

### Prédiction objective de l'effet des systèmes tactiques de communication et protection sur les performances de localisation sonore

Thomas Joubaud

#### ► To cite this version:

Thomas Joubaud. Prédiction objective de l'effet des systèmes tactiques de communication et protection sur les performances de localisation sonore. Acoustique [physics.class-ph]. Conservatoire national des arts et metiers - CNAM, 2017. Français. NNT: 2017CNAM1132. tel-01694013

### HAL Id: tel-01694013 https://theses.hal.science/tel-01694013

Submitted on 26 Jan 2018  $\,$ 

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# le cnam

École doctorale Sciences des Métiers de l'Ingénieur

Laboratoire de Mécanique des Structures et des Systèmes Couplés

### THÈSE DE DOCTORAT

présentée par : Thomas JOUBAUD

soutenue le : 15 septembre 2017

pour obtenir le grade de : Docteur du Conservatoire National des Arts et Métiers

Discipline / Spécialité : Mécanique, génie mécanique, génie civil / Mécanique

### Prédiction objective de l'effet des systèmes tactiques de communication et protection sur les performances de localisation sonore

#### THÈSE dirigée par

M. GARCIA Alexandre

#### RAPPORTEURS

M. DUCOURNEAU Joël Mme. MEUNIER Sabine

#### PRÉSIDENT DU JURY

M. PAQUIER Mathieu

#### EXAMINATEURS

M. LANGRENNE Christophe Mme. NICOL Rozenn M. VOIX Jérémie Mme. ZIMPFER Véronique Professeur, CNAM

Professeur, Université de Nancy Chargée de Recherche, LMA

Professeur, UBO LABSTICC

Ingénieur, co-encadrant, CNAM Ingénieure HDR, Orange Lab Professeur, ETSMTL Ingénieure, co-encadrante, ISL "J'essaie de cerner ma propre pensée. Et la parole, d'où vient la parole? Peut-être que les gens parlent sans arrêt, comme les chercheurs d'or, pour trouver la vérité. Ils éliminent tous les mots qui n'ont pas de valeurs, et pour finir ils en trouvent un, tout seul. Or un seul mot, tout seul, c'est déjà le silence."

Regular Barbary, A Backward Glance On A Travel Road (extrait de Le petit soldat, Jean-Luc Godard)

### Remerciements

Le travail de recherche présenté dans ce mémoire a été réalisé à l'Institut francoallemand de recherches de Saint-Louis (ISL) au sein du groupe Acoustique et Protection du Combattant (APC), et mené en collaboration avec le Laboratoire de Mécanique des Structures et des Systèmes Couplés (LMSSC) du Conservatoire National des Arts et Métiers (CNAM) à Paris.

Je tiens tout d'abord à remercier Véronique ZIMPFER de m'avoir permis de réaliser cette thèse et de m'avoir encadré pendant ces trois ans. Nos discussions ont permis de guider mes travaux et ton aide sur les différents dispositifs du laboratoire m'a été précieuse. Je pense que les tests subjectifs que tu nous fais passer régulièrement vont finir par me manquer.

Je souhaite ensuite remercier Alexandre GARCIA d'avoir accepté de diriger ma thèse depuis Paris. Tes conseils m'ont permis à la fois d'avancer dans mon projet et de ne pas sortir du sujet. Ce mémoire et mes articles ont aussi pu profiter de ta capacité à détecter les fautes que j'y ai laissées. Je remercie aussi Christophe LANGRENNE d'avoir participé à l'encadrement de ma thèse et de m'avoir aidé pour les mesures sur la sphère de reproduction sonore. Mes remerciements vont aussi à tous les membres du CNAM que j'ai eu le plaisir de rencontrer.

Je tiens à remercier Pierre NAZ de m'avoir accueilli au sein du groupe APC à l'ISL et de m'avoir apporté une précieuse aide sur les questions administratives. Je remercie bien sûr aussi l'ensemble de l'équipe APC, la bonne ambiance et les discussions à la pausecafé ayant grandement contribué à la vitesse à laquelle ces trois années sont passées. Je tiens particulièrement à remercier les autres doctorants avec qui j'ai partagé mon bureau : Johanna, Guillaume, Matthias et Floriane. Pour finir, je remercie l'équipe de badminton de l'ISL et le reste du personnel de l'institut que j'ai pu rencontrer.

