaker).

– Conclusion –

Ce chapitre présente donc les principales techniques de spatialisation; à commencer par l'écoute à l'aide de casques audio, puis en utilisant des systèmes de diffusions par haut-parleurs. L'ensemble des techniques, hormis WFS trop coûteuse, ont un intérêt dans l'étude de la méthode STAR.

La technique binaurale est utilisée par le biais du rendu binaural, permettant de passer de signaux spatialisés par les différentes méthodes multi-voies à des signaux deux voies audibles simplement au casque.

Les techniques VBAP et HOA, méthodes validées et stables, sont utilisées comme base de comparaison et de validation pour de nombreux tests.

Enfin, le principe des méthodes transaurales est utilisé directement dans la méthode STAR, afin de recréer les indices acoustiques synthétiques.

Le chapitre précédent présentait l'audition humaine et son fonctionnement. Ce chapitre permet de comprendre les différentes méthodes de spatialisation, qui ont chacune un intérêt et des limitations intrinsèques à leur fonctionnement qui seront prises en compte dans les parties suivantes.

Ce chapitre présente brièvement les méthodes de spatialisation utilisées comme outils de comparaison dans cette thèse.

Deuxième partie

Localisation sonore perceptive

À propos de la partie – localisation sonore perceptive –

La méthode STAR se veut être une approche perceptive, que ce soit dans les procédés de spatialisation ou dans ceux de localisation. Avant de se lancer dans les processus de spatialisation, qui sont le cœur de la méthode STAR, il est intéressant de se pencher sur les méthodes de localisation. Afin de localiser objectivement un son en utilisant la méthode perceptive, les enregistrements doivent être effectués à l'aide de micros placés au fond des oreilles de l'auditeur, ou à l'aide d'un mannequin – comme par exemple le mannequin KEMAR du SCRIME présenté en Annexe 11.8.1 –. Dans le cas où les sons sont enregistrés, dans ce travail, ils seront toujours acquis de cette manière.

Cette partie se divise en deux chapitres. Le premier présente la méthode de localisation perceptive pour l'azimut, développée lors de travaux précédents, mais pas testée. Des tests ont été effectués en conditions anéchoïques en utilisant des bases de données HRTF, et réelles, en utilisant des BRIRs, provenant de bases de données mises à disposition dans l'état de l'art ou celle du SCRIME enregistrée au cours de ce travail de thèse.

Le second chapitre présente, quant à lui, les travaux effectués afin de mettre au point une méthode de localisation perceptive pour l'élévation, utilisant une approche inédite en utilisant une méthode type Shazam sur les noeuds et pics – indices spectraux détenant l'information de l'élévation dans les HRTFs – afin de retrouver une élévation provenant d'un enregistrement lambda. Cette méthode montre des résultats encourageants mais nécessite des mises au point.

Enfin, dans cette partie, il n'est pas présenté de travail sur la localisation de la distance. En effet, comme expliqué dans la section 1.2.3, ce processus de localisation effectué par l'homme nécessite des informations et traitements qui ne sont pas disponibles dans les conditions d'utilisation de la méthode STAR ou contraire à sa philosophie.

Cette partie sur la localisation sonore perceptive est un apport important de ce travail de thèse ; les nombreux tests ont permis de valider et de comprendre les limitations de la méthode de localisation en azimut existante. La recherche sur une méthode de localisation en élévation perceptive est totalement nouvelle – bien qu'utilisant les travaux sur les noeuds et pics de lida – et présente des résultats encourageants.

Chapitre 3

Méthode de localisation perceptive pour l'azimut

– Introduction –

Ce premier chapitre décrit la méthode de localisation sonore perceptive (basée sur les indices acoustiques) de la méthode STAR.

Comme toute méthode, il est indispensable de valider la méthode de localisation : c'est ce qui est décrit dans la seconde partie de ce chapitre.

3.1 La méthode

Les indices acoustiques, dont principalement les indices interauraux (ILDs et ITDs), introduits Sec. 1.4, servent de base pour la méthode de localisation perceptive, proposée par Mouba and Marchand (2006) et Mouba et al. (2008).

3.1.1 Rappel sur les modèles d'indices acoustiques utilisés dans la méthode STAR

Les indices acoustiques ITDs et ILDs, représentant les différences de temps et d'intensité reçues par les deux oreilles de l'auditeur sont les indices utilisés pour la localisation de l'azimut d'un son. Strutt (Lord Rayleigh) (1907) a notamment développé la théorie selon laquelle les ILDs sont plus efficaces pour les hautes fréquences, *a contrario* des ITDs plus efficaces pour la localisation de sons basses fréquences. La physique des ondes de son côté implique que les hautes fréquences sont plus sensibles aux atténuations sélectives dans l'air, tout en ayant un signal comportant une ambiguïté de phase ; alors que les basses fréquences sont moins sensibles aux atténuations, mais n'ont pas d'ambiguïtés sur la phase (Blauert (1997)).

Notre méthode de localisation utilise ces propriétés afin d'obtenir la localisation la plus précise.

Pour rappel, les Eqn. (3.1, 3.2) sont les ILD et ITD synthétiques du modèle STAR.

$$\text{ILD}_{\text{model}}(\theta, f) = \alpha(f) \sin(\theta) \quad (3.1)$$

$$\text{ITD}_{\text{model}}(\theta, f) = \beta(f) r \sin(\theta) / c \quad (3.2)$$

Avec les coefficients α et β (Fig. 3.1) provenant d'un apprentissage de l'ensemble des sujets de la base CIPIC (Algazi et al. (2001)), et contenant les informations relatives aux têtes et oreilles des sujets.

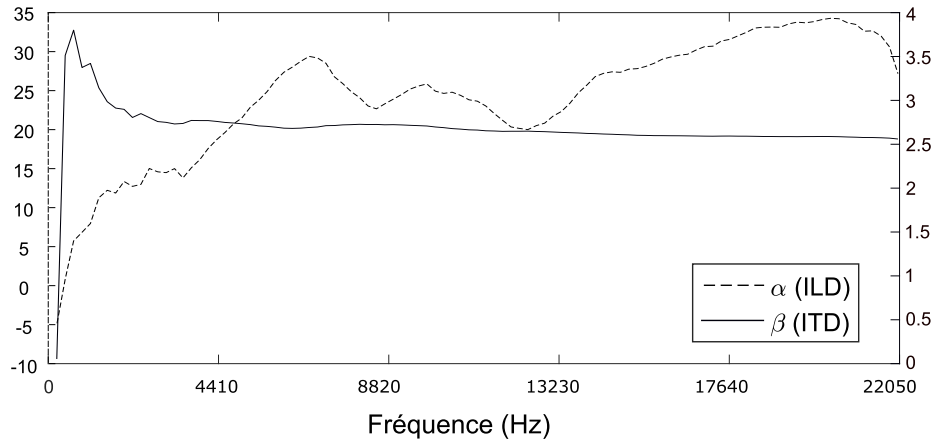


FIGURE 3.1 – Coefficients α et β , obtenus par correspondance au modèle via la méthode des moindres carrés, et contenant l'ensemble des informations relatives aux têtes et oreilles de chaque sujet de la base, et ceci pour toutes les fréquences et tous les azimuts.

À partir des signaux gauche (L) et droit (R), il est possible de calculer les ILDs (Eqn. 3.3) et ITDs (Eqn. 3.4).

$$\text{ILD}(f) = 20 \log_{10}(|L(f)/R(f)|) \quad (3.3)$$

$$\text{ITD}_p(f) = \frac{1}{2\pi f} (\angle(L(f)/R(f)) + 2\pi p) \quad (3.4)$$

Le coefficient p signifiant que la phase est déterminée modulo 2π .

On dispose alors d'un modèle d'ITDs et d'ILDs dépendant de la fréquence, ainsi que d'ITDs et ILDs calculées à partir d'enregistrements. La Sec. 3.1.2 présente la méthode utilisée afin de retrouver l'azimut à l'aide du modèle et des enregistrements.

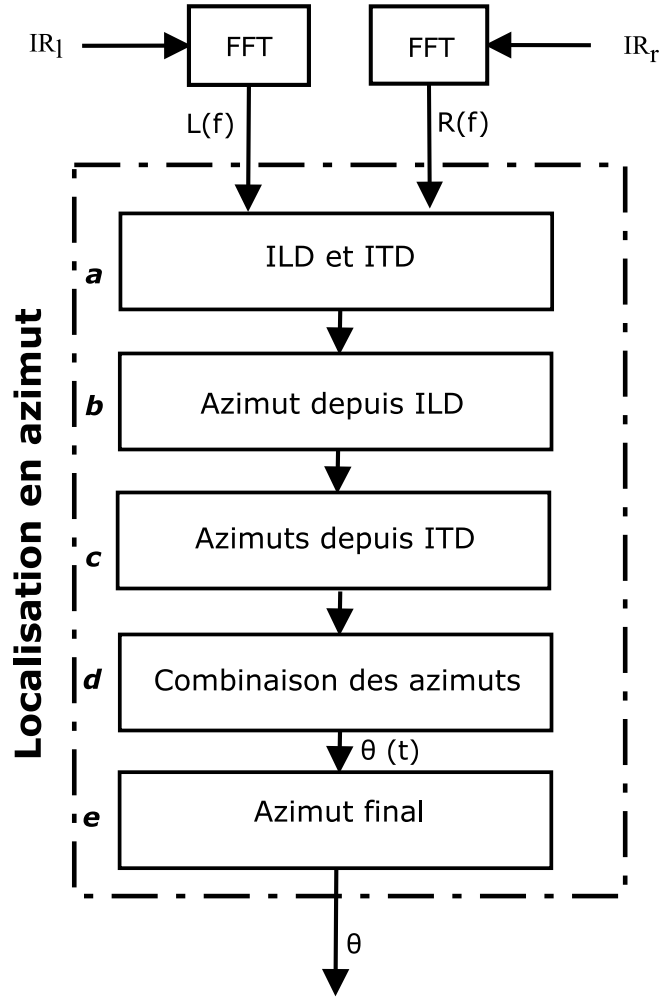
3.1.2 Le processus de localisation

Le but du modèle introduit Sec. 1.4 et 3.1.1 est de recréer les indices acoustiques, en fonction de l'azimut de la source, de la fréquence, et en utilisant les coefficients *alpha* et *beta* (Fig. 3.1).

Cette méthode est la même que proposée par Harald Viste (Viste (2004)), mais utilisant des modèles simplifiés.

La méthode de localisation repose sur l'hypothèse qu'avec un enregistrement effectué de manière similaire à la localisation humaine (intra-oreille ou à l'aide d'un mannequin), les modèles permettront d'estimer θ qui est la seule variable inconnue.

La Fig. 3.2 résume le processus de localisation.

FIGURE 3.2 – *Processus de localisation perceptif.*

Partie a : Les ITDs et ILDs réelles sont calculées grâce aux Eqns. (3.3) et (3.4).

Partie b : Un azimuth est estimé à partir des ILDs, grâce au facteur α connu. Ainsi, en inversant l'Eqn. (3.1), on obtient l'Eqn. (3.5) permettant de déduire θ_{ILD} à partir des ILDs.

$$\theta_{ILD}(f) = \arcsin(ILD(f)/\alpha) \quad (3.5)$$

Partie c : D'autres azimuths sont estimés, de manière similaire à la partie b, cette fois en inversant l'Eqn. (3.2), on obtient l'Eqn. (3.6), avec cependant le facteur d'ambiguïté p qui subsiste.

$$\theta_{ITD,p}(f) = \arcsin(ITD_p(f)c/(\beta r)) \quad (3.6)$$

Partie d : Ainsi le $\theta_{ITD,p}$ qui se trouve le plus proche de θ_{ILD} est considéré comme l'azimut final $\theta(f)$, en utilisant l'Eqn. (3.7).

$$\theta(f) = \theta_{ITD,m} \text{ with } m = \operatorname{argmin}_p |\theta_{ILD}(f) - \theta_{ITD_p}(f)| \quad (3.7)$$

Partie e : En théorie, une même source sonore devrait donner tout le temps le même résultat, quelles que soient les fréquences. En pratique, la présence de bruit et de réverbération dissipe l'énergie. C'est pourquoi

3.2. VALIDATION DE LA MÉTHODE DE LOCALISATION PERCEPTIVE

l'azimut estimé θ est au final défini comme étant celui du pic de l'histogramme de distribution d'énergie (Mouba and Marchand (2006)), comme montré sur la Fig. 3.3¹.

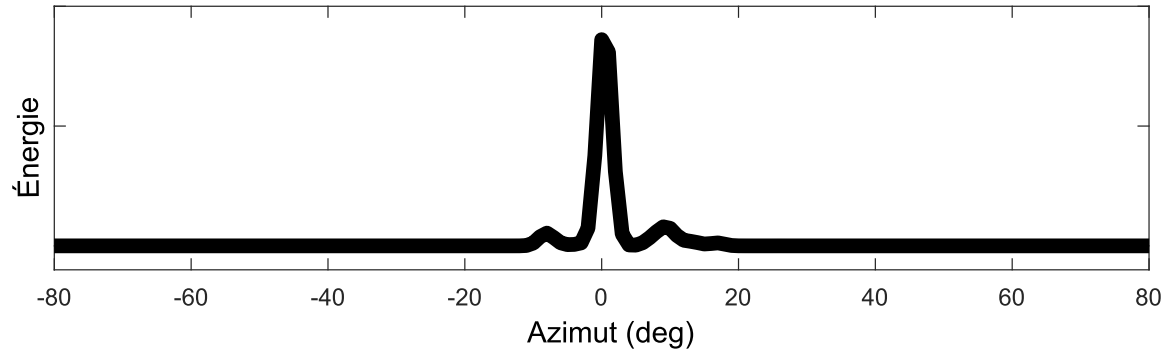


FIGURE 3.3 – *Histogramme de l'azimut dans des conditions idéales (en utilisant la base CIPIC). L'énergie est très concentrée autour de la position de la source sonore (ici $\theta = 0^\circ$).*

3.2 Validation de la méthode de localisation perceptive

3.2.1 Étude en conditions idéales

Le but de cette section est d'étudier la résistance de la méthode présentée en Sec. 3.1.2, dans différents contextes. La première étape afin d'étudier la résistance de cette méthode utilise les données provenant de bases de données. Nous utilisons donc la base CIPIC présentée dans la Sec.1.3, contenant les HRIRs (dans des conditions anéchoïques) de 45 sujets (dont KEMAR avec oreilles larges), et de nombreux azimuts.

Ces conditions sans écho ni bruit sont les conditions idéales pour la localisation sonore.

Ainsi la figure 3.4 présente les résultats d'estimation de notre méthode de localisation pour cinq azimuts et l'ensemble des sujets de la base CIPIC.

1. Ici, l'énergie de l'histogramme est représentée par le nombre d'itérations où chaque azimut est détecté dans la source.

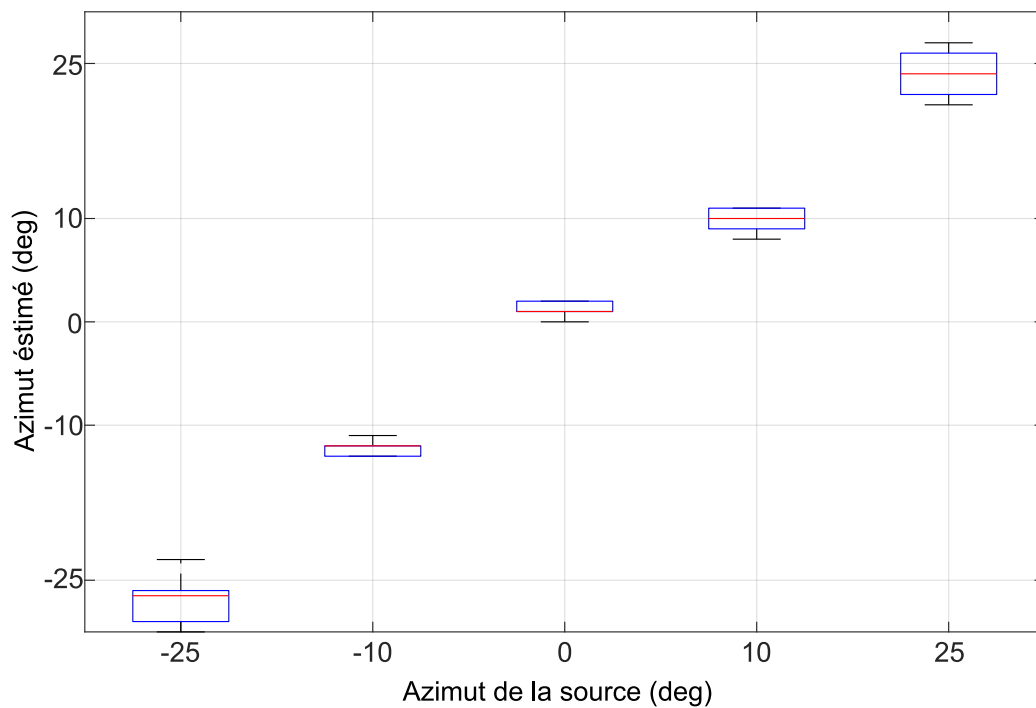


FIGURE 3.4 – *Précision de localisation de l'azimut en conditions idéales (anéchoïques), ceci pour cinq azimuts, les 45 sujets de la base CIPIC ainsi que 25 élévations.*

La Fig. 3.5 présente les résultats uniquement pour un individu (ici le mannequin KEMAR larges oreilles), et toutes les élévations. En prenant un individu séparé, on observe que l'écart type¹ est plus important, tout en ayant des outliers.

1. Les boîtes à moustaches ne donnent pas directement la visualisation de l'écart type, mais l'écart interquartile; celui-ci permet cependant de se faire une bonne idée de ce que serait l'écart type.

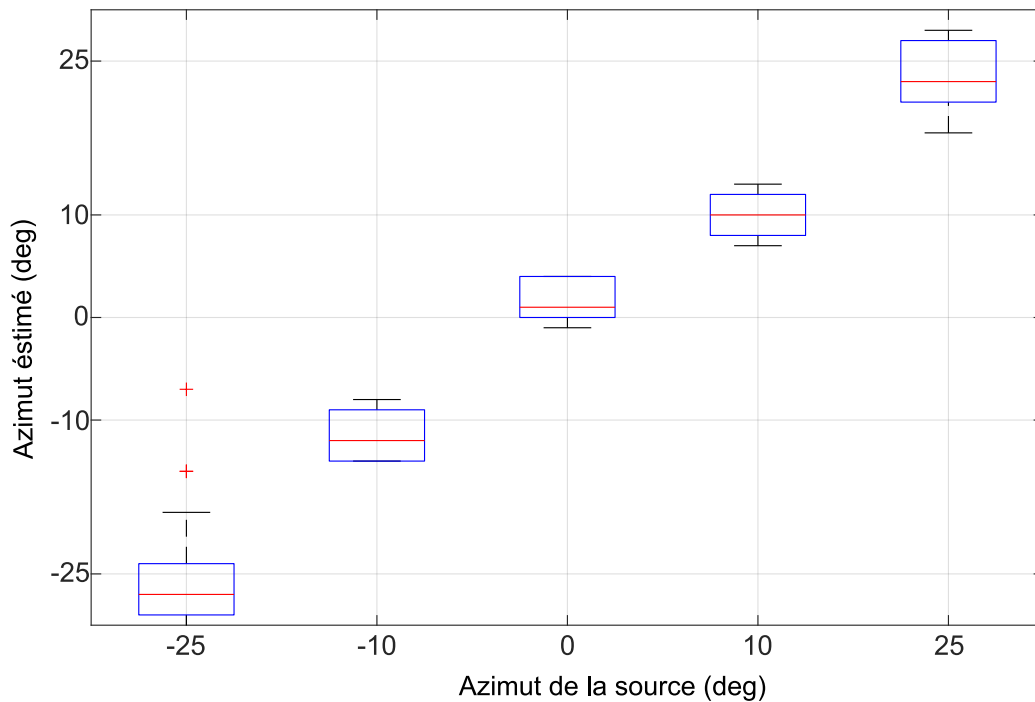


FIGURE 3.5 – Précision de localisation de l’azimut en conditions idéales (anéchoïques), pour le mannequin KEMAR grandes oreilles provenant de la base CIPIC ainsi que 25 élévations.

Les erreurs sur ces figures sont plus faibles que celles attendues pour les performances de la localisation en azimut de l’homme sec. 1.2.2, ce qui s’explique facilement par les conditions anéchoïques de la base CIPIC. De plus, le procédé de localisation mené sur un grand nombre de sources sonores permet d’assurer la robustesse de la méthode.

La méthode de localisation est donc performante et robuste dans les conditions anechoïques.

3.2.2 Étude en conditions réelles

La Sec. 3.2.1, permet de valider la méthode de localisation en conditions anéchoïques. Bien qu’idéale, cette configuration n’est pas réaliste. En effet, les chambres anéchoïques sont inexistantes si l’on sort du contexte scientifique. La littérature propose un certain nombre de bases de données BRIR, qui correspondent, elles, à la réalité d’une pièce (avec réverbération), introduites Sec.1.3.2. Dans le cadre de cette étude nous utilisons la base de l’institut de communication de Rostok. Les BRIRs sont enregistrées avec un mannequin KEMAR grandes oreilles, en effectuant une rotation de la tête de $\pm 80^\circ$ avec un pas de 2° . La pièce de dimension $5\text{ m} \times 5.75\text{ m}$, et 3 m de hauteur, est composée de murs avec fenêtres, d’une porte en bois, et une moquette peu épaisse sur le sol. Afin de simuler différentes configurations de pièces, des éléments absorbants sont disposés. Les configurations considérées ici sont :

- 1 : aucun élément absorbant ;
- 2 : éléments absorbants large fréquence sur les murs et devant les fenêtres ;
- 3 : éléments absorbants sur les murs ;
- 4 : élément absorbant pyramidal de profondeur 7 cm en ajout.

La Fig. 3.6 montre la précision de localisation à partir de notre modèle pour les différentes configurations de pièces, ainsi que pour les enregistrements effectués au sein du SCRIME. La ligne en pointillés rouges représente les résultats attendus.

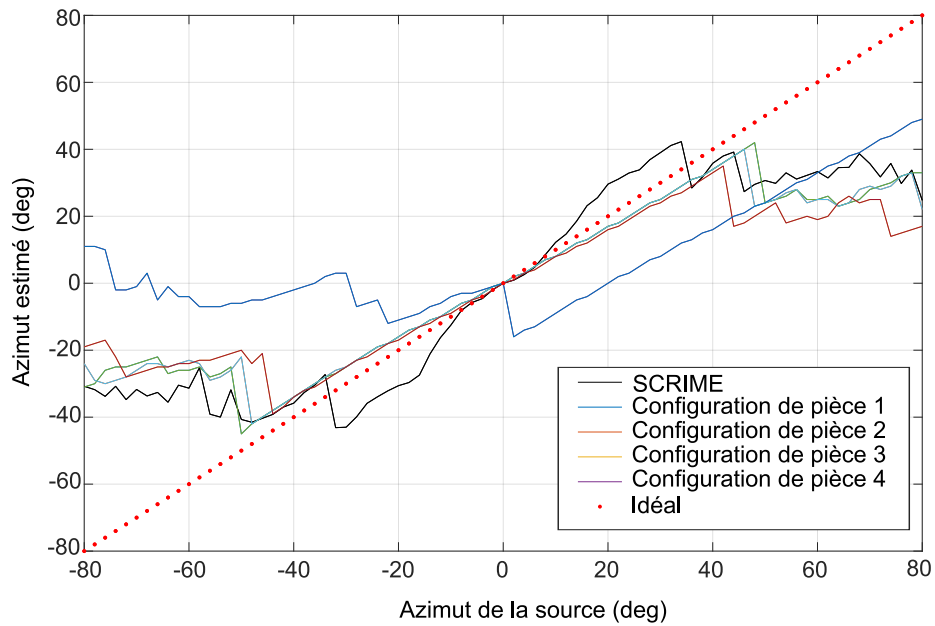


FIGURE 3.6 – Erreur d'estimation pour la localisation de l'azimut dans différentes configurations de pièces (BRIRs). Les résultats sont du même ordre que l'idéal dans l'intervalle $-40^\circ + 40^\circ$.

Les estimations d'azimut suivent bien dans l'ensemble l'idéal, pour les azimuts compris dans la fourchette $\pm 40^\circ$. Au-delà, un biais est visible, qui peut être expliqué par des réverbérations. En effet, si la source réverbérée possède une plus grande énergie que la source directe, alors la source réverbérée sera considérée comme l'azimut estimé. La Fig. 3.7 montre l'exemple de ce biais, la méthode de localisation aurait dû trouver 48 degrés, mais le pic d'énergie de la réverbération présente à 25 degrés est plus important, créant une confusion entre l'image source et la source réelle.

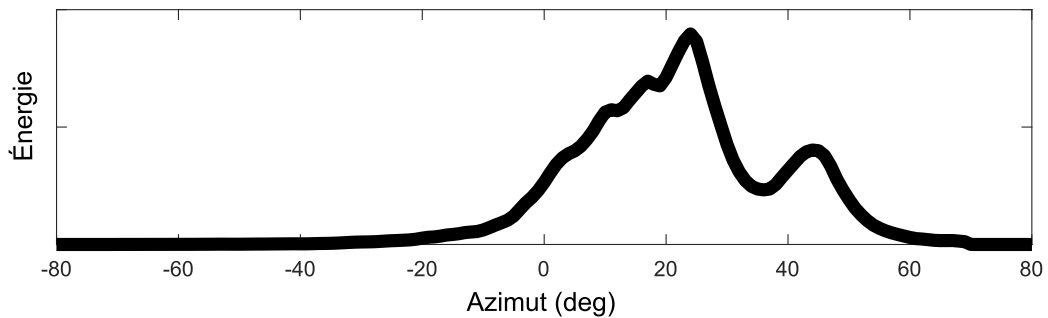


FIGURE 3.7 – Histogramme de répartition de l'énergie de la localisation de l'azimut dans des conditions réelles (première configuration de la base BRIR). L'exemple montre le résultat de la localisation pour une source réelle présente à 48° , l'énergie est dissipée et la source réverbérée (environ 25°) possède plus d'énergie que la source réelle.

Cependant, malgré cette confusion sur les angles importants, dans la fourchette $\pm 40^\circ$, la méthode de localisation fonctionne correctement et coïncide avec la précision de localisation humaine (Sec. 3.1.2).

La Fig. 3.8 montre les erreurs d'estimation pour chaque pièce et les 41 azimuts compris dans la fourchette $\pm 40^\circ$.

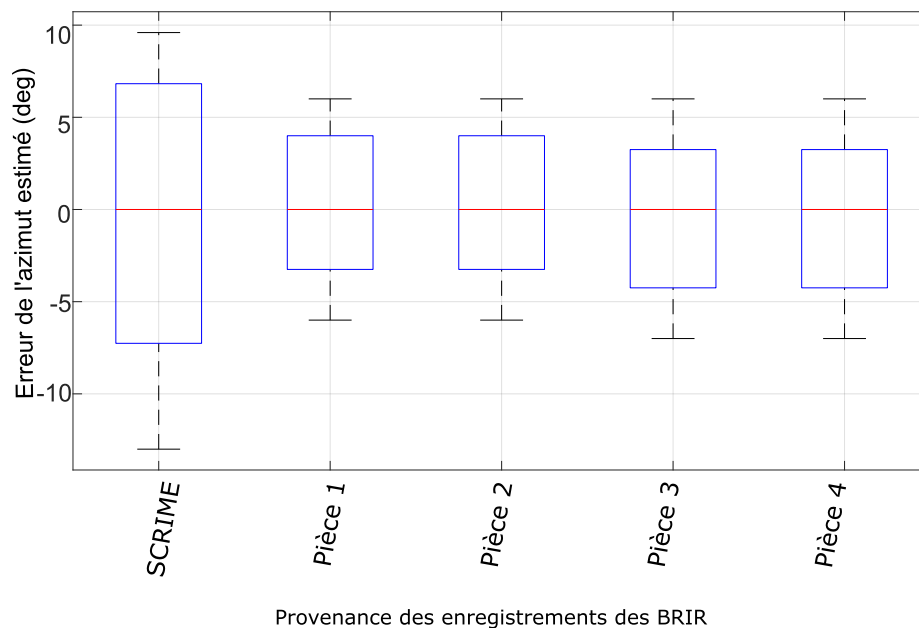


FIGURE 3.8 – Erreur d'estimation pour la localisation de l'azimut dans différentes configurations de pièces (BRIRs). Il est à noter que la pièce 1 génère de nombreux outliers, qui ne sont pas visibles sur ce tracé.

Ces résultats montrent que l'estimation ne possède pas de biais (erreur moyenne de 0); mais aussi un écart type comparable aux performances (confirmant ainsi les constatations de la Fig. 3.6). Ceci confirme que le modèle est performant y compris dans des conditions réelles.

Ces figures apportent également un second intérêt, en comparant les différentes bases BRIR avec les enregistrements effectués au SCRIME, montrant une bonne similitude.

– Conclusion –

Après avoir présenté la méthode de localisation perceptive, ce chapitre valide cette méthode en conditions anéchoïques/idéales dans un premier temps puis réelles dans un second temps, avec une limite pour les angles importants en conditions réelles due au biais causé par les réverbérations.

Aussi, la méthode de localisation perceptive STAR pour l'azimut fonctionne et est validée; elle sera donc utilisable comme outil objectif de caractérisation lors de la spatialisation de sources sonores.

Ce chapitre présente la localisation sonore perceptive pour l'azimut, d'après les travaux de Viste (2004) et Mouba and Marchand (2006). La méthode simplifiée pendant ce travail de thèse est dans un premier temps présentée, ainsi que les validations effectuées. Ces travaux sont publiés dans Méaux and Marchand (2020b), ainsi que dans Méaux and Marchand (2021b). Cette méthode est utilisée à des fins d'évaluation objective des méthodes de spatialisation.

Chapitre 4

Une approche perceptive pour la localisation sonore en élévation

– Introduction –

La localisation perceptive pour l'élévation est une approche nouvelle, et plus complexe que la localisation en azimuth.

Iida et al. (2007) dans leurs travaux de recherche proposent un modèle permettant de retrouver l'élévation en fonction des principaux nœuds et pics (NP), théorie intéressante, puisque les pics et nœuds peuvent être traités comme des indices spectraux.

L'objectif de ce chapitre est de présenter les travaux de recherche menés afin d'obtenir une méthode de localisation se basant sur ces nœuds et pics, utilisant notamment une technique inédite, couplant l'approche Shazam avec la théorie des nœuds et pics.

Les résultats sont donnés en considérant la base CIPIC dans son ensemble, ou le sujet 4 pour des tests isolés.

4.1 Origine des pics et nœuds

Les nœuds et pics correspondent aux maximums et minimums locaux sur les HRTFs. Takemoto et al. (2012) montrent que l'effet le plus important sur les HRTFs relève de la forme des oreilles (le reste du corps jouant une influence uniquement pour les fréquences en dessous de 5 kHz); et que celles-ci peuvent à elles seules permettre de retrouver les principaux pics et nœuds codant l'élévation (et donc rendant la localisation en élévation possible).

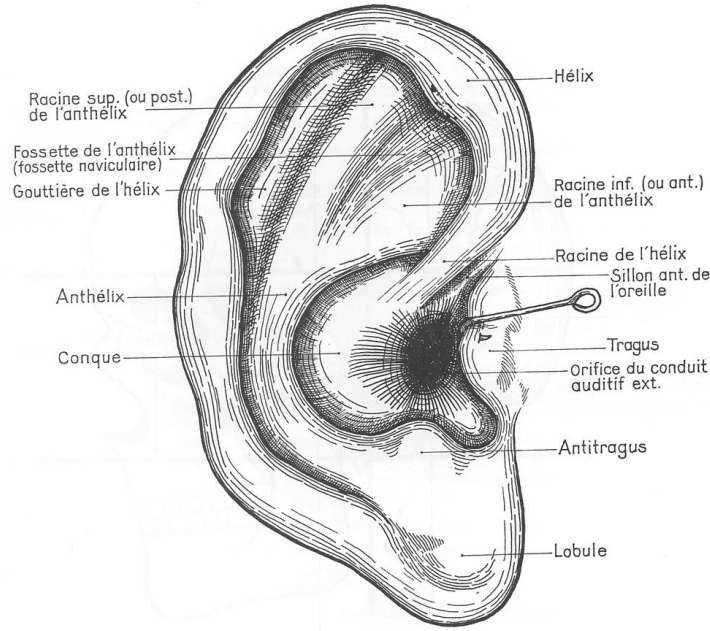


FIGURE 4.1 – Schéma du pavillon de l'oreille externe (*docplayer.fr*).

Des expériences permettant d'identifier l'origine des nœuds et pics ont été menées (Shaw (1997); Kahana and Nelso (2005)), montrant que l'origine des pics et nœuds provient de résonances se produisant dans les pavillons des oreilles externes.

Concernant le pic P1 qui nous intéressera par la suite, celui-ci provient du premier mode de résonance généré par la profondeur de la cavité de la conque. La fréquence correspond à l'inverse de la longueur d'onde, et à $1/4$ de la profondeur de cette cavité.

Contrairement à la génération des pics, constants lors des changements d'élévations, la génération des nœuds est un processus plus compliqué.

Raykar and Duraiswami (2005) ont émis l'hypothèse que les nœuds sont générés par des interférences entre la source sonore directe et la source sonore réverbérée par la conque. Une seconde hypothèse générée par Takemoto et al. (2012) est que différents nœuds dépendant de l'élévation de la source sonore sont générés avec des phases différentes, et que le nœud se forme à l'entrée du conduit auditif.

Dans ces deux hypothèses, la position des nœuds provient d'interférences entre la source directe et les réverbérations se produisant dans l'oreille, et ils dépendent de l'élévation.

4.2 Étude d'un modèle existant

4.2.1 Le modèle développé par Iida

Le modèle donné par Iida et al. (2007) correspond aux Eqn. 4.1 et 4.2, et permet de modéliser l'évolution des deux nœuds principaux (N1 et N2) en fonction de l'élévation. Le pic P1 est fixe. Le triplet N1N2P1 est suffisant d'après ce modèle pour la localisation en élévation.

$$f_{N1}(\phi) = 5.77 \cdot 10^{-5} \phi^4 - 2.41 \cdot 10^{-2} \phi^3 + 2.79 \phi^2 + 4.79 \cdot 10^1 \phi + 6.06 \cdot 10^3 \quad (4.1)$$

$$f_{N2}(\phi) = 2.35 \cdot 10^{-6} \phi^4 - 2.98 \cdot 10^{-3} \phi^3 + 4.85 \cdot 10^{-1} \phi^2 + 2.3 \cdot 10^1 \phi + 8.52 \cdot 10^3 \quad (4.2)$$

La Fig.4.2, est tracée à partir des Eqn. 4.1 et 4.2.

4.2. ÉTUDE D'UN MODÈLE EXISTANT

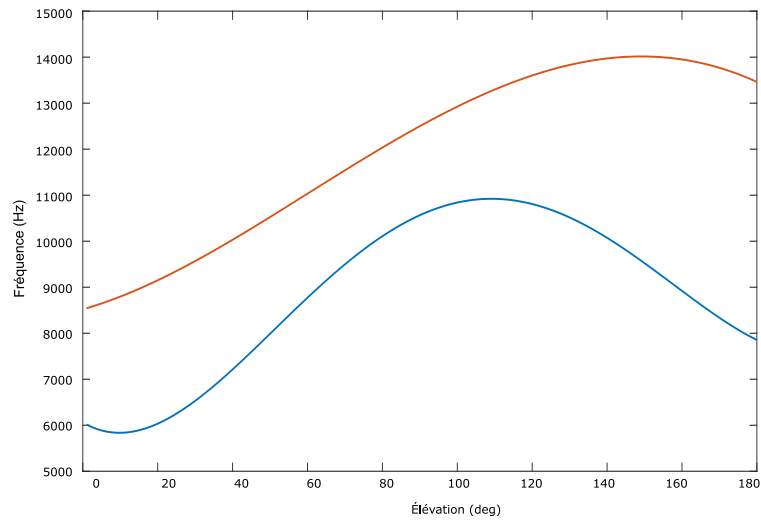


FIGURE 4.2 – Tracé du modèle d'après Iida et al. (2007) utilisé pour connaître la position des nœuds en fonction de l'élévation (N1 en bleu et N2 en rouge); la troisième composante permettant la localisation verticale étant le premier pic P1, constant aux alentours de 3500 Hz.

4.2.2 Application du modèle sur la base CIPIC

Pour le travail de cette thèse, il a été choisi la base CIPIC, comme présenté dans la Sec. 1.3.2. Cette section va comparer le modèle de Iida avec la base CIPIC.

Comparaison du modèle et de la moyenne des HRTFs de la base CIPIC

La figure 4.3 montre l'ensemble des HRTFs de la base CIPIC pour différentes élévations (ici la moyenne de tous les sujets). Sur ces HRTFs sont tracés le pic P1 (en noir) ainsi que N1 (en bleu) et N2 (en rouge).

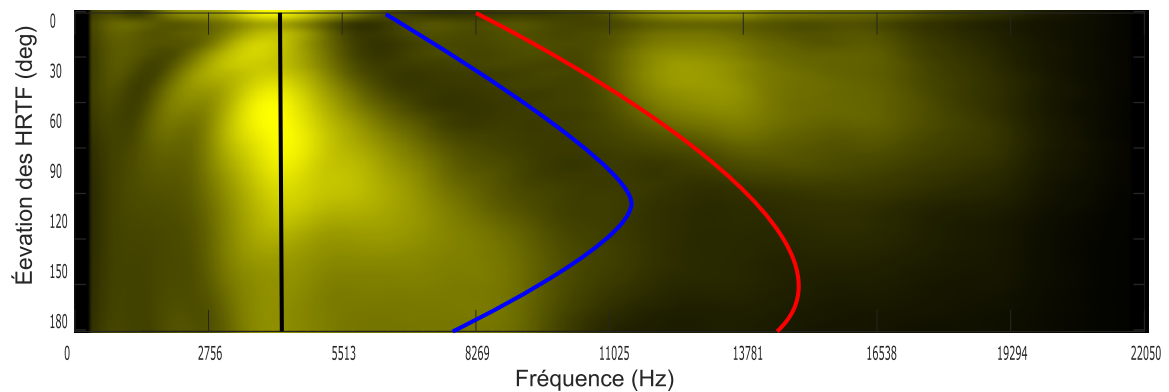


FIGURE 4.3 – Ensemble des HRTFs de la base CIPIC, avec P1 (ligne noire), N1 (ligne bleu) et N2 (ligne rouge) représentés.

En regardant la Fig. 4.3, le pic P1 est bien visible (en jaune), et ne semble pas varier en fonction de l'élévation, ainsi qu'un nœud variant avec l'élévation (noir suivant le tracé bleu). Cependant il est difficile ici de distinguer deux nœuds. Une hypothèse est que l'utilisation de la moyenne des HRTFs a lissé ces nœuds, ce qui n'est pas en accord avec un système permettant de localiser l'élévation à partir d'un modèle de pic et nœuds (il faut en effet que cela fonctionne pour tout individus, et donc pour la moyenne des individus).

Méthode de détection des nœuds et pics sur la base CIPIC

Afin de trouver les nœuds et pics, le procédé décrit par Iida (2017) est mis en place.

Procédé de détection des nœuds et pics :

1. Chargement des HRIRs gauches et droites en provenance de la base CIPIC.
2. Calcul des Early HRIRs.
 - (a) Calcul d'une fenêtre de Blackman Harris (largeur 15 ms).
 - (b) Détection de la position d'amplitude maximale de l'HRIR.
 - (c) 'Clipping' de la fenêtre centré sur le maximum de l'amplitude des HRIRs.
3. Passage en fréquentiel via une FFT (Transformée de Fourier Rapide).
4. Passage du binaural provenant des HRIRs gauches et droites en monaural (*les résultats étant quasiment identiques en prenant uniquement l'oreille gauche ou droite, il a donc été choisi d'utiliser un signal monaural*).
5. Suppression des micro-variations par un lissage sur 4 échantillons (équivalent à 400 Hz).
6. Détection de P1, P2 *n'étant pas utile pour la localisation, il n'est pas détecté*.
 - (a) Restriction de la zone de recherche dans l'intervalle $[3000 : 5000] Hz^1$, comprenant l'ensemble des pics possibles de la base CIPIC.
 - (b) Le pic d'amplitude maximale sur cet intervalle est défini comme étant P1.
7. Détection de N1 et N2.
 - (a) Restriction de la zone de recherche $[5500 : 11000] Hz^1$ pour N1 et $[7500 : 16000] Hz^1$ pour N2.
 - (b) Sélection de N2 comme étant celui d'amplitude maximale (*à condition que des différences significatives se retrouvent entre les pics d'après le travail de Moore et al. (1989)*). Dans le cas inverse sélection du N2 de fréquence maximale.
 - (c) N1 est déterminé comme étant le nœud suivant P1 et étant à distance raisonnable de N2 (seuil défini à 1500 Hz) *Les choix de N1 et N2 sont faits de manière à obtenir le plus souvent les bons N1 et N2*

Utilisation de la méthode de détection sur un sujet de la base

La méthode décrite Sec. 4.2.2 est appliquée sur un sujet de la base CIPIC. La Fig. 4.4 trace les HRTFs une fois traitées du sujet 4 pour 6 élévations différentes et $\theta = 0^\circ$. Cette figure compare les pics et nœuds trouvés en suivant la méthode de détection, les valeurs proposées par le modèle, ainsi que les plus proches candidats possibles du modèle.

Sur cette figure, on observe que le pic P1 est quasiment constant, coïncidant avec la théorie. Cependant, les nœuds et pics, bien que quasiment tout le temps détectés comme ceux également les plus proches du modèle sont assez éloignés, ce qui pose problème pour retrouver l'élévation.

1. Les valeurs données pour les zones de recherche proviennent du travail de Yan et al. (2014).

4.2. ÉTUDE D'UN MODÈLE EXISTANT

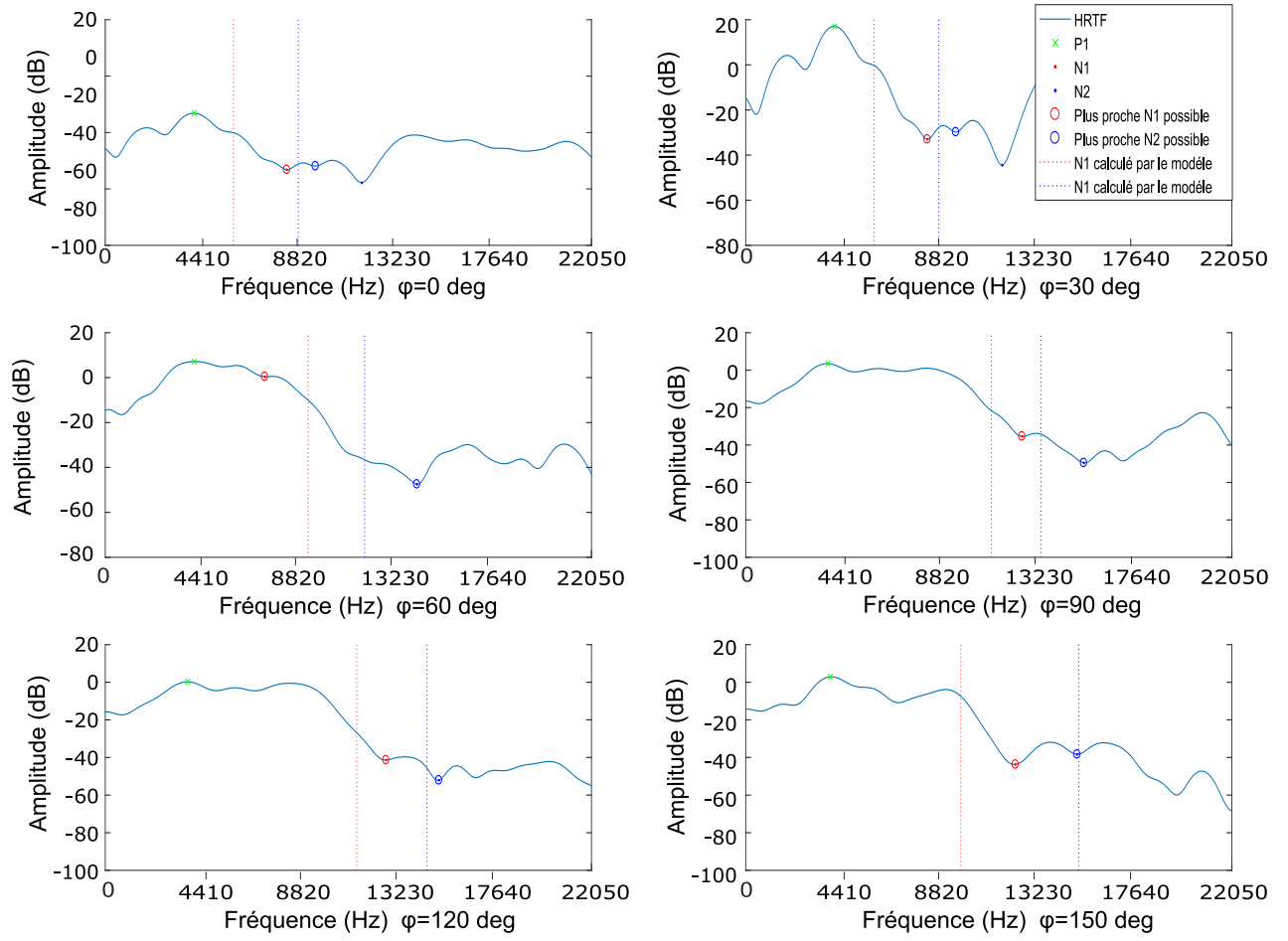


FIGURE 4.4 – Visualisation des HRTFs traitées pour 6 élévations, pour le sujet 4, en azimut $\theta = 0^\circ$.

En prenant en compte l'ensemble des élévations, on obtient la Fig. 4.5.

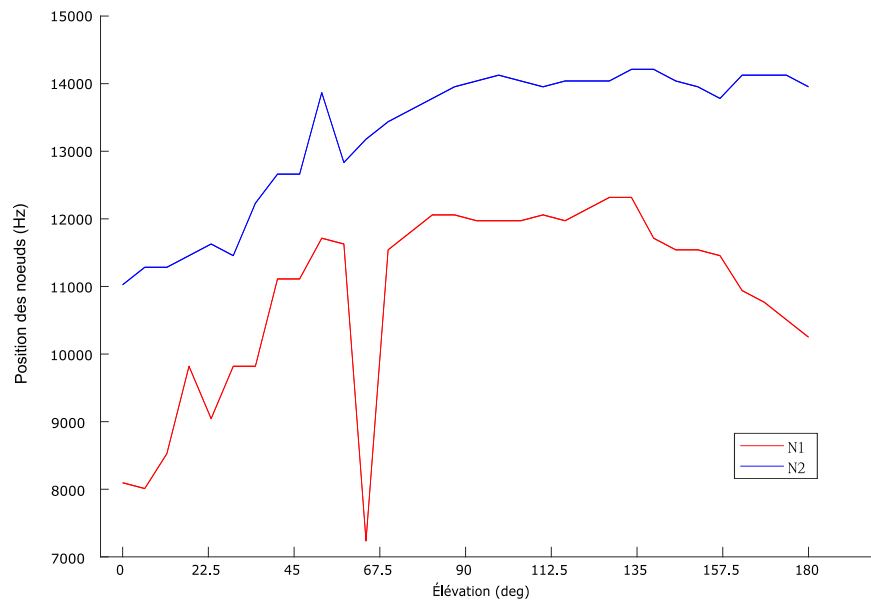


FIGURE 4.5 – Évolution des positions de N1 et N2 détectés en fonction de l'élévation. On observe une erreur pour l'élévation 60° .

Sur cette Fig. 4.5, l'allure correspond à l'allure du modèle en Fig. 4.2. Il faut tout de même noter que les valeurs sont sensiblement différentes, et que l'on a un décrochage important pour l'élévation 60° . Ce décrochage est également visible sur la Fig. 4.4, où il n'y a pas de nœuds proches du modèle. De plus, bien que le sujet 4 donne de plutôt bons résultats, ceci n'est pas le cas de tous les sujets. En effet, l'allure se retrouve sur une majorité de cas mais pour certains le tracé semble aléatoire, probablement par une récurrence du phénomène présenté ici pour $\phi = 60^\circ$.

L'évolution de N1 et N2 donnée par la méthode proposée Sec. 4.2.2 pour l'ensemble des sujets de la base est présentée sur la Fig. 4.6. Sur cette figure n'apparaît pas P1, ce pic étant stable avec une valeur d'environ 4000 Hz.

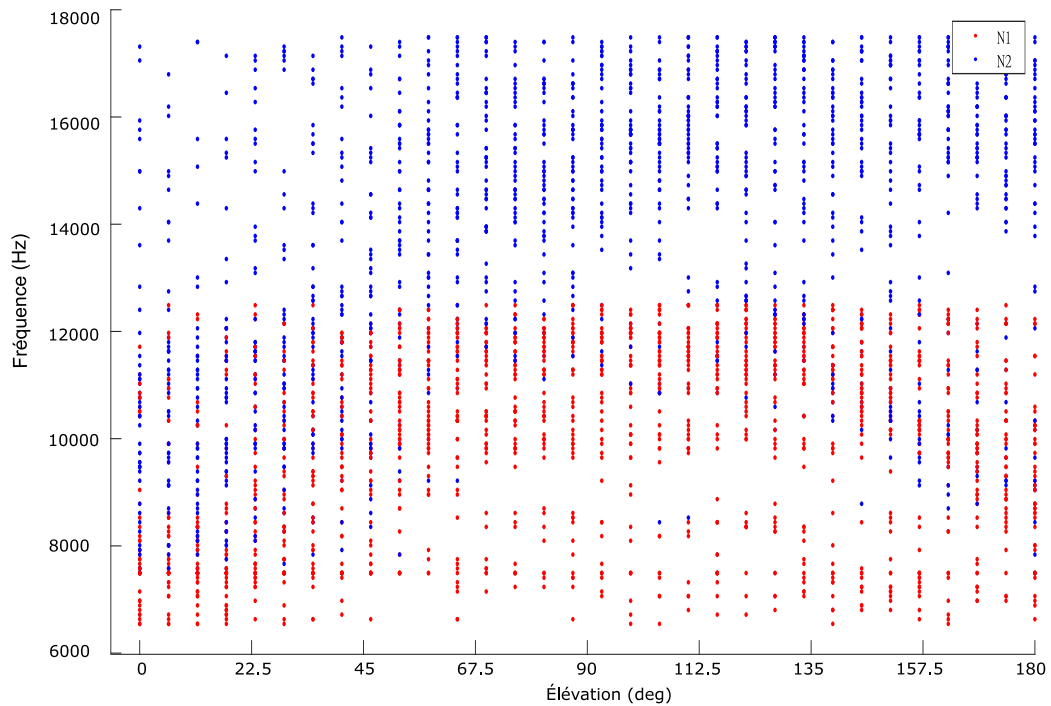


FIGURE 4.6 – Nœuds détectés sur l'ensemble des élévations et pour tous les sujets. Pour chaque élévation, il y a ainsi 45 points tracés pour N1 et pour N2.

La Fig. 4.6 met en exergue ce qui a été annoncé précédemment. Bien que l'allure des nœuds soit globalement respectée, ni une stabilité suffisante, ni les valeurs ne permettent de retrouver l'élévation des sujets de la base CIPIC à partir du modèle.

Étude des nœuds et pics les plus proches du modèle

N'étant pas possible de retrouver l'élévation via la méthode de localisation des nœuds et pics, une approche différente est de visualiser dans un second temps les nœuds les plus proches du modèle (comme déjà proposé dans la Fig. 4.4). La Fig. 4.7 trace de manière similaire à la Fig. 4.6 l'ensemble des nœuds N1 et N2 pour les 45 sujets de la base CIPIC. Cependant, au lieu d'utiliser la méthode de localisation présentée Sec. 4.2.2, N1 et N2 sont sélectionnés comme étant le plus proche possible de ceux attendus par le modèle.

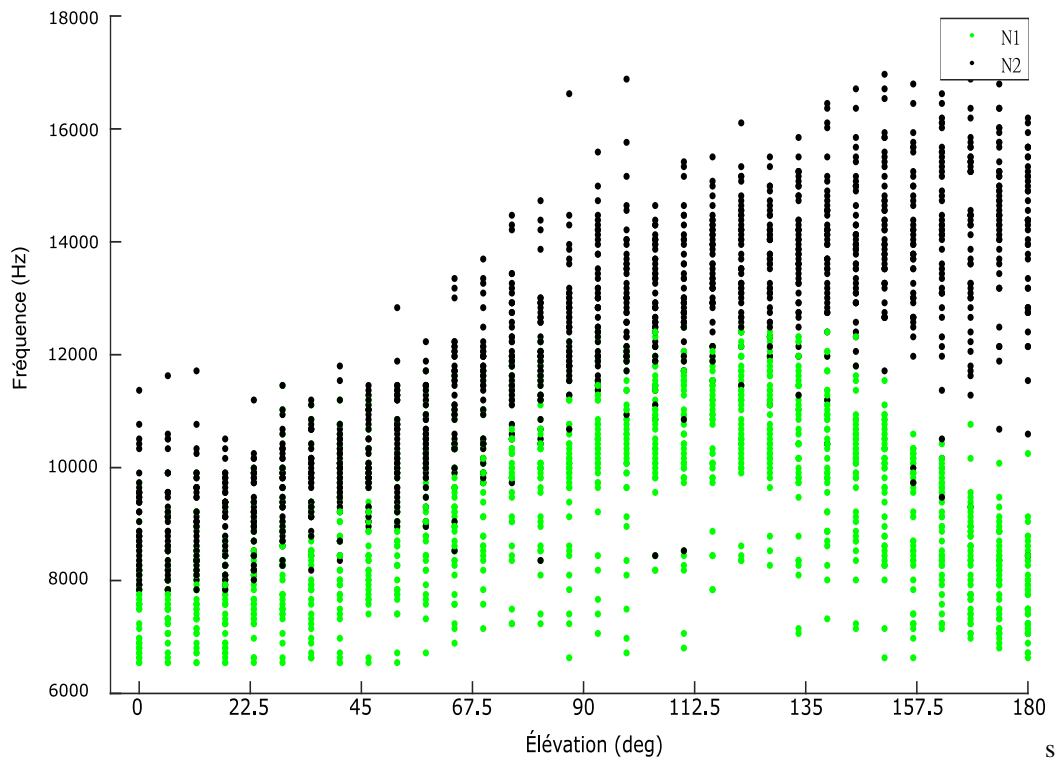


FIGURE 4.7 – Répartition des nœuds détectés au plus proche de ceux attendus par le modèle.

En procédant de cette seconde manière, l'évolution des nœuds proposés par le modèle est bien visible, et respecte mieux les valeurs attendues. Cependant, pour effectuer une localisation perceptive en élévation, il n'est pas possible de procéder ainsi (ne connaissant évidemment pas la position attendue des nœuds). De plus, même au plus près du modèle, une grande variabilité existe, qui est confirmée dans le travail de Yan et al. (2014), dont les valeurs pour la base CIPIC sont résumées dans le Tab. 4.1.

	Moyenne	Minimum	Maximum
P1	4095	3187	5340
N1	7545	5771	10939
N2	10384	7494	16107

TABLE 4.1 – Valeurs moyennes, maximales et minimales des pics et nœuds de la base CIPIC, d'après le travail de Yan et al. (2014) (valeurs en Hz).

Les données de ce tableau coïncident avec les différentes observations menées précédemment. En effet, une grande variabilité inter-sujet est visible, due au fait que chaque sujet possède des HRTFs potentiellement très différentes. À noter aussi que le travail de Yan et al. (2014) donne les résultats pour plusieurs bases de données d'HRTFs, présentant toutes des résultats similaires à la base CIPIC.

Bien que variables, il est prouvé que les pics et nœuds sont à l'origine de la localisation en élévation (présenté en Sec. 1.2.4). De plus, le travail de la Sec. 4.2 montre que le modèle de Iida est suivi par la majorité des sujets, même si la variabilité des sujets et l'identification des nœuds posent problème pour l'utiliser à des fins de localisation. C'est en se basant sur ces observations, et dans la volonté d'obtenir une méthode de localisation en élévation perceptive robuste qu'est née l'approche présentée dans la Sec. 4.3.

4.3 Utilisation d'une approche Shazam pour la localisation perceptive de l'élévation

Dans cette section, le but est d'appliquer une méthode similaire à la méthode Shazam proposée par Wang (2006), connue pour identifier des musiques sur une base de données. L'idée est d'appliquer cette méthode aux HRTFs, en se servant de la base CIPIC afin de créer une table d'apprentissage contenant la fréquence des pics, nœuds et différents couples nœud/pics, et ceci pour toutes les élévations. Ainsi, en théorie, à partir d'une HRTF inconnue, il est possible via la table d'apprentissage de retrouver l'élévation en utilisant les fréquences des différents pics et nœuds détectés sur cette HRTF inconnue.

L'idée générale de la méthode Shazam est de créer et stocker dans une table des marqueurs uniques obtenus sur des données (par exemple de morceaux de musique). Ainsi, lors d'une analyse d'un échantillon inconnu (partie d'une chanson), les marqueurs attribués de la même manière à l'échantillon sont comparés à ceux contenus dans la table et permettent de retrouver les données correspondantes (le titre de la chanson dans le cas de musiques). Le même procédé est appliqué sur les HRTFs, en utilisant les positions des pics et nœuds en tant que marqueurs.

En appliquant ceci au cas de l'étude présentée ici, l'ensemble des fréquences possibles est coupé en intervalles (dont la largeur est déterminée de manière judicieuse). Ensuite, pour chaque sujet et élévation, un marqueur est déterminé (un pic par exemple). La fréquence de chaque marqueur est alors utilisée pour ranger dans la bonne case de la table l'élévation connue correspondante. Dans le cas d'une table à plusieurs dimensions, chaque fréquence pour le premier marqueur possède, non pas les élévations pouvant lui être associées, mais une seconde table contenant l'ensemble des fréquences du second marqueur, et les élévations sont stockées dans cette seconde table. Le processus inverse utilise des HRTFs d'élévations inconnues. On récupère les marqueurs (fréquence du pic par exemple), et on se rapporte à la table afin d'obtenir les élévations possibles pour cette fréquence. Ceci fonctionne de manière similaire pour des paires de marqueurs, mais cette fois-ci en utilisant la table en deux dimensions.

4.3.1 Construction de la table

Étude du nombre de nœuds et pics détectés.

Une première investigation concerne le nombre de pics et nœuds détectés sur la base CIPIC en fonction de l'élévation, afin de vérifier que le nombre de données sera suffisant.

La Fig. 4.8 montre le nombre de pics et de nœuds en fonction de l'élévation. Le nombre de nœuds et pics détectés est assez important pour créer une table d'apprentissage correspondant à environ 4 nœuds et 2 pics possibles par sujet et par élévation. De plus, le nombre de nœuds et pics est constant quelle que soit l'élévation et ne devrait donc pas inclure de biais (il y a plus de nœuds détectés que de pics).

4.3. UTILISATION D'UNE APPROCHE SHAZAM POUR LA LOCALISATION PERCEPTIVE DE L'ÉLÉVATION

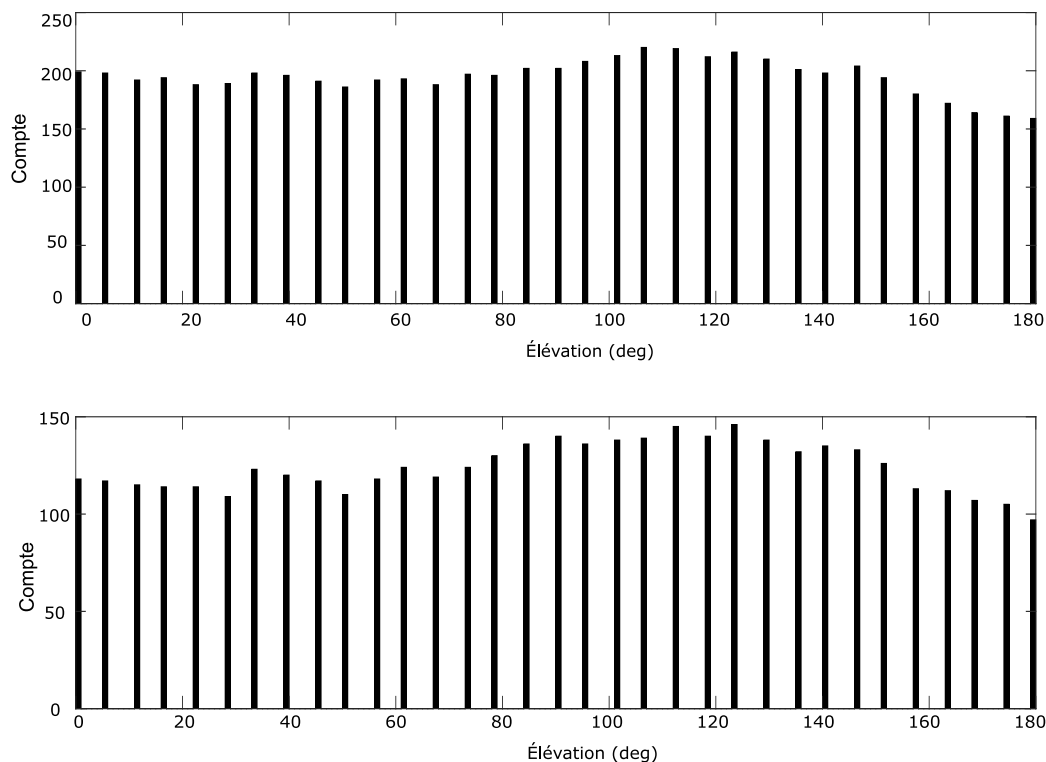


FIGURE 4.8 – Répartition du nombre de nœuds (premier graphique) et pics (second graphique) en fonction de l'élévation.

De plus, en effectuant la détection des pics candidats N1 et N2 avec la méthode se rapprochant le plus du modèle (proposée en Sec. 4.2.2, on obtient 98 N1 sans candidat et 84 N2 sans candidat (sur $45 \times 35 = 1575$ HRTFs testées), soit 6% pour N1 et 5% pour N2. Le pic P1 possède tout le temps un candidat.

Obtenir une table Shazam compacte

Avec cette méthode, il est important d'obtenir une table d'apprentissage compacte, permettant une correspondance lors d'une recherche dans la table quelle que soit la fréquence. En effet, si les pas de la table sont trop faibles, il risque d'y avoir des 'trous', empêchant de faire correspondre certaines données. Dans le cas inverse, si le pas est trop important, il y aura des collisions (présentées en Sec. 4.3.1), diminuant la précision de la méthode, voire, dans les cas extrêmes, ne produisant plus qu'un unique résultat.

Les Fig. 4.9 et 4.10¹, montrent différents pas pour P et N, deux pas arbitraires de 10 et 110 et le pas le plus juste possible. Afin de trouver cette valeur, le pas est incrémenté jusqu'à ce que la table n'ait plus de trous. Cette première valeur de pas sans trous est considérée comme optimale. Ces valeurs sont, dans cet exemple, de 86 pour N et 87 pour P; qui sont les plus faibles valeurs de tables compactes. Il est intéressant de se rendre compte que la répartition pour les nœuds est bien plus étendue et ne présente pas spécialement de pics lorsque l'on considère toutes les élévations, alors que les pics montrent une répartition bien moins homogène et regroupée autour de la valeur de P1 théorique.

1. Les valeurs de la table sont limitées pour les pics et nœuds, l'échelle des figures est donc adaptée et n'est pas comprise entre 0 et 22050 Hz (majorité des figures de ce travail).

4.3. UTILISATION D'UNE APPROCHE SHAZAM POUR LA LOCALISATION PERCEPTIVE DE L'ÉLÉVATION

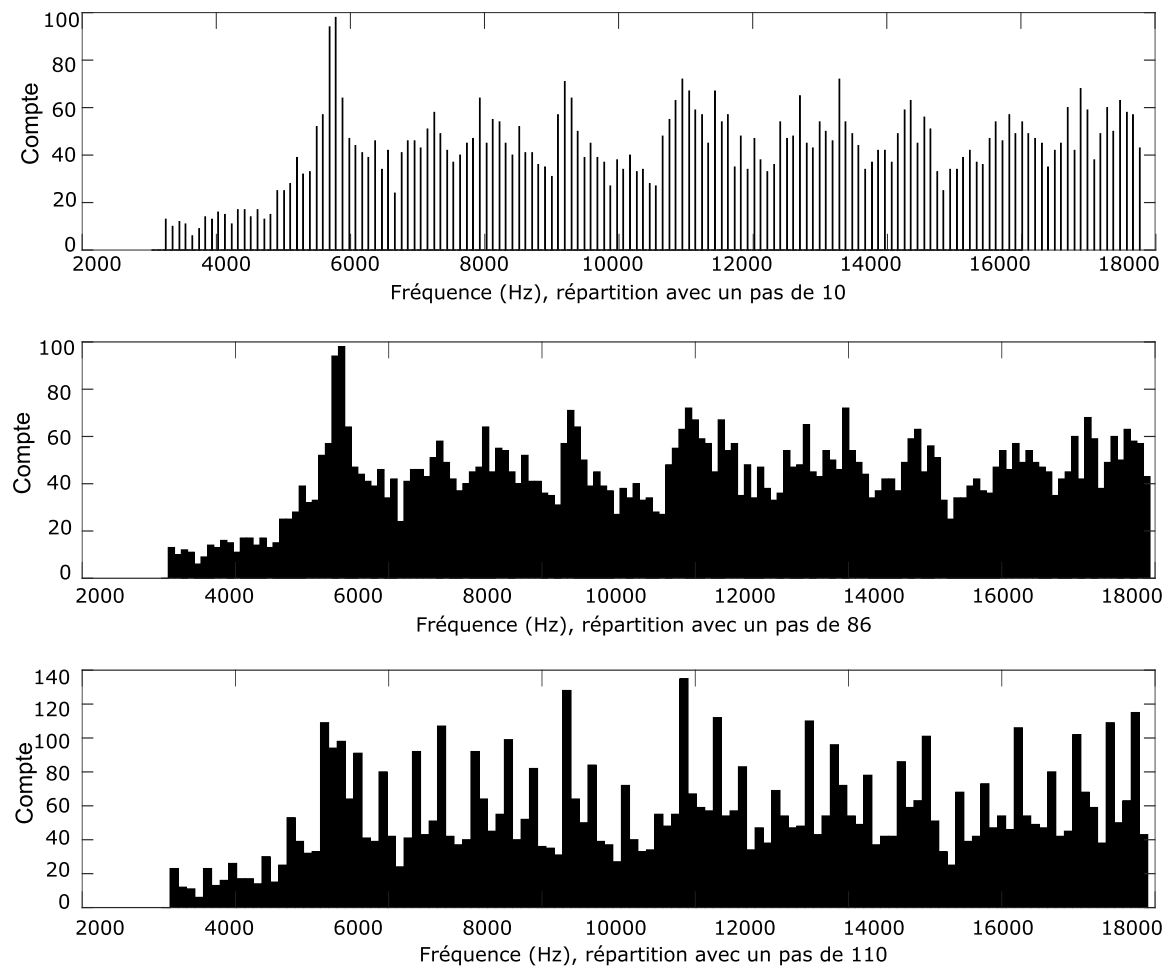


FIGURE 4.9 – La table Shazam compacte pour les nœuds de la base CIPIC. Le pas idéal est de 86 Hz.

4.3. UTILISATION D'UNE APPROCHE SHAZAM POUR LA LOCALISATION PERCEPTIVE DE L'ÉLÉVATION

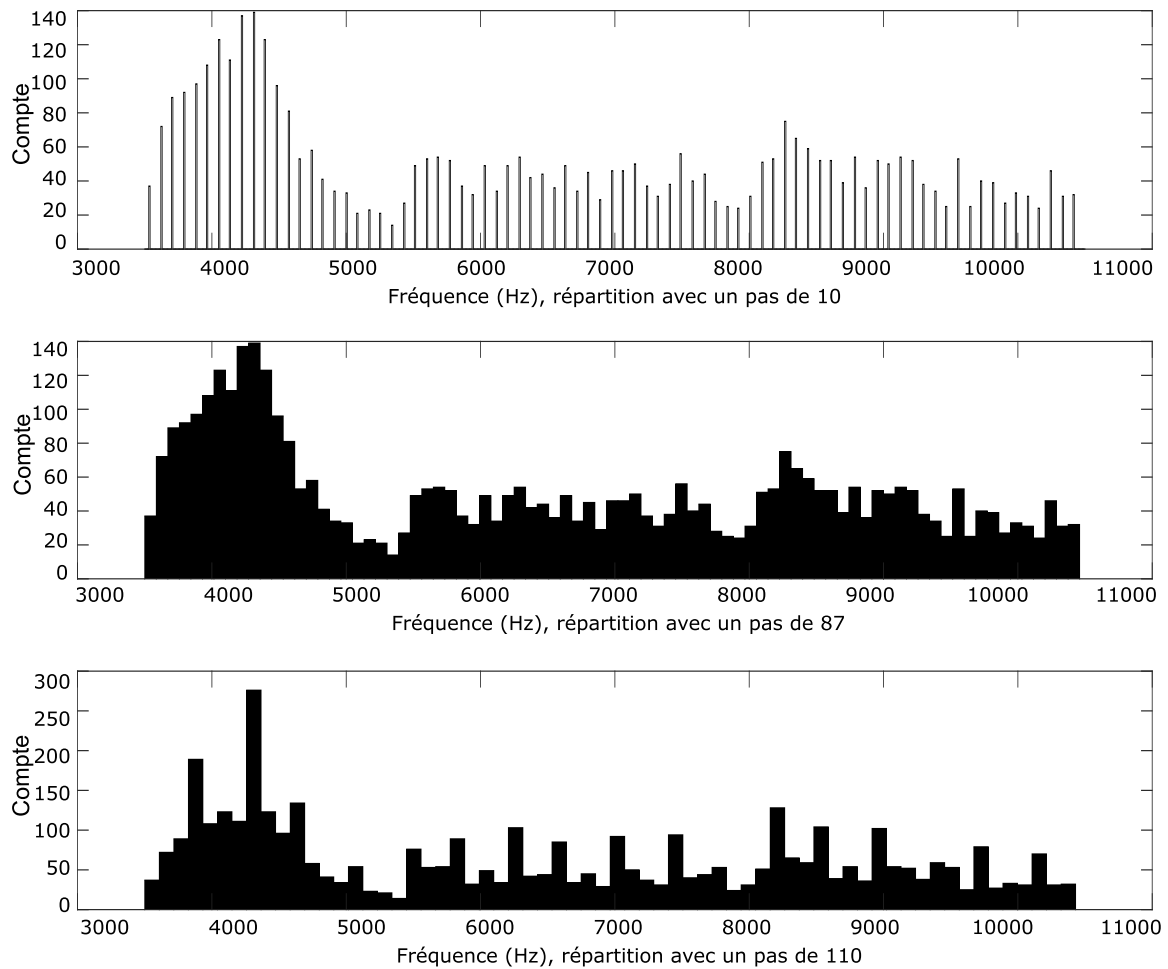


FIGURE 4.10 – La table Shazam compacte pour les pics de la base CIPIC. Le pas idéal est de 87 Hz.

Quand apparaissent les collisions

Pour un bon fonctionnement de la méthode Shazam, il faut un nombre important de détections (table compacte) avec peu de collisions (ajout d'erreurs, donnant pour une même entrée de nombreuses données différentes). La Fig. 4.11 propose des exemples de tables avec collisions, avec sur le même intervalle de fréquence jusqu'à 700 élévations stockées, et qui sont variées (la collision sur la table pour les pics étant similaire à celle pour les nœuds, elle n'est pas présentée ici).

4.3. UTILISATION D'UNE APPROCHE SHAZAM POUR LA LOCALISATION PERCEPTIVE DE L'ÉLEVATION

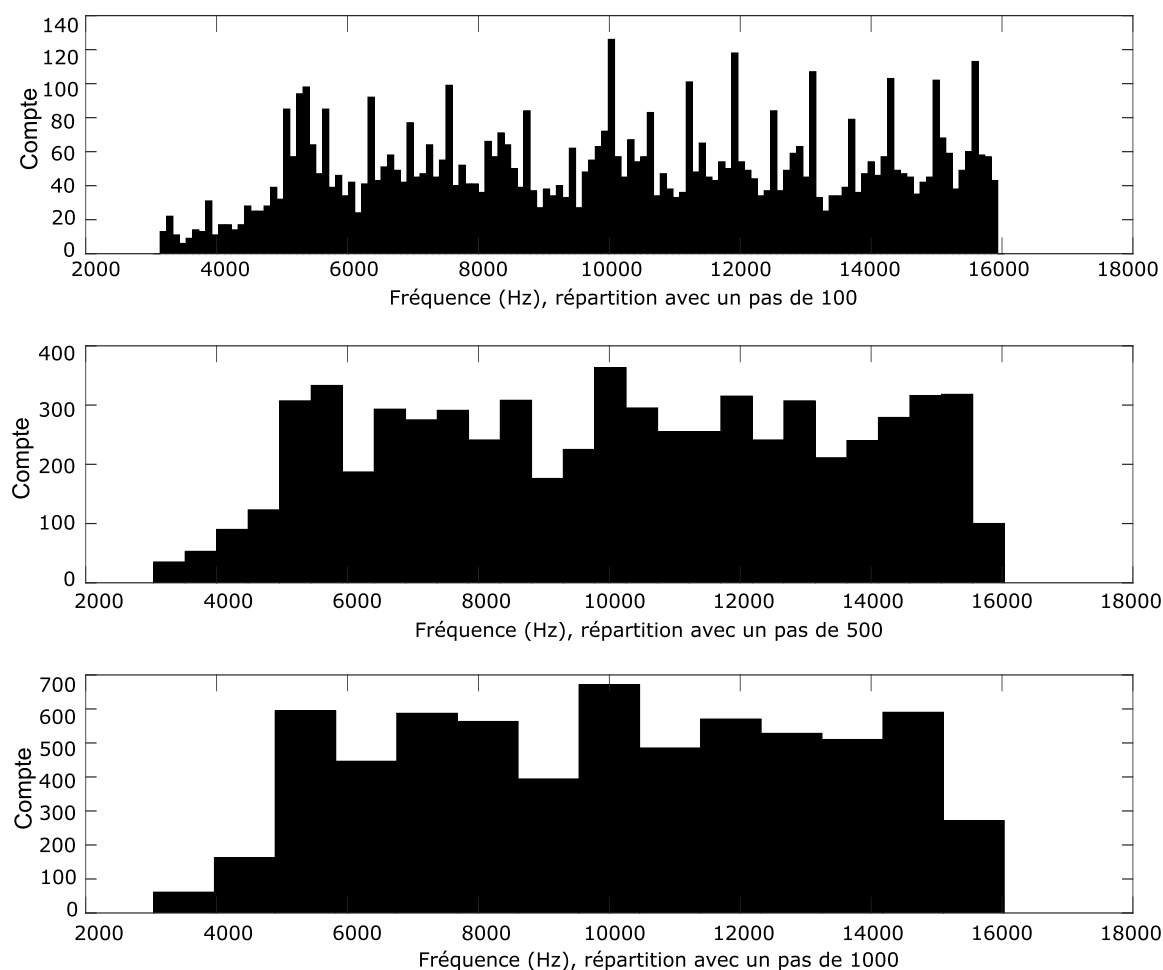


FIGURE 4.11 – *Histogrammes de table Shazam de collisions pour les nœuds.*

4.3.2 Utilisation de la table pour la détection de l'élévation

En utilisant les pas présentés dans la Sec. 4.3.1, les tables de nœuds et pics sont créées. Cette partie va les utiliser et discuter leurs performances afin de retrouver l'élévation à l'aide d'une HRTF connue puis inconnue. À noter que des tables N1 N2 ont été ajoutées aux tables de positions de N, contenant la valeur la plus proche pour leurs constructions et limitant la recherche aux valeurs possibles lors de la recherche.

Cas d'un sujet contenu dans la base

La Fig. 4.12 montre les différentes combinaisons de pics et nœuds testées afin de retrouver l'élévation d'un sujet appris dans la base (ici le sujet test est le sujet 4 de la base CIPIC pour l'élévation 90°).

Le sujet étant appris dans la base, il devrait être retrouvé. Cependant, il est possible que celui-ci soit 'noyé' dans la masse d'autres sujets et élévations, et que la bonne élévation ne ressorte pas. C'est très souvent le cas si l'on regarde des données isolées comme uniquement un pic et un nœud. C'est pourquoi la méthode Shazam utilise des combinaisons. En utilisant des rapports entre différents points, le nombre de marqueurs est en effet grandement augmenté. Le cas de notre sujet montre parfaitement ceci sur la Fig. 4.12, en essayant de retrouver l'élévation avec une donnée unique, la bonne valeur (élévation 90°) est présente mais noyée dans la masse. Grâce à une combinaison, le nombre de correspondances fausses est drastiquement réduit. On voit ainsi que les combinaisons PN, NP, N1P, N2P, N1N2 et NN produisent le bon résultat, et ceci largement au-dessus des fausses détections.

4.3. UTILISATION D'UNE APPROCHE SHAZAM POUR LA LOCALISATION PERCEPTIVE DE L'ÉLEVATION

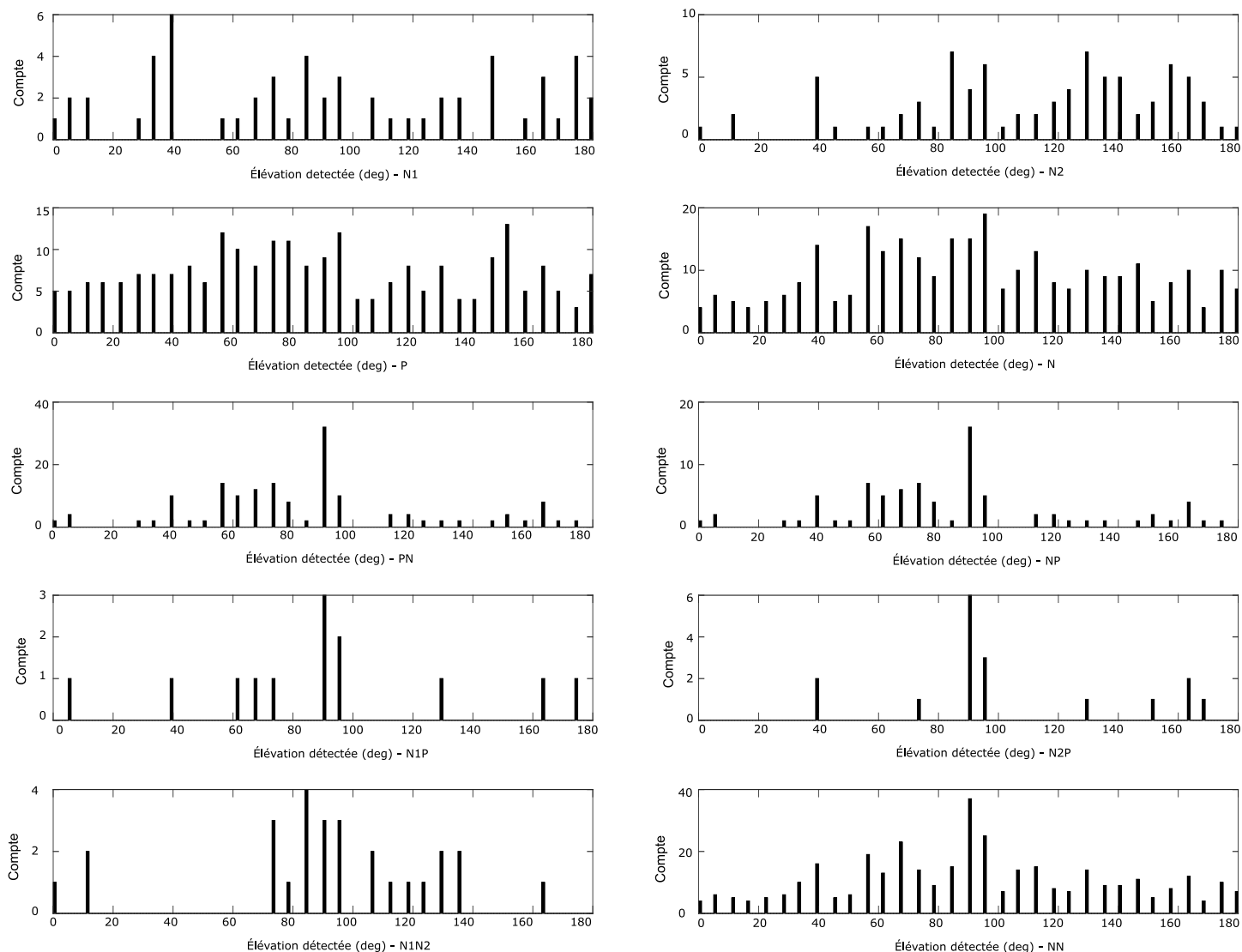


FIGURE 4.12 – *Histogrammes pour un sujet appris (ici sujet 4, élévation 90°).*

Il n'est donc pas nécessaire de regarder toutes les données comme présentées sur la Fig. 4.12, on se contentera dans la suite de cette étude des conglomerats de combinaisons (résultats additionnés).

La Fig. 4.13 montre les différents conglomerats considérés, à savoir N1P, N2P, NP ; N1P, N2P, NP, NN, PN ; et NP, PN, NN.

Ces conglomerats semblent fournir des résultats identiques.

4.3. UTILISATION D'UNE APPROCHE SHAZAM POUR LA LOCALISATION PERCEPTIVE DE L'ÉLEVATION

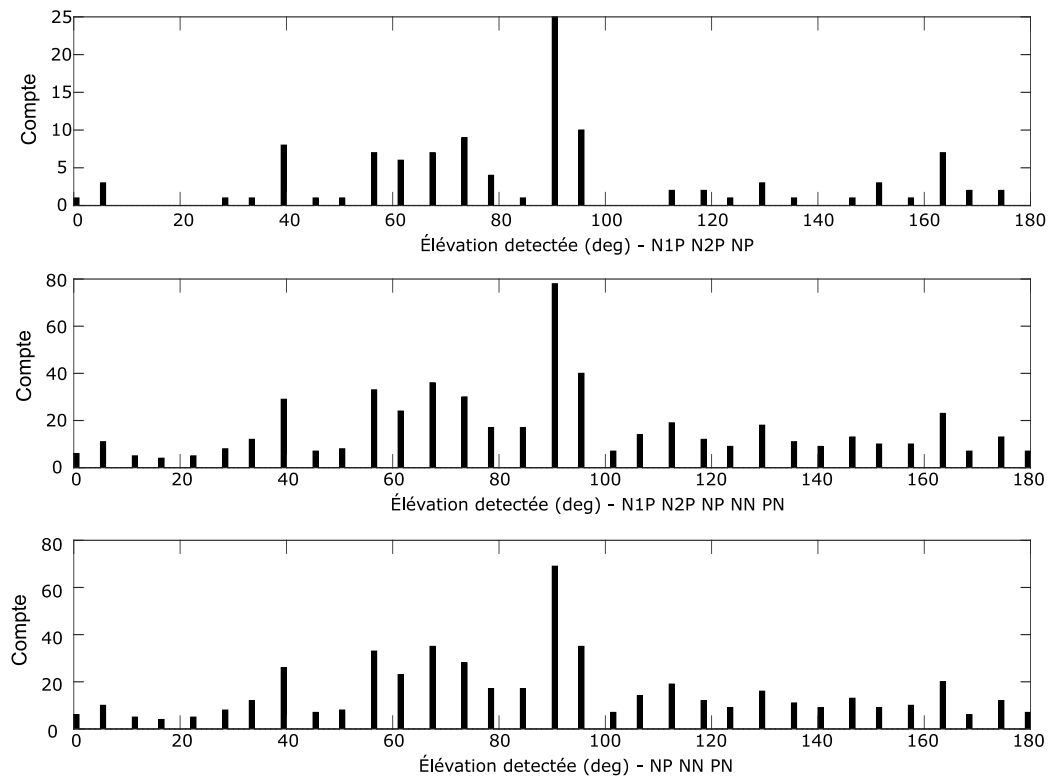


FIGURE 4.13 – *Histogrammes des conglomerats de couples NP pour un sujet appris (ici sujet 4, élévation 90°).*

La Fig. 4.14 présente l'évolution de la détection en fonction de l'élévation, dans le cas d'un sujet contenu dans la base. Cette figure présente en noir la bonne détection (compte dans la base de l'élévation réelle), en pointillés bleus la plus mauvaise détection, qui correspond à la valeur du pic le plus élevé et ne correspondant pas à l'élévation testée. Enfin la ligne verte montre la moyenne des valeurs des autres pics.

4.3. UTILISATION D'UNE APPROCHE SHAZAM POUR LA LOCALISATION PERCEPTIVE DE L'ÉLEVATION

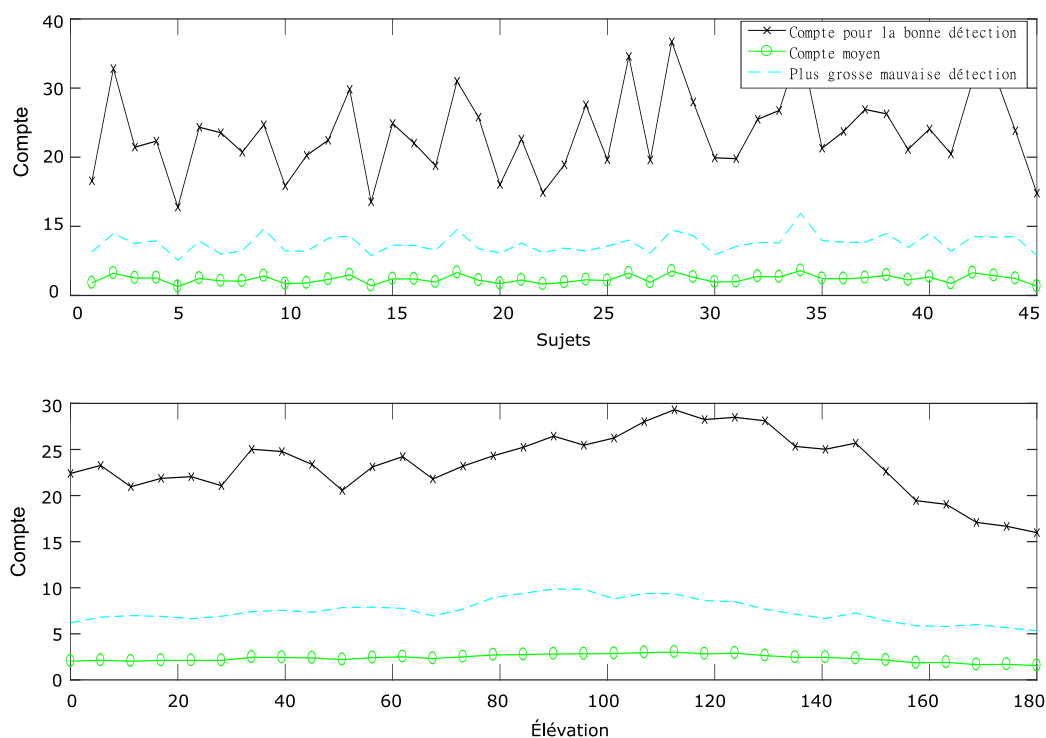


FIGURE 4.14 – Synthèse des détections sur l'ensemble de la base pour des sujets appris, en considérant le conglomérat NIP N2P NP.

Cette synthèse montre que la bonne détection est très supérieure à la moyenne ainsi qu'au second plus grand pic de l'histogramme¹, et ceci que l'on regarde les sujets ou bien l'élévation. On notera aussi que l'on a seulement 1.18 % (sur l'ensemble des élévations et sujets) des cas où le pic le plus élevé ne correspond pas à l'élévation testée.

La méthode proposée ici, dans le cas d'un sujet connu est donc valable et utilisable.

Cas d'un sujet non contenu dans la base

Pour ces tests, la table Shazam est réapprise sans prendre en compte le sujet dont l'on souhaite déterminer les élévations relatives aux HRTFs. La Fig. 4.15 montre les résultats. Bien que le pic à 90° soit présent, la moyenne des autres valeurs est très proche, et de plus, un pic d'une mauvaise détection est supérieur.

1. Nous appelons second plus grand pic de l'histogramme la deuxième élévation détectée, la première pouvant être bonne ou pas. Ainsi, si la première élévation est la bonne détection en regardant l'écart avec la seconde, on voit si cette détection est robuste (gros écart). Et dans le cas inverse, on peut observer si le second pic est le bon.

4.3. UTILISATION D'UNE APPROCHE SHAZAM POUR LA LOCALISATION PERCEPTIVE DE L'ÉLEVATION

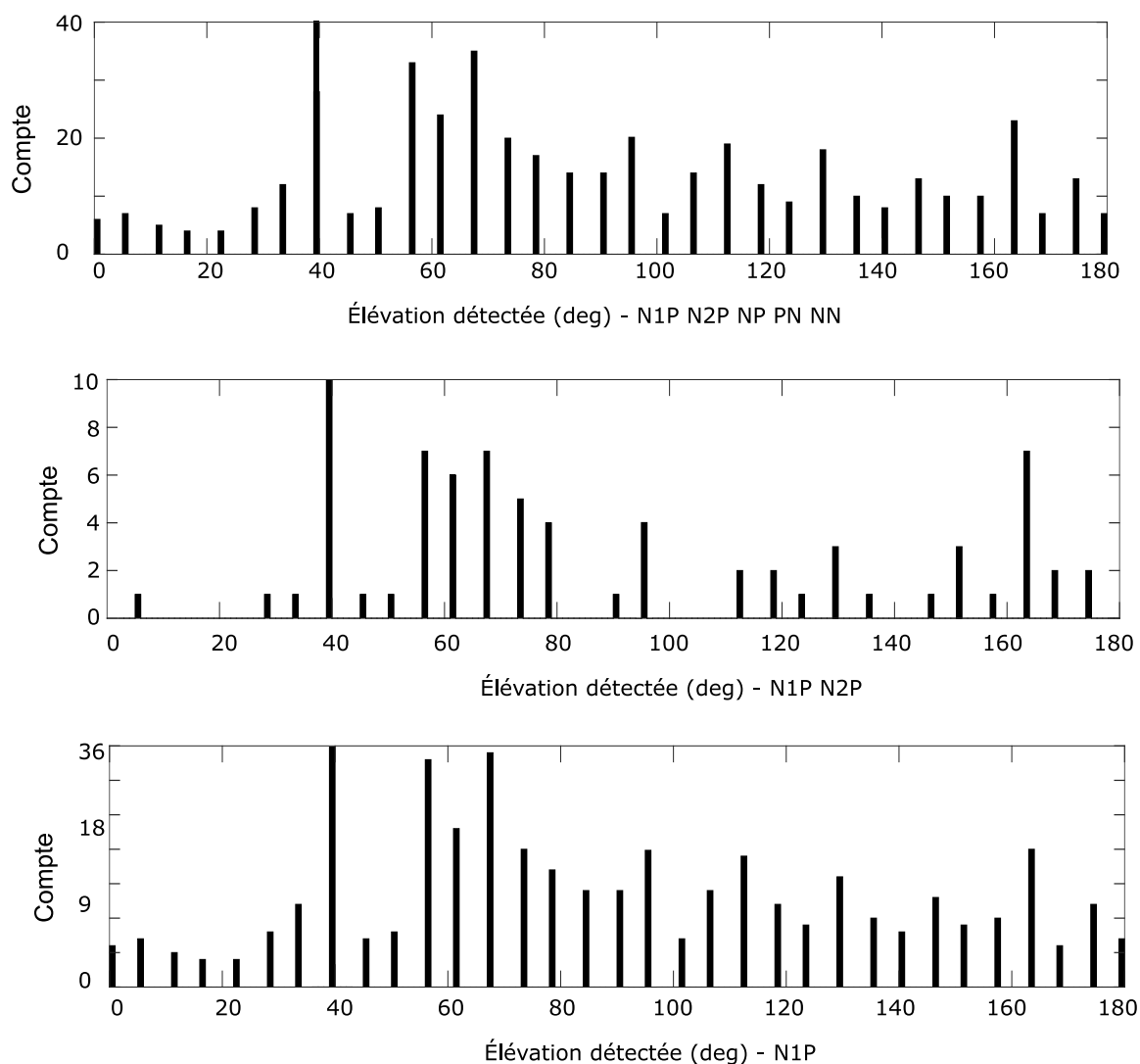


FIGURE 4.15 – *Histogrammes des conglomerats de couples NP pour un sujet non appris (ici sujet 4, élévation 90°).*

La Fig. 4.16, synthétise les détections pour tous les sujets et toutes les élévations. On constate en effet que la plus mauvaise détection est toujours plus importante que la bonne détection, et, de plus, que la bonne détection est tout juste supérieure à la moyenne des pics. Plus généralement, l'occurrence d'une mauvaise détection est d'environ 99 %, rendant impossible l'utilisation de cette méthode sans la connaissance des HRTFs du sujet.