Ce travail n'aurait pas non plus été possible sans l'ensemble des auditeurs volontaires qui ont accepté de participer à mes tests de localisation sonore. Mes sincères remerciements vont à chacune de ces personnes qui ont eu le courage d'affronter près de deux heures de bruits blancs leur tournant autour dans une cabine sombre.

Je souhaite enfin remercier les personnes qui m'ont entouré pendant ma thèse, les chevaliers de Mulhouse, les anciens de l'ECM à Paris et les amis de Bretagne. Je remercie spécialement l'équipe MB pour les quelques jours de répétition que nous avons réussi à faire et qui m'ont permis de souffler. Pour finir, je remercie profondément ma famille qui m'a soutenu pendant toute la durée de mes études.

### Résumé

Dans de nombreuses situations civiles ou militaires, la protection de l'audition du personnel est cruciale. La perception et l'interprétation de l'environnement sonore par l'auditeur doivent cependant être préservées. Les Systèmes Tactiques de Communication et Protection (TCAPS) sont des protections auditives qui, à la fois, protègent suffisamment les oreilles de l'auditeur contre les bruits dangereux, et préservent l'intelligibilité, permettant ainsi la communication vocale à bas niveau. Des études précédentes ont cependant démontré que les TCAPS continuent de détériorer la perception de l'environnement sonore de l'auditeur, en particulier sa capacité à localiser les sources sonores. Sur le plan horizontal, la dégradation des indices acoustiques empêchant, en temps normal, l'auditeur de confondre les sources avant et arrière, en est la principale explication. Dans ce travail de thèse, une expérience subjective de localisation sonore est conduite avec six TCAPS : deux bouchons d'oreille passifs, deux bouchons d'oreilles actifs et deux casques actifs. Si aucun protecteur ne permet de retrouver les performances de la condition d'écoute normale, l'expérience permet de classifier les TCAPS du point de vue de la localisation sonore : les performances des bouchons passifs sont meilleures que celles des bouchons actifs, et les casques actifs engendrent les plus mauvaises performances. Dans le cadre de la conception et de l'évaluation des TCAPS, une méthode prédisant leur dégradation des performances de localisation sonore, basée sur des mesures électroacoustiques, serait plus adaptée que des expériences comportementales très chronophages. Dans ce contexte, deux méthodes basées sur les Fonctions de Transfert Relatives à la Tête (HRTF) mesurées sur tête artificielle sont étudiées : un processus d'appariement et un réseau de neurones à trois couches. Ils sont optimisés pour reproduire les performances de localisation humaine en condition d'écoute normale. Les méthodes sont ensuite appliquées aux HRTF mesurées avec les six TCAPS, et prédisent des probabilités de localisation en fonction de la position. Comparé aux résultats de l'expérience subjective, le réseau de neurones prédit des performances réalistes avec les bouchons d'oreille, mais surestime les erreurs avec les casques. Le modèle d'appariement prédit correctement les performances de localisation. Toutefois, la vraisemblance de ses distributions de probabilité avec les observations subjectives demeure plus faible que celle du réseau de neurones. Pour finir, les deux méthodes développées dans cette étude sont indépendantes de la tête artificielle utilisée, et peuvent être utilisées pour évaluer non seulement des prototypes de TCAPS, mais aussi des prothèses auditives.

**Mots clés :** Protection auditive, localisation sonore, psychoacoustique, HRTF, ITD, ILD, indices spectraux, réseau de neurones

### Abstract

#### Objective prediction of the effect of tactical communication and protective systems on sound localization performance

In many civilian or military situations, hearing protection is of major importance. The listener's acoustical situational awareness must however also be preserved. Tactical Communication and Protective Systems (TCAPS) are hearing protection devices that sufficiently protect the listener's ears from hazardous sounds and preserve speech intelligibility, thus allowing low-level speech communication. However, previous studies demonstrated that TCAPS still deteriorate the listener's situational awareness, in particular the ability to locate sound sources. On the horizontal plane, this is mainly explained by the degradation of the acoustical cues normally preventing the listener from making frontback confusions. In the present PhD work, a behavioral sound localization experiment is conducted with six TCAPS: two passive and two active earplugs, and two active earmuffs. The performance in open ear condition is not retrieved with any protector, but the experiment ranks the TCAPS by type : passive earplugs lead to better performance than active earplugs, and active earmuffs induce the worst performance. As part of TCAPS development and assessment, a method predicting the protector-induced degradation of the sound localization capability, and based on electroacoustic measurements, would be more suitable than time-consuming behavioral experiments. In this context, two methods based on Head-Related Transfer Functions (HRTFs) measured on an artificial head are investigated : a template-matching model and a three-layer neural network. They are optimized to fit human sound localization performance in open ear condition. The methods are applied to the HRTFs measured with the six TCAPS, providing position-dependent localization probabilities. Compared with the behavioral results, the neural network predicts realistic performances with earplugs, but overestimates errors with earmuffs. The template-matching model predicts human performance well. However, the likelihood of the resulting probability distributions with the behavioral observations is lower than that of the neural network. Finally, both methods developed in this study are independent of the artificial head used, and can be applied to assess not only TCAPS prototypes, but also hearing aids.