4.3. UTILISATION D'UNE APPROCHE SHAZAM POUR LA LOCALISATION PERCEPTIVE DE L'ÉLEVATION

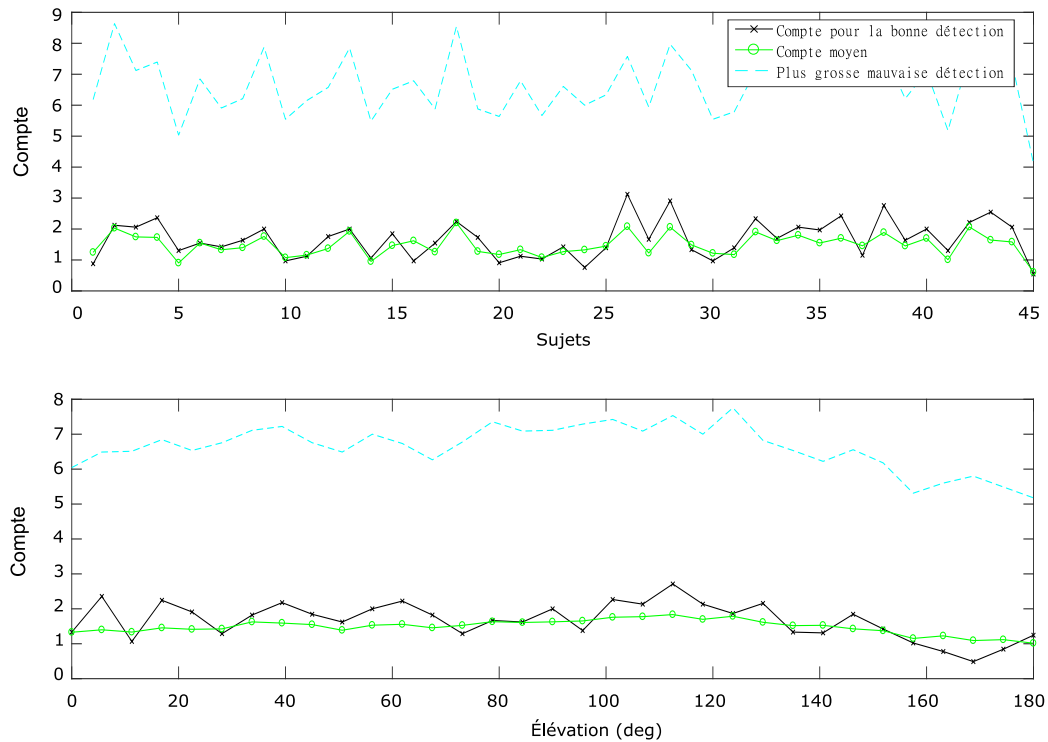


FIGURE 4.16 – Synthèse des détections sur l'ensemble de la base pour des sujets non appris, en considérant le conglomérat N1P N2P NP.

4.3.3 Cas de l'utilisation d'HRTFs théoriquement identiques

Une des utilisations souhaitées de cette méthode était une calibration automatique de l'algorithme de spatialisation, en localisant automatiquement la position des différents haut-parleurs. On a vu que si les HRTFs du sujet n'étaient pas apprises, il n'était pas possible de retrouver l'élévation via cette méthode ; mais que, dans le cas de données apprises, la méthode fonctionnait et était robuste. Or, de nombreuses bases de données proposent les HRTFs provenant de mannequins, dont par exemple le mannequin KEMAR. La question qui se pose donc est de savoir si cette méthode de localisation perceptive pour l'élévation fonctionnerait pour des HRTFs similaires ; cas par exemple d'un enregistrement en conditions réelles via le mannequin KEMAR.

La Fig. 4.17, montre les tests effectués avec l'enregistrement d'un mannequin KEMAR dans le studio du SCRIME (qui diffère donc légèrement des HRTFs qui sont apprises dans la base).

4.3. UTILISATION D'UNE APPROCHE SHAZAM POUR LA LOCALISATION PERCEPTIVE DE L'ÉLEVATION

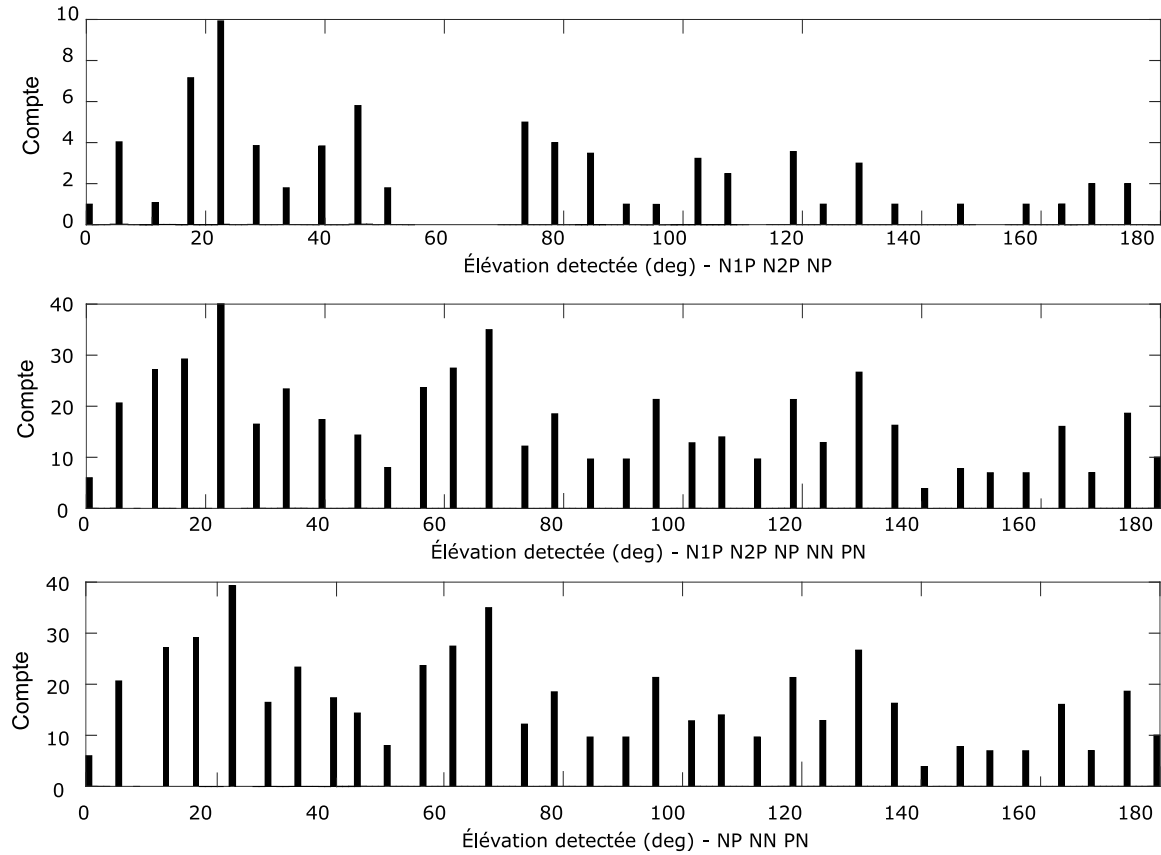


FIGURE 4.17 – Utilisation de la méthode Shazam sur les HRTFs du SCRIME, enregistrées pour $\theta = 0^\circ$ et $\phi = 25^\circ$.

La Fig.4.17 montre de bons résultats¹. Lors de l'utilisation d'un mannequin KEMAR pour l'enregistrement et de l'utilisation des HRTFs de ce mannequin pour l'apprentissage dans la base, la méthode proposée pour la localisation perceptive de l'élévation semble fonctionner.

1. Le même test pour l'élévation $\phi = 40^\circ$ donne des résultats similaires mais légèrement moins démarqués.

– Conclusion –

Le premier point de ce chapitre s'est porté sur le modèle de Iida permettant théoriquement de retrouver, à l'aide d'un modèle de nœuds et pics présents dans les HRTFs, l'élévation de la source. Cette méthode ne fonctionnant pas en pratique avec la base CIPIC considérée, ce qui est dû notamment à la grande variabilité de ces nœuds et pics, nous avons été amenés à tester une méthode innovante.

Cette méthode est la combinaison des théories provenant du modèle de Iida avec une technique similaire à celle de la méthode Shazam. Une table permet d'apprendre dans un tableau les élévations associées aux pics, nœuds et ensembles pics/nœuds considérés. Il résulte de cette méthode de très bons résultats lorsque le sujet testé a également servi à la construction du tableau. Cette méthode ne fonctionne cependant pas dans le cas où le sujet n'est pas compris dans le panel utilisé pour l'apprentissage. Un dernier test encourageant permet de voir que la méthode semble fonctionner pour des HRTFs proches, avec dans notre exemple l'utilisation d'HRTFs d'un mannequin KEMAR pour l'apprentissage dans la base et d'enregistrements BRIRs avec ce même mannequin, provenant des studios du SCRIME.

Cette méthode possède donc de nombreuses perspectives; avec dans un premier temps une amélioration de la détection de pics et nœuds, une augmentation du nombre de sujets composant la base; et dans un second temps une sélection des sujets ayant des caractéristiques physiologiques proches du sujet, ou encore la spécialisation d'HRTFs. Aussi, cette méthode pourrait par la suite être utilisée comme la méthode de localisation en élévation, afin de caractériser objectivement des sons spatialisés.

Ce chapitre présente une idée innovante pour la localisation sonore perceptive en élévation, tout en s'appuyant sur les travaux de Iida et al. (2007). Bien que présentant des résultats encourageants, la méthode n'est pas actuellement suffisamment robuste pour avoir été publiée et nécessite encore des travaux de recherche. C'est une des perspectives qui s'ouvrent à la fin de cette thèse.

Troisième partie

Spatialisation sonore perceptive

À propos de la partie spatialisation sonore perceptive

Cette partie, qui est le coeur de la méthode STAR, discute de la spatialisation sonore perceptive. Le but est de respecter la philosophie souhaitée pour STAR – facilité de mise en œuvre, peu coûteuse en calculs, et utilisant une approche perceptive et générique.

Le premier chapitre de cette partie présentera la spatialisation sonore perceptive pour l'azimut, dont les modèles synthétiques ont déjà été présentés précédemment ; avec une discussion sur chaque étape.

Le second chapitre présentera la spatialisation sonore perceptive pour l'élévation, reposant sur les indices acoustiques spectraux et plus précisément les principaux nœuds et pics que l'on retrouve dans les HRTFs ; et utilisant des modèles afin de conserver l'aspect synthétique de la méthode.

Le troisième et dernier chapitre de cette partie présente la spatialisation sonore perceptive pour la distance. Contrairement aux autres, il est compliqué d'obtenir une spatialisation 'brute', puisque le procédé choisi pour STAR spatialise l'éloignement des sources relativement (entre deux sources ou sur la même source en fonction du temps). Là aussi, l'utilisation de modèles est appliqué pour respecter le côté synthétique de la méthode.

Les tests de plusieurs domaines de validations des méthodes, étant plus conséquents que ceux menés pour la localisation et représentant une partie importante de cette thèse, seront présentés dans la dernière partie.

La spatialisation en azimut est proposée en partant du travail de thèse de Joan Mouba (Mouba (2009)), Nous avons ici poursuivi et testé cette dimension, mais aussi éliminé des erreurs détectées, avec discussions et études sur différents facteurs. Concernant l'élévation et la distance, l'élévation est un apport à part entière de cette thèse, et la distance reprend et développe les travaux de thèse de Joan Mouba.

La spatialisation STAR en 3D

Les prochains chapitres décrivent en détail la façon dont les différentes dimensions sont spatialisées dans la méthode STAR. Cette section est à part et a pour but d'introduire le procédé de spatialisation de STAR 3D dans son ensemble.

La Fig. 4.18 permet une synthèse de la méthode STAR.

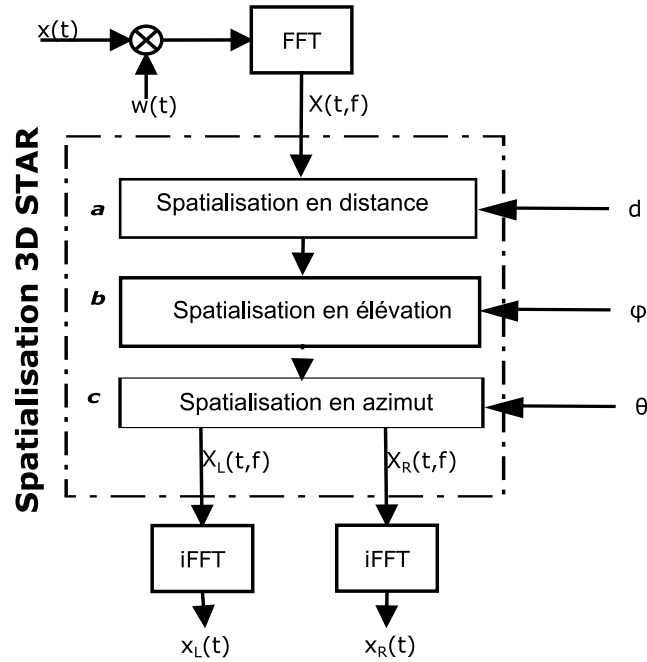


FIGURE 4.18 – *Processus complet de spatialisation de la méthode STAR.*

Partie a : Le signal mono est spatialisé en fonction de la distance d , suivant le procédé expliqué au chapitre 7.

Partie b : Le signal mono est spatialisé en élévation en fonction de ϕ . Dans le cas de plusieurs couronnes de haut-parleurs, les deux couronnes entourant la source sont utilisées, en appliquant VBAP et STAR, comme expliqué dans le chapitre 6.

Partie c : Le (ou les dans le cas de plusieurs couronnes) signal mono est spatialisé en azimuth sur les deux haut-parleurs encadrant la source tel que présenté dans le chapitre 5.

Chapitre 5

La spatialisation en azimuth

– Introduction –

Ce premier chapitre présente la première des trois dimensions de spatialisation de la méthode STAR : l'azimut, dont la spatialisation fut en premier lieu développée par Mouba and Marchand (2006); Mouba et al. (2008), puis améliorée dans les travaux de Méaux and Marchand (2019, 2021a).

Ce chapitre présente la méthode STAR pour la spatialisation de l'azimut, tout en discutant des différents éléments constitutifs.

5.1 Principe général de la spatialisation de l'azimut de la méthode STAR

La méthode STAR est conçue pour s'appliquer à différentes sources sonores (ou objets spatiaux). Chaque source constitue un signal sonore à jouer à une position donnée pour un temps donné.

Cette méthode consiste à effectuer une annulation croisée des chemins entre les deux haut-parleurs et les oreilles des auditeurs (d'une manière transaurale (sec. 2.2.3)), en utilisant des chemins acoustiques provenant de modèles (donc synthétiques (sec. 5.2)). Le modèle de chemins acoustiques est basé sur les indices acoustiques utilisés par le système humain pour la localisation (Sec. 1.2.2). Le but étant de donner aux auditeurs l'impression d'une spatialisation dans l'espace de chaque source, sans pour autant essayer de reconstituer l'onde acoustique correspondante, telle que souhaité pour une approche perceptive.

La Fig. 5.1 présente une configuration type dans laquelle la méthode STAR est utilisée. Nous utilisons dans ce genre de configuration le paradigme par paire proposé par Chowning (1971), consistant à choisir uniquement les deux haut-parleurs les plus proches de la source sonore; correspondant à un premier haut-parleur à la droite de la source et un second à sa gauche pour la spatialisation de l'azimut. Ce fonctionnement est similaire à la méthode VBAP (Sec. 2.2.2) en deux dimensions. Aussi, quand la source provient exactement de la position d'un haut-parleur, le signal est directement joué par ce haut-parleur sans recourir à un processus de spatialisation (la source sonore étant alors placée exactement au bon endroit de sa spatialisation dans l'espace).

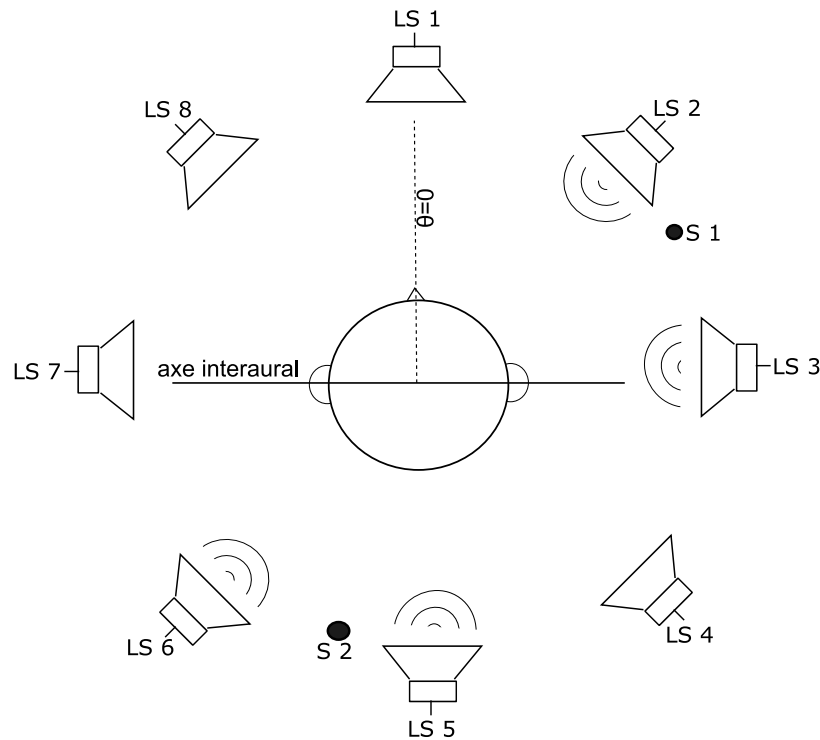


FIGURE 5.1 – Configuration octophonique, représentant huit haut-parleurs. Dans cet exemple, deux sources $S1$ et $S2$, et quatre haut-parleurs actifs : $S1$ utilisant les haut-parleurs $LS2$ et $LS3$, et $S2$ utilisant les haut-parleurs $LS5$ et $LS6$.

La Fig.5.2 présente le fonctionnement général de la méthode STAR, dont les traitements s'effectuent dans le domaine spectral. Chaque source est ainsi passée en fréquentiel à l'aide d'une transformée de Fourier à court terme (STFT), utilisant une transformée de Fourier rapide (FFT), puis traitée et distribuée sur les différents haut-parleurs, dont le signal est obtenu via une FFT inverse. Ainsi, pour n sources et m hauts-parleurs, $n + m$ FFTs sont calculées (par exemple 6 dans le cas illustré dans la Fig.5.1). Pour les calculs, la méthode utilise une fenêtre de Hann et des trames d'échantillons de taille 1024, avec $F_e = 44100$ Hz, et en utilisant un chevauchement de 50%.

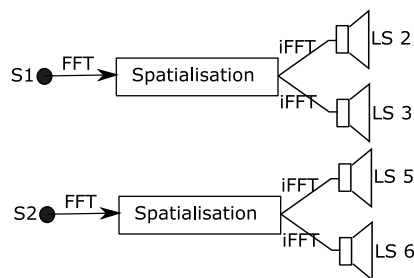


FIGURE 5.2 – Principe général de spatialisation de la méthode STAR.

Les basses fréquences dans ce genre de traitements peuvent entraîner certains problèmes. Avec la STFT, par exemple, des clics pour des sources mobiles apparaissent lorsqu'il n'y a pas assez de périodes du signal contenues dans la fenêtre w . Pour remédier à ces problèmes potentiels, un filtre est appliqué sur les fréquences inférieures à 150Hz avant d'appliquer la spatialisation. Ces basses fréquences sont ensuite réintroduites de manière égale dans toutes les enceintes (il aurait été possible également d'utiliser un caisson de basse, choix qui n'a pas été retenu dans nos expérimentations). Ce procédé n'influe en rien la spatialisation, les humains

localisant très mal les basses fréquences (Blauert (1997)).

5.2 Chemins acoustiques synthétiques

Cette section reprend la construction des chemins acoustiques synthétiques de la méthode STAR.

5.2.1 Les chemins acoustiques de la méthode STAR

La méthode STAR considère donc les chemins acoustiques allant de la source sonore vers les oreilles des auditeurs. Ces chemins acoustiques sont représentés dans le domaine spectral par des fonctions de transfert et pour la méthode STAR calculés à partir d'un modèle d'indices acoustiques. Ces mêmes chemins sont les HRTFs dans le cas d'enregistrements en condition anéchoïque et les BRIRs pour les autres conditions.

Ces modèles d'indices sont présentés de manière complète dans la Sec. 1.4. Pour rappel succinct, ils sont calculés en utilisant l'Eqn. 5.1.

$$\begin{aligned} \text{ILD}(\theta, f) &= \alpha(f) \sin(\theta) \\ \text{ITD}(\theta, f) &= \beta(f)r \sin(\theta)/c \end{aligned} \quad (5.1)$$

qui utilise les coefficients α et β , redonnés sur la Fig. 5.3, obtenus par correspondance aux HRTFs de la base CIPIC en utilisant la méthode des moindres carrés.

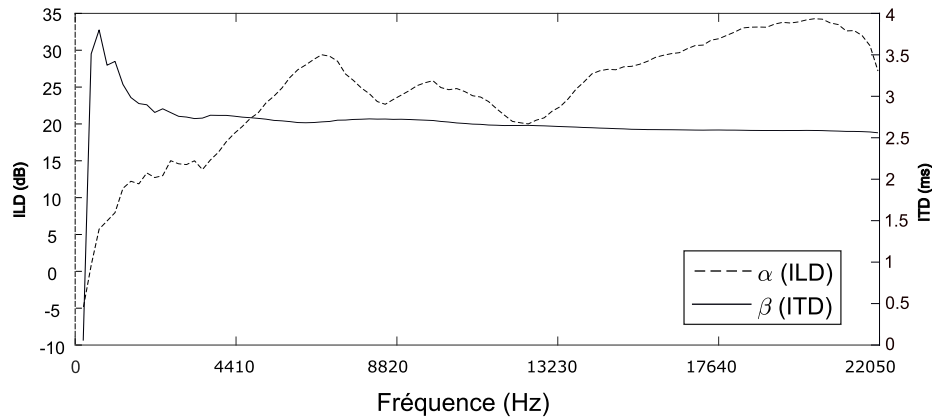


FIGURE 5.3 – Coefficients α et β , obtenus par correspondance au modèle via la méthode des moindres carrés, et contenant l'ensemble des informations relatives aux têtes et oreilles de chaque sujet de la base, et ceci pour toutes les fréquences et tous les azimuts.

On obtient alors les ITDs (Fig.1.16) et ILDs (Fig.1.12) synthétiques, continues dépendant uniquement des coefficients α , β et de l'azimut θ .

À partir de ces indices acoustiques synthétiques, l'Eqn. 5.2 propose une méthode pour obtenir des chemins acoustiques, en se basant sur le fait que les HRTFs sont symétriques (observés en Sec. 1.3.3).

$$\begin{aligned} H_L &= 10^{+\Delta_a(f)/2} \cdot e^{+i\Delta_\phi(f)/2} \\ H_R &= 10^{-\Delta_a(f)/2} \cdot e^{-i\Delta_\phi(f)/2} \end{aligned} \quad (5.2)$$

où $\Delta_a(f)$ et $\Delta_\phi(f)$ sont donnés dans l'Eqn. 5.3, permettant une simplification de l'écriture des chemins synthétiques.

$$\begin{aligned} \Delta_a(f) &= \text{ILD}(\theta, f)/20 \\ \Delta_\phi(f) &= \text{ITD}(\theta, f) \cdot 2\pi f \end{aligned} \quad (5.3)$$

5.2. CHEMINS ACOUSTIQUES SYNTHÉTIQUES

Alors H_L et H_R sont les chemins entre la source et l'oreille gauche et entre la source et l'oreille droite respectivement.

La Fig. 5.4 propose une étude plus précise de la symétrie des HRTFs, en utilisant la localisation d'un nœud, élément caractéristique des HRTFs (Iida et al. (2007)), ceci en faisant varier l'azimut pour une élévation $\phi = 0^\circ$, et en regardant l'amplitude du nœud pour l'oreille gauche et l'oreille droite.

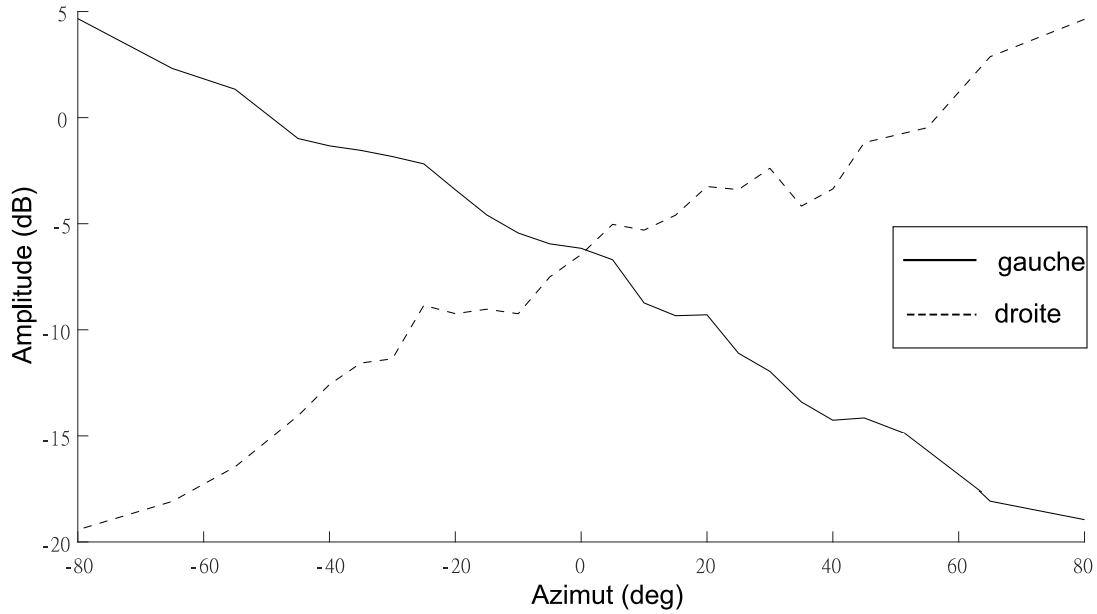


FIGURE 5.4 – Second nœud (N_2) en fonction de l'azimut θ , et pour une élévation $\phi = 0^\circ$. La symétrie des HRTFs est visible ici.

5.2.2 Discussion sur les chemins acoustiques

Les Fig. 5.5 et 5.6 présentent l'erreur quadratique entre les HRTFs réelles de la base CIPIC, et celles calculées synthétiquement (présentés en Sec. 5.2 à partir des ITDs et ILDs synthétiques présentées en Sec. 1.4).

Cette erreur est globalement du même ordre de grandeur que lors de la comparaison d'HRTFs individuelles, avec cependant une augmentation pour les valeurs extrêmes. Ces valeurs ne seront en pratique jamais atteintes dans la méthode STAR (l'angle maximal étant de $1/2$ de l'angle formé par les haut-parleurs entourant la source, pour atteindre un angle de 80° , il faudrait un écartement des haut-parleurs de 160° , ce qui n'a aucun intérêt avec cette méthode de spatialisation).

Ces deux figures permettent donc de valider le choix des chemins synthétiques, dont l'erreur avec les HRTFs réelles est globalement du même ordre de grandeur qu'entre HRTFs individualisées.

5.2. CHEMINS ACOUSTIQUES SYNTHÉTIQUES

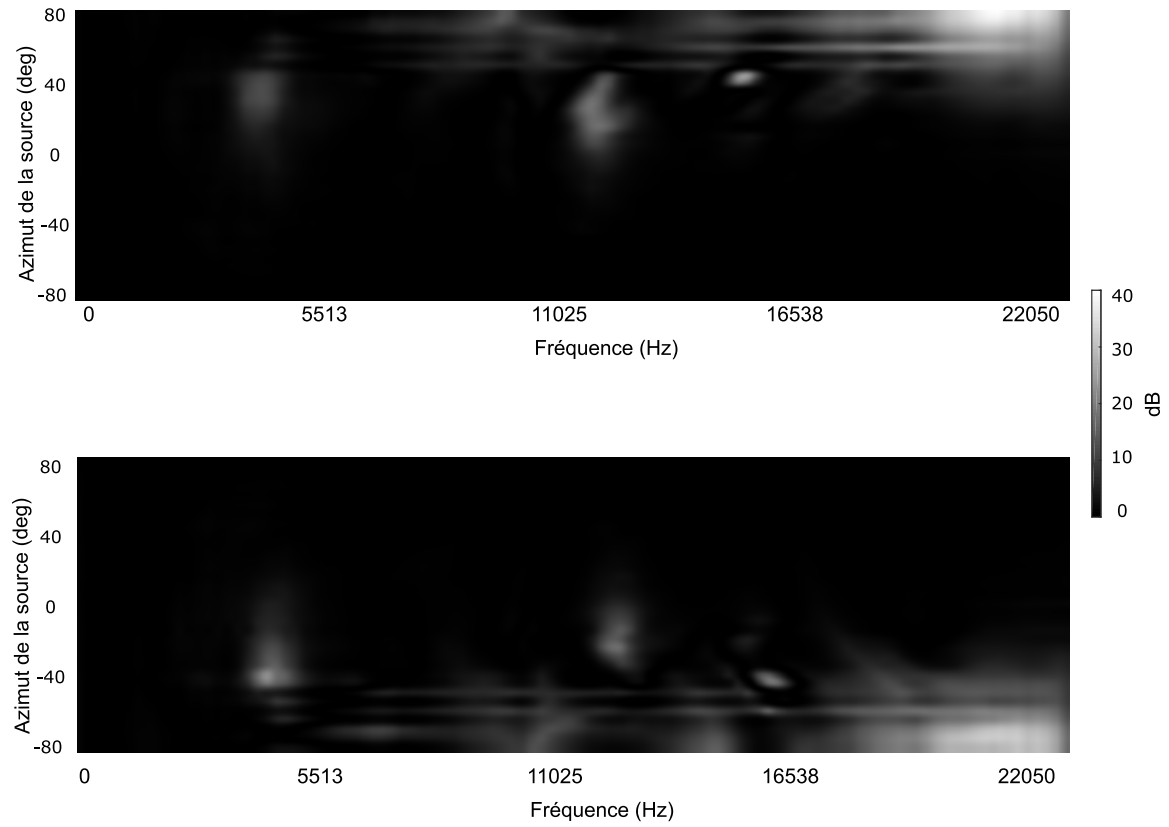


FIGURE 5.5 – *Erreur quadratique entre les HRTFs réelles du mannequin KEMAR Large Pinnae et les chemins acoustiques synthétiques de la méthode STAR, pour une élévation $\phi = 0^\circ$. L'erreur, hors valeurs extrêmes, est du même ordre que les erreurs entre HRTFs individualisées.*

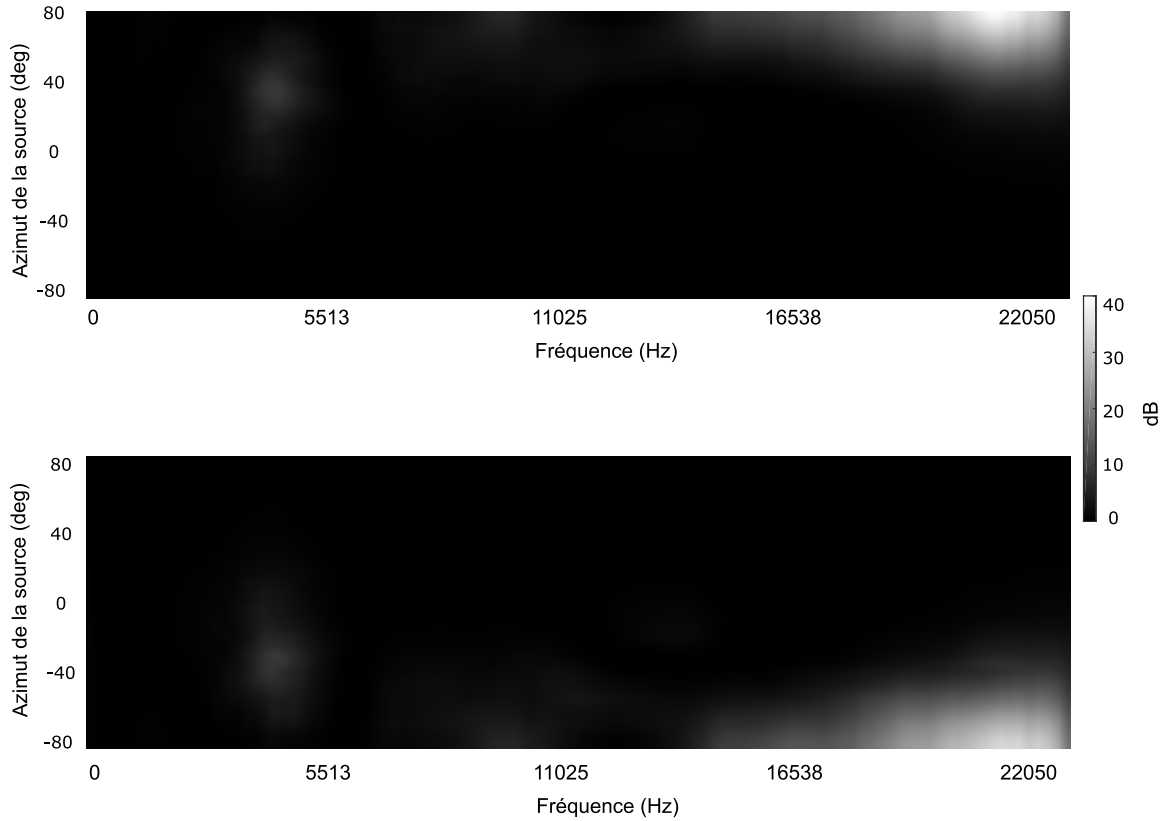


FIGURE 5.6 – Erreur quadratique entre la moyenne des HRTFs de la base CIPIC et les chemins acoustiques synthétiques de la méthode STAR, pour une élévation $\phi = 0^\circ$. L'erreur, hors valeurs extrêmes, est du même ordre que les erreurs entre HRTFs individualisées.

Tout comme dans la Sec. 1.3.3, l'erreur sur les données de KEMAR a un aspect plus chaotique, et celle sur la moyenne est plus lisse. Cependant, ici, les ordres de grandeurs des extrêmes sont les mêmes.

Il est intéressant d'effectuer un rapprochement également avec les Fig. 1.7 et 1.8 qui présentent l'énergie des HRTFs pour KEMAR ainsi que pour l'ensemble des données de la base CIPIC. Contrairement à ces figures, si l'on traçait l'énergie via les chemins synthétiques, on ne verrait pas de pic d'énergie, les chemins étant parfaitement symétriques (découlant des ITDs et ILDs). Cependant, l'énergie des Fig. 1.7 et 1.8 correspond dans l'ensemble, et la présence de ces pics n'est pas un souci pour la localisation/spatialisation en azimut (ne dépendant ni de l'azimut ni de la fréquence). Cette observation appuie cependant l'approche effectuée pour la spatialisation en élévation Sec. 6.

5.3 Le principe de transauralité dans la méthode STAR

La méthode STAR pour l'azimut repose essentiellement sur le principe transaural abordé en Sec. 2.2.3. Comme présenté sur la Fig. 2.6, le but est de reconstruire les chemins virtuels H_L et H_R (entre la source virtuelle et l'auditeur) à partir des chemins acoustiques réels (entre les haut-parleurs et l'auditeur).

Ainsi, pour un son s , les sons enregistrés devraient respecter l'équation introduite précédemment 2.5.

Nous avons vu dans la section 5.2.1 que l'on pouvait, grâce au modèle de la méthode STAR, obtenir simplement n'importe quel chemin acoustique. Ainsi, il est possible de retrouver les coefficients K_L et K_R à appliquer sur les haut-parleurs, permettant de reproduire les chemins acoustiques synthétiques dépendants eux-mêmes des ITDs et ILDs. Ainsi, on peut produire un son spatialisé de manière perceptive, grâce à l'intégration de ces indices acoustiques.

La Fig. 5.7 permet une synthèse de la méthode STAR pour la spatialisation en azimut

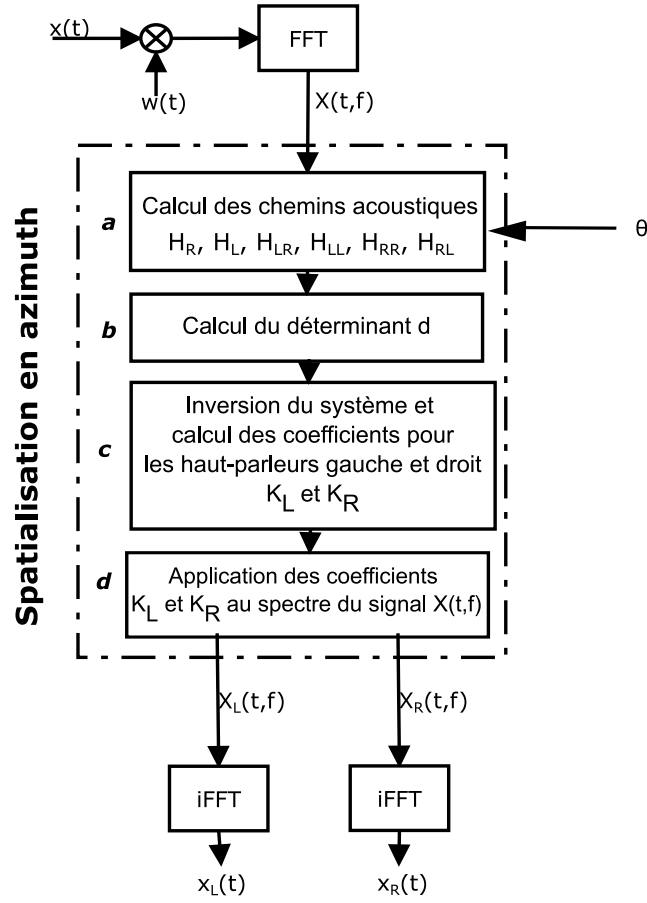


FIGURE 5.7 – Le processus de spatialisation STAR pour l'azimut.

Partie a : Chaque chemin acoustique est calculé (comme montré dans la Fig. 2.6 de la section 2.2.3) en utilisant la procédure de calcul énoncée dans la section 5.2 présentant les chemins acoustiques, et utilisant l'Eqn. 5.2.

Partie b : Le déterminant du système est calculé en utilisant l'Eqn. 5.4.

$$d = H_{LR} \cdot H_{RL} - H_{LL} \cdot H_{RR} \quad (5.4)$$

Partie c : Le système est inversé à l'aide du déterminant, permettant de trouver les coefficients à l'aide des Eqn. 5.5.

$$\begin{aligned} K_L &= (H_R \cdot H_{RL} - H_L \cdot H_{RR})/d \\ K_R &= (H_L \cdot H_{LR} - H_R \cdot H_{LL})/d \end{aligned} \quad (5.5)$$

Partie d : Les coefficients ainsi calculés sont appliqués sur les spectres gauche et droit des haut-parleurs (Eqn. 5.6).

$$\begin{aligned} X_L(t, f) &= K_L(f) \cdot X(t, f) \\ X_R(t, f) &= K_R(f) \cdot X(t, f) \end{aligned} \quad (5.6)$$

5.4 Le déterminant

En pratique, l'application de la méthode n'est pas aussi simple, et différents points sont à prendre en considération. Le premier de ces points concerne le déterminant qui joue un rôle essentiel dans la méthode STAR.

5.4.1 Le déterminant de la méthode STAR

Le déterminant d (Eqn. 5.4) possède certaines contraintes afin qu'il n'y ait pas d'incidents. La première de ces contraintes est que sa valeur doit être suffisamment importante. En effet, si d se rapproche trop de 0, le système sera mal conditionné, pouvant faire exploser les valeurs des coefficients. Dans la méthode STAR, la valeur minimale pour garantir la stabilité du système est de 0.01.

De plus, comme montré dans l'équation 5.4, le déterminant dépend uniquement des chemins acoustiques des haut-parleurs (donc de la position de ces haut-parleurs relativement aux oreilles).

Évidemment, un problème apparaît si les deux haut-parleurs sont à la même position, mais ce cas de figure ne peut pas avoir d'occurrence dans la pratique. Cependant, dans le cas où l'on dispose d'un nombre extrêmement important de haut-parleurs, il est possible que deux haut-parleurs soient très proches. La Fig. 5.8 représente la norme du déterminant en fonction de l'espacement de deux haut-parleurs. La partie en rouge correspond aux valeurs du déterminant inférieur à la valeur minimale choisie de 0.01. On voit alors que l'angle minimal entre deux haut-parleurs est de 2° .

En pratique, un écartement aussi faible ne sera jamais le cas de la méthode STAR, qui, contrairement à la méthode WFS (Sec. 2.2.1), est conçue pour des configurations de haut-parleurs espacés.

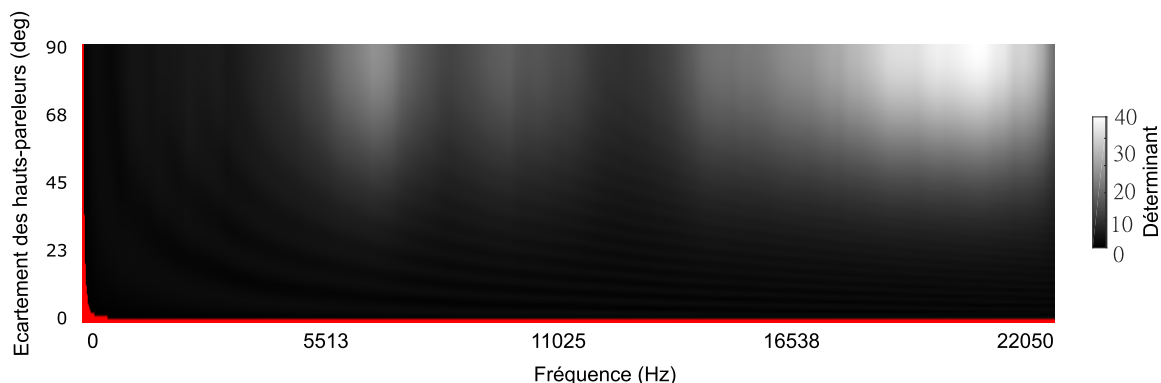


FIGURE 5.8 – Norme du déterminant en fonction de l'écartement des haut-parleurs et de la fréquence. En rouge, les valeurs n'assurant pas la stabilité du système (< 0.01). L'espacement minimal des haut-parleurs est donc de 2° .

Un autre problème survient dans les basses fréquences, comme visible sur la Fig. 5.8, et où les valeurs minimales du déterminant sont rencontrées pour des espacements bien supérieurs (environ 20°). Ce cas de figure ne pose pas de problème, les basses fréquences en dessous de 150 Hz étant filtrées et diffusées sur l'ensemble des haut-parleurs (Sec. 5.1).

Le dernier problème survient lorsque le système de diffusion sonore provoque une symétrie entre les chemins synthétiques. Un déterminant nul apparaît alors sur la paire de haut-parleurs situés symétriquement via l'axe transaural, cas par exemple des haut-parleurs 2 et 3 ainsi que 6 et 7 dans la configuration expérimentale présentée sur la Fig. 10.5, problème plus largement discuté dans la section 5.4.2. Une des solutions est d'effectuer une rotation des azimuts de référence et ainsi se retrouver dans une configuration ne présentant pas de problèmes (5.1), solution choisie pour les premières expérimentations.

Cependant, cette technique ne convient pas à la philosophie souhaitée pour la méthode STAR, celle-ci devant être robuste, et facile de mise en œuvre. Aussi, le choix a été fait de placer l'azimut de référence au centre des haut-parleurs concernés. Ainsi, les déterminants dépendent uniquement des fréquences et non plus de

5.4. LE DÉTERMINANT

l'azimut. La Fig. 5.9 montre un exemple de ce déterminant exempté de problème hormis dans les basses fréquences filtrées. Le choix de cet azimut de référence n'affecte pas la qualité de spatialisation de la méthode STAR, les indices acoustiques étant correctement conservés.

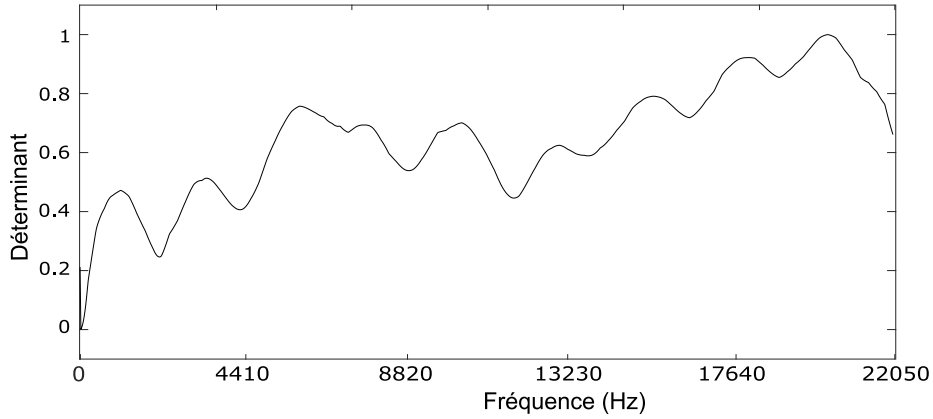


FIGURE 5.9 – Norme du déterminant pour l'azimut de référence $\theta = 0^\circ$.

5.4.2 Discussion sur le déterminant

La Fig. 5.10 montre le déterminant de la méthode initiale avant correction, dans la configuration produisant l'erreur. On distingue sur cette figure différentes bandes qui correspondent chacune à une paire de haut-parleurs, les deux bandes noires correspondant aux configurations symétriques par rapport à l'axe transaural et entraînant un déterminant nul.

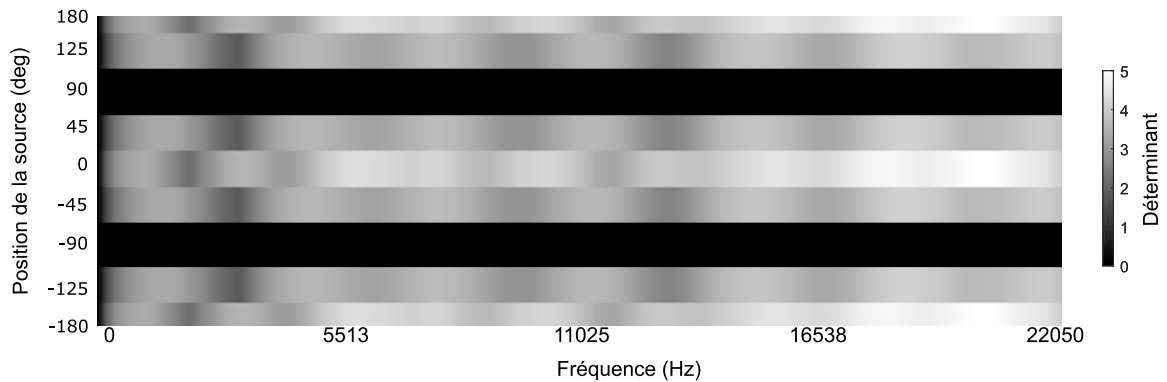


FIGURE 5.10 – Déterminant initial avec une erreur sur les deux bandes jouées par les haut-parleurs de part et d'autre de l'axe transaural. Ici, les haut-parleurs sont placés au niveau de chaque changement de bande, avec l'axe transaural passant par 90° et -90° , provoquant une symétrie entre les deux haut-parleurs les entourant et ainsi un déterminant nul.

Bien que visuellement compréhensible, et observé dans les expérimentations, il est important de comprendre ce coefficient nul, ainsi que de vérifier que d'autres configurations ne poseraient pas de problème. Pour cela, on choisit d'étudier le système, et plus particulièrement de le résoudre pour les résultats à proscrire. Les eq 5.7 à 5.15 déterminent les solutions pour que le déterminant soit nul, et donnent comme unique solution $\sin(\theta_g) = -\sin(\theta_d)$.

Le discriminant ne devant pas être nul, en résolvant l'Eqn.5.7, on trouve les cas particuliers rendant

5.4. LE DÉTERMINANT

instable la méthode.

$$d = H_{LR}.H_{RL} - H_{LL}.H_{RR} = 0 \quad (5.7)$$

Ainsi :

$$\begin{aligned} H_{LR}.H_{RL} - H_{LL}.H_{RR} &= 0 \\ H_{LR}.H_{RL} &= H_{LL}.H_{RR} \end{aligned} \quad (5.8)$$

En se reportant aux Eqn. 5.2 on obtient alors :

$$\begin{aligned} 10^{\Delta_a(f)/2} \cdot e^{i\Delta_\phi(f)/2} \cdot 10^{-\Delta_a(f)/2} \cdot e^{-i\Delta_\phi(f)/2} &= 10^{\Delta_a(f)/2} \cdot e^{i\Delta_\phi(f)/2} \cdot 10^{-\Delta_a(f)/2} \cdot e^{-i\Delta_\phi(f)/2} \\ 10^{ILD(\theta,f)/2} \cdot e^{iITD(\theta,f/2)} \cdot 10^{-ILD(\theta,f)/2} \cdot e^{-iITD(\theta,f/2)} &= 10^{ILD(\theta,f)/2} \cdot e^{iITD(\theta,f/2)} \cdot 10^{-ILD(\theta,f)/2} \cdot e^{-iITD(\theta,f/2)} \end{aligned} \quad (5.9)$$

On sait que les haut-parleurs ont le même angle, aussi on pose :

$$\begin{aligned} \theta_{LR} &= \theta_{LL} = \theta_g \\ \theta_{RL} &= \theta_{RR} = \theta_d \end{aligned} \quad (5.10)$$

Permettant d'obtenir :

$$\begin{aligned} 10^{ILD(\theta_g,f)/2} \cdot e^{iITD(\theta_g,f/2)} \cdot 10^{-ILD(\theta_d,f)/2} \cdot e^{-iITD(\theta_d,f/2)} &= 10^{ILD(\theta_g,f)/2} \cdot e^{iITD(\theta_g,f/2)} \cdot 10^{-ILD(\theta_d,f)/2} \cdot e^{-iITD(\theta_d,f/2)} \\ 10^{ILD(\theta_g,f)-ILD(\theta_d,f)} \cdot e^{i[ITD(\theta_g,f)-ITD(\theta_d,f)]} &= 10^{ILD(\theta_g,f)-ILD(\theta_d,f)} \cdot e^{i[ITD(\theta_g,f)-ITD(\theta_d,f)]} \end{aligned} \quad (5.11)$$

De plus, comme $ILD = \alpha(f)\sin(\theta)$ et $ITD = \beta(f).r.\sin(\theta)/c$

$$\begin{aligned} 10^{\alpha(f)\sin(\theta_g)-\alpha(f)\sin(\theta_d)} \cdot e^{i\beta(f)r\sin(\theta_g)/c-i\beta(f)r\sin(\theta_d)/c} &= 10^{\alpha(f)\sin(\theta_g)-\alpha(f)\sin(\theta_d)} \cdot e^{i\beta(f)r\sin(\theta_g)/c-i\beta(f)r\sin(\theta_d)/c} \end{aligned} \quad (5.12)$$

Recherchant une égalité, on peut considérer les exponentielles comme des constantes qui s'annulent et ainsi :

$$\begin{aligned} 10^{\alpha(f)\sin(\theta_g)-\alpha(f)\sin(\theta_d)} &= 10^{\alpha(f)\sin(\theta_d)-\alpha(f)\sin(\theta_g)} \\ 10^{\alpha(f)\sin(\theta_g)-\sin(\theta_d)+\sin(\theta_d)-\sin(\theta_d)} &= 0 \\ 10^{2\alpha(f)\sin\theta_g+\sin\theta_d} &= 0 \end{aligned} \quad (5.14)$$

Et on obtient ainsi la solution, le déterminant étant égal à 0 dans le cas où l'Eqn 5.15 est vérifiée.

$$\sin(\theta_g) = -\sin(\theta_d) \quad (5.15)$$

Ceci justifie l'utilisation d'une approche fixant l'azimut de référence au centre des deux haut-parleurs, donnant la Fig. 5.11. En procédant de la sorte, on sort du contrôle strict de la configuration, et on utilise les positions réelles des haut-parleurs. De plus, ce processus s'intègre parfaitement à la méthode STAR pouvant utiliser plus facilement différentes configurations, et reposant sur le principe que les auditeurs peuvent bouger

5.5. LES COEFFICIENTS

la tête et ne pas être à des positions fixes lors de l'écoute. Thurlow et al. (1967) montrent dans leurs travaux que l'on effectue une rotation pour la localisation d'un son d'en moyenne 30° pour les hautes fréquences et 40° pour les basses fréquences, permettant de justifier le choix de l'azimut de référence entre les deux haut-parleurs (l'auditeur étant libre d'effectuer une légère rotation pour se mettre dans le bon référentiel).

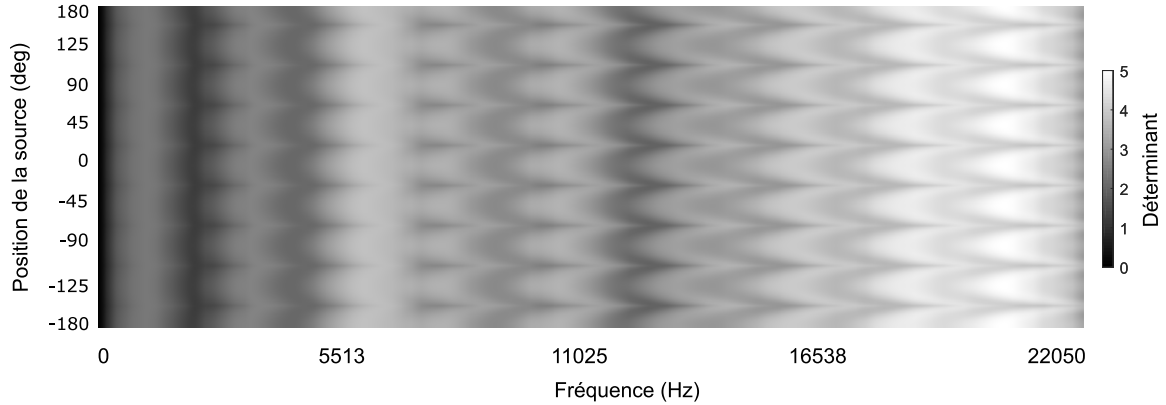


FIGURE 5.11 – Déterminant une fois la méthode corrigée.

On voit sur cette figure que le déterminant n'est jamais égal à 0 ; seules les basses fréquences posent problèmes. De plus, en regardant entre les haut-parleurs, une progression dans les valeurs du déterminant est visible, contrairement à la méthode initiale où les valeurs étaient constantes et dépendaient uniquement des fréquences.

5.5 Les coefficients

Le second point à observer est le déterminant qui joue un rôle fondamental dans la méthode STAR. En effet, ce sont ces coefficients qui sont appliqués aux sources sonores devant être envoyées aux haut-parleurs.

5.6 Les coefficients de la méthode STAR

Le calcul des coefficients K_L et K_R est décrit dans les Eqn. (5.5), correspondant à la dernière étape de spatialisation de la méthode STAR.

Aussi, même si le déterminant du système est correct, il est important de vérifier que les coefficients qui en découlent sont sensés. Par exemple, une forte valeur de ces coefficients peut être dangereuse, pouvant provoquer une saturation des haut-parleurs, et de plus, contrairement à VBAP et HOA, les coefficients de la méthode STAR sont complexes.

La Fig. 5.12 montre le module des coefficients droit et gauche en fonction de la position d'une source sonore virtuelle. Les valeurs sont très majoritairement comprises entre 0 et 1.4, et ne dépassent jamais 1.82. Aussi, un risque de saturation existe, bien que faible, si les sources sonores spatialisées sont puissantes.

5.6. LES COEFFICIENTS DE LA MÉTHODE STAR

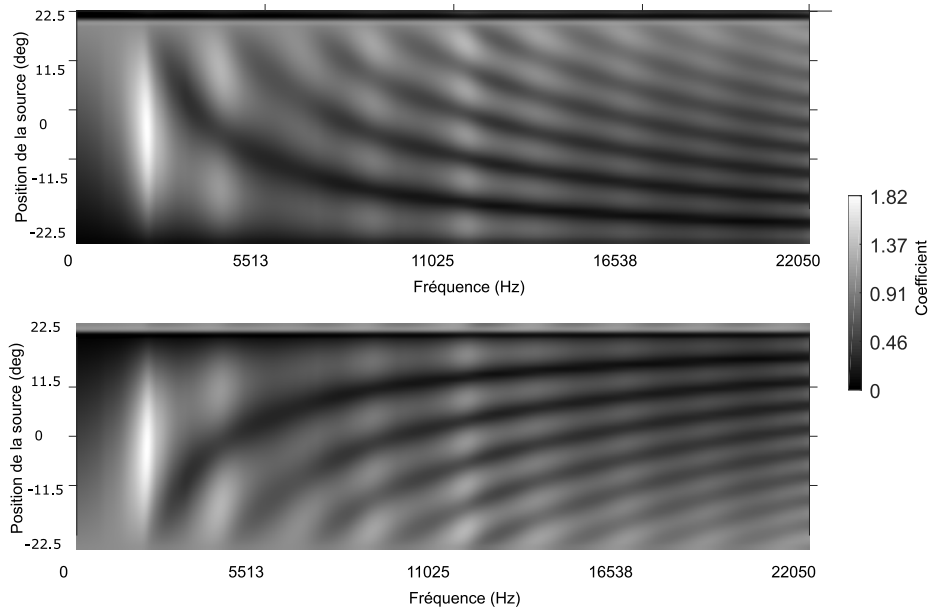


FIGURE 5.12 – Amplitude des coefficients K_L (graphique du haut) et K_R (graphique du bas) en fonction de la fréquence et de l'azimut θ de la source sonore virtuelle.

5.6.1 Discussion sur les coefficients

Les méthodes STAR et VBAP sont proches sur de nombreux points, tel que le paradigme de prendre les deux haut-parleurs entourant la source. Il est donc intéressant dans un premier temps de comparer les coefficients des deux méthodes. La Fig. 5.13 montre les coefficients de VBAP et STAR pour spatialiser une source sonore virtuelle à un azimut $\theta = 0^\circ$, correspondant au centre de deux haut-parleurs dans la configuration expérimentale choisie (Fig. 10.5), les deux haut-parleurs étant placés à $\pm 22.5^\circ$.

Bien que les coefficients de la méthode STAR soient complexes et spectraux-dépendants (variant en fonction de la fréquence), la Fig. 5.13 montre que les coefficients des deux méthodes restent proches (et ceci quelle que soit la position de la source sonore). Les coefficients de VBAP sont constants et ne dépendent pas de la fréquence.

5.6. LES COEFFICIENTS DE LA MÉTHODE STAR

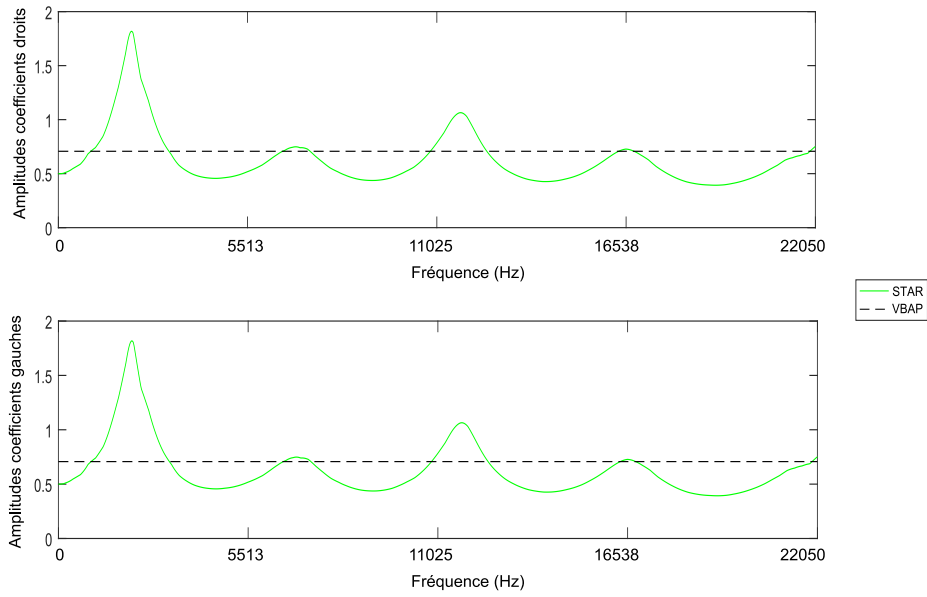


FIGURE 5.13 – Coefficients VBAP (en pointillé noir) et STAR (ligne verte). La comparaison est donnée pour $\theta = 0^\circ$, cas où les coefficients gauches et droits sont identiques.

Les Fig. 5.12 et 5.13 mettent en avant le côté sinusoïdal et symétrique du modèle d'ITDs et d'ILDs utilisé afin de calculer les coefficients dans la première étape de la méthode.

Il est aussi intéressant de comparer les coefficients de la méthode d'origine et de la méthode STAR. Ainsi, la Fig. 5.14 montre le coefficient de la méthode initiale, et la Fig. 5.15 de la méthode en fixant l'azimut de référence comme étant au centre des deux haut-parleurs.

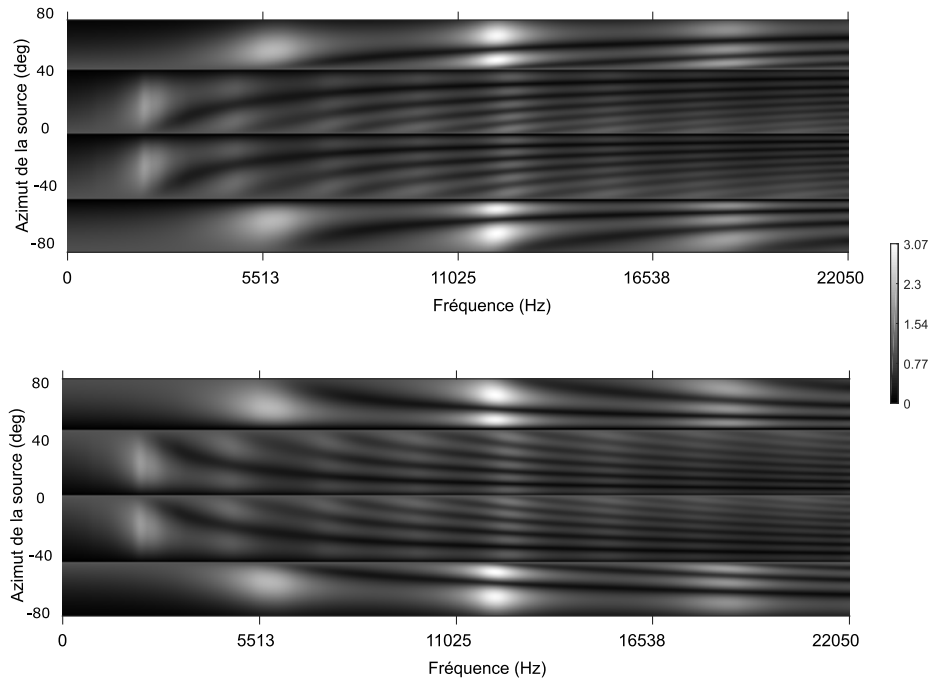


FIGURE 5.14 – Coefficients initiaux. (Gauche pour le graphique du haut et droit pour le graphique du bas).

La Fig. 5.14 présente un défaut majeur, qui est l'importance des coefficients. En effet, ceux-ci montent

5.6. LES COEFFICIENTS DE LA MÉTHODE STAR

jusqu'à 3 et sont en grande partie supérieurs à 2, notamment lorsque l'azimut de la source virtuelle augmente. Ce défaut peut produire des saturations en fonction du contenu audio spatialisé.

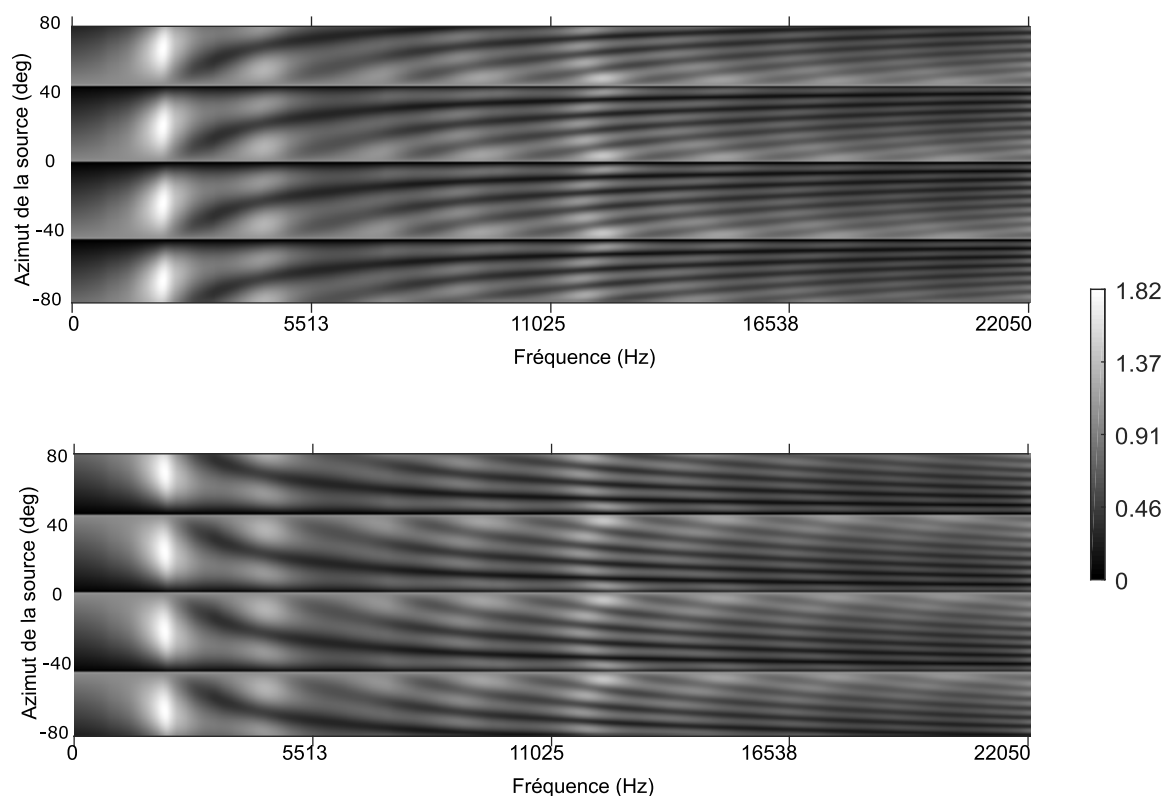


FIGURE 5.15 – Coefficients avec la méthode adaptée. (Gauches pour le graphique du haut et droits pour le graphique du bas)

La Fig. 5.15 répète le schéma présenté dans la Fig. 5.12, les observations sont donc identiques.

– Conclusion –

Ce chapitre présente la méthode de spatialisation STAR pour l'azimut, mise en place par les modèles d'indices acoustiques vus dans la première partie. Il a également mis en avant et justifié les choix effectués dans la méthode STAR, notamment par une analyse du déterminant et des coefficients.

La méthode de spatialisation nécessite maintenant des validations qui seront présentées dans la partie IV. La méthode de spatialisation en azimut décrite, le prochain chapitre présente la spatialisation STAR pour l'élévation.

Ce chapitre présente la technique de spatialisation STAR née de l'évolution de la méthode proposée par Mouba and Marchand (2006). Cette nouvelle méthode STAR a été présentée dans la publication Méaux and Marchand (2019) puis dans la revue de l'AES (Méaux and Marchand (2021a)).

Chapitre 6

La spatialisation en élévation

– Introduction –

Le chapitre précédent propose la méthode de spatialisation de la méthode STAR pour l'azimut. Ce chapitre présente la spatialisation en élévation. La littérature et des expérimentations ont montré que l'élévation est localisée à l'aide d'indices spectraux, et, bien qu'ayant un aspect sujet-dépendant, il existe des indices monauraux, constitués notamment de pics et nœuds globaux.

L'idée de la méthode STAR est de façonner les HRTFs afin de recréer les pics et nœuds responsables de cette localisation. Cette méthode utilisant les indices spectraux contenus dans les HRTFs permet à la méthode STAR de garder la même philosophie, indices spectraux (méthode perceptive) et en utilisant des modèles (méthode synthétique).

La première partie de ce chapitre va présenter les modèles synthétiques, puis la seconde partie le processus de spatialisation.

L'élévation représente un processus plus compliqué, notamment pour la localisation, nécessitant un apprentissage par l'homme, et est présentée dans la section 1.2.4. Comme énoncé précédemment, des études ont prouvé que les pics et nœuds présents dans les HRTFs jouent un rôle important pour cette spatialisation (Butler and Belendiuk (1977); Musicant and Butler (1984)). Hebrank and Wright (1974) ont notamment introduit l'influence de pics et nœuds spécifiques à la localisation en élévation. Plus récemment, Iida et al. (2007) ont proposé un modèle pour la localisation, introduit et discuté dans le chapitre 4). Dans la méthode STAR, il est proposé d'utiliser les deux premiers nœuds (dont la fréquence dépend de l'élévation ϕ), ainsi que du premier pic (dont la fréquence est fixe).

6.1 Les modèles utilisés pour la spatialisation en élévation

La spatialisation en élévation utilise le modèle proposé par Iida, montrant que des nœuds spécifiques servent à la localisation du son au-delà de 4 kHz, et dont la position fréquentielle et la profondeur dépendent de l'élévation.

Le modèle d'Iida, rappelé dans l'Eqn 6.1, donne la position des deux nœuds principaux en fonction de l'élévation, la position des pics étant fixe.

$$\begin{aligned} f_{N1}(\phi) &= 5.77 \cdot 10^{-5} \phi^4 - 2.41 \cdot 10^{-2} \phi^3 \\ &\quad + 2.79 \phi^2 + 4.79 \cdot 10^1 \phi + 6.06 \cdot 10^3 \\ f_{N2}(\phi) &= 2.35 \cdot 10^{-6} \phi^4 - 2.98 \cdot 10^{-3} \phi^3 \\ &\quad + 4.85 \cdot 10^{-1} \phi^2 + 2.3 \cdot 10^1 \phi + 8.52 \cdot 10^3 \end{aligned} \tag{6.1}$$

6.1. LES MODÈLES UTILISÉS POUR LA SPATIALISATION EN ÉLÉVATION

La méthode STAR va utiliser ce modèle afin de connaître la position des pics et nœuds, ce qui servira à les positionner en fonction de l'élévation mais aussi pour l'apprentissage de la profondeur et de la largeur de ces pics et nœuds.

La localisation en élévation dépendant de la profondeur et largeur de ces pics et nœuds, la même démarche utilisée par STAR repose sur un apprentissage de ces largeurs/profondeurs, de manière similaire au processus utilisé pour les coefficients α et β (servant à la construction des ILDs et ITDs synthétiques).

La Fig. 6.1 montre l'évolution en fonction de l'élévation de la largeur et de la profondeur des nœuds N1 et N2 ainsi que du pic P1. Les nœuds et pics sélectionnés sont ceux les plus proches du modèle (Eqn. 6.1), tel que discuté dans la Sec. 4.2.2, dont les valeurs sont suffisantes pour la détection (Moore et al. (1989)).

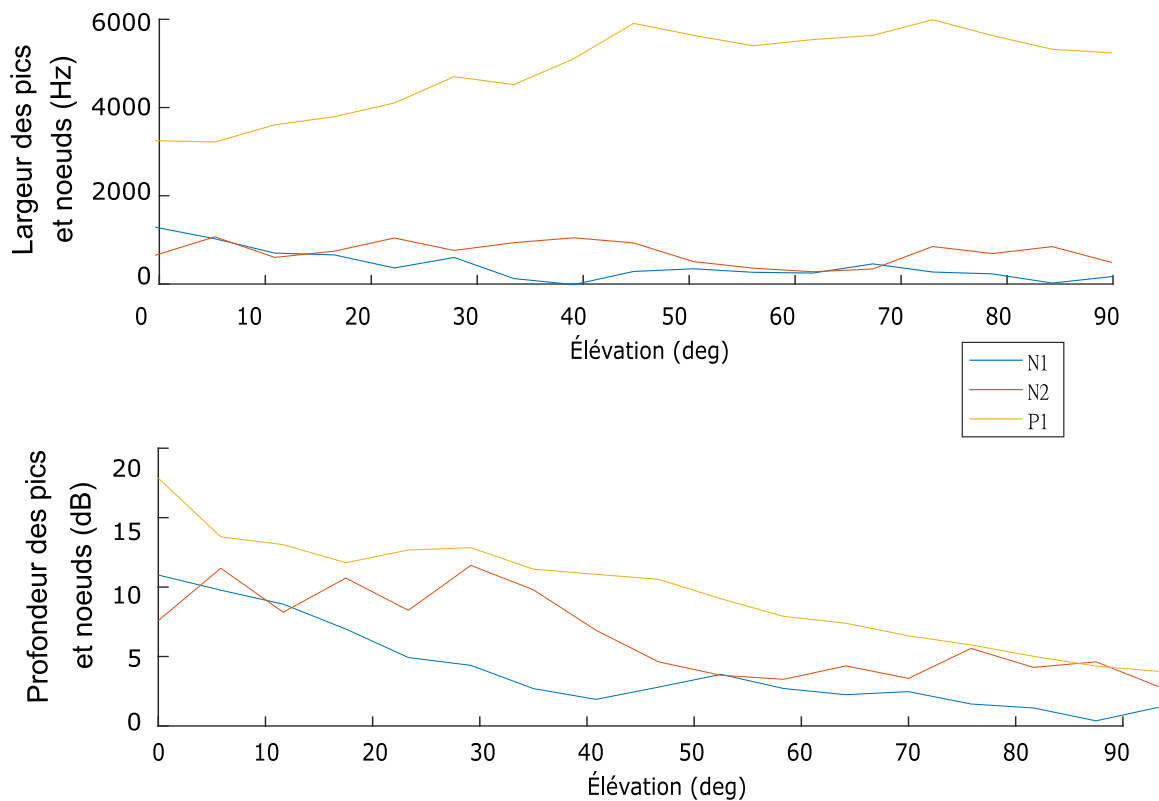


FIGURE 6.1 – *Largeur (Hz) et profondeur (dB) des nœuds et pics moyens provenant de la base CIPIC, en fonction de l'élévation.*

À partir de ces valeurs brutes, et en utilisant la méthode des moindres carrés, nous obtenons alors des polynômes, les Eqn. 6.2 à 6.7 donnent ces différents polynômes, avec W_x correspondant à la largeur en Hz et H_x à la hauteur en dB, dépendant de l'élévation ϕ (exprimée en degrés).

$$W_{N1}(\phi) = -5.86 \cdot 10^{-3} \phi^3 + 9.77 \cdot 10^{-1} \phi^2 - 5.18 \cdot 10^1 \phi + 1.78 \cdot 10^3 \quad (6.2)$$

$$W_{N2}(\phi) = 4.76 \cdot 10^{-3} \phi^3 + 6.44 \cdot 10^{-1} \phi^2 + 1.99 \cdot 10^1 \phi + 1.27 \cdot 10^3 \quad (6.3)$$

$$W_{P1}(\phi) = -6.75 \cdot 10^{-3} \phi^3 + 4.22 \cdot 10^{-1} \phi^2 + 3.56 \cdot 10^1 \phi + 3.25 \cdot 10^3 \quad (6.4)$$

$$H_{N1}(\phi) = -2.9 \cdot 10^{-5} \phi^3 + 5.51 \cdot 10^{-3} \phi^2 - 3.69 \cdot 10^{-1} \phi + 1.15 \cdot 10^1 \quad (6.5)$$

$$H_{N2}(\phi) = 5.38 \cdot 10^{-5} \phi^3 - 7.15 \cdot 10^{-3} \phi^2 + 1.66 \cdot 10^{-1} \phi + 8.81 \quad (6.6)$$

$$H_{P1}(\phi) = -1.10 \cdot 10^{-5} \phi^3 - 1.47 \cdot 10^{-3} \phi^2 - 1.73 \cdot 10^{-1} \phi + 1.52 \cdot 10^1 \quad (6.7)$$

La Fig. 6.2 illustre ces polynômes.

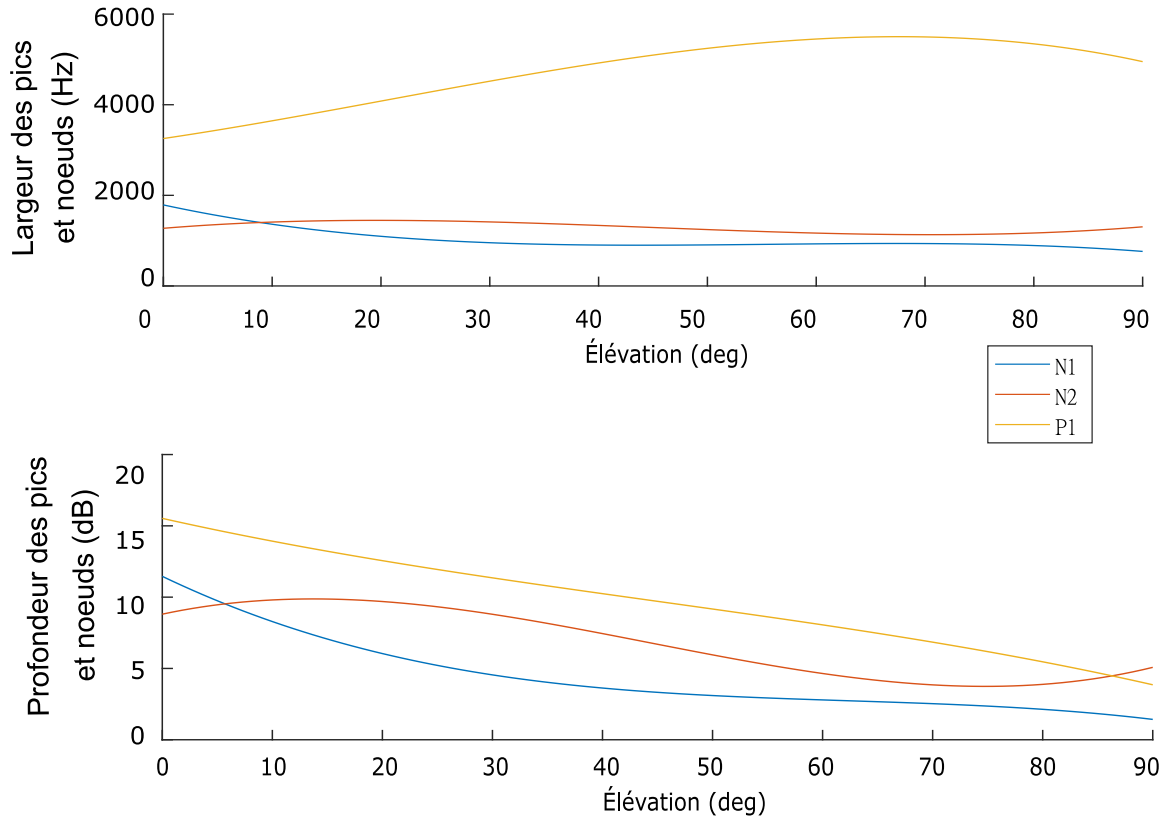


FIGURE 6.2 – Courbe polynomiale représentant l'évolution de la largeur et de la hauteur des pics et des nœuds en fonction de l'élévation. Les données utilisées pour l'apprentissage proviennent de la base de données CIPIC.

6.2 La méthode de spatialisation en élévation STAR

Les différents modèles utiles pour la spatialisation en élévation présentés section 6.1 permettent de proposer le système de spatialisation en élévation complet (Fig.6.3).

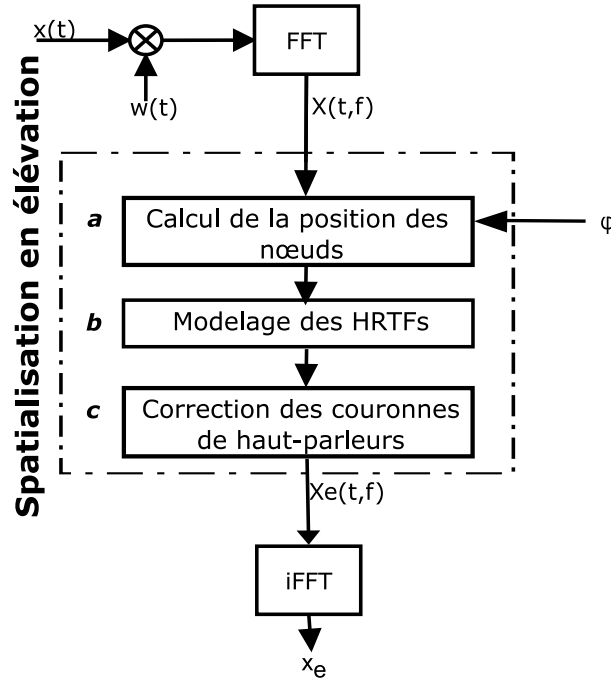


FIGURE 6.3 – Le processus de spatialisation STAR pour l'élévation.

Partie a : Calcul de la position des nœuds en fonction de l'élévation à spatialiser, en utilisant les équations 6.1.

Partie b : Modelage des HRTFs en ajoutant le coefficient de mise en forme (Fig. 6.4, ligne pointillée verte), calculées à l'aide de la fonction Gaussienne (Eqn. 6.8).

$$\begin{aligned}
 S_c(\phi) = & H_{P1}(\phi) \cdot e^{-(f-f_{P1})^2/(0.5 \cdot W_{P1}(\phi)^2)} \\
 & - H_{N1}(\phi) \cdot e^{-(f-f_{N1}(\phi))^2/(0.5 \cdot W_{N1}(\phi)^2)} \\
 & - H_{N2}(\phi) \cdot e^{-(f-f_{N2}(\phi))^2/(0.5 \cdot W_{N2}(\phi)^2)}
 \end{aligned} \tag{6.8}$$

Partie c : Modelage des HRTFs en soustrayant le coefficient de mise en forme correspondant à l'élévation des haut-parleurs (Fig. 6.4 ligne rouge), calculées de manière similaire à la partie b.

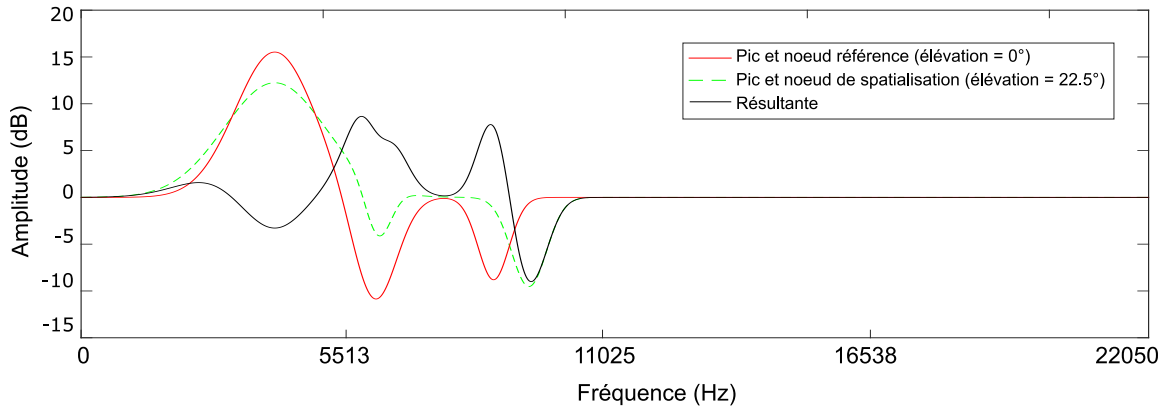


FIGURE 6.4 – Coefficients de mise en forme des HRTFs calculées en utilisant les paramètres provenant des modèles apprises sur la base CIPIC, et du modèle de lida, pour une élévation de 22.5 degrés.

6.2. LA MÉTHODE DE SPATIALISATION EN ÉLÉVATION STAR

En pratique, pour chaque couronne de haut-parleurs, le coefficient résultant de ceux de la source sonore virtuelle (Fig. 6.4 ligne noire) et des haut-parleurs est directement calculé et appliqué.

Lorsque la configuration de diffusion sonore dispose de plusieurs haut-parleurs, le processus de spatialisation sonore précédent est couplé avec la méthode VBAP permettant ainsi de tirer les avantages des deux méthodes, augmentant la précision et la sensation d'élévation en ajoutant des coefficients d'élévation. De plus, la sensation de l'élévation sera d'autant plus grande que la couronne supérieure sera élevée. La méthode STAR possède cependant l'avantage, comme démontré dans les tests subjectifs associés (Chap. 11), de fonctionner sur une seule couronne de haut-parleurs.

Actuellement, seuls des tests préliminaires ont été réalisés afin de vérifier le fonctionnement de la méthode STAR couplée à la méthode VBAP, et d'autres tests doivent être effectués.

Enfin, afin d'optimiser le procédé de spatialisation de STAR, la spatialisation de l'élévation est effectuée sur le signal monaural avant que celui-ci ne soit spatialisé en azimut. Cette démarche est validée par les tests proposés dans le chapitre 8 montrant l'indépendance des différents paramètres.

– Conclusion –

Ce chapitre ajoute ainsi la spatialisation de l'élévation, donnant la seconde dimension de spatialisation pour la méthode STAR.

Cette spatialisation de l'élévation s'effectue à l'aide de modèles, basés sur des indices spectraux, gardant ainsi la démarche de STAR d'être une méthode synthétique et perceptive. Ces choix apportent, de plus, la grande flexibilité recherchée par la méthode, permettant en effet une spatialisation en élévation sur des dispositifs ne disposant que d'une unique couronne de haut-parleurs ; et ceci sans empêcher l'utilisation de dispositifs plus complexes (plusieurs couronnes) permettant d'améliorer la sensation d'élévation.

La méthode de spatialisation de l'élévation nécessite maintenant des validations, tout comme la spatialisation de l'azimut, lesquelles sont présentées dans la partie IV. La méthode de spatialisation en élévation décrite et venant compléter la spatialisation en azimut, le prochain et dernier chapitre présente la dernière dimension de spatialisation de la méthode STAR, la distance.

La spatialisation en élévation de la méthode STAR repose sur les travaux de Iida et al. (2007), et est un axe nouveau apporté par cette thèse. L'idée fut introduite dans les travaux de Méaux and Marchand (2021a), puis développée et publiée dans un second article d'extension Méaux and Marchand (2021b).

Chapitre 7

La spatialisation en distance

– Introduction –

Pour une méthode de spatialisation en 3D, et après avoir présenté la spatialisation de l'élévation et de l'azimut, il est nécessaire de positionner l'éloignement d'une source sonore.

Malheureusement, et comme discuté précédemment dans le chapitre concernant la localisation d'un son par l'homme, ce processus est compliqué, et utilise certains paramètres que nous ne pouvons pas contrôler dans les conditions d'utilisation de STAR, tels que le rapport champ direct et réverbéré, l'utilisation d'un signal connu, etc.

D'autres paramètres sont cependant contrôlables, tels que le contenu spectral qui est utilisé pour les distances très proches (utilisant notamment l'ILD qui est déjà implémentée), ou pour les champs lointains avec l'absorption du son dans l'air (dépendant de la fréquence et fonctionnant pour des distances supérieures à 15 m). Aussi, même s'il n'est pas possible d'attribuer un éloignement brut à la source sonore, il est possible, notamment grâce à l'intensité sonore qui joue un rôle primordial dans la sensation de distance (un son faible apparaîtra lointain, et un son fort proche), de spatialiser les sources sonores les unes par rapport aux autres (ou la même source sonore avec des variations de distances dans le temps ¹).

Ainsi, la perte de pression acoustique sera utilisée pour spatialiser l'éloignement des sources entre elles, couplée avec l'absorption dans l'air, dont des modèles permettent l'exploitation.

La première partie de ce chapitre présente le modèle d'absorption du son dans l'air utilisé, et la seconde le processus de spatialisation en distance.

a. Simulant alors un déplacement (éloignement ou rapprochement) au cours du temps.

7.1 Le modèle d'absorption du son dans l'air utilisé

Pour la spatialisation en distance de la méthode STAR et ses domaines d'application, il est possible d'utiliser la perte de niveau acoustique couplée à l'absorption du son dans l'air, dépendant de la fréquence et de la distance. La perte de pression acoustique revient à une perte de dB chaque fois que la distance double, ce qui représente un processus assez simple. L'absorption du son dans l'air sur un bruit blanc (possédant toutes les fréquences), prenant seulement ce paramètre en compte (pas de notion de perte de pression acoustique, réverbération, etc.). L'absorption des fréquences aiguës est bien plus importante que celle des fréquences graves, ce qui est bien visible pour les grandes distances (courbes verte, violette et jaune), et se ressent également sur des distances moins importantes. C'est un phénomène bien connu, expliquant par exemple

7.2. LA MÉTHODE DE SPATIALISATION EN DISTANCE DE LA MÉTHODE STAR

pourquoi la voix d'un homme porte plus loin que celle d'une femme. De plus, le contenu fréquentiel d'un son provenant d'une source lointaine contiendra moins de hautes fréquences qu'un son provenant d'une source proche, ce qui est exploité par notre cerveau dans la localisation de la distance.

Bass et al. (1984) ont proposé dans leur travail un modèle d'absorption dans l'air, présenté dans la Fig. 7.1, déterminant l'atténuation en dB en fonction des fréquences et de la distance. La Fig. 7.1 propose ces atténuations pour des distances de la source comprises entre 0 et 100 m, et permet de voir l'atténuation importante dans les hautes fréquences (60 dB pour des fréquences de plus de 22000 Hz à 100 m).

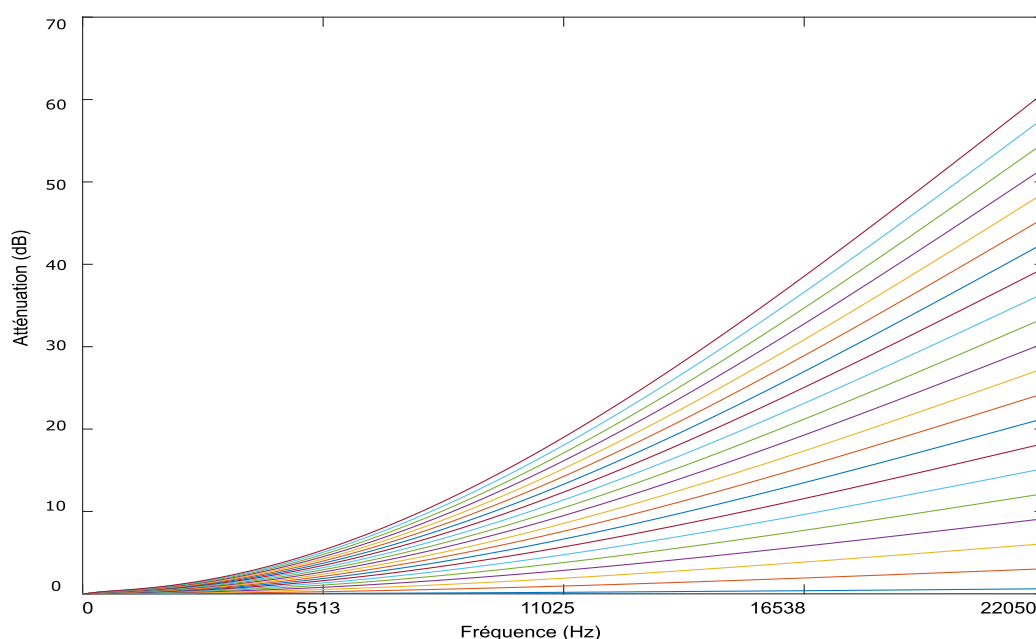


FIGURE 7.1 – *Modèle d'atténuation du son par l'air en fonction de la fréquence et de la distance (entre 0 et 100 m). Chaque courbe représente une distance, la moins élevée représentant $d = 0$ m (confondue avec l'axe des abscisses) et la plus élevée $d = 100$ m*

En regardant ce modèle, on observe que les atténuations pour les distances faibles (< 15 m) sont faibles, et expliquent pourquoi le contenu fréquentiel est utilisé pour des distances plus importantes.