**Keywords :** Hearing protection, sound localization, psychoacoustic, HRTF, ITD, ILD, spectral cues, neural network

## Table des matières

Ta	Table des matièresix				
$\mathbf{Li}$	Liste des tableaux xi				
Ta	able	des fig	gures x	viii	
$\mathbf{Li}$	iste (	des acı	ronymes	xix	
Ν	ome	nclatu	re	xxi	
In	trod	luction	L Contraction of the second	1	
I	Co	ntexte	de l'étude	<b>5</b>	
	1	Lutte	contre le bruit dans les domaines civil et militaire	5	
		1.1	Réglementation	5	
		1.2	Protections auditives conventionnelles	7	
		1.3	$Mesure(s) \ de \ l'atténuation \ des \ protections \ auditives \ \ . \ . \ . \ . \ .$	8	
		1.4	Perte d'insertion de protecteurs conventionnels	10	
2 Protection auditive et <i>situational awareness</i>		ction auditive et <i>situational awareness</i>	12		
		2.1	Systèmes Tactiques de Communication et Protection	13	
		2.2	Performances de localisation sonore	19	
II	Eff	et des	TCAPS sur les indices de localisation sonore	23	
	1	Méca	nismes de la localisation sonore	23	
		1.1	Différence interaurale d'intensité	24	
		1.2	Différence interaurale de temps	26	
		1.3	Limitation de la théorie duplex	30	
		1.4	Indices spectraux	31	
	2	Head	Related Transfer Functions	37	
		2.1	Protocole de mesure	38	
		2.2	Estimation des différences interaurales de temps et d'intensité	42	

	2.3	Modèle auditif de lissage des HRTF	46
	2.4	Indices de localisation sonore avec TCAPS	47
	2.5	Conclusion sur les HRTF mesurées avec les TCAPS	55
IIIEx	périenc	e subjective de localisation sonore	57
1	Protoc	cole expérimental	57
	1.1	Auditeurs	57
	1.2	Dispositif expérimental	58
	1.3	Procédure	59
	1.4	Interprétation des erreurs de localisation	60
	1.5	Analyse des données	61
2	Analy	se des résultats	62
	2.1	Performances de localisation sonore	62
	2.2	Étude des erreurs	68
	2.3	Temps de réponse	72
	2.4	Effets de l'âge et du genre	78
	2.5	Étude du ressenti des auditeurs	80
3	Conclu	usion	82
IV Pré	dictior	des performances de localisation	85
1	Modèl	e d'appariement	86
	1.1	Principe	86
	1.2	Mesure de similarité spectrale	86
	1.3	Développement du modèle d'appariement	87
	1.4	Optimisation des paramètres	92
	1.5	Conclusion sur le modèle d'appariement	97
2	Réseau	u de neurones	98
	2.1	Modélisation	98
	2.2	Apprentissage du réseau	101
	2.3	Application à la localisation sonore	102
	2.4	Conclusion sur le réseau de neurones	105
3	Perfor	mance des modèles	105
	3.1	Étude des matrices de probabilité	106
	3.2	Analyse des métriques	112
	3.3	Vraisemblance des modèles	115
	3.4	Conclusion sur les performances des modèles	116
4	Correc	tion du modèle d'appariement	117
	4.1	Révision du principe du modèle	117
	4.2	Dépendance en azimut des paramètres	121