7.2 La méthode de spatialisation en distance de la méthode STAR

En utilisant le modèle proposé en Sec. 7.1, couplé à la perte de pression acoustique, on obtient le processus de spatialisation en distance de la méthode STAR (Fig.7.2).

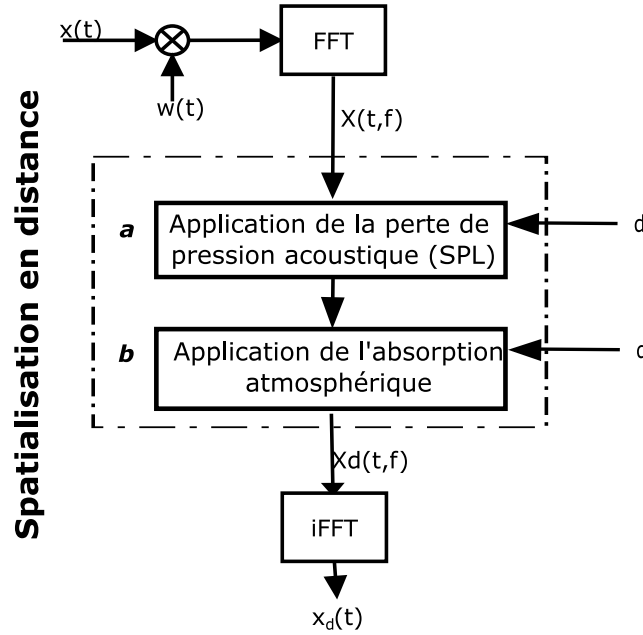


FIGURE 7.2 – Le processus de spatialisation STAR pour la distance.

Partie a : L'effet principal de la distance sur le signal source (ou spectre) est une diminution du niveau de pression acoustique (SPL). Cette diminution est de 6 dB à chaque fois que la distance double. En pratique, une perte de seulement 4,25 dB est plus réaliste Zahorik and Wightman (2001a) et s'applique sur l'Eqn. 7.1 par le coefficient A dont la valeur est fixée à 4.25.

Partie b : Le second effet pris en compte par la méthode STAR est l'absorption du son par l'air qui dépend de la fréquence et de la distance. Cette absorption est simulée par l'application sur le spectre du coefficient D (Eqn. 7.1)

$$D(d) = 10^{-d\alpha_d/20} \cdot d^{-A/20 \log_{10}(2)} \quad (7.1)$$

avec α_d le coefficient proposé par Bass et al. (1984) pour l'atténuation due à l'absorption atmosphérique.

– Conclusion –

La spatialisation de la distance dans la méthode STAR clôture la spatialisation 3D, en ajoutant la dernière composante. Contrairement à l'élévation et l'azimut, la distance est majoritairement une spatialisation relative entre les différents objets sonores spatialisés, ce qui est suffisant pour le champ d'utilisation de la méthode STAR. Les aspects perceptif et synthétique sont cependant présents via l'utilisation du modèle d'absorption du son dans l'air.

Cette dimension ajoutée nécessite, tout comme les autres dimensions, des tests de validation présentés dans la partie IV. De plus, la méthode STAR possédant les trois dimensions de spatialisation, des tests de validation sur l'immersion 3D seraient également nécessaires à effectuer dans le futur.

La spatialisation en distance de la méthode STAR est développée en parallèle de la partie sur l'élévation, et est également un axe nouveau apporté par la thèse, bien qu'introduit dans les travaux de Mouba (2009). Cette spatialisation étant introduite par les travaux de Méaux and Marchand (2021a), puis développée et publiée dans un second article d'extension Méaux and Marchand (2021b).

Quatrième partie

Les tests de validation de la méthode STAR

À propos de la partie les tests de validation de la méthode STAR

La partie précédente présente les différentes phases de spatialisation de la méthode STAR. Celles-ci sont cohérentes avec les différentes théories et la philosophie souhaitée. Il est cependant nécessaire de valider cette méthode de spatialisation. Afin de mener la validation, deux groupes de méthodes d'évaluation s'offrent à nous ; des méthodes objectives et des méthodes subjectives. Les méthodes d'évaluation sont dites objectives lorsqu'il est directement possible de mesurer la spatialisation – cas par exemple lorsqu'on peut mesurer, via une test, la localisation en azimuth d'une source précédemment spatialisée. Les méthodes d'évaluation subjective proviennent de résultats de tests d'écoute effectués sur des panels d'auditeurs qui donnent leurs 'avis' ou 'ressentis'. Les résultats, dépendant de chaque sujet, sont donc subjectifs, bien que la répétabilité avec un protocole expérimental rigoureux et concis permette de dire qu'ils sont représentatifs.

Le premier chapitre de cette partie concerne l'indépendance de paramètres comprenant l'azimut, établie de manière objective. Cette étude permet de valider le choix pour la méthode STAR de spatialiser les différentes dimensions séparément.

Le second chapitre étudie une cartographie d'indices acoustiques – plus précisément les ILDs – de manière également objective ; ces ILDs provenant de mesures à l'aide du mannequin KEMAR ou simulées à partir d'HRTFs.

Le troisième chapitre présente des tests comparatifs subjectifs effectués entre les trois méthodes de spatialisation, et permettant de comparer des méthodes existantes avec la méthode STAR.

Enfin, le dernier chapitre porte sur la validation de l'élévation et de la distance, de manière également subjective, n'ayant pas encore de tests objectifs pour mesurer ces dimensions. Couplé à cette validation, le test d'indépendance entre l'élévation et la distance est effectué.

Chapitre 8

Évaluation objective de l'indépendance de l'élévation et de la distance avec l'azimut

– Introduction –

Ce chapitre présente un pré-requis important à la méthode de spatialisation STAR présentée dans les chapitres précédents. En effet, STAR traite les trois dimensions de spatialisation séparément, se basant sur le fait que la distance et l'élévation utilisent des indices monauraux, et que l'azimut utilise des indices binauraux.

Ce chapitre décrit les différents tests effectués afin de prouver cette indépendance, dans le cadre où il est possible d'évaluer objectivement cette indépendance. Une évaluation est possible pour les couples azimut/élévation et azimut/distance, grâce à la méthode de localisation perceptive présentée dont les performances sont prouvées, et donnant des mesures objectives. La dernière indépendance entre la distance et l'élévation est testée au cours des mêmes tests que ceux destinés à l'évaluation de ces paramètres, présentée dans la partie des tests subjectifs (dépendant donc de l'avis d'un panel d'auditeurs).

Ainsi, ce chapitre présentera dans un premier temps les études d'indépendance entre l'azimut et l'élévation, puis dans un second temps entre la distance et l'azimut ; avec pour chaque cas des mesures réelles d'une part, et des mesures simulées à l'aide de la méthode STAR d'autre part.

Dans ce chapitre, l'ensemble des mesures est effectué au SCRIME, studio utilisé essentiellement par des musiciens et ayant d'assez bonnes performances sonores, sans pour autant être physiquement contrôlé. Le studio dispose de 18 enceintes Genelec 8030 disposées en trois couronnes de haut-parleurs, a une surface de 40 m², avec un des murs avec une porte en bois, le mur opposé possédant trois grandes fenêtres en bois. L'ensemble des murs est équipé de panneaux acoustiques ; le sol est recouvert d'un tapis fin. Les sources sonores utilisées sont des bruits blancs de longueur $n = 2^{18}$ échantillons, échantillonnés à 44100 Hz, et n'utilisant que le haut-parleur situé en face du mannequin KEMAR avec de grands lobes d'oreilles. La distance des haut-parleurs est de 2,6 m pour les tests sur l'élévation et varie pour les distances. Le mannequin est fixé à un système présenté dans l'annexe 11.8.5 permettant de tourner facilement le mannequin sur les plages $\pm 80^\circ$ tous les 2° , mais aussi de changer l'élévation en modifiant l'axe vertical par inclinaison. L'ensemble de ces mesures est rangé dans une base nommée SCRIME, puis comparée aux différentes bases BRIR existantes en Sec. 1.3.2.

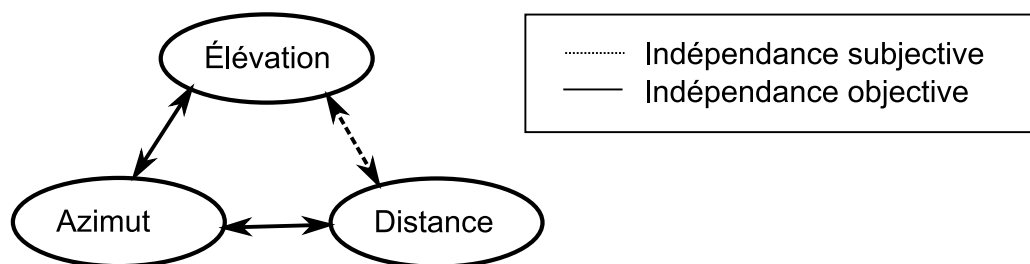


FIGURE 8.1 – Les trois paramètres de spatialisation de STAR et leurs méthodes d'évaluation d'indépendance.

De plus, ce chapitre présente les tests d'indépendance objectifs, rendus possibles par la méthode de localisation perceptive présentée dans le chapitre 3. Ces indépendances objectives, comme montré dans la Fig. 8.1, concernent d'une part l'azimut avec la distance, et d'autre part l'élévation avec la distance. N'ayant pas de moyen d'évaluer objectivement l'indépendance entre la distance et l'élévation, nous avons eu recours à des tests subjectifs afin d'évaluer une indépendance entre l'ensemble des paramètres ; cette indépendance est montrée dans la Sec. 11.2.

8.1 Indépendance entre l'azimut et l'élévation

8.1.1 Étude sur l'élévation en conditions anéchoïques

La première étude considère l'indépendance entre l'azimut et l'élévation. Une première observation intéressante est faite dans le cadre d'une situation en chambre anéchoïque, en utilisant les HRTFs de la base CIPIC afin de simuler l'élévation, puis de localiser cette élévation avec la méthode de localisation présentée dans le Chap. 3 (en utilisant donc le même procédé que pour la validation de la méthode).

La Fig. 8.2 montre la localisation de la méthode pour les élévations $[-45 : 90]$ degrés, ceci pour différents azimuts et sur l'ensemble des sujets de la base CIPIC.

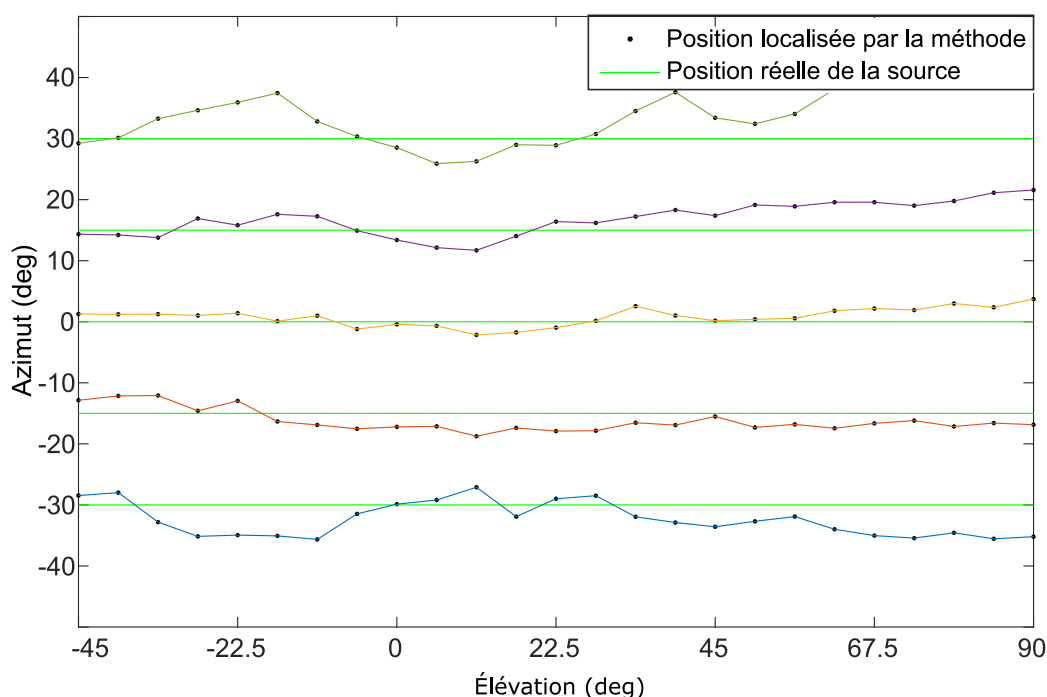


FIGURE 8.2 – Localisation de l'azimut pour 5 positions, 43 sujets de la base CIPIC et 25 élévations.

8.1. INDÉPENDANCE ENTRE L'AZIMUT ET L'ÉLEVATION

En condition anéchoïque, l'élévation ne semble pas avoir d'influence sur l'azimut, observation confirmée par la Fig. 3.4 qui montre que l'erreur est faible et constante quelle que soit l'élévation.

8.1.2 Étude sur l'élévation en conditions réelles

Les conditions anéchoïques ne satisfaisant pas la philosophie de STAR, dont le domaine d'application correspond à des conditions réelles, il est nécessaire d'utiliser des BRIRs au lieu des HRTFs. N'ayant pas de BRIRs satisfaisantes pour l'élévation dans la bibliographie, la base SCRIME a été construite, dans les conditions d'enregistrement introduites en début de chapitre. Le mannequin fixé sur le haut d'une chaise de bureau à l'aide d'une puissante charnière fixe est incliné, simulant ainsi l'élévation. En étant incliné, les rotations du mannequin restant permises, il est ainsi possible d'acquérir l'ensemble des azimuts. Un biais est cependant présent lorsque le mannequin est incliné, dû à la taille du mannequin, et dépendant des azimuts. Replacer la tête au bon endroit pouvant engendrer trop d'erreurs expérimentales, il a été choisi d'appliquer une correction *a posteriori*, utilisant la taille du mannequin, ainsi que les angles verticaux et horizontaux connus.

La Fig. 8.3 présente le dispositif KEMAR lors d'un enregistrement simulant une position verticale de 40° , visible sur le compas non solidaire du support ¹.



FIGURE 8.3 – Inclinaison du mannequin pour simuler l'élévation (ici 40°).

En utilisant ainsi la base de données créée par ces enregistrements, on obtient les Fig. 8.4 et 8.5, qui montrent la position localisée de la source sonore utilisant la méthode de localisation perceptive (Chap. 3), ainsi que l'erreur obtenue sur l'ensemble des azimuts et pour cinq élévations.

1. Visible sur la photographie présentée dans les annexes.

8.1. INDÉPENDANCE ENTRE L'AZIMUT ET L'ÉLEVATION

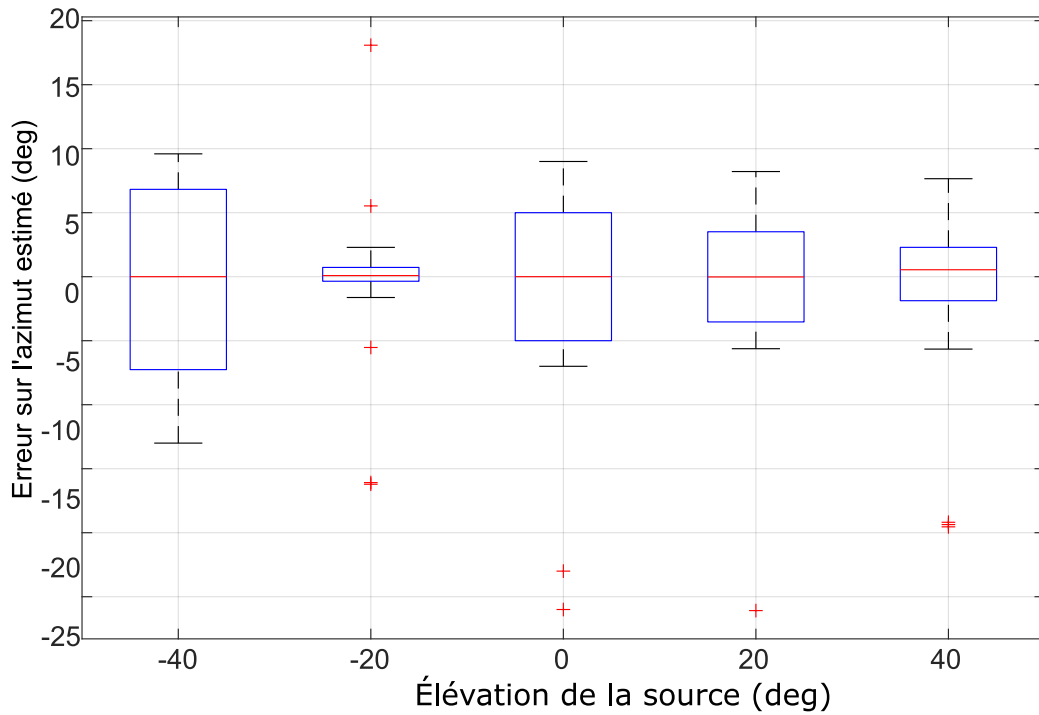


FIGURE 8.4 – Erreur d'estimation de l'azimut pour différentes élévations en utilisant les BRIRs du SCRIME. L'erreur est cohérente avec la précision humaine et ne dépend pas de l'élévation.

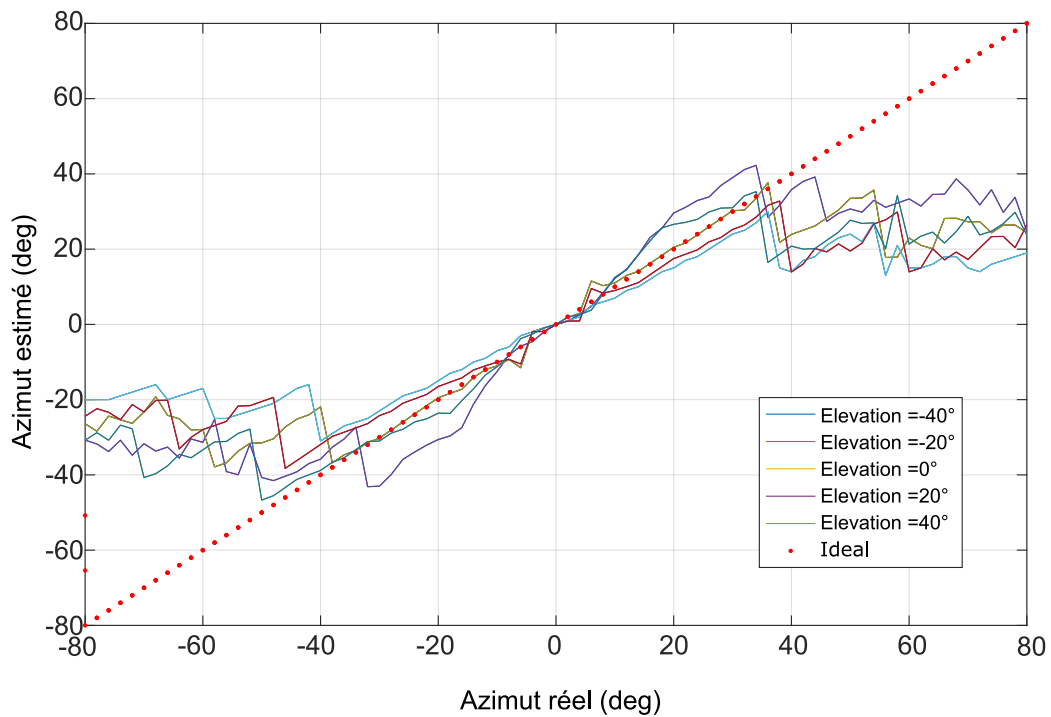


FIGURE 8.5 – Erreur d'estimation de l'azimut pour différentes élévations en utilisant les BRIRs du SCRIME. Les résultats suivent l'idéal dans la plage $\pm 40^\circ$, et ceci quelle que soit l'élévation.

Ces deux figures permettent de confirmer les observations effectuées lors de l'étude en conditions ané-

8.1. INDÉPENDANCE ENTRE L'AZIMUT ET L'ÉLEVATION

choïques (Sec. 8.1.1); l'élévation n'influe pas sur la localisation de l'azimut, l'erreur de localisation étant très similaire quelle que soit l'élévation.

8.1.3 Étude sur l'élévation simulée

La dernière étape est de montrer que l'indépendance entre l'azimut et l'élévation est également présente lors des simulations d'élévation en utilisant la méthode STAR. Aussi les Fig. 8.6 et 8.7 montrent cette étude, cette fois en laissant le mannequin fixe et en simulant l'élévation.

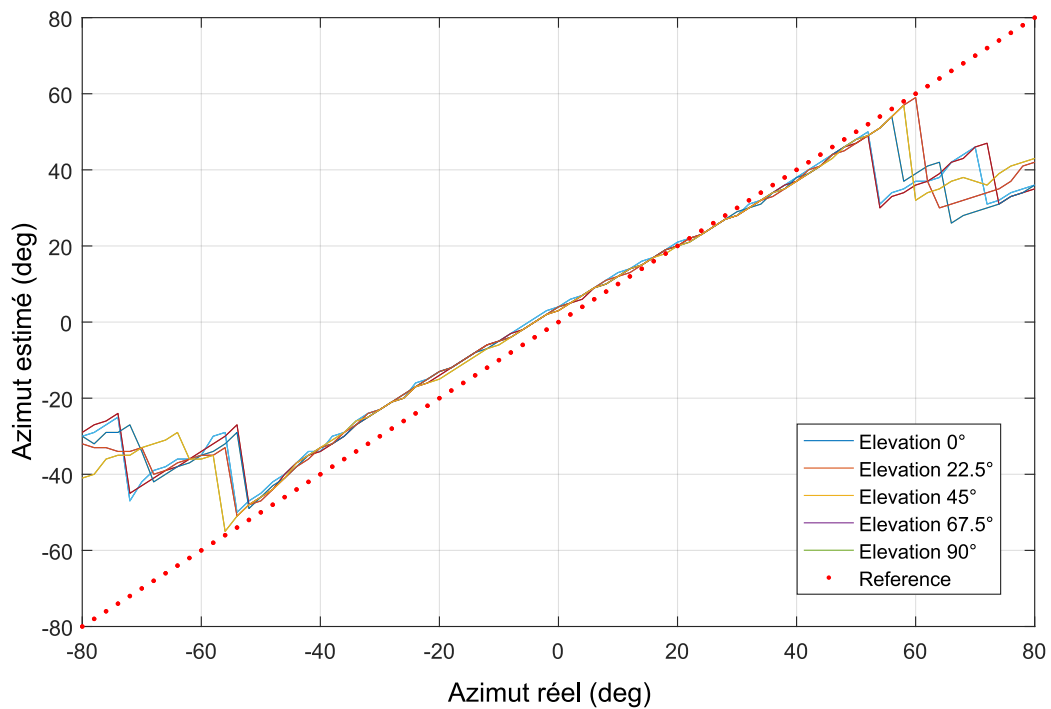


FIGURE 8.6 – Erreur d'estimation de l'azimut pour différentes élévations simulées. L'erreur est cohérente avec la précision humaine et ne dépend pas de l'élévation.

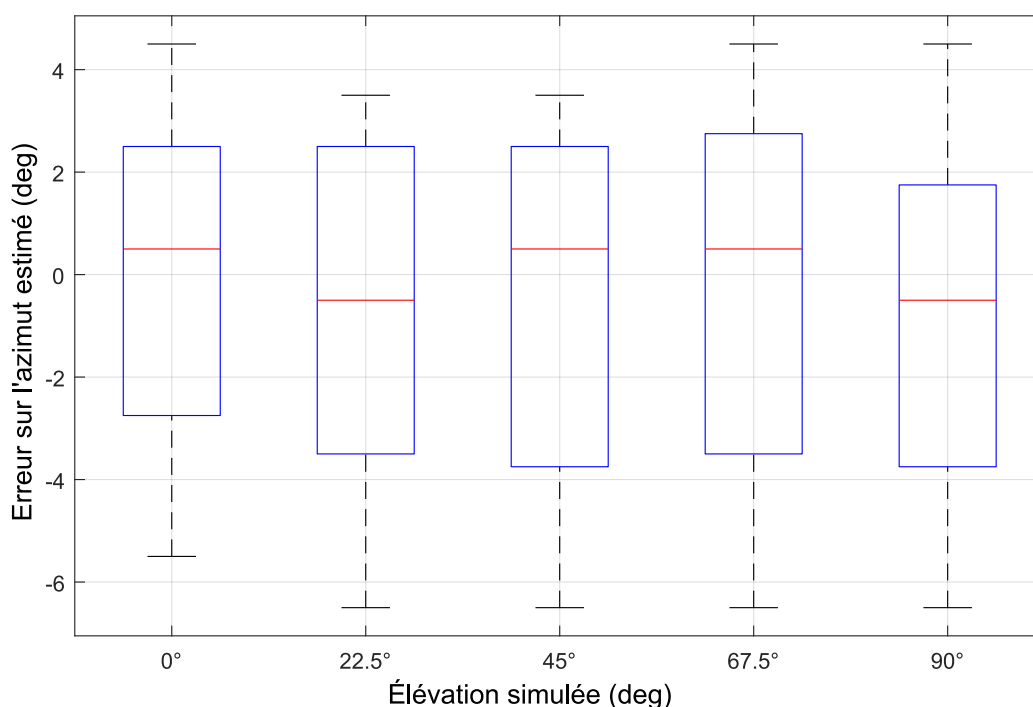


FIGURE 8.7 – Erreur d'estimation de l'azimut pour différentes élévations simulées. Les résultats suivent l'idéal dans la plage $\pm 40^\circ$, et ceci quelle que soit l'élévation.

Les résultats sont similaires (voire meilleurs) que ceux obtenus avec l'élévation réelle (dus à la diminution de réverbération), confirmant définitivement l'indépendance entre l'élévation et l'azimut, et ainsi permettant de valider le processus de localisation de la méthode STAR.

8.2 Indépendance entre l'azimut et la distance

Contrairement à l'élévation, il n'existe pas de base HRTF présentant des différences de distances importantes pour des sources éloignées (ce qui est logique, comme introduit dans la section sur la localisation de la distance (Sec. 1.2.3), le contenu spectral influant sur la localisation de la distance pour les champs proches (< 1 m). C'est pour cela que, contrairement à l'étude de l'indépendance entre l'élévation et l'azimut, il n'y a pas dans cette section d'étude en condition anéchoïque.

8.2.1 Étude sur la distance réelle

Pour étudier l'indépendance entre l'azimut et la distance, il faut donc enregistrer, pour différentes distances et azimuts, une source sonore connue, pour ensuite via la méthode de localisation (Chap. 3) retrouver l'azimut, et ainsi en déduire l'indépendance ou non de ces paramètres. Afin d'obtenir les meilleurs résultats, il a été choisi de se mettre en conditions réelles, et en champs semi-libres. Certaines bases BRIR telles que AIR-database qui proposent des BRIRs enregistrées dans des couloirs ont été écartées, étant des cas trop particuliers (réverbérations très importantes). Dans cette optique, une première approche a été faite en plein air (Fig. 8.8). Les tests ainsi menés n'ont pas donné de résultats concluants, le moindre vent provoquant d'importantes erreurs sur la localisation (ce qui est logique, la méthode ne faisant pas de séparation de sources, l'énergie provenant de la source virtuelle provoquée par le vent contre le micro d'une des oreilles étant très importante).

8.2. INDÉPENDANCE ENTRE L'AZIMUT ET LA DISTANCE



FIGURE 8.8 – Exemple des tests d'enregistrement dans des conditions extérieures (ici pour un azimuth $\theta = 80^\circ$ et une distance $d = 2\text{m}$).

Aussi, pour la distance, les mesures sont prises au studio du SCRIME dans la même pièce que les acquisitions pour l'élévation (Fig. 8.9). Le mannequin est positionné à 1, 2, 4, et 7 mètres (la distance maximale possible sans être trop proche des murs étant limitée par la taille de la pièce). De plus, dans la plupart des conditions d'utilisation standard de STAR, une distance de 7 m des haut-parleurs est suffisante.



FIGURE 8.9 – Exemple d'enregistrement effectué pour les tests d'indépendance de la distance dans le studio du SCRIME. Ici pour une distance $d = 2\text{m}$ et un azimuth $\theta = 0^\circ$.

Les niveaux de pression acoustique ont été enregistrés à l'aide d'un microphone Schoeps MK4 à capsule unique.

les Fig. 8.10 et 8.11 montrent les résultats de la localisation de l'azimut en faisant varier la distance. L'erreur reste faible dans une grande plage d'azimuts.

Le Tab. 8.1 renseigne les différents niveaux acoustiques en fonction de la distance, qui ne diminuent pas au-delà de - 86 dB, ce qui est suffisant pour la localisation (Joyce and John (2004)). Le Tab. 8.2 renseigne la perte de pression acoustique entre les différentes distances considérées, montrant que les SPLs ne décroissent

8.2. INDÉPENDANCE ENTRE L'AZIMUT ET LA DISTANCE

pas de manière linéaire de 6 dB chaque fois que la distance double, ce qui est concordant avec la notion que nous ne sommes pas en champ libre (hormis pour les distances de 1 à 2 m, très proches des haut-parleurs et donc du champ libre). Cette observation permet notamment de justifier la valeur plus réaliste de perte de pression acoustique de 4.25 dB déterminée par Kolarik et al. (2016), et utilisée dans la spatialisation de la distance (Chap. 7).

Distance (m)	1	2	4	7
Niveau acoustique (dB)	- 71	- 77	- 82	- 86

TABLE 8.1 – Niveaux acoustiques pour les différentes distances considérées, enregistrées avec un microphone Schoeps MK4 à capsule unique.

Distance (m)	1 to 2	2 to 4	4 to 7
Perte de SPL (dB)	- 6	- 5	- 4

TABLE 8.2 – Perte de pression acoustique entre les différentes positions considérées.

Il est intéressant également de noter le niveau de bruit du studio lors de l'enregistrement qui est d'environ - 91 dB, valeur assez proche des - 86 dB que l'on obtient pour la distance de 7 m. Ceci peut expliquer la présence de l'erreur (Fig. (8.10)).

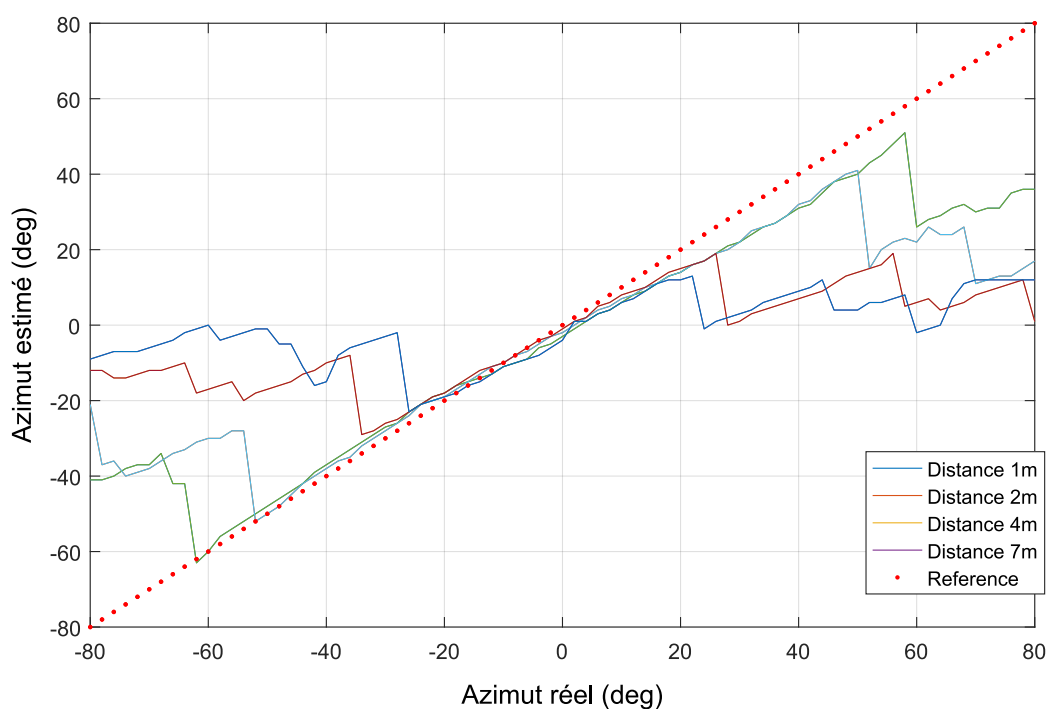


FIGURE 8.10 – Estimation de l'azimut pour différentes distances. Les résultats suivent assez bien l'idéal. On voit aussi que plus la distance est importante et plus le décrochage se trouve dans des azimuts faibles.

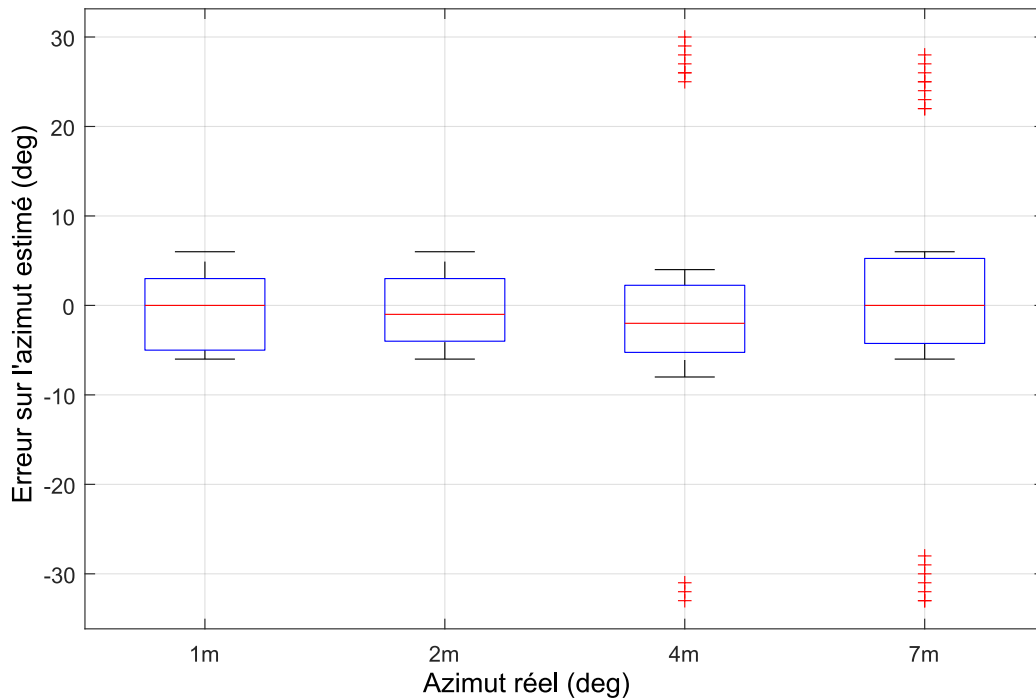


FIGURE 8.11 – Estimation de l'azimut pour différentes distances. L'erreur obtenue concorde avec la précision de localisation de l'humain. De nombreux outliers sont présents lorsque l'on est dans des distances plus importantes. Cependant, la moyenne et l'écart type ne dépendent pas de la distance.

En se basant sur ces observations, on déduit que la distance a un effet sur la localisation en azimut, de par les réverbérations qui ont un impact d'autant plus important que la distance est grande. Cependant, en considérant la moyenne et l'écart type, seule l'erreur engendrée est plus importante. La distance et l'azimut sont donc indépendants l'un de l'autre.

8.2.2 Étude sur la distance simulée

L'étude sur les données réelles (Sec. 8.2.1) valide ainsi que la distance n'a pas d'effet lié à la localisation de l'azimut, et ainsi l'indépendance des paramètres. Tout comme pour l'élévation (Sec. 8.1), il est important d'également étudier le cas où les distances mesurées ne sont pas réelles mais simulées.

Les Fig. 8.12 et 8.13 montrent les résultats de cette étude. De plus, étant simulées, il est facile de considérer des distances bien plus importantes (15 m).

8.2. INDÉPENDANCE ENTRE L'AZIMUT ET LA DISTANCE

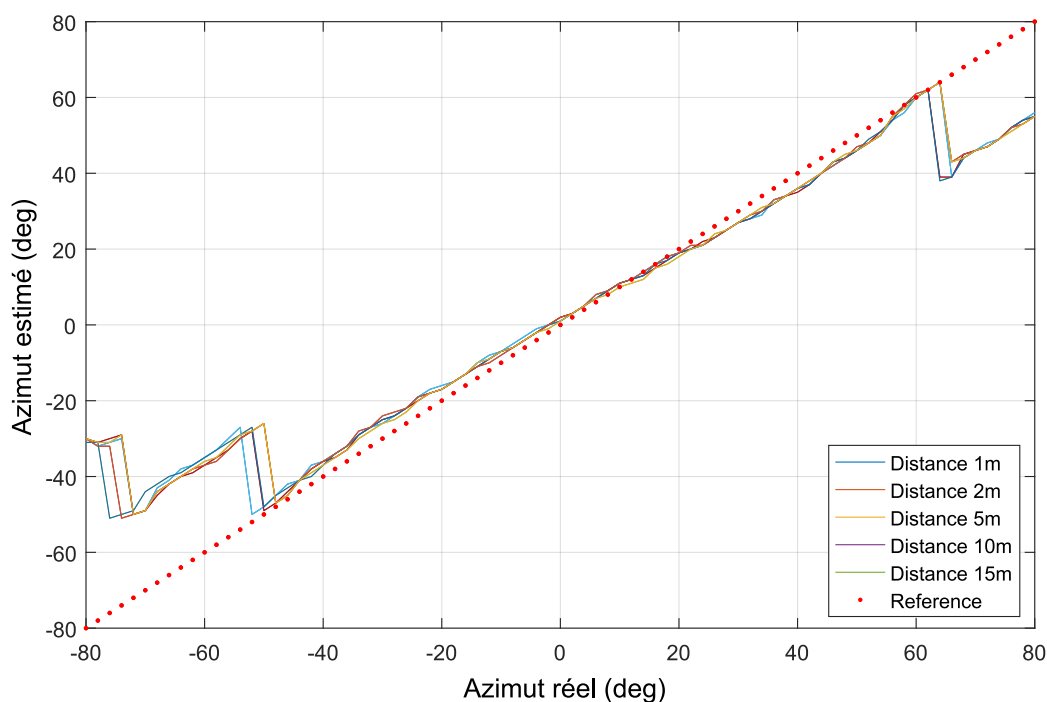


FIGURE 8.12 – Estimation de l'azimut pour différentes distances simulées. Les résultats suivent bien l'idéal dans une large plage d'azimuts $\pm 60^\circ$. La distance n'influe pas sur la détection de l'azimut, les paramètres sont donc indépendants.

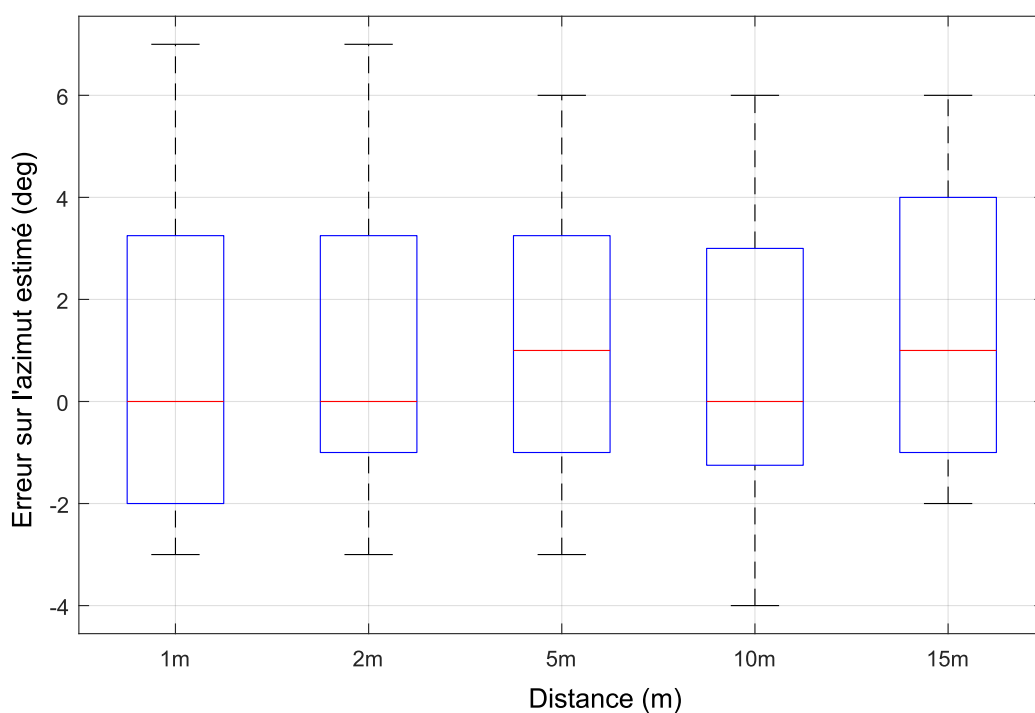


FIGURE 8.13 – Estimation de l'azimut pour différentes distances simulées. L'erreur concorde avec la précision de localisation du son par l'homme et ne dépend pas de la distance. Aucun outlier n'est présent contrairement aux résultats pour des distances réelles. Les paramètres sont donc indépendants.

8.2. INDÉPENDANCE ENTRE L'AZIMUT ET LA DISTANCE

Les résultats sont meilleurs que ceux mesurés en conditions réelles, supprimant également l'ensemble des outliers. Ceci s'explique par le fait que lors des simulations, l'énergie est moins distribuée en lien avec la distance plus courte parcourue par l'onde sonore ; l'enregistrement est également moins sujet aux réverbérations (les décrochages n'apparaissent que vers 60°). On remarque également qu'en distance simulée, il n'y a aucun rapport entre la distance et l'azimut, confirmant que les deux paramètres sont indépendants.

– Conclusion –

L'indépendance entre l'azimut et la distance, ainsi qu'entre l'azimut et l'élévation ont été prouvés dans ce chapitre ; et ceci que ce soit dans le cas de données réelles ou simulées. Les conditions simulées donnent, de plus, de meilleurs résultats que les conditions réelles.. Cette étude d'indépendance est importante pour justifier le procédé de spatialisation de la méthode STAR, qui effectue de manière indépendante la spatialisation des différentes dimensions.

La dernière indépendance à valider est l'indépendance entre la distance et l'élévation, qui, pour le moment, ne peut être testée que subjectivement (ne disposant pas de méthode objective de localisation ni en distance ni en élévation), et qui profitera des tests de performances de la spatialisation de ces deux dimensions.

L'indépendance entre azimut et distance ainsi qu'entre azimut et élévation étant prouvée, il est possible de procéder aux différents tests présentés dans la suite de cette partie.

L'indépendance des paramètres est devenue un point important pour la méthode STAR, dont le processus spatialise les dimensions les unes après les autres. Pour cela, au cours de la thèse, l'indépendance entre les trois paramètres de spatialisation est évaluée ; pour l'indépendance entre élévation/distance et azimut, les tests d'évaluation sont objectifs. Les résultats ont été largement développés dans les travaux de Méaux and Marchand (2020b), mais aussi repris dans une vision plus globale dans la revue de l'AES Méaux and Marchand (2021b).

Chapitre 9

Évaluation objective de la méthode STAR par la cartographie des ILDs

– Introduction –

Maintenant que l'indépendance des paramètres est validée de manière objective¹, le procédé de spatialisation de la méthode STAR est justifié. Cependant, bien que justifié, rien ne permet de juger des performances de la méthode. Des tests subjectifs, utilisant l'avis d'auditeurs, ont été menés. Il est aussi intéressant de regarder si on peut étudier les performances de la méthode STAR de manière objective, avec les outils dont nous disposons.

Grâce au mannequin KEMAR, il est possible d'enregistrer des sons tels qu'un auditeur les recevrait, et, grâce aux bases de données HRTF, de simuler ces sons. Il est alors possible d'exploiter les résultats recueillis de manière objective. La distance et l'élévation étant complexes dans leur exploitation, et n'ayant pas de moyen de les mesurer (absence de méthode de localisation), il n'est pas possible d'effectuer une étude objective pour ces dimensions. Pour l'azimut, cette étude est possible, l'azimut étant, de plus, le paramètre primordial (le plus important de la localisation humaine). Cette étude objective est rendue possible notamment grâce à l'exploitation des indices interauraux qui sont faciles d'exploitation.

Ce chapitre va utiliser ces indices, et en particulier les ILDs, plus simples d'utilisation car ne dépendant pas de la phase, afin d'évaluer objectivement trois méthodes (HOA, VBAP, STAR); et ceci en produisant des cartographies des ILDs. Ces cartographies, en plus de proposer une matière d'exploitation brute, peuvent également être comparées. Les méthodes HOA et VBAP étant validées dans la littérature, les performances de STAR peuvent être justifiées grâce à une comparaison avec ces méthodes.

a. ou le sera pour l'indépendance distance et élévation dans le chapitre suivant.

Ce chapitre compare les cartographies des ILDs (et plus précisément les erreurs d'ILDs intrinsèque à chaque méthode) pour différentes méthodes de spatialisation et différentes positions d'auditeurs. L'erreur d'ILDs consiste en la différence entre les ILDs considérées (provenant de mesures ou de simulations), et de la référence d'ILD, provenant du mannequin KEMAR à grandes oreilles de la base CIPIC (Algazi et al. (2001)), supposée au centre. En théorie, plus l'erreur est faible, et plus la méthode est précise. Pour chaque configuration (source, haut-parleurs, et auditeurs), nous considérerons la moyenne des ILDs dans le temps et la fréquence.

Pour chaque simulation et mesure, nous utilisons un système octophonique de haut-parleurs, avec un

9.1. CARTOGRAPHIE DES DIFFÉRENCES D'INTENSITÉ INTERAURALES (ILDs) PAR SIMULATIONS

bruit blanc durant 6 s, les exemples proposés sont réalisés avec une source localisée à 30° ¹, et avec le mannequin KEMAR à grandes oreilles (que ce soit dans les simulations ou dans les mesures réelles) pour simuler l'auditeur.