	5	Correct	tion du réseau de neurones	. 128
		5.1	Définition des améliorations	. 129
		5.2	Étude des corrections	. 131
	6	Discuss	sion	. 134
		6.1	Résultats finals	. 134
		6.2	Effet des TCAPS sur les performances de localisation sonore	. 135
		6.3	Limitations	. 136
		6.4	Conclusion	. 138
$\mathbf{V}$	Vali	dation	de la modélisation et application	139
	1	Indépe	ndance vis-à-vis de la tête artificielle	. 139
		1.1	Dispositif expérimental du CNAM	. 140
		1.2	Optimisation du modèle d'appariement avec le KEMAR	. 142
		1.3	Optimisation du réseau de neurones avec le KEMAR	. 144
		1.4	Comparaison des têtes artificielles	. 147
		1.5	Conclusion sur l'indépendance des modèles à la tête artificielle	. 152
	2	Applica	ations	. 153
		2.1	Amélioration des performances d'un casque talk-through	. 153
		2.2	Étude des performances de bouchons à atténuation uniforme	. 156
		2.3	Prédiction de la dégradation des performances de localisation créée	
			par des aides auditives	. 159
		2.4	Conclusion sur la validation des modèles de prédiction	. 167
Co	onclu	sion		169
	1	Résulta	ats	. 169
	2	Perspe	ctives	. 171
Bi	bliog	raphie		173
A	nnex	es		189
$\mathbf{A}$	Syst	èmes c	le coordonnées polaires verticales et interaurales	189
в	La f	onctio	n gammatone et sa transformée de Fourier	193
$\mathbf{C}$	Esti	mation	des paramètres d'une distribution de von Mises	197
D	Déc	omposi	ition en composantes latérale et polaire	199
$\mathbf{E}$	Mat	rices d	e probabilité de localisation sonore	203
Gl	ossai	re		209

## Liste des tableaux

II.1	Inter-Subject Spectral Difference (ISSD) (en $dB^2$ ) estimée pour chacun des six systèmes tactiques de communication et protection
III.1	Nombres moyens de localisations exactes, de localisations voisines et d'erreurs pour les 7 conditions d'écoute (sur 80)
IV.1	Erreur de décomposition/reconstruction des probabilités de l'expérience subjective dans chaque condition d'écoute
IV.2	Écarts-types $\sigma_{ITD}$ et $\sigma_{\delta}$ du modèle d'appariement revu optimisés pour chaque condition d'écoute
IV.3	Nombre d'itérations nécessaires à l'apprentissage du réseau de neurones en fonction de sa sortie théorique et du seuil d'apprentissage $E. \ldots 130$
V.1	Paramètres des trois modèles d'appariement optimisés à partir des DTF mesurées sur la tête artificielle ISL et sur le mannequin KEMAR 142
V.2	Métriques prédites par le modèle d'appariement et le réseau de neurones dans le cadre de l'amélioration du casque P6
V.3	Comparaison des métriques prédites par le modèle d'appariement et le réseau de neurones avec différents bouchons d'oreille passifs

# Table des figures

1.1	Photographies de protections individuelles contre le bruit conventionnelles.	8
I.2	Principe de la mesure de la réduction de bruit et de la perte d'insertion	9
I.3 Photographies du dispositif de mesure de la perte d'insertion d'un		
	tection auditive sur tête artificielle en champ diffus. $\hdots$	11
I.4	Perte d'insertion de trois protecteurs conventionnels en fonction de la	
	fréquence	12
I.5	(a) Photographie du bouchon 3 M^ Combat Arms^ - Génération 4.1,	
	(b) Perte d'insertion en champ diffus pour 5 niveaux extérieurs différents	14
I.6	(a) Photographie du bouchon SureFire EP4 Sonic Defenders Plus, (b)	
	Perte d'insertion en champ diffus pour 5 niveaux extérieurs différents .	15
I.7	(a) Photographie du système Nacre QuietPro <sup>®</sup> , (b) Perte d'insertion en	
	champ diffus pour 5 niveaux extérieurs différents	16
I.8	(a) Photographie du système Silynx Clarus <sup>™</sup> , (b) Perte d'insertion en	
	champ diffus pour 5 niveaux extérieurs différents	17
I.9	(a) Photographie du casque $3M^{TM}$ Peltor <sup>TM</sup> ComTac <sup>TM</sup> XP, (b) Perte	
	d'insertion en champ diffus pour 5 niveaux extérieurs différents	18
I.10	(a) Photographie du casque MSA Supreme Pro, (b) Perte d'insertion	
	en champ diffus pour 5 niveaux extérieurs différents	19
II.1	Systèmes de coordonnées sphériques utilisés dans cette étude	24
II.2	Différence interaurale d'intensité mesurée en fonction de la fréquence	
	pour une source sonore située à $10^\circ,45^\circ$ et $90^\circ$ sur le plan horizontal	25
II.3	Exemple de différence interaurale d'intensité en fonction de l'azimut $\ .$	26
II.4	Exemple de différence interaurale de temps en fonction de l'azimut	27
II.5	Différence interaurale de temps calculée en millisecondes en fonction de	
	l'azimut pour un son pur à 250 Hz, à 1 kHz, pour une source complexe	
	2-4 kHz et pour le modèle de Woodworth	28
II.6	Représentation de deux cônes de confusion différents et caractérisés par	
	les angles $\theta_R$ pour le rouge, et $\theta_B$ pour le bleu	31

II.7	Distributions de la pression aux fréquences M1 à M5 dans la conque et	
	dans le conduit auditif. Figure tirée de (Shaw & Teranishi, 1968)	32
II.8	Distributions de la pression pour les modes de la conque et du pavillon	
	lorsque le conduit auditif est bouché. Figure tirée de (Shaw, 1997)	33
II.9	Amplification du signal reçu à l'oreille par diffraction du champ acous-	
	tique sur la sphère rigide en fonction de la fréquence	33
II.10	Répartition des bandes directionnelles de Blauert (1969) en fonction de	
	la fréquence.	34
II.11	Exemple de spectres reçus aux oreilles et présentant un creux qui se	
	décale en fréquence en fonction de l'azimut	35
II.12	Photographie du dispositif de mesure d'HRTF de l'ISL	39
II.13	Schéma du dispositif expérimental. La tête est placée au centre du cercle	
	constitué de 16 sources sonores	40
II.14	Schéma de principe de la mesure de fonction de transfert	40
II.15	Spectre des haut-parleurs, identifiés par leur position, après égalisation	
	du niveau global.	41
II.16	Estimation de l'ITD par étude du seuil des HRIR droite et gauche	45
II.17	ITD estimée par la méthode du maximum d'intercorrélation	45
II.18	Fonctions de transfert des 28 gammatones du banc de filtres utilisé pour	
	simuler la tonotopie cochléaire	47
II.19	Exemple de lissage de DTF par la banque de filtres auditifs gammatones.	48
II.20	ITD estimée en fonction de l'azimut à partir des mesures d'HRTF avec	
	et sans TCAPS.	49
II.21	ILD estimée en fonction de l'azimut à partir des mesures d'HRTF avec	
	et sans TCAPS.	51
II.22	ILD estimée à partir d'un modèle de tête sphérique pour lequel la dis-	
	tance entre le point de mesure et l'oreille varie de 0 cm à 4 cm	52
II.23	Indices spectraux avant-arrière ipsilatéraux et contralatéraux en fonc-	
	tion de la fréquence pour chaque TCAPS	53
III.1	Photographies du dispositif expérimental pour les expériences subjec-	
	tives de localisation sonore	58
III.2	Disposition des sources sonores : l'auditeur est placé au centre du cercle	
	constitué de 16 sources sonores	59
III.3	Schéma du dispositif expérimental	59
III.4	Interprétation des erreurs de localisation dans le cas où la source est à	
	la position d'azimut $22,5^{\circ}$	61
III.5	(a) Nombre de localisations exactes pour chaque condition d'écoute, (b)	
	Répartition des nombres moyens de localisations exactes en fonction de	
	l'azimut	63