Les ILDs de références sont calculées en utilisant l'équation 1.8 présentée dans la Sec. 1.4.2, à partir des signaux binauraux obtenus par convolutions simples du signal source (bruit blanc) avec les HRIRs de la base de données CIPIC pour l'azimut désiré (30°). Pour ce calcul des ILDs, l'équation 1.8 est utilisée avec les données provenant d'une paire de transformées de Fourier à court terme ainsi qu'une fenêtre de Hann de taille $n = 2048$ échantillons, ainsi qu'un chevauchement de 50%, pour les sons (binauraux) enregistrés à une fréquence d'échantillonnage $F_e = 44100$ Hz.

9.1 Cartographie des différences d'intensité interaurales (ILDs) par simulations

La première partie de ce chapitre est consacrée à l'étude des cartographies d'ILDs simulées (provenant d'HRTFs de la base CIPIC).

9.1.1 Protocole de simulation

Le rendu binaural

Pour la simulation, la méthode de spatialisation est exécutée *in silico*, et les sorties des huit haut-parleurs de la configuration octophonique doivent être transformées en signaux binauraux afin de calculer les ILDs utilisées pour les cartographies.

Le rendu binaural est une technique permettant d'obtenir un signal binaural à partir de configurations telles que la nôtre, notamment pour une écoute au casque ; et permettant donc le calcul d'indices acoustiques.

Différentes techniques sont proposées dans la littérature afin d'obtenir un rendu binaural à partir d'un signal multi-canaux, particulièrement utile pour des méthodes, telles que HOA, utilisant l'ensemble des haut-parleurs à leur disposition. Dans le cas de l'étude objective mettant en comparaison les trois techniques de spatialisation, la plus simple de ces méthodes est choisie par équité². Cette méthode prend en compte les chemins acoustiques entre chacun des haut-parleurs et les oreilles gauche et droite, comme exposé dans la Fig. 9.1.

1. La source est située entre deux haut-parleurs sans être en leur centre, permettant ainsi d'éviter toute configuration particulière.

2. Permettant ainsi de limiter le risque d'influences venant de cette méthode.

9.1. CARTOGRAPHIE DES DIFFÉRENCES D'INTENSITÉ INTERAURALES (ILDS) PAR SIMULATIONS

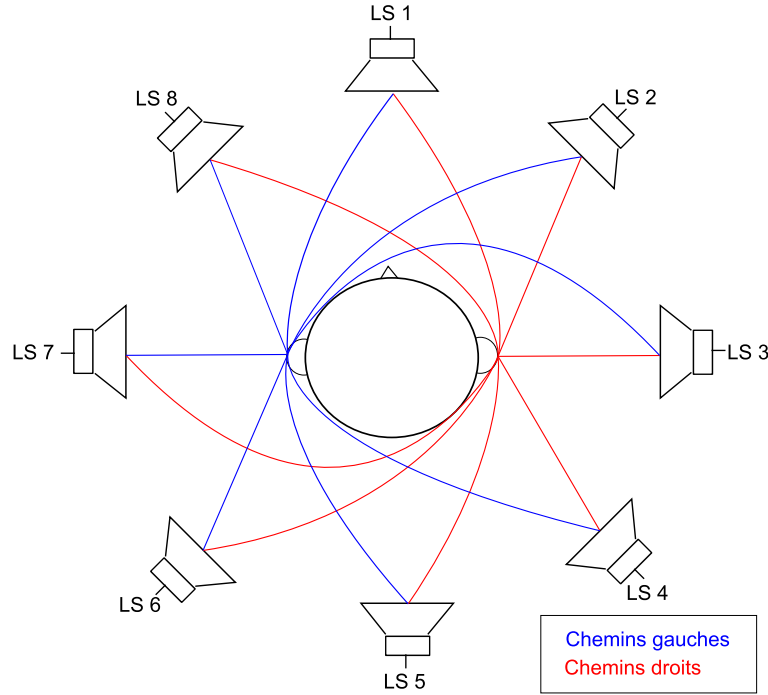


FIGURE 9.1 – Chemins acoustiques pour la configuration octophonique.

Ces chemins sont en théorie les HRIRs d'une source placée à la position de chaque haut-parleur. Ainsi, pour une configuration octophonique de 8 haut-parleurs, nous avons l'Eqn. 9.1 :

$$s_{l,r} = \sum_{n=1}^N \text{HRIR}_{l,r}(\theta_n) * s_n \quad (9.1)$$

où $s_{l,r}$ désigne le côté gauche ou droit d'un signal binaural, $\text{HRIR}_{l,r}(\theta)$ étant l'HRIR gauche ou droit pour un azimut θ , s_n le signal joué par le haut-parleur n , et $*$ la convolution. De plus, pour des haut-parleurs répartis de manière homogène en cercle, on a $\theta_n = (n - 1) \cdot 2\pi/N$ (radians). Les HRIRs proviennent de la base CIPIC sans interpolation.

La simulation

Afin de considérer les différentes positions d'auditeurs, la simulation utilise une grille de 100×100 , couvrant une surface d' $1m^2$, et calcule chaque position sur cette grille pour le mannequin KEMAR (virtuel), regardant devant lui (vers l'azimut $\theta = 0$).

Pour la source sonore localisée en azimut 30° , un angle relatif (Fig.9.2) est calculé pour chaque position de la grille¹, puis chaque méthode de spatialisation est appliquée. Un rendu binaural (Sec. 9.1.1) est alors appliqué permettant de calculer les ILDs grâce à l'Eqn. 1.8.

1. Cette démarche utilise le fait que la distance n'a pas d'influence sur la localisation en azimut.

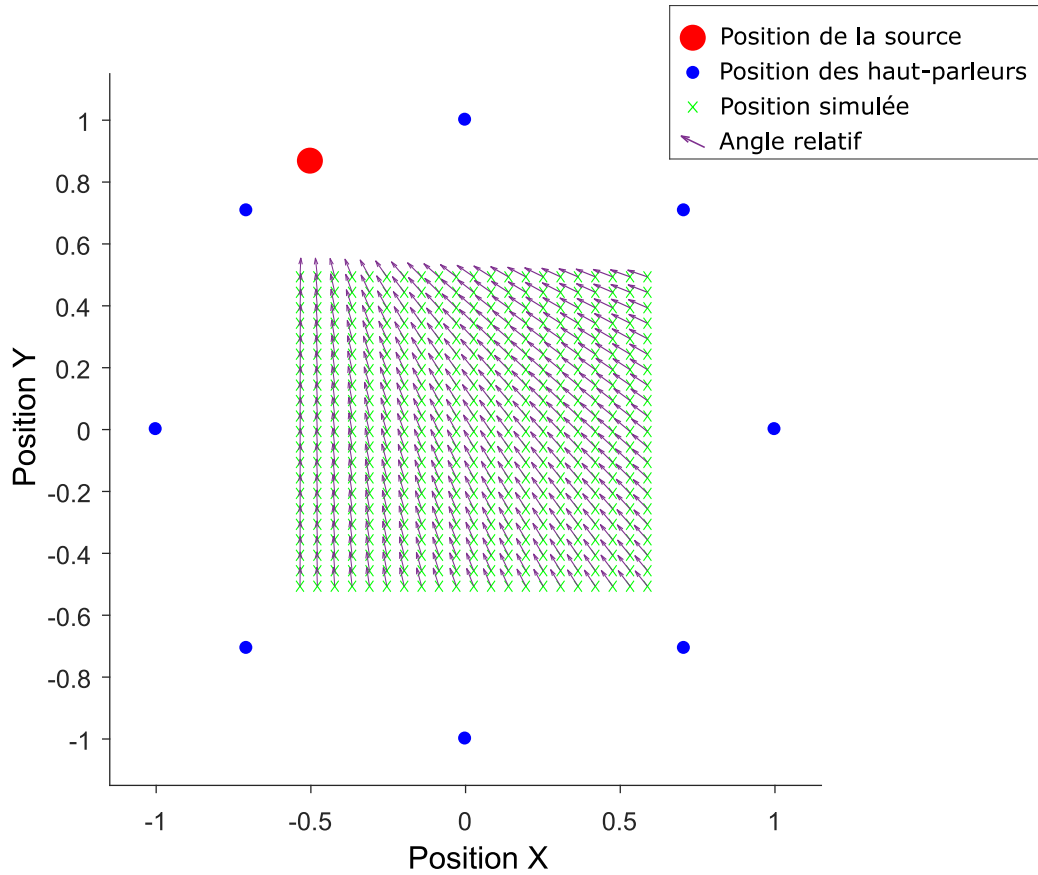


FIGURE 9.2 – Maillage simplifié utilisé pour la simulation de cartographies des ILDs.

9.1.2 Résultat des simulations

Les résultats de ces simulations pour les trois méthodes sont tracés sous forme de cartes dans les Fig. 9.3, 9.4 et 9.5; montrant l'erreur moyenne de l'ILD pour les méthodes de spatialisation considérées, avec la grille de 100×100 représentant les différentes positions de l'auditeur, une configuration de haut-parleurs octophonique régulière, et une source sonore située en $\theta = 30^\circ$.

La première constatation sur l'aspect de ces trois cartographies est que la méthode HOA est plus chaotique (donc très dépendante de la position des auditeurs), alors que la méthode VBAP semble être la plus harmonieuse, la méthode STAR donnant des résultats intermédiaires.

Nous voyons également un effet de diagonale (ligne noire orientée globalement entre la source et le centre), phénomène normal étant donné qu'un déplacement le long de cette ligne affecte uniquement la distance et pas l'angle relatif impactant les ILDs.

Enfin, la dernière information importante est que toutes les méthodes semblent avoir une erreur faible au centre de la configuration (approchant 0, et représentée sur les cartographies par la couleur noire).

La figure 9.6 représente l'évolution en fonction de l'azimut pour une plage d'azimuts $\theta = [10; 160]^\circ$, et ceci pour les trois méthodes, permettant de confirmer les observations précédentes. Cette figure permet également de montrer que la méthode VBAP présente des changements plus brutaux que les deux autres méthodes lorsque les azimuts varient.

9.1. CARTOGRAPHIE DES DIFFÉRENCES D'INTENSITÉ INTERAURALES (ILDS) PAR SIMULATIONS

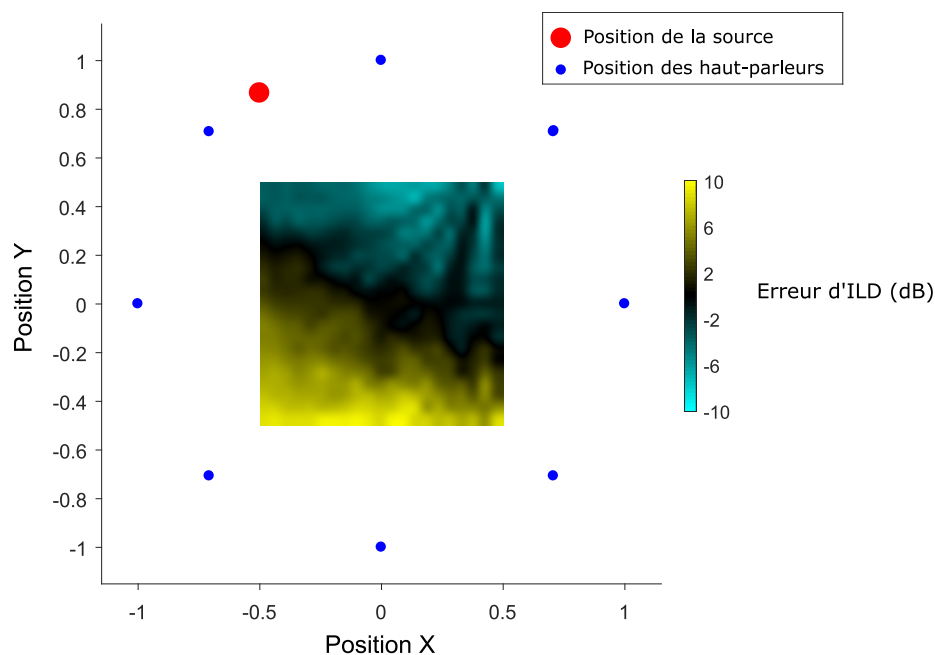


FIGURE 9.3 – Cartographie des ILDs utilisant la méthode de simulation pour une configuration octophonique régulière, une source placée à $\theta = 30^\circ$, et un maillage de 100×100 . — Méthode de spatialisation HOA.

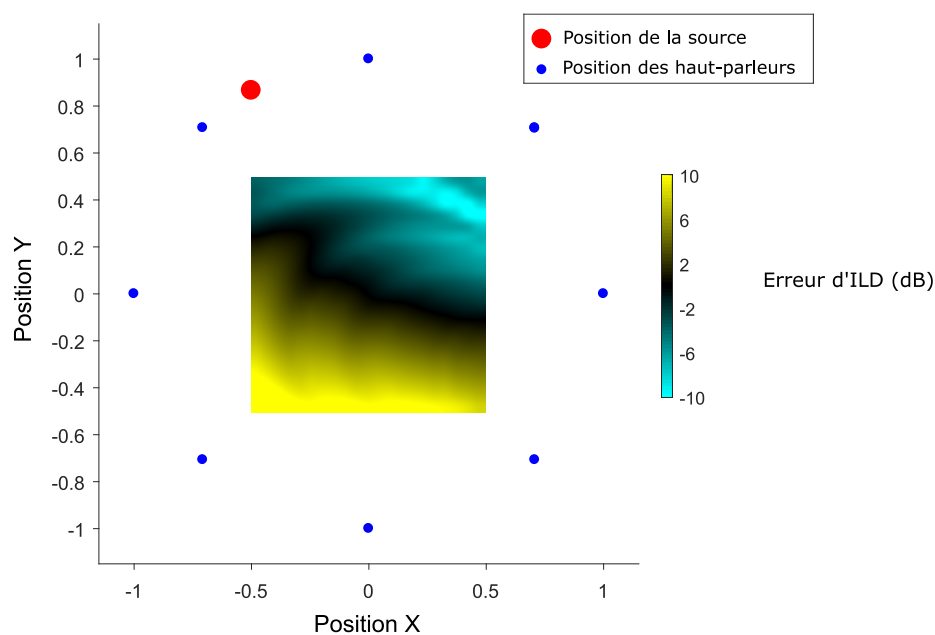


FIGURE 9.4 – Cartographie des ILDs utilisant la méthode de simulation pour une configuration octophonique régulière, une source placée à $\theta = 30^\circ$, et un maillage de 100×100 . — Méthode de spatialisation STAR.

9.1. CARTOGRAPHIE DES DIFFÉRENCES D'INTENSITÉ INTERAURALES (ILDS) PAR SIMULATIONS

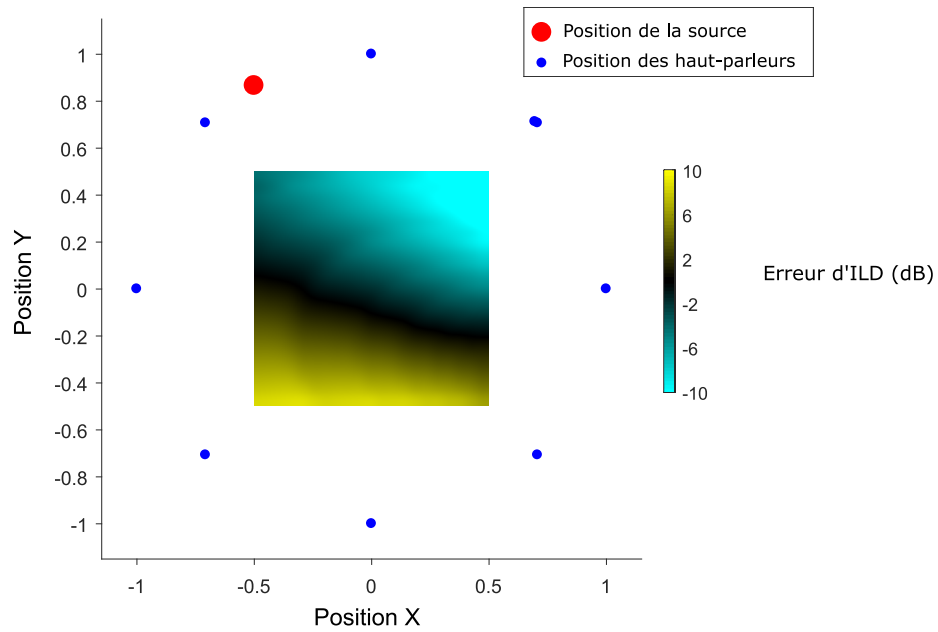


FIGURE 9.5 – Cartographie des ILDs utilisant la méthode de simulation pour une configuration octophonique régulière, une source placée à $\theta = 30^\circ$, et un maillage de 100×100 . — Méthode de spatialisation VBAP.

9.1. CARTOGRAPHIE DES DIFFÉRENCES D'INTENSITÉ INTERAURALES (ILDS) PAR SIMULATIONS

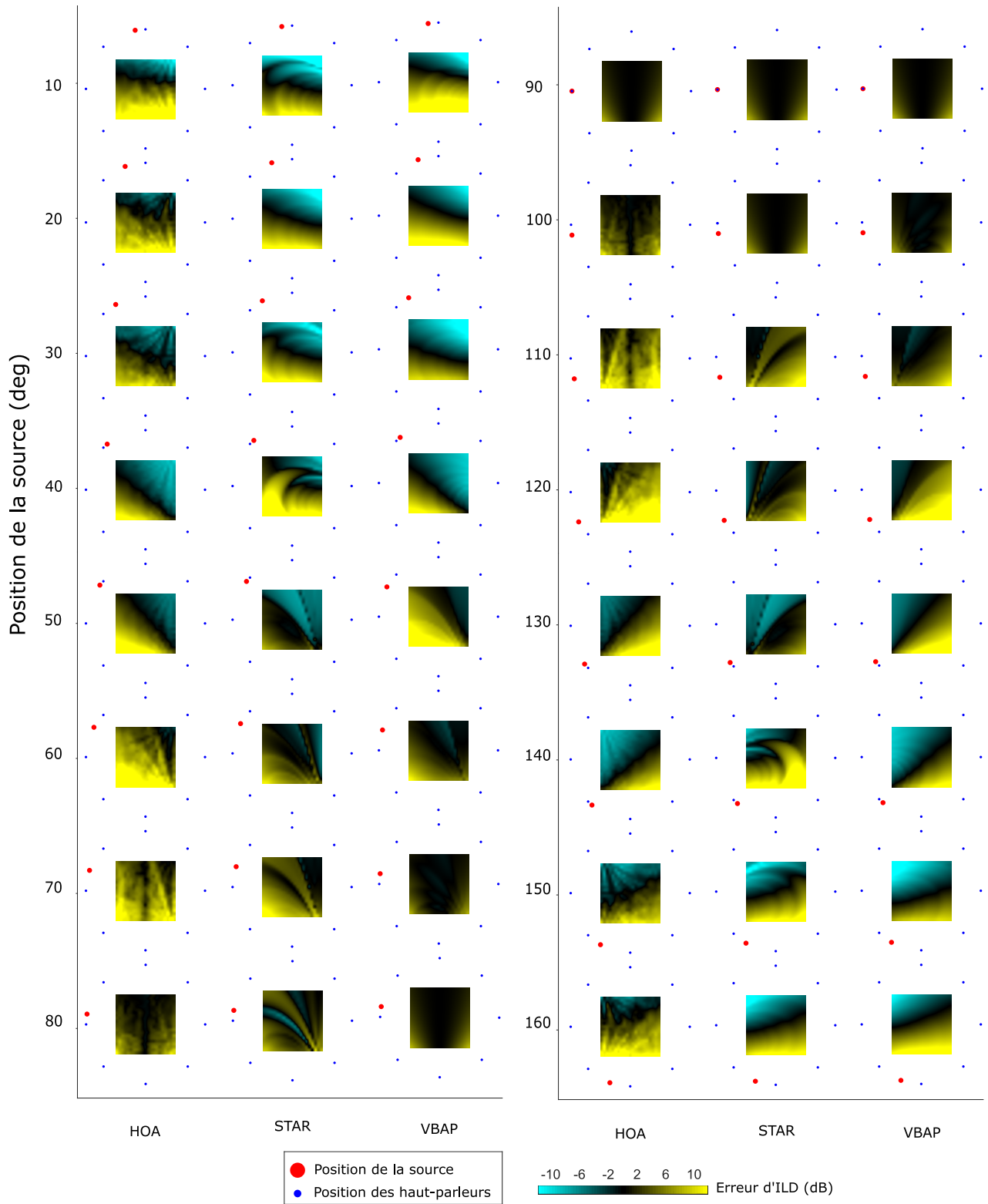


FIGURE 9.6 – Évolution des cartographies des ILDs pour les trois méthodes de simulation, une configuration octophonique régulière, un maillage de 100×100 , et un azimut θ allant de 10° à 160° avec un pas de 10° .

9.2 Cartographie des différences d'intensité interaurales (ILDs) réelles

La section précédente présente les travaux effectués pour la simulation de cartographies d'ILDs. Cette partie reprend les mêmes principes et outils, mais en utilisant cette fois-ci des mesures réelles effectuées au sein du SCRIME à l'aide du mannequin KEMAR.

9.2.1 Protocole expérimental

Les mesures sont effectuées dans le dôme du SCRIME, à l'aide d'un mannequin KEMAR à grandes oreilles, et un système ¹ permettant de déplacer le mannequin sur des marqueurs placés à l'avance afin d'obtenir une grille de 5×5 carrés de 40 cm de côté, couvrant une surface totale de $4m^2$ (Fig. 9.7). Cela imite la position de chaque auditeur dans un certain public autour du centre. La figure 9.8 montre ce public (de mannequins KEMAR), le centre étant affiché en gras.



FIGURE 9.7 – Mise en place expérimentale pour la prise de mesures des ILDs au sein du dôme du SCRIME.

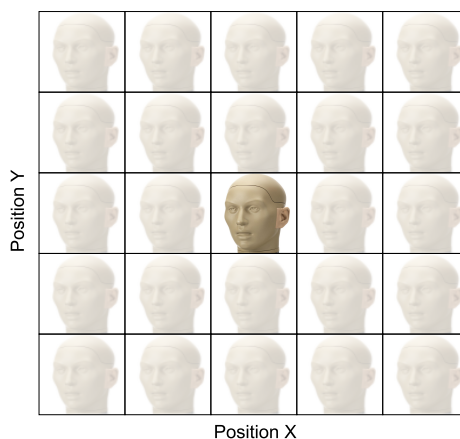


FIGURE 9.8 – Simulation d'un public par la présence de 25 mannequins KEMAR, la position centrale est représentée en gras.

9.2.2 Résultats des expérimentations

Bien que la cartographie expérimentale tende à confirmer les observations faites avec les simulations, la réverbération semble avoir un effet comparable aux différences des méthodes de spatialisation elles-mêmes ².

1. Le matériel utilisé a été introduit précédemment, et est également présenté plus en détail en annexe.

2. Il est donc difficile de savoir si les observations sont dues aux réverbérations ou aux méthodes de spatialisation.

9.2. CARTOGRAPHIE DES DIFFÉRENCES D'INTENSITÉ INTERAURALES (ILDS) RÉELLES

De plus, le maillage n'étant que de dimensions 5×5 pour que les expériences soient faisables, la résolution est bien plus faible que lors des simulations. Les graphiques des Fig. 9.9, 9.10 et 9.11 sont interpolés. Les tableaux de données correspondant aux mesures sont donnés dans les Tab. 9.2, 9.1 et 9.3.

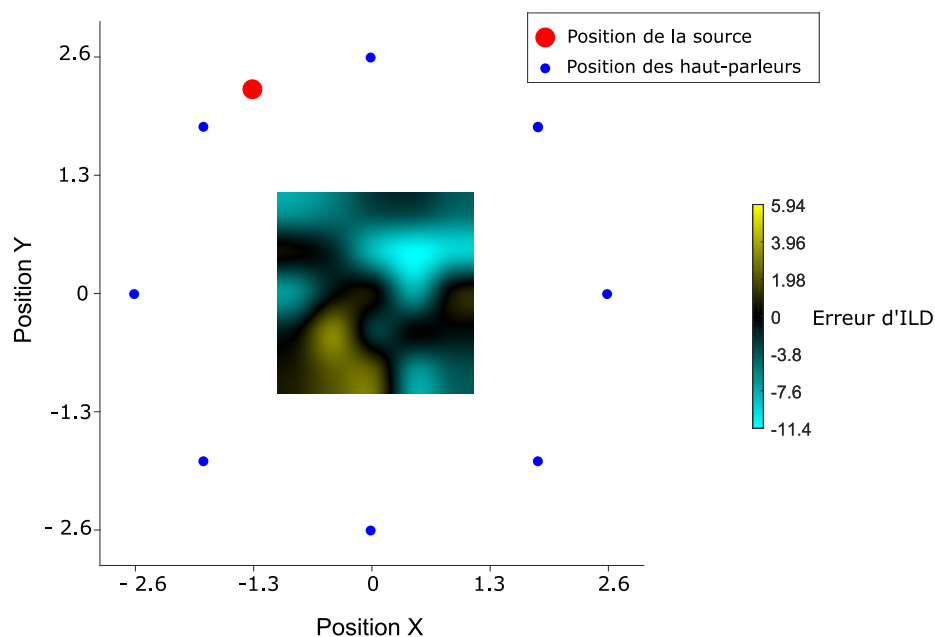


FIGURE 9.9 – Cartographie des ILDs utilisant la méthode expérimentale pour une configuration octophonique régulière, une source placée à $\theta = 30^\circ$, et un maillage de 5×5 . — Méthode de spatialisation STAR.

Y \ X	2.6	1.3	0	-1.3	-2.6
2.6	-1.1	0.3	6.2	-0.3	6.6
1.3	-4.0	5.9	0.5	1.8	5.2
0	-4.1	2.4	-0.3	8.3	3.1
-1.3	6.4	0.1	7.4	11.4	2.9
-2.6	3.9	0.5	-0.5	9.4	4.5

TABLE 9.1 – Erreur des ILDs (en dB) utilisant la méthode expérimentale pour une configuration octophonique régulière, une source placée à $\theta = 30^\circ$, et un maillage de 5×5 . La position centrale est représentée en gras — Méthode de spatialisation STAR.

9.2. CARTOGRAPHIE DES DIFFÉRENCES D'INTENSITÉ INTERAURALES (ILDS) RÉELLES

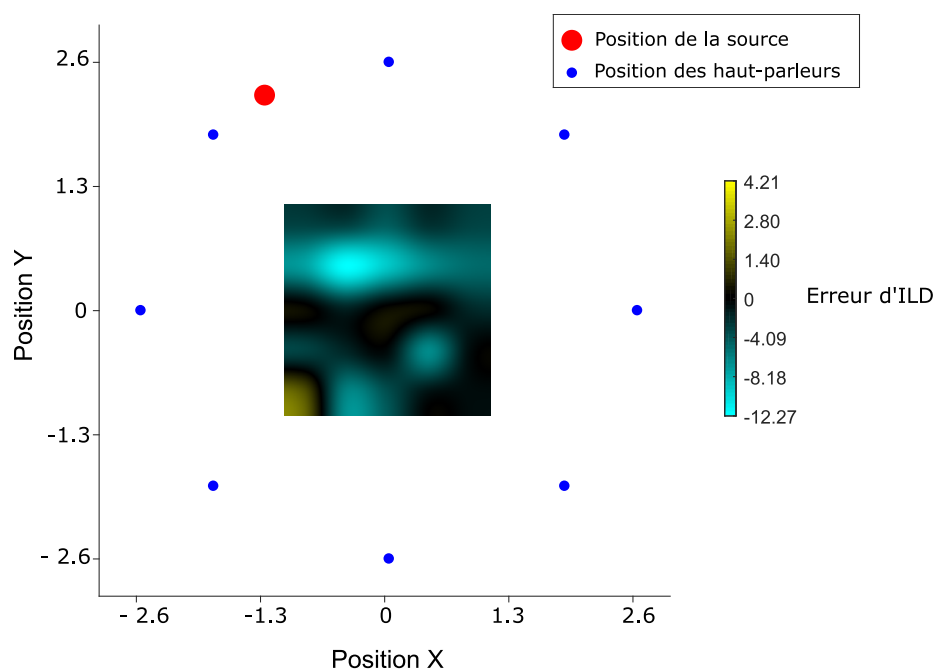


FIGURE 9.10 – Cartographie des ILDs utilisant la méthode expérimentale pour une configuration octophonique régulière, une source placée à $\theta = 30^\circ$, et un maillage de 5×5 . — Méthode de spatialisation HOA.

Y\X	2.6	1.3	0	-1.3	-2.6
2.6	-0.5	-0.6	-1.9	-5.1	-3.7
1.3	-0.7	-6.6	0.4	-6.6	-2.3
0	-3.9	-0.5	0.8	-9.6	-4.4
-1.3	-6.2	-1.8	-1.5	-12.2	-2.9
-2.6	4.2	-3.1	0.5	-7.6	-3.1

TABLE 9.2 – Erreurs des ILDs (en dB) utilisant la méthode expérimentale pour une configuration octophonique régulière, une source placée à $\theta = 30^\circ$, et un maillage de 5×5 . La position centrale est représentée en gras — Méthode de spatialisation HOA.

9.2. CARTOGRAPHIE DES DIFFÉRENCES D'INTENSITÉ INTERAURALES (ILDS) RÉELLES

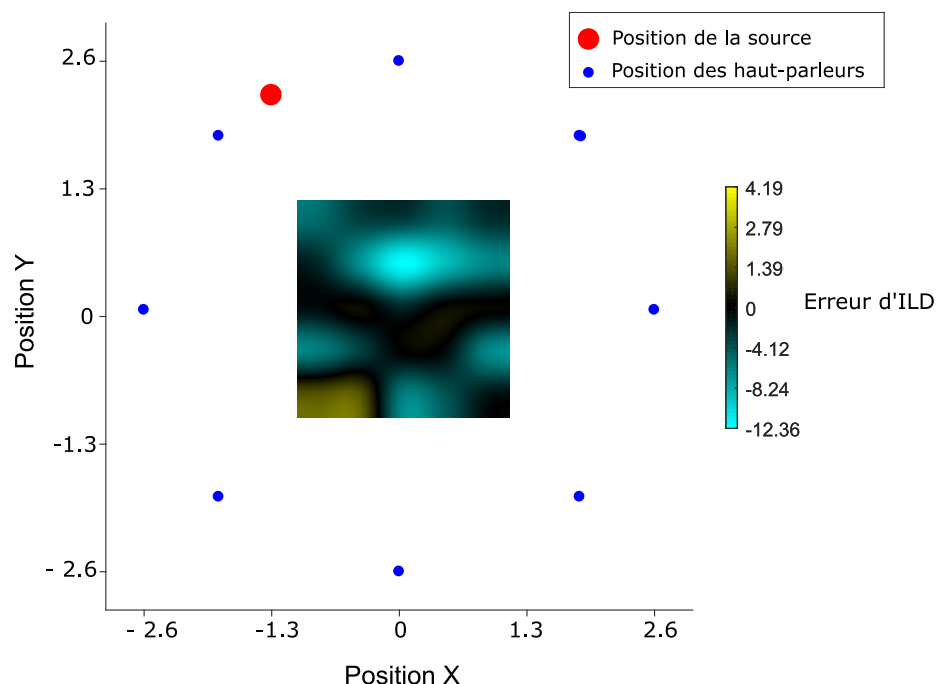


FIGURE 9.11 – Cartographie des ILDs utilisant la méthode expérimentale pour une configuration octophonique régulière, une source placée à $\theta = 30^\circ$, et un maillage de 5×5 . — Méthode de spatialisation VBAP.

Y\X	2.6	1.3	0	-1.3	-2.6
2.6	-1	-6.5	0.3	-6.6	-2.3
1.3	-4.1	-0.5	0.4	-9.2	-4.3
0	-6.0	-0.0	-1.8	-12.3	-2.9
-1.3	4.2	-3.2	0.6	-6.9	-3.1
-2.6	4.0	-6.0	-0.5	-1.9	-5.6

TABLE 9.3 – Erreurs des ILDs (en dB) utilisant la méthode expérimentale pour une configuration octophonique régulière, une source placée à $\theta = 30^\circ$, et un maillage de 5×5 . La position centrale est représentée en gras — Méthode de spatialisation VBAP.

Le Tab. 9.4 propose un résumé des différentes valeurs au centre des cartographies expérimentales et simulées. Il apparaît clairement que la méthode STAR est la méthode avec la meilleure reconstruction d'ILDs (possédant l'erreur la plus petite), ceci comparé aux deux autres méthodes ; cette observation restant vraie à la fois dans les cartographies simulées et expérimentales.

Les méthodes VBAP et HOA tentent de reconstruire les ondes acoustiques au centre du dispositif, et donc ne prennent pas en compte la présence de la tête de l'auditeur contrairement à la méthode STAR, ce qui explique probablement ses bonnes performances.

9.2. CARTOGRAPHIE DES DIFFÉRENCES D'INTENSITÉ INTERAURALES (ILDS) RÉELLES

	HOA	STAR	VBAP
Erreur des ILD simulées	0.7	0.0	-2.4
Erreur des ILD expérimentales	0.8	-0.3	-1.8

TABLE 9.4 – Erreur des ILDs au centre du dispositif, pour les tests expérimentaux ainsi que les simulations, ceci pour les trois méthodes comparées, et une source située en azimuth $\theta = 30^\circ$.

D'autres tests ont été menés en ne considérant que des bandes spectrales, et différents décodeurs HOA ; les observations présentées dans ce chapitre restent valides quelle que soit la configuration.

– Conclusion –

Ce chapitre sur la cartographie d'indices acoustiques et plus spécifiquement des ILDs clôture les tests objectifs. Nous avons montré que la méthode STAR est au moins aussi bonne que les méthodes prises en référence, HOA et VBAP, de par son comportement en simulation : erreur très faible au centre, aspect assez doux dans l'évolution de la cartographie ; mais aussi de par ses valeurs d'erreur des ILDs, plus faibles que celles des méthodes de référence à la fois en simulations et en conditions réelles. Ces observations sont logiques et encourageantes, la notion de méthode perceptive de STAR reposant sur la recreation de ses indices acoustiques.

Ce chapitre permet donc d'ores et déjà de montrer l'intérêt et les performances de la méthode STAR pour la spatialisation en azimuth, qui peuvent maintenant être complétés avec les études subjectives.

La méthode STAR reposant sur les indices acoustiques, l'idée de faire des cartographies des ILDs a été adoptée afin de comparer objectivement les différentes méthodes dans des configurations d'utilisation souhaitée pour la méthode STAR. Les résultats concluants ont été publiés à la conférence de DAFx 2020 (Méaux and Marchand (2020a)), puis repris dans leur globalité dans la revue de l'AES (Méaux and Marchand (2021a)).

Chapitre 10

Évaluation subjective de comparaison de méthodes

– Introduction –

Les tests objectifs de la méthode STAR sont présentés dans les deux précédents chapitres. Ces tests apportent des observations et résultats nécessaires; il est opportun maintenant de les compléter avec des tests dits subjectifs.

Le premier de ces deux tests, présenté dans ce chapitre, est une comparaison de méthodes (HOA, STAR et VBAP), dans différentes configurations (source ponctuelle, dynamique, polyphonique), et ceci en conditions réelles (salle de concert, salle de classe), après avoir effectué un pré-test permettant d'identifier les personnes non sensibles à la spatialisation, pour pouvoir les exclure avant l'exploitation des résultats. Chacun des tests consiste en un questionnaire qu'un panel d'auditeurs doit remplir. En exploitant les réponses de ces auditeurs, on obtient alors des 'scores' pour les différentes méthodes, permettant de les comparer les unes aux autres.

Ce chapitre présente les différents protocoles de tests subjectifs effectués sur notre panel d'auditeurs, ainsi que les résultats obtenus. Ils permettent de conforter les observations menées dans le chapitre précédent. La spatialisation de l'azimut ainsi validée, le prochain chapitre concernera la validation pour l'élévation et la distance.

10.1 Protocoles des tests perceptifs de comparaison des méthodes de spatialisation

Cette section décrit les différents tests effectués; le pré-test, et les trois tests statique, dynamique et polyphonique. Pour chacun de ces trois tests, une source sonore spatialisée est jouée en utilisant les trois méthodes de spatialisation, auxquelles est ajoutée une 'ancré'. Cette ancre consiste en une version monophonique des sources sonores (utilisant l'ensemble des fréquences).

L'intensité des sources sonores est ajustée afin que le volume soit le même pour chaque méthode. Les 4 sources ainsi spatialisées sont attribuées aléatoirement à différentes lettres (A, B, C ou D), et la séquence (A, B, C, D) est jouée puis répétée.

Les auditeurs sont ensuite questionnés pour chaque méthode, et ont la consigne de répondre aléatoirement dans le cas où ils ne connaissent pas la réponse (cas de figure se présentant souvent avec l'ancré). Les auditeurs sont aussi libres d'écrire des commentaires pour chacun des tests.

10.1.1 Tests préliminaires

Avant de procéder aux tests nous intéressant pour la comparaison des différentes méthodes, un test préliminaire est effectué. Celui-ci vise les auditeurs non experts des domaines sonores, et qui ne sont pas non plus habitués au son spatial. Ce test préliminaire est également considéré comme un "échauffement", les sujets pouvant écouter la sonorisation et se concentrer/découvrir la spatialisation sonore. De plus, ce pré-test permet d'identifier les sujets non réceptifs à la spatialisation¹, et donc de les exclure du panel de réponses lors de l'exploitation des résultats.

Pour ce pré-test, 4 chants d'oiseaux² sont joués 4 fois chacun, à des positions différentes. Pour chaque exemple, le premier chant représente la référence, dont la même position est jouée aléatoirement dans l'un des trois exemples sonores suivants, l'auditeur devant le retrouver³. Aussi, une position est correcte (celle de la référence), et les chants autres sont spatialisés loin de la position du son référence, afin que cet exercice ne soit pas trop compliqué pour les personnes non expertes.

10.1.2 Tests statiques

Pour le test statique, un extrait musical est joué par un saxophone à un azimut fixe, spatialisé à l'aide de différentes méthodes. On demande alors aux sujets deux choses : la première est de localiser cette source unique, en plaçant la lettre correspondant à l'extrait concerné sur un cercle de référence (voir Fig. 10.1); la seconde, pour évaluer la qualité du son, d'évaluer chaque extrait correspondant à une méthode sur une échelle de notation de type MUSHRA (mus (2015)) (Fig. 10.2).

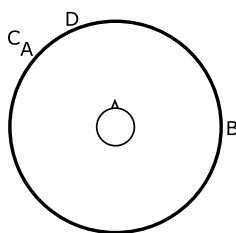


FIGURE 10.1 – Exemple du cercle de référence permettant aux auditeurs de positionner les différents extraits joués.



FIGURE 10.2 – Exemple de l'échelle de notation permettant d'attribuer un score aux différents extraits joués.

10.1.3 Tests dynamiques

L'objectif du test dynamique est de comparer les méthodes de spatialisation dans le cas d'une source mobile. Pour cela, une trajectoire circulaire est créée (avec une orientation directe) sur une musique percussive (tambours). Une fois de plus, on demande deux choses aux auditeurs : de choisir la trajectoire qu'ils trouvent la meilleure parmi les 8 trajectoires possibles (Fig. 10.3); puis, comme précédemment, d'évaluer la qualité des extraits sonores.

1. Cas par exemple de personnes pouvant souffrir d'un trouble de l'audition.
2. Les chants d'oiseaux sont intéressants du fait que ce sont des stimuli connus de tout le monde et que l'on localise fréquemment (afin de voir l'oiseau chantant). De plus, il existe une grande variété de types de chants (allant du chant aigu et sifflant du rapace, à la voix rauque des anatidés (oie, cygne), et passant par les chants des pinsons, mésanges etc.).
3. Le pré-test propose ainsi 12 localisations de chants d'oiseaux.

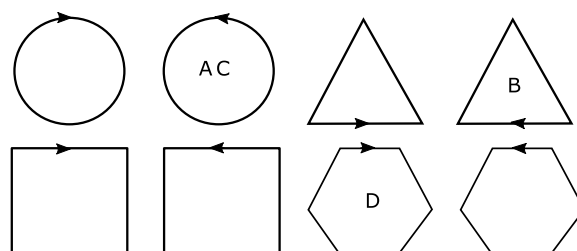


FIGURE 10.3 – Choix parmi les 8 trajectoires possibles, la réponse correcte étant le cercle (dans le sens de rotation direct).

10.1.4 Tests polyphoniques

Le troisième test est un test sur un extrait polyphonique. Une chanson musicale pop (jazz) est spatialisée, utilisant les chanteurs et instruments (batterie, basse, saxophone, guitare, claviers) en tant qu'objets audio spatiaux, c'est-à-dire des sources individuelles réparties dans l'espace avec des positions choisies par l'ingénieur ayant effectué le mix artistique. De plus, cet extrait musical a l'avantage d'être riche, présentant une voix chantée ayant des dynamiques différentes, divers instruments, et se terminant a cappella. Pour cet extrait, il est demandé aux auditeurs d'évaluer trois paramètres, tous sur des échelles de notation (Fig. 10.2) :

- la qualité sonore (comme dans les deux tests précédents),
- l'immersion,
- l'intelligibilité (clarté).

10.2 Dispositif Expérimental

Pour ces tests, il a été choisi d'utiliser une configuration similaire à celle d'expériences menées précédemment (Sec. 10.3), ceci afin de permettre d'utiliser l'ensemble des résultats. La configuration est donc une configuration octophonique, placée dans une salle de classe possédant une réverbération modérée (Fig. 10.4).



FIGURE 10.4 – Photographie du dispositif expérimental mis en place dans la salle de classe.

La Fig. 10.5 décrit le dispositif expérimental présenté dans la photographie précédente (Fig. 10.4). Les haut-parleurs LS 1 à LS 8 sont les haut-parleurs actifs, les B représentent 4 pièges (haut-parleurs inactifs, pour augmenter artificiellement la complexité de configuration¹), et 9 sièges sont placés au milieu de cette configuration (S 1 à S 9).

1. Il était impossible de masquer les haut-parleurs, et la vue des haut-parleurs peut avoir une influence sur les réponses des auditeurs.

10.3. RÉSULTATS DES TESTS PERCEPTIFS DE COMPARAISON DES MÉTHODES DE SPATIALISATION

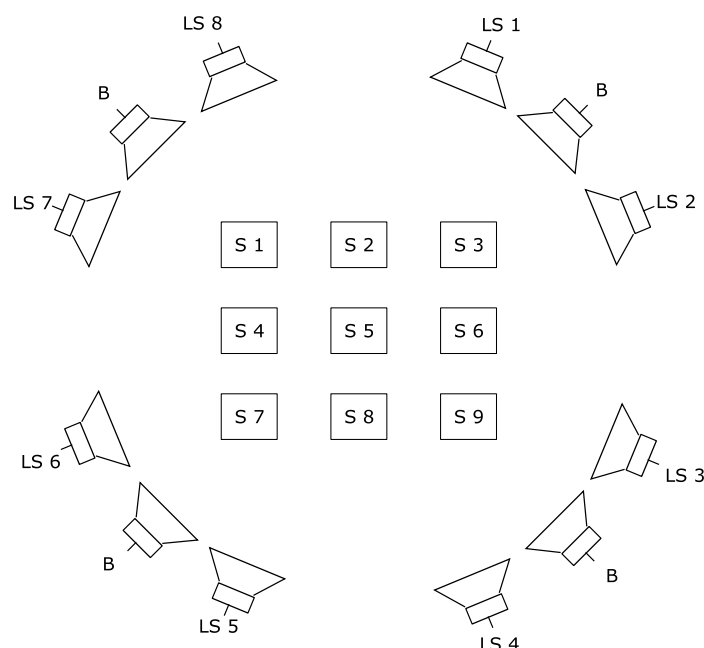


FIGURE 10.5 – *Dispositif expérimental (utilisé en 2019), disposant de 8 haut-parleurs actifs, (LS 1 à LS 8), 4 inactifs utilisés comme pièges (B), et en leur centre 9 places assises (S1 à S9).*

Le panel d’auditeurs est représenté par 32 personnes¹ (l’expérience est réalisée 4 fois), composé presque exclusivement d’amateurs ou de néophytes en musique (seulement 1 professionnel musicien dans ces tests), avec principalement des étudiants et des personnels universitaires. Le pré-test exclut uniquement 2 personnes (personnes n’ayant pas réussi à trouver la référence pour au minimum la moitié des 4 passages). Aussi, le panel final est composé de 30 personnes (8 femmes et 22 hommes), âgés de 17 à 49 ans, dont 23 de moins de 25 ans.

10.3 Résultats des tests perceptifs de comparaison des méthodes de spatialisation

En 2015, lors du festival Électrocution pour la musique électro-acoustique se déroulant à Brest, les expérimentations précédentes ont été menées par un groupe d’étudiants de Master (Aude Besnard, Anaëlle Marsollier, Clément Ruffini, and Aurianne Skybyk) avec une configuration octophonique placée dans une salle de concert assez réverbérante (ancienne usine de béton), avec un public constitué de compositeurs, d’ingénieurs son, et d’autres personnes avec une majorité de professionnels de la musique habitués au son spatial (29 sujets au total). Cette configuration était exactement celle utilisée pour la diffusion des concerts du festival. En tout, cette expérience a été répétée en 3 occasions, lors du festival Électrocution pour la musique, par les étudiants de master; puis au cours de la thèse dans un premier temps dans une salle de classe, puis lors des JIM.

La conclusion de ces expériences montrait à l’époque que la méthode STAR avait un grand sweet spot, obtenait de meilleurs résultats pour le test dynamique (VBAP donnant les plus mauvais résultats, avec de nombreux hexagones choisis au lieu de cercles, entraînant des sauts entre haut-parleurs), et était préféré pour le test polyphonique, bien que le timbre sonore ait parfois été jugé comme ayant une "qualité nasale". Ainsi, les résultats étaient assez prometteurs, le problème de timbre ayant été résolu, celui-ci provenant de grandes valeurs des coefficients pour les hautes fréquences et produisant un effet de filtre passe-haut.

1. Un second panel d’auditeurs a également été soumis aux mêmes tests au cours des JIM 2019 avec une configuration et des résultats similaires.