III.6	(a) Nombre de localisations voisines pour chaque condition d'écoute, (b)	
	Répartition des nombres moyens de localisations voisines en fonction de	
	l'azimut	65
III.7	Représentation des valeurs moyennes des nombres de localisations exactes	
	et voisines dans chaque condition d'écoute	67
III.8	Représentations des nombres d'erreurs avant-arrière suivant la condition	
	d'écoute.	70
III.9	Représentations des nombres d'erreurs grossières suivant la condition	
	d'écoute.	71
III.10	Temps de réponse moyens des auditeurs pour chaque condition d'écoute	
	(a) dans le cas de localisations exactes, (b) pour les localisations voisines,	
	et (c) pour les erreurs de localisation. $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$	74
III.11	Distribution du temps de réponse moyen pour chaque type de TCAPS	
	en fonction de l'azimut de la source, (a) pour les localisations voisines,	
	(b) pour les erreurs avant-arrière, et (c) pour les erreurs grossières	76
III.12	Nombre de localisations exactes et temps de réponse moyens pour chaque	
	condition d'écoute et en fonction de l'âge	78
III.13	Nombre de localisations exactes et temps de réponse moyens pour chaque	
	condition d'écoute et en fonction du genre	79
III.14	Représentation des valeurs moyennes des nombres de localisations exactes	
	réels et estimés dans chaque condition d'écoute	81
III.15	Représentation des valeurs moyennes des nombres de localisations exactes	
	estimés et de la difficulté ressentie dans chaque condition d'écoute	82
IV.1	Schéma de principe du modèle d'appariement.	88
IV.2	Pondérations appliquées aux indices de similarité gauche et droit en	
	fonction de l'azimut.	90
IV.3	Schéma de la projection sur la dimension latérale depuis la dimension	
	azimutale	93
IV.4	Optimisation des paramètres liés aux ITD et ILD pour le modèle d'ap-	
	pariement : (a) $\sigma_{ITD}$ , (b) $\sigma_{ILD}$ et (c) $w_{ITD}$	95
IV.5	Optimisation des paramètres liés aux indices spectraux pour le modèle	
	d'appariement : (a) $\sigma_{\delta}$ et (b) $w_{\delta}$	96
IV.6	Comparaison de la fonction sigmoïde de Baumgartner et al. (2014), de	
	la fonction gaussienne correspondante ( $\sigma_{\delta} = 0,16$ ) et de la fonction	
	gaussienne obtenue au cours du processus d'optimisation ( $\sigma_{\delta}=0,1).$	97
IV.7	Schéma d'un perceptron à $m$ entrées.	98
IV.8	Graphe de la fonction sigmoïde pour différentes pentes $a$	100
IV.9	Schéma d'un réseau de neurones à une couche couchée	100

IV.10	Probabilités de localisation de l'expérience subjective pour chaque couple
	de positions cible et test, dans les sept conditions d'écoute. $\ldots$ 107
IV.11	Sorties du modèle d'appariement pour chaque couple de positions cible
	et test, dans les sept conditions d'écoute. $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $108$
IV.12	Sorties du réseau de neurones pour chaque couple de positions cible et
	test, dans les sept conditions d'écoute
IV.13	Probabilités de localisation voisine, d'erreurs avant-arrière, grossières,
	gauche-droite, mixte et précision de localisation estimées pour l'expé-
	rience subjective, le modèle d'appariement et le réseau de neurones, dans
	chaque condition d'écoute
IV.14	Log-vraisemblance des modèles de localisation calculée à partir des ob-
	servations de l'expérience subjective, dans chaque condition d'écoute 116
IV.15	Schéma de principe du modèle d'appariement modifié
IV.16	Optimisation des paramètres du modèle d'appariement. Composante
	latérale : (a) $\sigma_{ITD}$ , (b) $w_{lat}$ . Composante polaire : (c) $\sigma_{\delta}$
IV.17	(a) Optimisation de la pondération $w_{lat}$ . (b) Écart-type de l'ITD $\sigma_{ITD}$
	optimisé en fonction de l'azimut de la source
IV.18	Écart-type relatif aux indices spectraux $\sigma_{\delta}$ optimisé en fonction de l'azi-
	mut de la source
IV.19	Composante latérale des observations subjectives et de la sortie des
	modèles d'appariement en condition d'écoute normale 125
IV.20	Composantes polaires avant en fonction de l'azimut de la source pour
	la condition d'écoute normale
IV.21	Étude des performances prédites par les différents modèles d'apparie-
	ment dans chaque condition d'écoute. $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $127$
IV.22	(a) Paramètre de concentration $\kappa$ obtenu en fonction de l'azimut à partir
	des résultats de l'expérience subjective en condition d'écoute normale.
	(b) Sortie théorique du réseau de neurones, constituée des distributions
	de von Mises correspondantes
IV.23	Étude des performances prédites en condition d'écoute normale par les
	différents réseaux de neurones
IV.24	Étude de la probabilité de localisation voisine et de la vraisemblance
	obtenues avec les différents réseaux de neurones dans chaque condition
	d'écoute
IV.25	Comparaison des performances de localisation et de la vraisemblance du
	modèle d'appariement et du réseau de neurones retenus, dans chaque
	condition d'écoute
V.1	Photographie du dispositif expérimental du CNAM