10.3.1 Tests statiques

La Fig. 10.6 propose les résultats des tests statiques, où les auditeurs ont dû localiser une source sonore fixe. Il apparaît clairement que le score de l'ancre (mono) est très mauvais, ce qui est normal, celle-ci étant jouée sur l'ensemble des haut-parleurs ; ce qui explique aussi pourquoi il est représenté sur quasiment toute la plage possible d'azimuts. Pour les trois méthodes, la moyenne est proche de la position réelle de la source virtuelle. Cependant, il apparaît clairement que les méthodes VBAP et STAR obtiennent de meilleurs scores que la méthode HOA (qui utilise toutes les enceintes, ce qui peut être un inconvénient pour une petite pièce, certains auditeurs assis relativement loin du sweet spot et/ou trop près d'une enceinte). Enfin, on observe que la méthode STAR présente la meilleure valeur moyenne avec cependant un écart type plus grand que celui de la méthode VBAP.

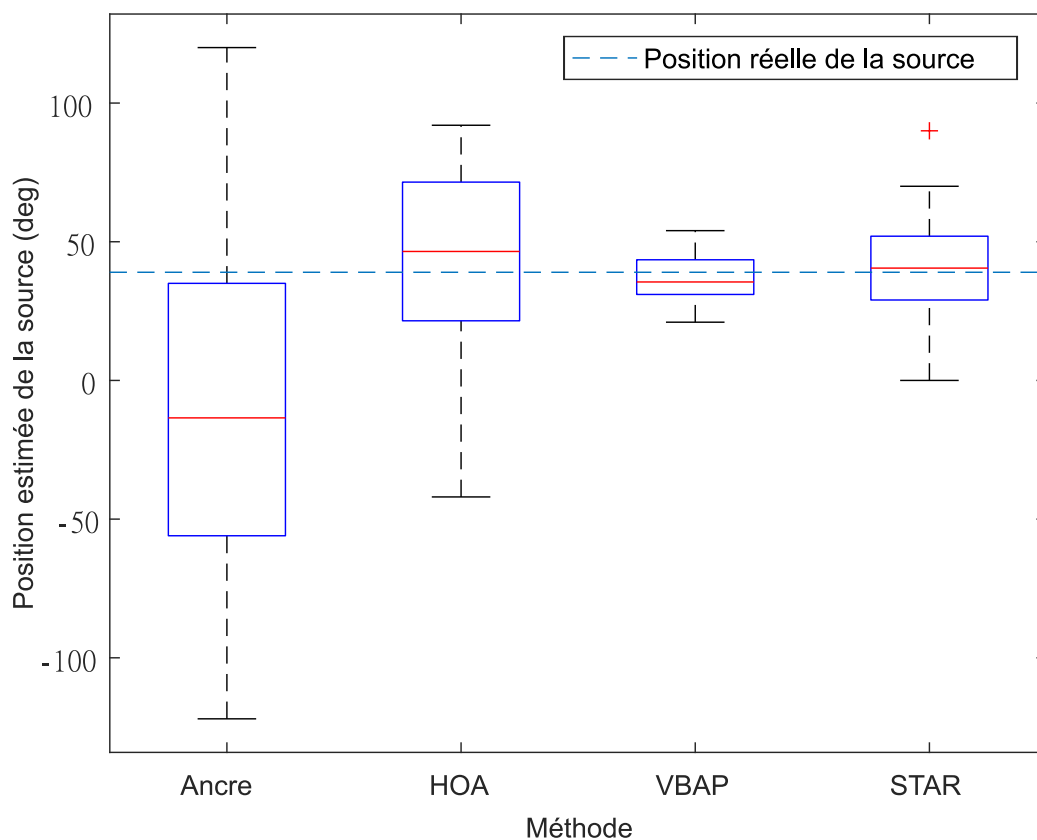


FIGURE 10.6 – Résultats des tests statiques. Position perçue par les auditeurs pour les différentes méthodes, et pour un azimut $\theta = 45^\circ$.

Une Analyse en Composantes Principales (PCA) menée sur les résultats des tests statiques, montre que les positions des auditeurs et leur perception de la position de la source sont corrélées. Plus précisément, l'auditeur a tendance à percevoir le son en direction du haut-parleur le plus proche, phénomène pas surprenant mais problématique.

Concernant la qualité perçue, les résultats (Fig. 10.7) sont assez surprenants, toutes les méthodes montrant des résultats comparables, avec une moyenne moyenne et un grand écart-type. Cela pourrait être un problème pour l'ancre, qui est mono mais possède une bande passante complète, le son étant donc probablement de qualité trop élevée pour une ancre. Une autre hypothèse est que les sujets ne savent ou ne peuvent pas évaluer la qualité spatiale.

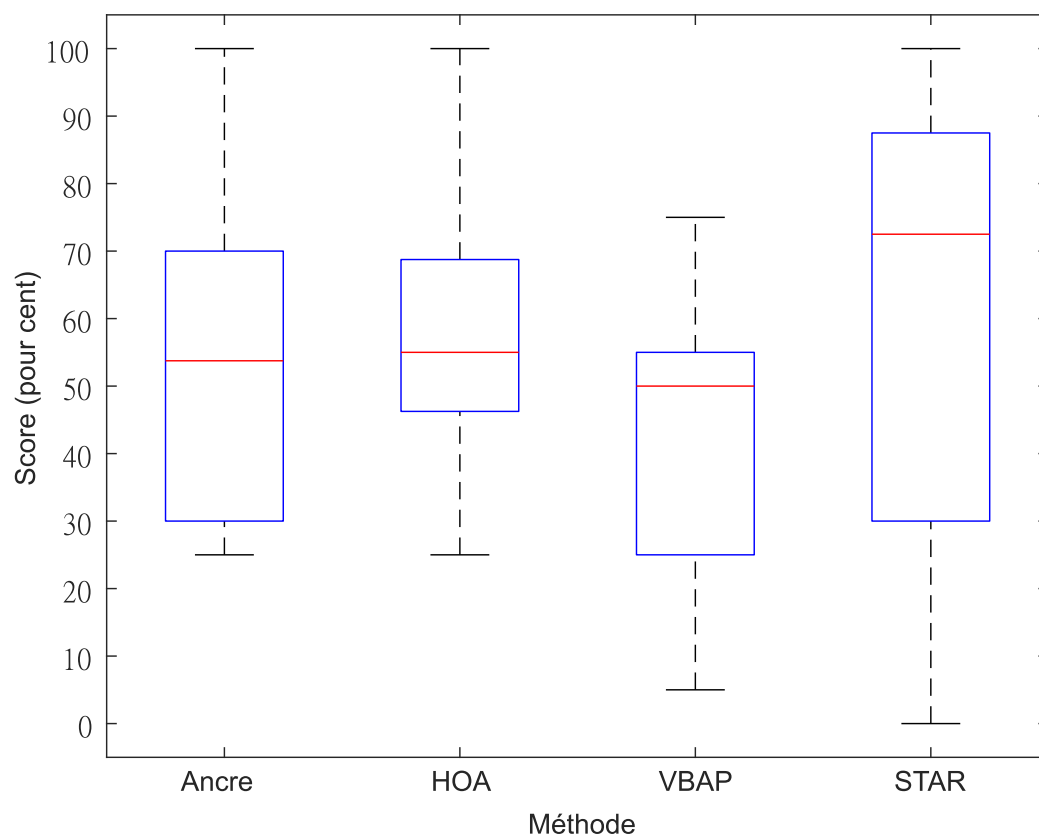


FIGURE 10.7 – Résultats des tests statiques pour la qualité estimée.

10.3.2 Tests dynamiques

La Fig. 10.8 montre les résultats des tests dynamiques, où les auditeurs ont dû reconnaître la trajectoire décrite par une source mobile (au cours du temps). L'ancre montre un comportement aléatoire, phénomène normal, n'ayant pas de trajectoire rendue en version monophonique, alors que les réponses sont toujours exigées. Pour toutes les autres méthodes, il y a une hésitation entre la trajectoire circulaire (correcte) et hexagonale (incorrecte). Dans l'ensemble, la méthode HOA semble mieux fonctionner, suivie des méthodes VBAP et STAR.

10.3. RÉSULTATS DES TESTS PERCEPTIFS DE COMPARAISON DES MÉTHODES DE SPATIALISATION

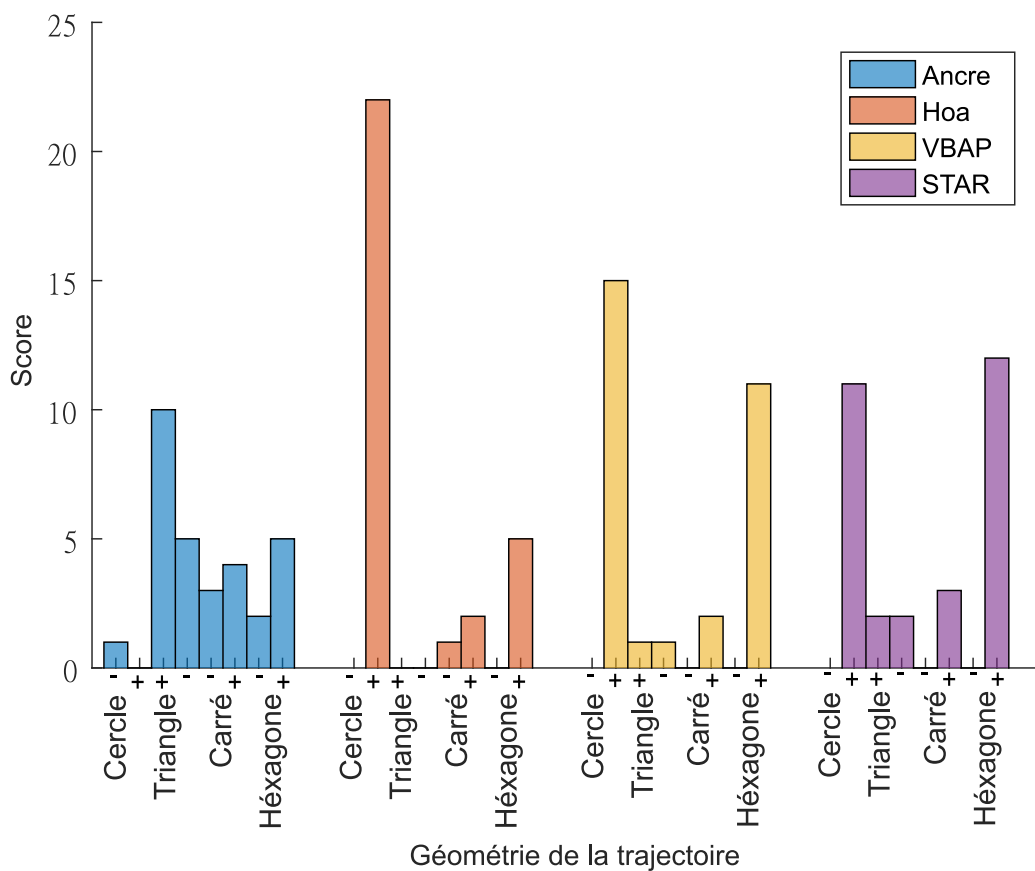


FIGURE 10.8 – Résultats des tests dynamiques pour les différentes méthodes. Le score de la trajectoire des sources perçues est ici représenté en pour cent - Tests réalisés en 2019.

Ces résultats sont surprenants, les tests réalisés en 2015 (Fig. 10.9) mettant la méthode STAR en avant avec un net avantage.

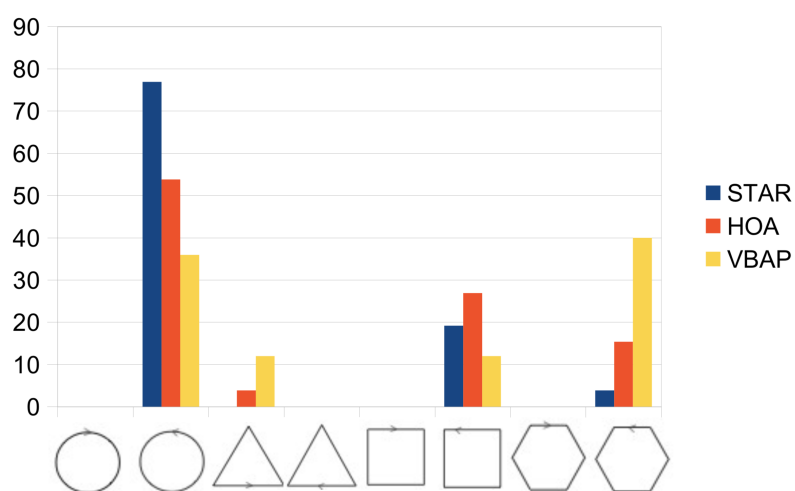


FIGURE 10.9 – Résultats des tests dynamiques pour les différentes méthodes. Le score trajectoire des sources perçues est ici représenté en pour cent - Tests réalisés en 2015.

10.3. RÉSULTATS DES TESTS PERCEPTIFS DE COMPARAISON DES MÉTHODES DE SPATIALISATION

Hormis les caractéristiques de la salle et la qualification du public, le seul changement entre 2015 et 2019 est le changement de référence de l'azimut, placé au centre des haut-parleurs, permettant entre autre de supprimer la "qualité nasale", et qui contribue peut-être à cette diminution de performance.

Concernant la qualité perçue, les résultats (Fig. 10.10) montrent que l'ancre a de bien moins bons résultats (l'ancre ne bougeant pas), et une quasi-égalité des trois méthodes.

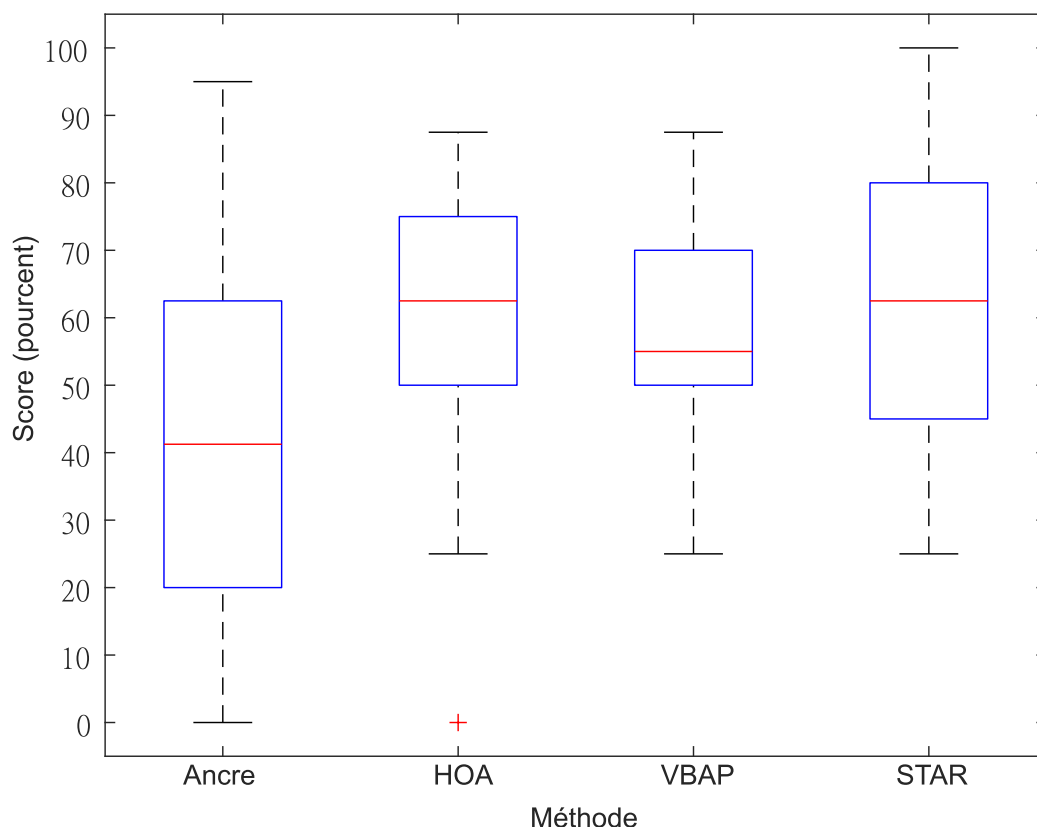


FIGURE 10.10 – Test dynamique, évaluation de la qualité des méthodes.

10.3.3 Tests polyphoniques

La Fig. 10.11 montre la qualité perçue dans le cas des tests polyphoniques. Cette fois, le score de l'ancre est statistiquement plus faible, mais les trois méthodes sont jugées donner des résultats également bons. Les résultats sont cohérents entre 2015 et 2019, même si en 2015 la méthode STAR fut légèrement préférée, mais de manière non statistiquement significative.

10.3. RÉSULTATS DES TESTS PERCEPTIFS DE COMPARAISON DES MÉTHODES DE SPATIALISATION

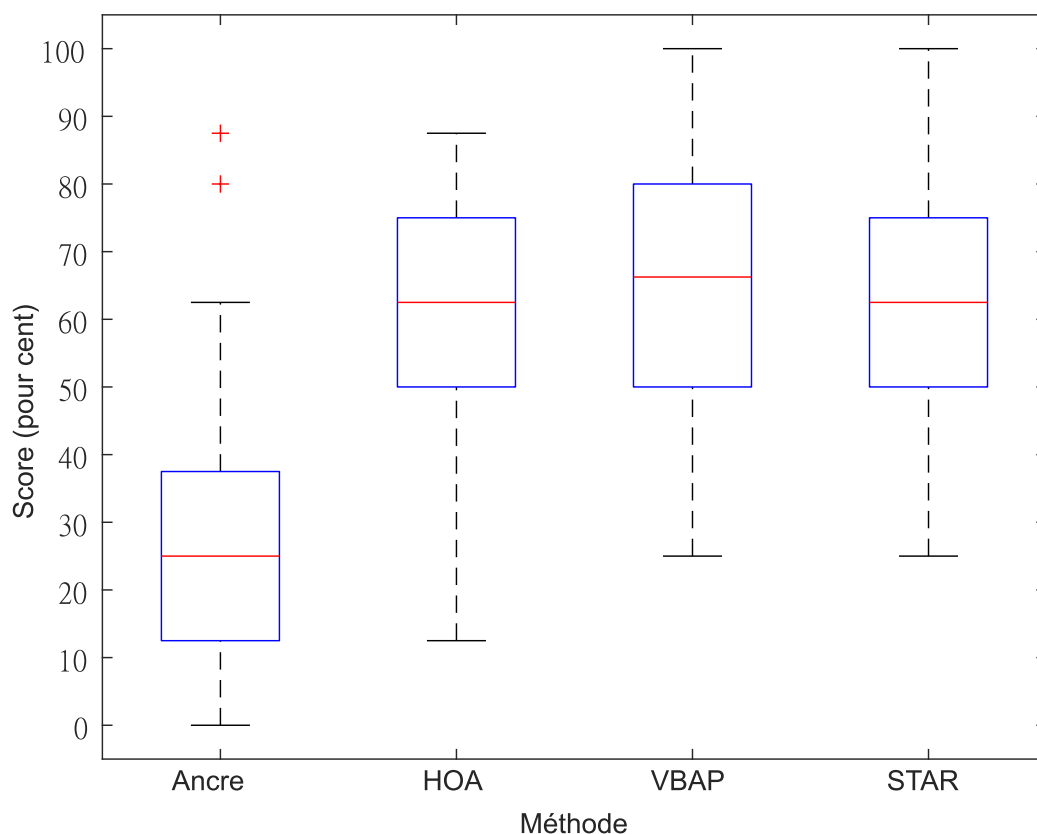


FIGURE 10.11 – *Tests polyphoniques, qualité du son ressentie.*

Contrairement aux tests de 2015, il a été demandé en 2019 l'immersion subjective ainsi que l'intelligibilité. La Fig. 10.12 montre les résultats de l'immersion perçue, qui est très similaire aux résultats sur la qualité perçue. Concernant l'intelligibilité (Fig. 10.13), la méthode STAR semble avoir quelques problèmes contrairement aux méthodes HOA et VBAP. Les coefficients de la méthode STAR étant spectraux et complexes, ils modifient également la phase dépendante de la fréquence, ce phénomène permettant de lisser les trajectoires peut également modifier le timbre des sources sonores.

10.3. RÉSULTATS DES TESTS PERCEPTIFS DE COMPARAISON DES MÉTHODES DE SPATIALISATION

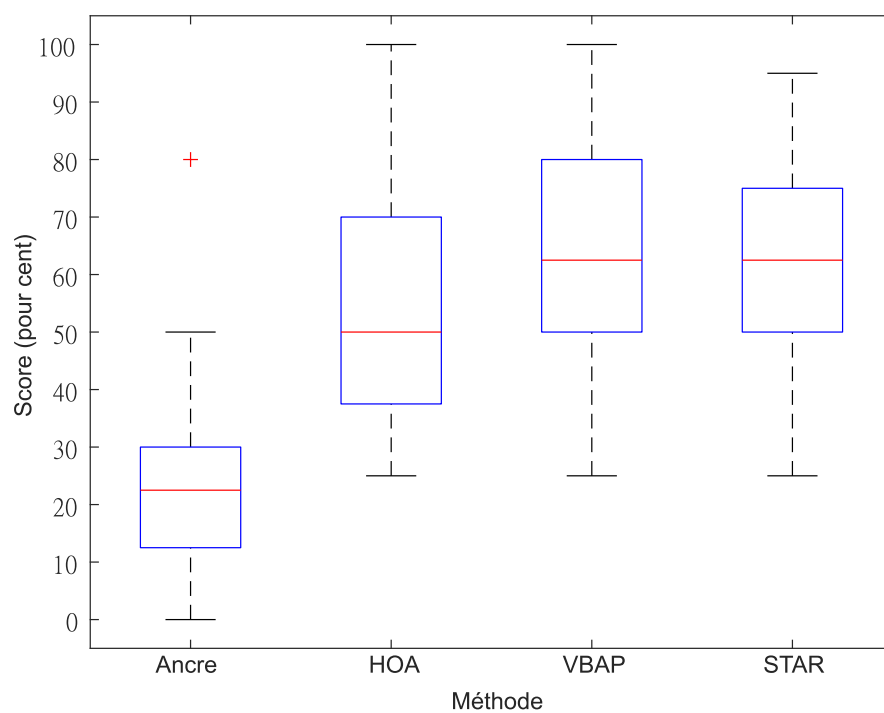


FIGURE 10.12 – Tests polyphoniques : immersion ressentie.

10.3. RÉSULTATS DES TESTS PERCEPTIFS DE COMPARAISON DES MÉTHODES DE SPATIALISATION

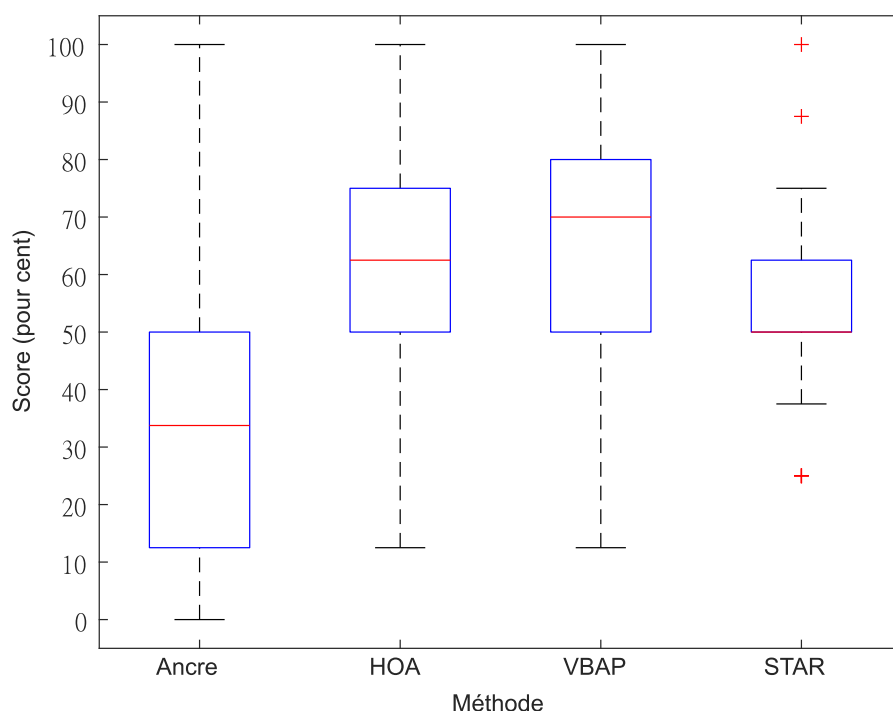


FIGURE 10.13 – Tests polyphoniques : intelligibilité.

– Conclusion –

Ces tests comparatifs subjectifs de méthodes menés de manière subjective permettent de montrer que les trois méthodes mises en concurrence sont équivalentes, chacune d'entre elles pouvant ressortir légèrement favorite aux deux autres en fonction de ce qui est testé. La méthode STAR est donc au minimum concurrentielle et avec des performances similaires aux méthodes VBAP et HOA.

Ces tests sont cependant menés uniquement sur l'azimut, et il conviendrait de reconduire des tests similaires avec l'élévation et la distance, implémentées depuis et présentés précédemment.

Les tests subjectifs de validation de la méthode STAR comparent trois grandes méthodes de spatialisation présentées précédemment. Les premiers tests conduits par Mouba and Marchand (2006), ont été reproduits avec la méthode STAR, dont les résultats sont présentés dans la conférence DAFx 2019 (Méaux and Marchand (2019)), ainsi que dans le journal de l'AES (Méaux and Marchand (2021a)).

Chapitre 11

Évaluation subjective pour la validation de la distance et de l'élévation, ainsi que leur étude d'indépendance

– Introduction –

Le chapitre précédent présente les tests subjectifs permettant de comparer les performances entre les différentes méthodes. Des tests similaires pour la méthode STAR en 3D seraient extrêmement utiles et font partie des perspectives possibles à ce travail de thèse.

Il est cependant nécessaire, avant d'effectuer ces tests, de confirmer que la spatialisation de l'élévation et de la distance fonctionnent correctement, ce qui a pu être effectué objectivement pour l'azimut, et n'est pas possible pour les composantes élévation et distance. C'est pourquoi un test subjectif a été mis au point afin de valider que la méthode STAR donne bien la sensation de distance et d'élévation.

Ce chapitre propose également l'étude d'indépendance entre l'élévation et la distance, effectuée au cours des mêmes tests. En effet, il est nécessaire dans les tests de mettre des 'pièges', la démarche suivie dans cette thèse utilise ces pièges afin de valider cette indépendance.

11.1 Protocole expérimental

Contrairement à l'azimut, pour lequel la méthode est validée à la fois subjectivement et objectivement ; il n'est actuellement pas possible de mesurer efficacement la distance et l'élévation de manière objective, même si des travaux sur ce sujet ont été menés (Iida et al. (2007)). Pour cette raison, les tests de validation pour l'élévation et la distance sont effectués de manière subjective en utilisant le protocole décrit dans cette partie.

Les expérimentations sont menées dans deux salles, à réverbération modérée (un salon et une salle de classe) afin d'obtenir des conditions réelles conformes aux conditions souhaitées pour l'utilisation de la méthode STAR.

Pour ce test, une seule enceinte Genelec 8030 est placée à 1 m de l'auditeur. Cinquante-cinq paires de bruits blancs, de durée 1 s et séparées de 0,5 s, sont jouées. Ces couples de bruits blancs sont spatialisés selon la méthode STAR en altitude et en distance. Le premier bruit est choisi à une distance/élévation aléatoire (plage 1 - 9 m et 0 - 80°), et le second est spatialisé à $\pm range$, (où *range* est également distribué selon les distances/élévations possibles) sur l'intervalle 1 - 9 m et 5 - 80°. L'ordre de passage des plages ainsi que le signe (plus ou moins élevé/distant) sont choisis de manière aléatoire.

11.2. RÉSULTATS DES TESTS DE VALIDATION DE LA MÉTHODE STAR POUR L'ÉLEVATION ET LA DISTANCE

Chaque auditeur (les yeux fermés) doit choisir si le second son est plus ou moins éloigné/élevé que le premier, la réponse est transcrite sur papier par un tiers, pour chaque réponse apportée. Avec cette méthode, l'auditeur est plus concentré que s'il devait garder les yeux ouverts afin de noter les réponses entre chaque paire de sons.

Pour chaque ensemble de 55 paires de sons, les cinq premières paires sont jouées afin de permettre une compréhension et une appréhension des tests, et ne sont pas prises en compte dans nos résultats. Les cinquante autres paires sont divisées en deux parties. La première partie permet la validation de la méthode et permet de proposer une comparaison entre deux altitudes/distances. La seconde partie a pour but de tromper l'auditeur en lui demandant de localiser le second son de la même manière, sachant que ce second exemple sonore est en réalité spatialisé à la même élévation/distance et seul le deuxième paramètre varie (celui qui n'est pas demandé)¹. Par exemple, il est demandé de localiser si le second son est plus ou moins distant, alors que les deux sources sonores sont spatialisées pour la même distance, et que seule l'élévation est changée. En utilisant ce protocole, les auditeurs ne devraient pas savoir quoi répondre. Les données devraient donc être décorréliées, montrant ainsi l'indépendance entre l'altitude et la distance.

Le panel d'auditeurs pour ce test est composé de 20 personnes, essentiellement d'amateurs ou de néophytes de la musique, 11 femmes et 9 hommes, âgés de 22 à 49 ans, dont 8 ont moins de 30 ans.

11.2 Résultats des tests de validation de la méthode STAR pour l'élévation et la distance

Les bonnes détections de la variation de l'altitude et de la distance pour les couples de bruits blancs spatialisés en distance et en altitude sont représentées sur les deux premiers tracés de la Fig. 11.1, et en trait plein vert sur le graphique Fig. 11.2. Les résultats d'ensemble de la Fig. 11.1 montrent une précision moyenne autour de 80 %. C'est un bon résultat, même en considérant les valeurs extrêmes qui varient considérablement, cela s'expliquant par le fait que certaines personnes sont moins sensibles et donnent alors des réponses au hasard².

1. L'auditeur ne sachant pas que le second paramètre varie, il évalue toujours le premier paramètre. Le second paramètre ne devant pas influencer sur le premier dans le cas d'une indépendance des paramètres.

2. Le positionnement des sources plus proches étant plus ressemblantes.

11.2. RÉSULTATS DES TESTS DE VALIDATION DE LA MÉTHODE STAR POUR L'ÉLEVATION ET LA DISTANCE

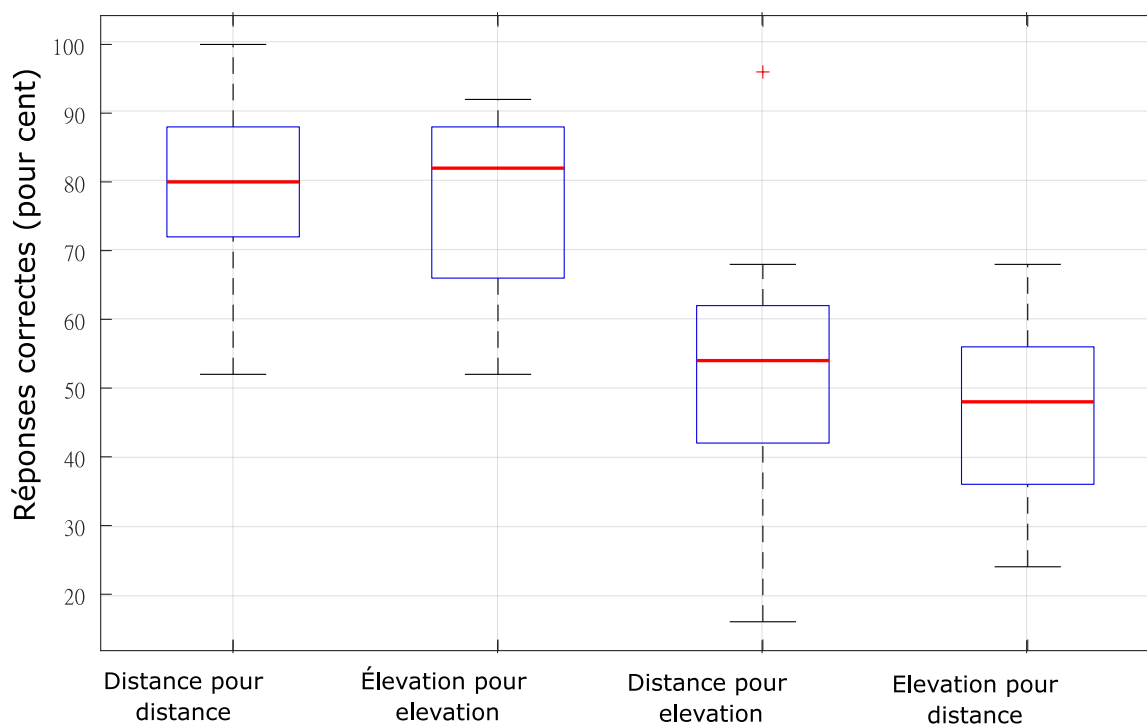


FIGURE 11.1 – Tests de validation de la méthode STAR pour l'élévation et la distance (les deux premières boîtes) et d'indépendance entre ces paramètres (les deux secondes boîtes).

Le second résultat intéressant est donné par la figure 11.2 qui révèle que plus la différence entre les deux spatialisations en élévation/distance est faible et plus l'erreur¹ sera grande. C'est un résultat cohérent avec le fonctionnement de l'audition humaine.

1. Ici, l'erreur est assimilée aux mauvaises réponses (ou à un mauvais pourcentage de bonnes réponses).

11.2. RÉSULTATS DES TESTS DE VALIDATION DE LA MÉTHODE STAR POUR L'ÉLEVATION ET LA DISTANCE

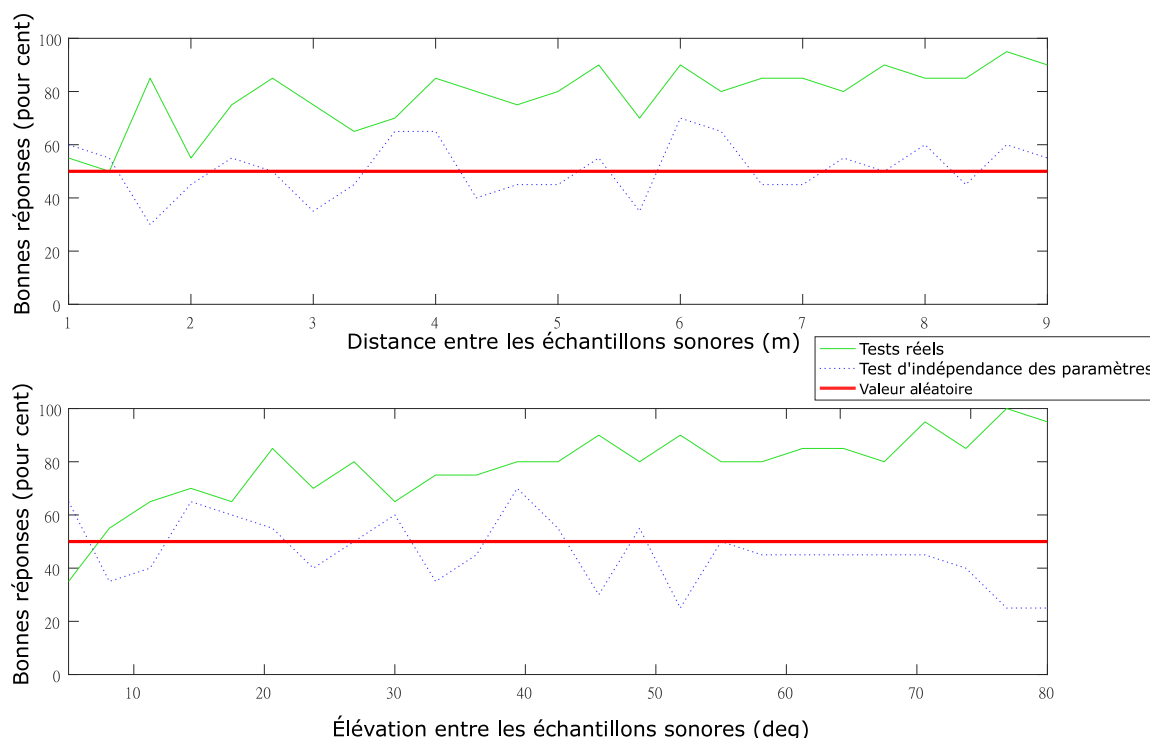


FIGURE 11.2 – Résultats de la validation de la méthode pour les paramètres distance (plage 0 - 9 m) et élévation (plage 0 - 80 °) et leur indépendance. Plus la distance/élévation entre les deux sources sont importantes et plus le pourcentage de bonnes réponses est grand (ligne verte). Lorsque le mauvais paramètre est évalué (ligne pointillée bleue), les résultats sont proches de l'aléatoire (ligne rouge).

Tous ces résultats sont cohérents avec la philosophie de la méthode STAR, procurant une sensation d'élévation et de distance, tout en étant facile de mise en œuvre (à l'aide d'une seule couronne de haut-parleurs) et en respectant les critères de l'audition humaine (importance de l'azimut, et sensation de distance et d'élévation).

De plus, comme présenté dans le chapitre 6, pour une configuration sur plusieurs couronnes de haut-parleurs, un couplage des méthodes VBAP et de STAR est alors utilisé pour la spatialisation de l'élévation. Des tests informels ont été menés (Fig. 11.3) en faisant varier le son d'une élévation $\phi = 0^\circ$ à $\phi = 90^\circ$. Ces tests permettent d'affirmer qu'il n'y a pas de phénomène perturbant audible. Il est cependant évidemment nécessaire d'effectuer des tests de comparaison de méthodes en 3D afin de justifier rigoureusement ces observations.

11.3. TEST D'INDÉPENDANCE SUBJECTIF POUR L'ÉLEVATION ET LA DISTANCE, COMPLÉMENTAIRE DE L'ÉTUDE OBJECTIVE DE LEURS INDÉPENDANCES AVEC L'AZIMUT



FIGURE 11.3 – Montage expérimental pour la vérification de la spatialisation de l'élévation avec les méthodes VBAP et STAR couplées lors de l'utilisation de configurations avec plusieurs couronnes de haut-parleurs. (Ici chaque couronne est représentée par un seul haut-parleur).

11.3 Test d'indépendance subjectif pour l'élévation et la distance, complémentaire de l'étude objective de leurs indépendances avec l'azimut

L'indépendance entre l'altitude et la distance utilise le même test que celui permettant d'évaluer la méthode STAR pour l'altitude et la distance. Les résultats sont présentés en Sec. 11.2. On s'attend à ce que la demande de détection de variation pour l'élévation en faisant varier la distance (et vice versa) soit décorrélée, montrant ainsi l'indépendance entre l'élévation et la distance.

Les résultats attendus sont bien les résultats obtenus, comme le montrent les deux derniers tracés de la Fig. 11.1, ainsi que la ligne pointillée bleue sur la figure 11.2. En effet, les résultats obtenus sont aléatoires comme attendus. Ceci nous permet de présumer que les paramètres sont bien indépendants.

11.3. TEST D'INDÉPENDANCE SUBJECTIF POUR L'ÉLEVATION ET LA DISTANCE, COMPLÉMENTAIRE DE L'ÉTUDE OBJECTIVE DE LEURS INDÉPENDANCES AVEC L'AZIMUT

– Conclusion –

Ce chapitre permet ainsi de compléter les différents tests effectués sur l'azimut en prenant en compte la distance ainsi que l'élévation.

Il y est montré que les sensations de distance et d'élévation sont bien respectées. De plus, au cours de ces tests, l'indépendance de ces deux dimensions est aussi mise en avant. Ce chapitre vient alors clôturer les différents tests de validation de la méthode STAR, en permettant d'assurer ses performances pour l'azimut, mais aussi que les sensations d'élévation et de distance sont bien présentes ; et ceci en respectant le domaine d'application souhaité pour la méthode.

Afin de finaliser complètement l'évaluation de la méthode, effectuer des tests subjectifs comparant la spatialisation en 3D des différentes méthodes pour l'élévation et la distance serait une bonne perspective pour compléter ces tests montrant que les sensations d'élévation et de distance sont respectées.

La distance et l'élévation introduites dans un premier article (Méaux and Marchand (2021a)), ainsi que leur indépendance, n'avaient alors pas pu être testées. Les tests présentés ici ont permis de valider ces deux paramètres et ainsi de fournir un article dans le journal de l'AES pour l'extension de la méthode STAR en 3D (Méaux and Marchand (2021b)), article en cours de soumission.

Cinquième partie

Conclusion et perspectives

CONCLUSION ET PERSPECTIVES

Cette thèse présente donc l'ensemble du travail mené afin d'obtenir une méthode de spatialisation et de localisation perceptive 3D. La synthèse de cette thèse est que l'objectif principal, à savoir obtenir une méthode de spatialisation complète, est obtenu. Les différents procédés de spatialisation proposés ont été testés et présentent de bons résultats (au moins aussi bons que les méthodes existantes). Il ressort néanmoins, notamment au cours des tests perceptifs, que la méthode semble plus performante en conditions réelles (alors qu'elle obtient de légèrement moins bons résultats en situations calibrées), ce qui est en accord avec les objectifs premiers de son développement. De plus, la méthode présente un intérêt unique de retranscrire une sensation d'élévation, y compris sur des systèmes possédant une unique couronne de haut-parleurs.

En plus de cette spatialisation maintenant efficiente, un travail sur la localisation perceptive a été mené ; une méthode pour la localisation en azimuth fonctionnant bien est notamment ici proposée et utilisée comme moyen de mesure objective. Une seconde méthode de localisation en élévation montre également des résultats encourageants.

Comme dans tout travail, de nombreuses améliorations et pistes sont à explorer. Concernant la spatialisation sonore 3D, des tests subjectifs sur des simulations 3D (similaires à ceux conduits pour l'azimuth) seraient intéressants ; et ceci en confrontant plusieurs types d'installations (plusieurs couronnes, une seule couronne, ou un dôme avec des haut-parleurs équi-répartis). Un ajout d'un facteur d'élargissement de source, ainsi que de réverbérations, sont également des pistes de développement intéressantes.

Pour ce qui est de la localisation perceptive, la distance semble un problème pour le moment impossible à résoudre (en respectant la philosophie de la méthode). Cependant, la localisation en élévation offre, quant à elle, des résultats encourageants. Une prise en compte d'un plus grand nombre de candidats, une amélioration de la détection de pics, une sélection d'HRTFs dans la base en fonction de données anthropomorphiques sont certaines pistes à creuser.

Pour conclure, ce travail de thèse produit un spatialisateur sonore 3D perceptif, avec une localisation pour auto-calibration possible en azimuth, le tout avec de bonnes performances en milieu réel et utilisant des configurations faciles de mise en œuvre. Des investigations supplémentaires ainsi que des recherches pourraient permettre d'obtenir un spatialisateur encore plus complet et autonome.

Sixième partie

Annexes

11.4 Liste des abréviations

AES	Audio Engineering Society
AIR	Aachen Impulse Response
ARI	Acoustics Research Institute
ASIO	Audio Stream Input Output
BRIR	Binaural Room Impulses Response
CIPIC	Center for Image Processing and Integrated Computing
CIT	Chiba Institute of Technology
DAFx	Digital Audio Effects
DRR	Direct-to-Reverberant Ratio (Rapport champ direct à champ réverbéré)
FFT	Fast Fourier Transform (Transformée de Fourier Rapide)
GNU	GNU's Not UNIX
GUI	Graphical User Interface
GIF	Graphics Interchange Format
HMS	Head Measurement System
iFFT	inverse Fast Fourier Transform (inverse de la Transformée de Fourier Rapide)
IHM	Interface Homme Machine
HOA	High Order Ambisonic
HRIR	Head Related Impulse Response
HRTF	Head Related Transfer Function (Fonction de transfert relative à la tête)
BRTF	Binaural Room Transfer Function
ILD	Interaural Level Difference (Différence interaurale de niveau ou différence interaurale d'intensité).
IRCAM	Institut de Recherche et Coordination Acoustique/Musique
ITD	Interaural Time Difference (Différence interaurale de temps)
JIM	Journées d'Informatique Musicale
JPEG	Joint Photographic Experts Group
KEMAR	Knowles Electronics Mannequin for Acoustics Research
LaBRI	Laboratoire Bordelais de Recherche en Informatique
L3i	Laboratoire Informatique, Image et Interaction
MADI	Multichannel Audio Digital Interface
MATLAB	MATrix LABoratory
MESM	Multiple Exponential Sweep Method
MIT	Massachusetts Institute of Technology
MLS	Maximum Length Sequence
MUSHRA	MUltiple Stimulus with Hidden Reference and Anchors
OSSIA	Open Software System for Interactive Applications
PCA	Principal Component Analysis (Analyse en Composantes Principales)
PDF	Portable Document Format
RIEC	Research Institute of Electrical Communication
ROP	Realistic Optimus Pro
SCRIME	Studio de Création et de Recherche en Informatique et Musiques Expérimentales
SPL	Sound Pressure Level (Niveau de pression acoustique)
STAR	Synthetic Transaural Audio Rendering
STFT	Short-Time Fourier Transform (Transformée de Fourier à court terme)
SVG	Scalable Vector Graphics
VBAP	Vector Base Amplitude Panning
WFS	Wave Field Synthesis (Hologophonie)

11.5 Symboles mathématiques utilisés

Beaucoup des symboles ci-dessous peuvent être trouvés en majuscules ou en minuscules. Dans ce cas, ces symboles en majuscules correspondent au signal dans le domaine spectral et ceux en minuscules au domaine

11.6. GUI-STAR, L'IHM D'ÉTUDE DE LA MÉTHODE STAR ET DE CRÉATION D'EXEMPLES SONORES

temporel.