V.2	Exemple de l'effet du lissage par liftrage cepstral sur l'amplitude d'une
	fonction de transfert
V.3	Composantes polaires avant, obtenues avec les DTF du KEMAR, en fonction de l'azimut de la source pour la condition d'écoute normale 143
V.4	Probabilités de localisation voisine et vraisemblances des trois modèles d'appariement du mannequin KEMAR pour chaque condition d'écoute. 144
V.5	Étude des performances prédites en condition d'écoute normale par les différents réseaux de neurones basés sur les DTF du KEMAR 146
V.6	Probabilités de localisation voisine et vraisemblances des cinq réseaux de neurones du mannequin KEMAR pour chaque condition d'écoute 147
V.7	Probabilités de localisation voisine, d'erreurs avant-arrière, grossières, gauche-droite, précision de localisation et vraisemblance estimées dans chaque condition d'écoute, pour l'expérience subjective et les modèles d'appariement obtenus avec la tête ISL et le mannequin KEMAR 149
V.8	Probabilités de localisation voisine, d'erreurs avant-arrière, grossières, gauche-droite, précision de localisation et vraisemblance estimées dans chaque condition d'écoute, pour l'expérience subjective et les réseaux de neurones obtenus avec la tête ISL et le mannequin KEMAR 151
V.9	Casque talk-through MSA Supreme <sup>®</sup> Pro : (a) modèle modifié avec un pavillon fermé, (b) modèle modifié avec un pavillon ouvert
V.10	Bouchons d'oreille à atténuation uniforme : (a) Etymotic $\text{ETY} \bullet \text{Plugs}^{\circ}$ ,
	(b) Etymotic $ER \bullet 20^{\circ}$ XS
V.11	Différence interaurale de temps estimée avec et sans audioprothèses 161
V.12	Différence interaurale d'intensité estimée avec et sans audioprothèses 162
V.13	Indices spectraux avant-arrière ipsi- et contralatéraux obtenus à partir d'HRTF mesurées avec et sans audioprothèses
V.14	Métriques $P_{loc}$ , $P_{sev}$ , $P_{fb}$ , $P_{lr}$ et $P_{acc}$ calculées avec et sans prothèses auditives, et obtenues par le modèle d'appariement et le réseau de neu-
	rones
A.1	Coordonnées polaires verticales (a) et interaurales (b)
B.1	Fonction gammatone avec une fréquence centrale à 1 kHz $\hdots$
B.2	Gain d'un filtre gammatone avec une fréquence centrale à 1 kHz $\ .$ 196
D.1	Décomposition de la probabilité au point d'azimut $\theta = 22,5^{\circ}$ en compo- santes latérale ( $\alpha = 22,5^{\circ}$ ) et polaire ( $\beta = 0^{\circ}$ )
D.2	Exemple de décomposition d'une distribution dans la dimension azimu- tale (en haut) en une composante latérale (à gauche) et une composante polaire (à droite).
	r (~ areaco)

E.1	Sorties du modèle d'appariement, optimisé pour la tête artificielle ISL,		
	pour chaque couple de positions cible (Azimut réel) et test (Azimut		
	testé), dans les sept conditions d'écoute		
E.2	Sorties du modèle d'appariement, optimisé pour le manne quin KEMAR, $% \left( {{{\rm{AR}}}} \right)$		
	pour chaque couple de positions cible (Azimut réel) et test (Azimut		
	testé), dans les sept conditions d'écoute		
E.3	Sorties du réseau de neurones, optimisé pour la tête artificielle ISL, pour		
	chaque couple de positions cible (Azimut réel) et test (Azimut testé),		
	dans les sept conditions d'écoute		
E.4	Sorties du réseau de neurones, optimisé pour le mannequin KEMAR,		
	pour chaque couple de positions cible (Azimut réel) et test (Azimut		
	testé), dans les sept conditions d'écoute		

## Liste des acronymes

#### ANOVA

Analysis Of Variance (analyse de variance).

#### $\mathbf{ANR}$

Active Noise Reduction (réduction active du bruit).

#### CNAM

Conservatoire National des Arts et Métiers.

#### DIR

Directional Impulse Response (réponse impulsionnelle directionnelle).

#### DTF

Directional Transfer Function (fonction de transfert directionnelle).

#### $\mathbf{HML}$

High, Middle, Low (cotes à haute, moyenne et basse fréquence).

#### HRIR

Head-Related Impulse Response (réponse impulsionnelle relative à la tête).

#### HRTF

Head-Related Transfer Function (fonction de transfert relative à la tête).