α	Facteur d'échelle pour les ILDs
β	Facteur d'échelle pour les ITDs
$\Delta_a(f)$	Coefficient de simplification pour les ILDs
$\Delta_\phi(f)$	Coefficient de simplification pour les ITDs
θ	Angle pour l'azimut
ϕ	Angle pour l'élévation
c	Célérité du son dans l'air (environ 335 m/s)
d	Distance
F_e	Fréquence d'échantillonnage
g	Facteur de gain (VBAP)
H	Désigne un chemin acoustique
H_L	Chemin acoustique arrivant de la source à l'oreille gauche
H_{LL}	Chemin acoustique entre le haut-parleur gauche et l'oreille gauche
H_{LR}	Chemin acoustique entre le haut-parleur gauche et l'oreille droite
H_R	Chemin acoustique arrivant de la source à l'oreille droite
H_{RL}	Chemin acoustique entre le haut-parleur droit et l'oreille gauche
H_{RR}	Chemin acoustique entre le haut-parleur droit et l'oreille droite
l	Désigne le côté gauche (left)
L	Signal audio enregistré par l'oreille gauche (spectre)
$\vec{l}$	Vecteur de direction des haut-parleurs (VBAP)
LS	Haut-parleurs
N	Noeuds
n, m	Désignent un nombre entier positif
o	Ordre
P	Pics
p	Vecteur de direction de la source virtuelle (VBAP)
r	Désigne le côté droit (right)
r	Rayon de la tête
R	Signal audio enregistré par l'oreille droite (spectre)
s	Source sonore
s	Sujet
$s_{l,r}$	Désigne la source gauche ou droite d'un signal binaural
s_n	Signal audio joué par le haut-parleur n
w	Fenêtre (utilisé pour le filtre de Hann dans STAR)

11.6 GUI-STAR, l'IHM d'étude de la méthode STAR et de création d'exemples sonores

GUI-STAR est le nom du logiciel d'étude de la méthode STAR. En effet, l'étude de la méthode STAR est complexe et a nécessité de pouvoir observer de multiples configurations. Il a donc été choisi de créer une IHM sur MatlabTM permettant de regrouper l'ensemble des scripts, exporter les images intéressantes, choisir les angles que l'on souhaite observer, les configurations, les sujets observés *etc.* Le nombre de possibilités d'observations est donc extrêmement important et l'IHM créé prend en compte toutes les mises à jour et adaptations de la méthode.

Les Fig. 11.4, 11.5 11.6, 11.7 et 11.8 donnent des exemples de cette interface. Ces figures montrent un tracé du déterminant, la visualisation des coefficients de la méthode VBAP comparés à la méthode STAR et des HRTFs de la base CIPIC. Ces exemples montrent certaines fonctionnalités telles que les informations associées aux tracés (importantes afin d'indiquer les intérêts de chaque visualisation de l'interface), le choix de l'échelle (cartésienne ou logarithmique), l'application de filtres, le choix de l'azimut, ou encore le choix du sujet considéré (pour les HRTFs). Ces exemples ne sont que des échantillons de ce que propose cette interface d'étude de la méthode STAR.

11.6. GUI-STAR, L'IHM D'ÉTUDE DE LA MÉTHODE STAR ET DE CRÉATION D'EXEMPLES SONORES

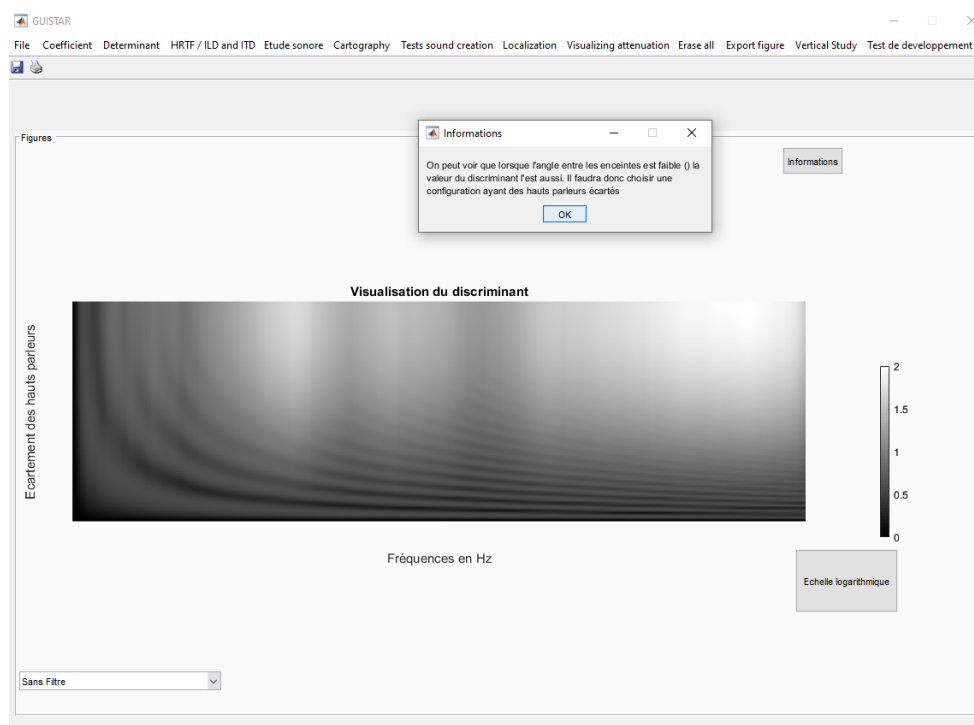


FIGURE 11.4 – Exemple de visualisation globale de l'interface GUI-STAR. Ici la visualisation du déterminant, avec les informations liées à la figure, le choix de l'échelle, ainsi que le filtre à appliquer.

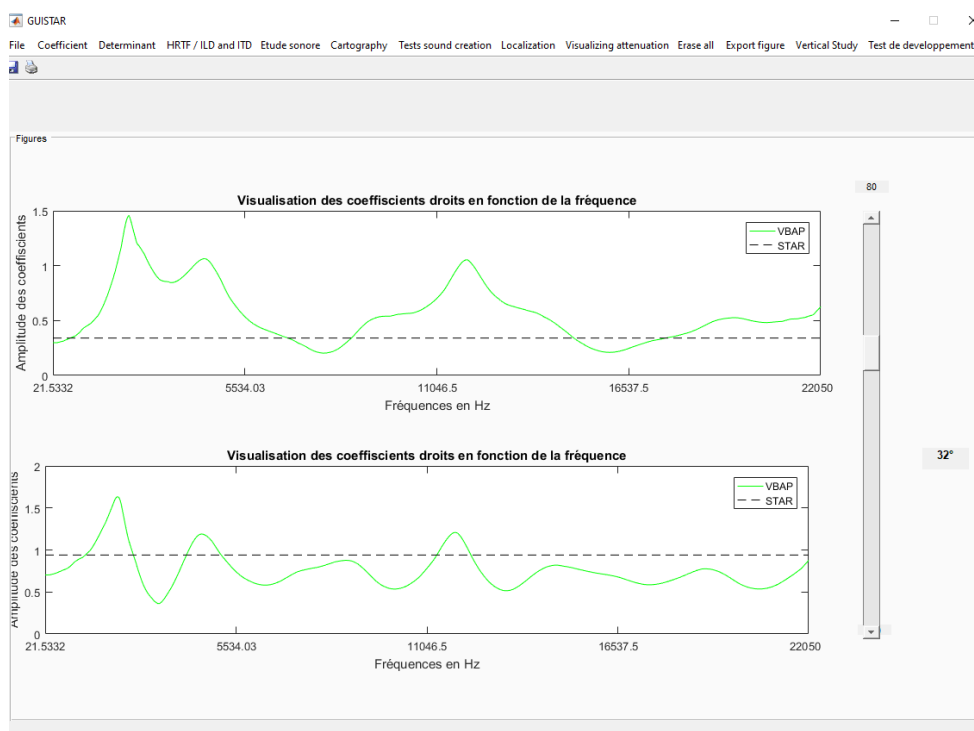


FIGURE 11.5 – Exemple de visualisation globale de l'interface GUI-STAR. Ici la comparaison des coefficients des méthodes STAR et VBAP en fonction de l'azimut modifiable via le slider sur le côté droit.

11.6. GUI-STAR, L'IHM D'ÉTUDE DE LA MÉTHODE STAR ET DE CRÉATION D'EXEMPLES SONORES

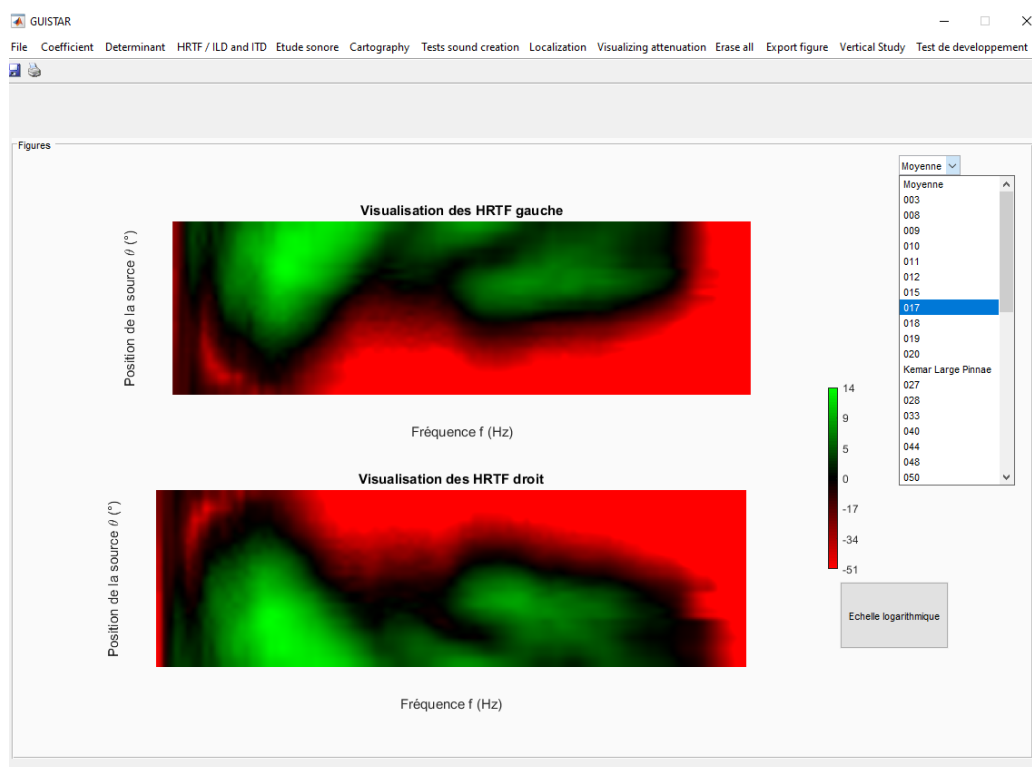


FIGURE 11.6 – Exemple de visualisation globale de l'interface GUI-STAR. Ici une visualisation des HRTFs avec le choix de l'échelle, ainsi que de n'importe quel sujet de la base CIPIC, y compris la moyenne.

L'intérêt de GUI-STAR, en plus de regrouper l'ensemble des scripts et de pouvoir les exécuter simplement et rapidement, est de pouvoir choisir de nombreuses options telles que la configuration des haut-parleurs (Fig. 11.5), ou encore l'exportation de sources sonores spatialisées (prenant en compte les paramètres choisis) (Fig. 11.6).

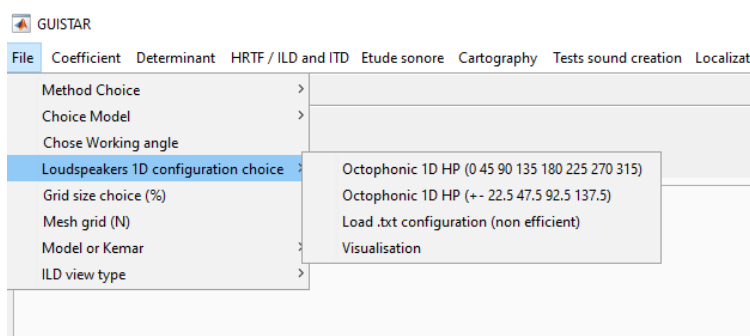


FIGURE 11.7 – Choix de la configuration de haut-parleurs dans GUI-STAR.

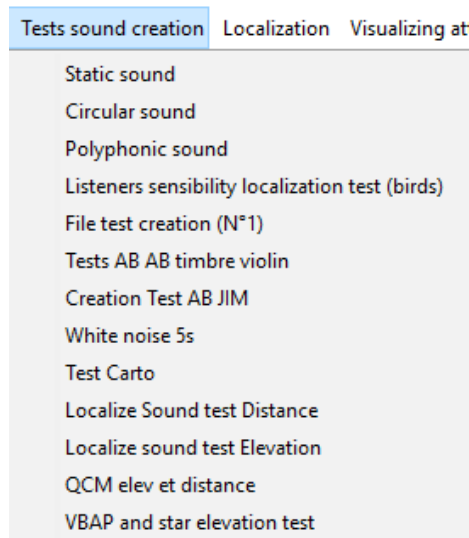


FIGURE 11.8 – Exportation de sons spatialisés avec GUI-STAR.

Voici ci-dessous une description des menus disponibles pour l'interface GUI-STAR.

1. Fichier (choix principaux s'appliquant à l'ensemble des tracés et exports de GUI-STAR)
 - (a) Choix de la méthode (permet de travailler sur les différentes solutions considérées pour la méthode STAR)
 - i. Méthode initiale
 - ii. En regardant la source sonore
 - iii. En regardant entre les deux haut-parleurs actifs
 - (b) Choix du modèle (permet de choisir le type de modèle d'indice acoustique utilisé)
 - i. Modèle d'ITD et d'ILD
 - A. ITD
 - $\sin(\theta) + \theta$
 - $\sin(\theta)$
 - θ
 - B. ILD
 - $\sin(\theta) + \theta$
 - $\sin(\theta)$
 - θ
 - θ (sans considérer le rayon de la tête)
 - C. Utilisation d'une HRTF réelle (en cours de développement)
 - ii. Choix de l'angle (ouvre une boîte de dialogue permettant de renseigner en degrés l'angle souhaité pour les figures le prenant en compte, telles que les cartographies par exemple)
 - iii. Configuration de haut-parleurs
 - Octophonique, haut-parleurs positionnés en (0, 45, 90, 135, 180, 225, 270, 315) degrés.
 - Octophonique, haut-parleurs positionnés en (22.5, 67.5, 112.5, 157.5, 202.5, 247.5, 292.5, 337.5) degrés.
 - Chargement d'une configuration (en cours de développement, fonctionne actuellement uniquement pour des configurations octophoniques)
 - Visualisation (permet de visualiser la configuration des haut-parleurs choisie, uniquement disponible pour la version 1D).
 - iv. Choix de la taille de la grille (pour les cartographies)

11.6. GUI-STAR, L'IHM D'ÉTUDE DE LA MÉTHODE STAR ET DE CRÉATION D'EXEMPLES SONORES

- v. Choix du maillage de la grille de simulation (pour les cartographies)
 - vi. Utilisation du modèle ou de KEMAR (sélectionne soit le modèle utilisé pour STAR, soit les données de KEMAR grandes oreilles de la base CIPIC)
 - Modèle
 - KEMAR
- 2. Coefficients (Différents tracés pour l'étude des coefficients de la méthode STAR)
 - (a) Visualisation des coefficients
 - i. Vue globale
 - ii. Vue uniquement pour un couple de haut-parleurs
 - iii. Vue en fonction de l'orientation de la tête
 - (b) Comparaison des coefficients de VBAP et STAR en fonction de l'angle
 - (c) Visualisation de l'énergie de VBAP et STAR en fonction de l'angle
 - (d) Visualisation des chemins acoustiques synthétiques
 - (e) Visualisation de l'énergie des chemins acoustiques
 - (f) Visualisation de l'erreur entre les différentes méthodes STAR
- 3. Déterminants (Différents tracés pour l'étude des déterminants de la méthode STAR)
 - (a) Visualisation du déterminant
 - (b) Vue du déterminant en fonction de l'angle entre les haut-parleurs
 - (c) Déterminant avec des sources sonores fixes
 - (d) Erreur sur le déterminant entre la méthode originelle et la méthode finale
 - (e) Visualisation d'un déterminant avec une source fixe en fonction de l'orientation de la tête
- 4. HRTF/ITD et ILD (Différents tracés pour l'étude des ITDs, ILDs et HRTFs utilisés pour la méthode STAR)
 - (a) HRTF
 - i. Visualisation des HRTFs de la base CIPIC
 - ii. Visualisation des HRTFs synthétiques (calculées à partir des modèles d'indices acoustiques)
 - iii. Erreur entre les HRTFs réelles et synthétiques
 - iv. Énergie des HRTFs
 - v. Visualisation 1D des HRTFs
 - (b) ILD et ITD synthétiques
 - (c) Coefficients α et β
- 5. Rendu binaural (permet de transformer un son spatialisé en octophonique en son binaural)
- 6. Cartographies qui regroupent l'ensemble des cartographies créées pour l'étude objective des cartographies de la méthode STAR)
 - (a) Visualisation de la grille de spatialisation
 - (b) Carte d'ILDs
 - Avec la méthode STAR
 - Avec la méthode VBAP
 - Avec la méthode HOA
 - (c) Exportation des cartographies tous les 5° permettant d'avoir une évolution de la cartographie (création de GIF par exemple)
 - (d) Résultats des tests expérimentaux au SCRIME
 - i. En regardant droit devant
 - Avec la méthode STAR
 - Avec la méthode VBAP

11.6. GUI-STAR, L'IHM D'ÉTUDE DE LA MÉTHODE STAR ET DE CRÉATION D'EXEMPLES SONORES

- Avec la méthode HOA
 - ii. En regardant vers la source sonore
 - Avec la méthode STAR
 - Avec la méthode VBAP
 - Avec la méthode HOA
 - iii. En regardant entre les haut-parleurs actifs
 - Avec la méthode STAR
 - Avec la méthode VBAP
 - Avec la méthode HOA
7. Création des tests sonores (Permet l'exportation de l'ensemble des fichiers sonores spatialisés)
 - (a) Test de son statique
 - (b) Test de son circulaire
 - (c) Test de son polyphonique
 - (d) Test préliminaire (des chants d'oiseaux)
 - (e) Création des fichiers de tests subjectifs globaux
 - (f) Tests de comparaison pour l'étude du timbre sur un extrait de morceau de violon
 - (g) Création des tests effectués pour les JIM 2019
 - (h) Création des échantillons à enregistrer pour la cartographie d'indices acoustiques
 - (i) Création des tests de validation de la distance
 - (j) Création des tests de validation de l'élévation
 8. Atténuation (Permet d'étudier le modèle d'atténuation utilisé pour la spatialisation de la distance)
 - (a) Visualisation du modèle de (Bass et al. (1984))
 - (b) Effet de l'atténuation sur un bruit blanc
 9. Tout effacer (Permet une ré-initialisation de l'interface)
 10. Exporter les figures (Permet d'exporter au format souhaité les figures tracées sur GUI-STAR)
 - (a) Format SVG
 - (b) Format PDF
 - (c) Format JPEG
 11. Étude de l'élévation concernant l'étude de l'élévation
 - (a) Visualisation de l'amplitude des nœuds et pics
 - Méthode de Iida
 - Nœuds et pics les plus proches du modèle
 - (b) Visualisation de la position fréquentielle des pics
 - Méthode de Iida
 - Nœuds et pics les plus proches du modèle
 - (c) Création de la table Shazam (crée toutes les possibilités considérées et les stocke dans un fichier .mat)
 - (d) Visualisation Shazam sur des données non apprises
 - (e) Visualisation Shazam sur des données apprises
 - (f) Visualisation Shazam sur des données non apprises mais HRTFs supposées identiques
 - (g) Visualisation de la table en fonction des pas choisis pour les nœuds
 - (h) Visualisation de la table en fonction des pas choisis pour les pics
 12. Test de développement (permet d'avoir les paramètres actifs de STAR souhaités pour la programmation de nouveaux scripts).

11.7 Les locaux et outils utilisés

Cette section d'annexe décrit les différentes salles où les enregistrements et tests audio ont été effectués.

11.7.1 Le dôme du SCRIME

La plupart des mesures effectuées pour les différents tests de la méthode STAR ont été réalisées au sein du dôme du SCRIME, dont une photographie panoramique est proposée sur la Fig. 11.9.



FIGURE 11.9 – Photographie panoramique du dôme du SCRIME et couronnes de haut-parleurs.

Le studio du SCRIME est une pièce rectangulaire de $40m^2$, recouverte au sol d'un tapis, ayant une porte en bois sur un côté et trois fenêtres sur le mur en face. Devant l'ensemble des murs sont disposés des panneaux acoustiques.

Le dôme en lui-même est constitué de plusieurs couronnes de haut-parleurs, bien visibles sur la Fig. 11.10. Ces couronnes sont constituées de 18 haut-parleurs Genelec 8040 accompagnés d'un Genelec 7050 en caisson de basse, contrôlés par un mac mini de 2014, avec un processeur de 2,8 GHz. Le dispositif sonore étant contrôlé par une carte son MADiface PRO et un convertisseur RME 32 canaux A/N.



FIGURE 11.10 – Couronnes de haut-parleurs.

Le dôme du SCRIME est utilisé couramment par des artistes compositeurs, l'acoustique de la pièce étant de bonne qualité bien que non contrôlée.

11.7.2 La salle de conférence et de diffusion Hémicyclia

Hémicyclia est une salle de conférence de $98m^2$ munie de plusieurs couronnes de haut-parleurs. Ceux-ci sont composés d'un ensemble hétéroclite que l'on peut retrouver sur le site du SCRIME.

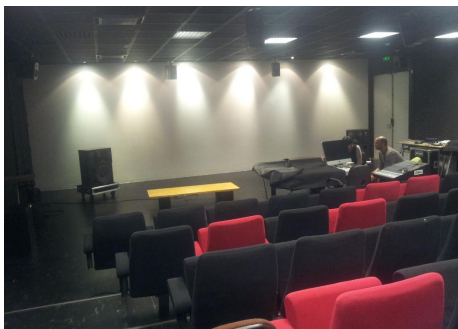


FIGURE 11.11 – *La salle de conférence et de diffusion du SCRIME Hémicyclia.*

Le fait que la salle soit moins calibrée et avec des enceintes hétéroclites, couplé au fait que pour ce travail il était préférable à des fins de cohérence de garder au maximum le même studio, font que les enregistrements effectués dans la salle Hémicyclia n'ont pas été exploités.

11.7.3 Salle de classe, salon et salle de concert

D'autres lieux ont été utilisés au cours de cette thèse : des salles de concerts (pour les tests subjectifs, lors des enregistrements de 2015 et les tests réalisés au cours des JIM 2019); une salle de classe présentée également pour les tests subjectifs (utilisée à la fois pour les tests subjectifs de comparaison des méthodes pour l'azimut de STAR, et pour les tests de validation de la spatialisation en élévation et en distance); et enfin, un salon personnel, notamment dû au contexte de crise sanitaire (ayant empêché durant une longue période l'accès aux installations du SCRIME). L'ensemble de ce panel de lieux d'enregistrements n'est pas dérangeant, mais au contraire intéressant afin de coller à la philosophie de STAR se voulant être une méthode versatile, utilisée en conditions réelles et simple de mise en œuvre.

11.8 Le matériel de diffusion et d'enregistrement

Cette section décrit brièvement les différents outils utilisés afin de diffuser et d'enregistrer les exemples sonores.

11.8.1 Le mannequin KEMAR

Le mannequin KEMAR est utilisé tout au long de ce travail de thèse. En effet, c'est un outil extrêmement utile, permettant d'obtenir facilement des enregistrements tels que ceux qui pourraient être obtenus avec un auditeur réel. Il est de plus extrêmement utile car contrôlé physiquement. Ainsi, les résultats obtenus par différents chercheurs sont en théorie identiques si les conditions expérimentales d'enregistrements le sont aussi.

La Fig. 11.12 montre le mannequin KEMAR utilisé pour les enregistrements du SCRIME.



FIGURE 11.12 – *Le mannequin KEMAR du SCRIME.*

Plusieurs micro/mannequin/type d'oreilles sont proposés pour les mannequins KEMAR. Le mannequin du SCRIME de référence 45BA dispose de deux microphones 40AG, est équipé de deux pavillons d'oreilles larges KB0065 (droite) et KB0066 (gauche), et est alimenté par un module d'alimentation 12AA.

11.8.2 Les cartes sons

En plus de la carte son du SCRIME utilisée lors de la diffusion d'échantillons sonores, différentes cartes sonores ont été utilisées :

- La carte MOTU ultralite Mk3 18 canaux pour les différents tests subjectifs de comparaisons de méthodes.
- La carte Son Scarlett 2i2 (Fig. 11.13) pour les tests de validation d'élévation et de distance, ainsi que pour l'ensemble des enregistrements via le mannequin KEMAR.
- Une carte Son Scarlett 4i4 pour la validation de la spatialisation des couronnes de haut-parleurs en couplant la méthode STAR avec la méthode VBAP.



FIGURE 11.13 – *Carte son Scarlett 2i2, utilisée pour l'ensemble des tests ne nécessitant pas plus de 2 haut-parleurs, ainsi que pour tous les enregistrements sur le mannequin KEMAR.*

11.8.3 Les micros

En plus des micros du mannequin KEMAR présentés dans la Sec. 11.8.1, il a été nécessaire d'utiliser des micros classiques, par exemple pour enregistrer les intensités sonores (voir le chapitre sur la spatialisation de la distance Sec. 8.2).

Dans ce cas de figure, le micro Schoeps MK4 (Fig. 11.14) à une capsule est utilisé.



FIGURE 11.14 – *Micro utilisé pour les enregistrements ne devant pas être perturbés par une morphologie humaine.*

11.8.4 Les haut-parleurs

Pour les haut-parleurs, le dôme du SCRIME présenté en Sec. 11.7.1 étant équipé de haut-parleurs Genelec 8040, il a été choisi de garder le même type de moyen de diffusion tout au long de ce travail de thèse. Aussi, le modèle Genelec 8030 (Fig. 11.15) n'est utilisé que pour les diffusions hors du dôme (ces modèles étant plus faciles pour le transport et la mise en place de par leur taille).



FIGURE 11.15 – *Enceinte Genelec 8030, utilisée pour les diffusions et tests sonores hors du studio du SCRIME.*

De plus, ces enceintes présentent l'avantage d'être assez peu colorées, ce qui est préférable pour nos études.

11.8.5 Les dispositifs créés pour les prises de son à l'aide du mannequin KEMAR

Pour les prises de mesure à l'aide du mannequin KEMAR, il est nécessaire de pouvoir orienter le mannequin, mais aussi de pouvoir le déplacer. Afin d'éviter des prises de mesure longues et engendrant des erreurs, des dispositifs de prises de mesure ont été mis au point au cours de cette thèse.

Dans un premier temps, la nécessité de possibilités de rotation et de déplacement a été rencontrée (prises de mesure en fonction de l'orientation du mannequin, et déplacements dans l'espace afin de procéder à la cartographie des ILDs). Un premier dispositif présenté sur la Fig. 11.16 est mis au point, constitué quasiment exclusivement de bois (présentant l'avantage d'être peu réverbérant). Ce dispositif est constitué d'une palette montée sur de solides pieds en bois au bout desquels sont placées des roues permettant le déplacement dans l'espace. En plus de ces roues, le plateau supportant le mannequin est découpé circulairement et fixé sur un axe de rotation, permettant de contrôler les différents azimuts.



FIGURE 11.16 – *Premier système de prises de mesure par le mannequin KEMAR, constitué d'une structure en bois, avec un axe de rotation et des roues pour le déplacement.*

Bien qu'utile et pratique dans un premier temps, ce premier dispositif a vite trouvé ses limites. En effet, le fait que l'axe de rotation soit fixé sur le bois engendre des vibrations lors des rotations et oblige à une attente relativement longue entre deux prises de son. De plus, il est impossible de procéder à des modifications de l'élévation. C'est pour cela que le second dispositif est mis en place. Un plateau est fixé au sommet d'une chaise de bureau, sur un support TV inclinable renforcé par un solide écrou. Le support TV permet de simuler des inclinaisons de $\pm 40^\circ$ ¹, une équerre solidaire du support (Fig. 11.17) fixe permet de visualiser l'inclinaison grâce à une aiguille mobile (fixée sur le support où repose le mannequin KEMAR). La chaise de bureau, quant à elle, est équipée d'un dispositif de roulement à billes permettant des rotations très précises et en douceur (ne nécessitant pas d'attente entre deux mesures). Enfin, l'ensemble est monté sur cinq roues en plastique permettant des déplacements dans l'espace.

Afin de limiter les réverbérations, des panneaux de mousse sont fixés sur les deux supports sous le mannequin.

1. Le mannequin est sanglé solidement au support du haut.



FIGURE 11.17 – *Second système de prises de mesure, plus efficace, constitué d'un support de chaise de bureau et un support TV inclinable solide.*

Ce second système s'est révélé très pratique, permettant de prendre rapidement de nombreuses mesures. Aussi, nombre des premières prises de mesure acquises avec le premier système ont été remplacées par d'autres plus précises et complètes prises à l'aide de ce second système.

11.8.6 Les logiciels utilisés

Cette dernière section d'annexe présente les principaux logiciels utilisés au cours de ce travail.

MatlabTM et Octave pour la recherche

Les premiers logiciels utilisés sont MatlabTM et Octave (Fig. 11.18). Ces deux logiciels sont des plateformes de calculs numériques et de programmations utilisées notamment pour leur rapidité et facilité de mise en œuvre. Le choix s'est en premier lieu porté sur le logiciel Octave pour sa version libre, puis, la qualité des graphiques ainsi que le GUI de MatlabTM ont fait que la version de MathWorks a ensuite été retenue.

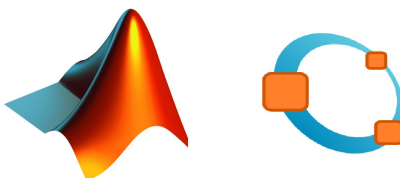


FIGURE 11.18 – *Logos des logiciels MatlabTM et GNU Octave.*

En parallèle des implémentations sous MatlabTM, des versions en C++ et Python ont été développées, plus performantes, et permettant d'envisager des intégrations dans des softs tels que Ossia Score développé par le SCRIME et proposant différentes méthodes de spatialisation (notamment différents décodeurs HOA).

Les logiciels audio

Deux logiciels audio ont été utilisés principalement au cours de cette thèse. Le premier est le logiciel Audacity, afin notamment de gérer les enregistrements audio via le mannequin. Le second est Ossia Score (SCRIME (2021)) développé par le SCRIME, permettant une gestion facilitée des pistes et événements. Il a notamment permis le contrôle de la Scarlett 4i4 à l'aide du protocole ASIO.

Bibliographie

- (2015). ITU-R BS.1116-3, methods for the subjective assessment of small impairments in audio systems including multichannel sound systems.
- Algazi, R., Duda, R., Thompson, D., and Avendano, C. (2001). The CIPIC HRTF database. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 99–102, New Paltz, New York.
- Bass, H., Sutherland, L., Piercy, J., and Evans, L. (1984). Absorption of sound by the atmosphere. *Physical acoustics : Principles and methods*, 17 :145–232.
- Bauck, J. (2001). A simple loudspeaker array and associated crosstalk canceler for improved 3d audio. *Journal of the Audio Engineering Society*, 49(1/2) :3–13.
- Bauer, B. B. (1961). Stereophonic earphones and binaural loudspeakers. *Journal of the Audio Engineering Society*, 9(2) :148–151.
- Berkhout, A., de Vries, D., and Vogel, P. (1993). Acoustic control by wave field synthesis. *Journal of the Acoustical Society of America*, 93(5) :2764–2776.
- Bertet, S. S. (2009). *Formats audio 3D hiérarchiques : caractérisation objective et perceptive des systèmes ambisonics d'ordres supérieurs*. PhD thesis, L'Institut National des Sciences Appliquées de Lyon, France.
- Blauert, J. (1969). Sound localization in the median plane. *Journal Acta Acustica united with Acustica*, 22 :205–213.
- Blauert, J. (1997). *Spatial Hearing*. MIT Press, Cambridge, Massachusetts, revised edition.
- Brungart, D. S. and Rabinowitz, W. M. (1999). Auditory localization of nearby sources. head-related transfer functions. *Journal of the Acoustical Society of America*, 106(3) :1465–1479.
- Brungart, D. S. and Scott, K. R. (1962). The effects of production and presentation level on the auditory distance perception of speech. *Journal of the Acoustical Society of America*, 110(1) :425–440.
- Burkhard, M. and Sachs, R. (1975). Anthropometric manikin for acoustic research. *Journal of the Acoustical Society of America*, 58(1) :214–222.
- Butler, R. (1969). Monaural and binaural localization of noise bursts vertically in median sagittal plane. *Journal of Auditory research*, 9(3) :230–235.
- Butler, R. A. and Belendiuk, K. (1977). Spectral cues utilized in the localization of sound in the median sagittal plane. *Journal of the Acoustical Society of America*, 61(5) :1264–1269.
- Chowning, J. M. (1971). The simulation of moving sound sources. *Journal of the Audio Engineering Society*, 19(1) :2–6.
- Daniel, J. (2001). *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. PhD thesis, Université Paris 6.

BIBLIOGRAPHIE

- de Recherche et Coordination Acoustique/Musique, I. (2021). Ircam database. "http://recherche.ircam.fr/equipes/salles/listen/system_protocol.html".
- Erbes, V., Geier, M., Weinzierl, S., and Spors, S. (2015). Database of single-channel and binaural room impulse responses of a 64-channel loudspeaker array. In *Journal of the Audio Engineering Society*, Warsaw, Poland.
- Gandemer, L., Parseihian, G., Bourdin, C., and Kronland-Martinet, R. (2018). Perception of surrounding sound source trajectories in the horizontal plane : A comparison of vbap and basic-decoded hoa. *Acta Acustica united with Acustica*, 104(2) :338–350.
- Gardner, M. B. (1969). Distance estimation of 0° or apparent 0°-oriented speech signals in anechoic space. *Journal of the Acoustical Society of America*, 45(1) :47–53.
- Gerzon, M. A. (1973). Periphony : With-height sound reproduction. *Journal of the Audio Engineering Society*, 21(1) :2–10.
- Gálvez, M. F. S. and Fazi, F. M. (2015). Loudspeaker arrays for transaural reproduction. In *International Congress on Sound and Vibration*, Florence, Italy.
- Hebrank, J. and Wright, D. (1974). Spectral cues used in the localization of sound sources on the median plane. *Journal of the Acoustical Society of America*, 56(6) :1829–1834.
- Iida, K. (2017). *Head-Related Transfer Function and Acoustic Reality*. Springer.
- Iida, K., Itoh, M., Itagaki, A., and Morimoto, M. (2007). Median plane localization using a parametric model of the head-related transfer function based on spectral cues. *Applied Acoustics*, 68(8) :835–850.
- Institute, A. R. (2021). The ari database. <https://www.oeaw.ac.at/en/isf/das-institut/software/hrtf-database>.
- Jeub, M., Schafer, M., and Vary, P. (2009). A binaural room impulse response database for the evaluation of dereverberation algorithms. In *Conference on Digital Signal Processing*, volume 16, pages 1–5.
- Joyce, V. and John, V. O. A. (2004). The influence of duration and level on human sound localization. *Journal of the Acoustical Society of America*, 115(4) :1705–1713.
- Kahana, Y. and Nelso, P. A. (2005). Numerical modelling of the spatial acoustic response of the human pinna. *Journal of Sound and Vibration*, 292 :205–213.
- Kearney, G., Gorzel, M., Rice, H., and Boland, F. (2015). Distance perception in interactive virtual acoustic environments using first and higher order ambisonic sound fields. *Journal Acta Acustica united with Acustica*, 98(1) :61–71.
- Kohlrausch, A. and van de Par, S. (2005). *Audio—Visual Interaction in the Context of Multi-Media Applications*, pages 109–138. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Kolarik, A. J., Moore, B. C. J., Zahorik, P., Cirstea, S., and Pardhan, S. (2016). Auditory distance perception in humans : a review of cues, development, neuronal bases, and effects of sensory loss. *Attention, Perception, and Psychophysics*, 78 :373–395.
- Laboratory, S. H. (2017). The cit hrtf database ver. 1.3. <http://www.iida-lab.it-chiba.ac.jp/HRTF/>.
- Letowski, T. and Letowski, S. (2011). Localization error : Accuracy and precision of auditory localization. *Advances in sound localization*, pages 55–78.
- Majdak, P., Balazs, P., and Laback, B. (2007). multiple exponential sweep method for fast measurement of head-related transfer functions. *Journal of the Audio Engineering Society*, 55(7/8) :623–637.
- Marchand, S. (2020). Une approche perceptive pour la spatialisation du son. *Revue Francophone d’Informatique et Musique*, (7-8). <https://revues.mshparisnord.fr:443/rfim/index.php?id=606>.

BIBLIOGRAPHIE

- Marentakis, G., Zotter, F., and Frank, M. (2014). Vector-base and ambisonic amplitude panning : A comparison using pop, classical, and contemporary spatial music. *Acta Acustica united with Acustica*, 100(5) :945–955.
- Mershon, D. H. and King, L. E. (1975). Intensity and reverberation as factors in the auditory perception of egocentric distance. *Journal of the Acoustical Society of America*, 18(6) :425–440.
- Moore, B. C. J. (2013). *An Introduction to the Psychology of Hearing*. Emerald.
- Moore, B. C. J., Oldfield, S. R., and Dooley, G. J. (1989). Detection and discrimination of spectral peaks and notches at 1 and 8 khz. *Journal of the Acoustical Society of America*, 85 :820–836.
- Mouba, J. (2009). *Manipulation spatiales de sons spectraux*. PhD thesis, École Doctorale de mathématique et d’Informatique de Bordeaux, France.
- Mouba, J. and Marchand, S. (2006). A source localization / separation / respatialization system based on unsupervised classification of interaural cues. In *International Conference on Digital Audio Effects (DAFx)*, pages 233–238, Montreal, Quebec, Canada.
- Mouba, J., Marchand, S., Mansencal, B., and Rivet, J.-M. (2008). Retrosnat : a perception-based system for semi-automatic diffusion of acousmatic music. In *Sound and Music Computing (SMC) Conference*, pages 33–40, Berlin, Germany.
- Musicant, A. D. and Butler, R. A. (1984). The influence of pinnae-based spectral cues on sound localization. *Journal of the Acoustical Society of America*, 75(4) :1195–1200.
- Méaux, E. and Marchand, S. (2019). Synthetic transaural audio rendering (STAR) : a perceptive approach for sound spatialization. In *International Conference on Digital Audio Effects (DAFx)*, pages 205–212, Birmingham, United Kingdom.
- Méaux, E. and Marchand, S. (2020a). Interaural cues cartography : Localization cues repartition for three spatialization methods. In *International Conference on Digital Audio Effects (DAFx)*, pages 258–264, Vienna, Austria.
- Méaux, E. and Marchand, S. (2020b). Sound source localization from interaural cues : Estimation of the azimuth and effect of the elevation. In *Forum Acusticum*, Lyon, France.
- Méaux, E. and Marchand, S. (2021a). Synthetic transaural audio rendering (star) : a perceptive 3d audio spatialization method. *Journal of the Audio Engineering Society*.
- Méaux, E. and Marchand, S. (2021b). Synthetic transaural audio rendering (star) : Extension to full 3d spatialization. *Journal of the Audio Engineering Society*. In submission.
- Møller, H. (1992). Fundamentals of binaural technology. *Applied Acoustics*, 36(3-4) :171–218.
- of Technology Media Lab Machine Listening Group, M. I. (1994). Hrtf measurements of a kemar dummy-head microphone. <https://sound.media.mit.edu/resources/KEMAR.html>.
- Pulkki, V. (1997). Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45(6) :456–466.
- Pulkki, V., Huopaniemi, J., and Huottilainen, T. (1996). Dsp tool for 8-channel audio mixing. pages 307–314.
- Raykar, V. C. and Duraiswami, R. (2005). Extracting the frequencies of the pinna spectral notches in measured head related impulse responses. *Journal of the Acoustical Society of America*, 118(1) :364–374.
- Risoud, M., Hanson, J.-N., Gauvrit, F., Renard, C., Lemesre, P.-E., Bonne, N.-X., and Vincent, C. (2018). Sound source localization. In *European Annals of Otorhinolaryngology, Head and Neck Diseases*, volume 135, pages 259–264.

BIBLIOGRAPHIE

- Sayers, B. M. (1964). Acoustic-image lateralization judgments with binaural tones. *Journal of the Acoustical Society of America*, 36(5) :923–926.
- Schroeder, M. and Atal, B. (1963). Computer simulation of sound transmission in rooms. In *Proceedings of the IEEE*, pages 536–537, New York, USA.
- SCRIME, O. (2021). Ossia score. <https://ossia.io/>.
- Shaw, E. (1997). *Acoustical Features of the Human External Ear, Binaural and Spatial Hearing in Real and Virtual Environments*. Psychology Press, robert gilkey, timothy r. anderson edition.
- Snow, W. B. (1955). Basic principles of stereophonic sound. *IRE Transactions on Audio*, AU-3(2) :42–53.
- Spors, S., Rabenstein, R., and Ahrens, J. (2008). The theory of wave field synthesis revisited. In *Audio Eng. Soc.*, page Paper 7358.
- Stevens, S. S. and Guirao, M. (1962). Loudness, reciprocity, and partition scales. *Journal of the Acoustical Society of America*, 34(9) :466–471.
- Strutt (Lord Rayleigh), J. W. (1877). Acoustical observations. *Philosophical Magazine*, 3(6) :456–464.
- Strutt (Lord Rayleigh), J. W. (1907). On our perception of sound direction. *Philosophical Magazine*, 13(74) :214–302.
- Takemoto, H., Mokhtari, P., Kato, H., and Nishimura, R. (2012). Mechanism for generating peaks and notches of head-related transfer functions in the median plane. *Journal of the Acoustical Society of America*, 132(6) :1465–1479.
- Thurlow, W. R., Mangels, J. W., and Runge, P. S. (1967). Head movements during sound localization. *Journal of the Acoustical Society of America*, 42(2) :489–493.
- Viste, H. (2004). *Binaural Localization and Separation Techniques*. PhD thesis, École Polytechnique Fédérale de Lausanne, Switzerland.
- von Hornbostel, E. M. and Wertheimer, M. (1920). Über die wahrnehmung der schallrichtung. In *Sitzungsber. Akad. Wiss*, Berlin.
- Wallach, H. (1940). The role of head movements and vestibular and visual cues in sound localization. *Journal of Experimental Psychology*, 27(4) :339–368.
- Wang, A. (2006). The shazam music recognition service. *Communications of the ACM*, 49 :44–48.
- Watanabe, K., Iwaya, Y., Suzuki, Y., Takane, S., and Sato, S. (2014). Dataset of head-related transfer functions measured with a circular loudspeaker array. *J. Acoust. Sci and Tech.*, 35(3) :159–165.
- Wightman, F. L. and Kistler, D. J. (1954). *Experimental Psychology*. Psychology Press, robert gilkey, timothy r. anderson edition.
- Winter, F., Wierstorf, H., Podlubne, A., Forgue, T., Manhes, J., Herrb, M., Spors, S., Raake, A., and Danès, P. (2016). Database of binaural room impulse responses of an apartment-like environment. In *Convention e-Brief of the Audio Eng. Soc.*
- Woodworth, R. S. and Schlosberg, H. (1954). *Experimental Psychology*. London : Methuen.
- Yan, X., Kazuhiro, I., and Yohji, I. (2014). Comparison in frequencies of spectral peaks and notches and anthropometric of pinnae between hrtf databases. Technical report, The Institute of Electronics Information and Communication Engineers.
- Young, P. T. (1931). The rôle of head movements in auditory localization. *Journal of Experimental Psychology*, 14(2) :95.

BIBLIOGRAPHIE

- Zahorik, P. and Wightman, F. L. (2001a). Loudness constancy with varying sound source distance. *Nature Neuroscience*, 4 :78–83.
- Zahorik, P. and Wightman, F. L. (2001b). Loudness constancy with varying sound source distance. *Journal Nature Neuroscience*, 4(1) :78–83.
- Zotter, F. and Frank, M. (2019). *Ambisonics A practical Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Springer Open.

Approche perceptive pour la spatialisation / localisation sonore 3D

La localisation sonore est le procédé utilisé par les êtres humains pour repérer un son dans l'espace. Afin de localiser ces sons, le cerveau traite l'information reçue, et crée des indices acoustiques. L'approche de la thèse pour la localisation sonore perceptive, reposant sur le travail d'Harald Viste pour la localisation de l'azimut, consiste à utiliser ces indices acoustiques dans un algorithme. L'algorithme initial est légèrement simplifié dans cette thèse, et testé dans des conditions réelles. De plus, une approche perceptive innovante pour la localisation de l'élévation est également présentée.

La spatialisation sonore est le procédé inverse, permettant de produire un son que l'on percevra à la position souhaitée dans l'espace. Du fait de l'impossibilité d'avoir un système de diffusion en tout point de l'espace, il est nécessaire de recourir à des algorithmes de spatialisation, permettant par exemple des diffusions via des haut-parleurs. L'approche perceptive de la thèse, basée sur le travail de Joan Mouba, est d'utiliser les indices acoustiques de la localisation sonore, dans ce travail en les créant dans les sources sonores spatialisées. Ce travail de thèse approfondit les recherches initiales, crée des outils pour aboutir à une proposition de méthode de spatialisation sonore perceptive 3D nommée STAR (Synthetic Transaural Audio Rendering), tout en validant la méthode par des tests rigoureusement menés.

Mots clefs : Spatialisation sonore 3D, Perceptif, localisation, indices acoustiques.

Perceptual approach for 3D sound spatialization / localization

Sound localization is the process used by humans to locate sound in space. In order to locate these sounds, the brain processes the information received, and creates acoustic cues. The thesis approach to perceptual sound localization, based on Harald Viste's work for azimuth localization, is to use these acoustic cues in an algorithm to locate a sound source. The initial algorithm is slightly simplified in this thesis, and tested in real conditions. In addition a perceptual approach for the location of the elevation is also presented.

Sound spatialization is the reverse process, making it possible to produce a sound that will be perceived at the position of the desired space. Due to the impossibility of having a broadcast system at any point in space, it is necessary to use spatialization algorithms, for example allowing broadcasts through loudspeakers. The perceptual approach of the thesis, based on the work of Joan Mouba, is to use the acoustic cues of sound localization, this time by creating them in spatialized sound sources. This thesis work deepens the initial research, notably proposes a 3D perceptual sound spatialization method called STAR (Synthetic Transaural Audio Rendering), while validating the method through tests. Keywords : 3D sound spatialization, perceptual, localisation, perceptual cues.

L3i

**Laboratoire Informatique, Image et Interaction Avenue
Michel Crépeau**

17042 LA ROCHELLE