#### $\mathbf{IL}$

Insertion Loss (perte d'insertion).

#### ILD

Interaural Level Difference (différence interaurale d'intensité).

#### ISD

Interaural Spectral Difference (différence interaurale spectrale).

#### $\mathbf{ISL}$

Institut franco-allemand de recherches de Saint-Louis.

#### ISSD

Inter-Subject Spectral Difference (différence spectrale inter-sujet).

#### ITD

Interaural Time Difference (différence interaurale de temps).

#### JND

Just-Noticeable Difference (différence juste perceptible).

#### KEMAR

Knowles Electronics Manikin for Acoustic Research (tête artificielle).

#### MIRE

Microphone In Real Ear (microphone dans l'oreille réelle).

#### $\mathbf{NR}$

Noise Reduction (réduction de bruit).

#### $\mathbf{NRR}$

Noise Reduction Rating (taux de réduction de bruit).

#### PICB

Protections Individuelles Contre le Bruit.

#### REAT

Real-Ear Attenuation at Threshold (atténuation réelle au seuil d'audition).

#### SNR

Single Number Rating (valeur unique d'atténuation).

#### TCAPS

Tactical Communication And Protective System (système tactique de communication et protection).

#### TFOE

Transfer Function of the Open Ear (fonction de transfert de l'oreille ouverte).

## Nomenclature

Le tableau suivant répertorie les différentes protections auditives utilisées dans cette étude, ainsi que les indices utilisés pour les désigner dans le texte :

Indice	Nom du protecteur	Type
P1	$3M^{TM}$ CombatArms <sup>TM</sup>	Bouchon non-linéaire
P2	$SureFire^{\circ}$ EP4 Sonic Defenders <sup>o</sup> Plus	Bouchon non-linéaire
<b>P3</b>	Nacre Quiet $\operatorname{Pro}^{^{\otimes}}$	Bouchon talk-through
<b>P</b> 4	Silynx Clarus <sup><math>TM</math></sup>	Bouchon talk-through
$\mathbf{P5}$	$3M^{TM}$ Peltor <sup>TM</sup> ComTac <sup>TM</sup> XP	Casque talk-through
<b>P6</b>	$\mathrm{MSA}~\mathrm{Supreme}^{^{\textcircled{0}}}~\mathrm{Pro}$	Casque talk-through

### Introduction

La protection de l'audition est d'importance majeure, que ce soit dans l'industrie ou bien sur le champ de bataille. Si, dans les domaines civils, le danger provient principalement de sources de bruit continu, les soldats sont aussi soumis à des bruits impulsionnels à très fort niveau (bruit d'armes à feu, explosions, ...). Quand le contrôle de la source de bruit n'est pas possible, le port de protections auditives individuelles est requis. Les différents bouchons d'oreille et casques qui ont été développés procurent l'atténuation nécessaire du bruit, en étant portés simultanément si besoin est. Cependant, la préservation de l'audition ne peut pas être prise en considération en excluant les problématiques de sécurité. Dans les années 90, Killion (1993) critique une politique du « Parvum Bonum, Plus Melius » (un peu c'est bien, plus c'est mieux) vis-à-vis de l'atténuation des protections auditives. En effet, une atténuation trop importante du bruit entraîne une surprotection de l'auditeur, qui l'empêche d'entendre correctement des sons importants. Par exemple, dans l'industrie, les travailleurs doivent pouvoir percevoir des signaux d'alarmes, ou de recul d'un véhicule. Dans le cas des militaires, les alarmes et menaces nécessitent aussi d'être détectées, ainsi que les ordres et, de manière plus générale, la parole. Ce dilemme entre santé et sécurité doit être résolu par les protections auditives de nouvelle génération.

Au-delà de la détection des signaux importants, l'ensemble des capacités permettant à l'auditeur de percevoir et d'interpréter son environnement sonore, appelé *situational awareness* en anglais, doit être conservé. Un système idéal de protection auditive devrait donc remplir plusieurs conditions :

- l'atténuation des bruits continus et impulsionnels à fort niveau;
- la préservation des sons à bas niveau, comme la parole;
- l'intelligibilité de la parole;
- le respect du contenu spectral de la source sonore pour permettre sa reconnaissance;
- la préservation de la capacité à localiser la position de la source sonore.

Ce type de protecteur idéal peut être grossièrement résumé à un système acoustiquement transparent, à l'exception de l'atténuation des bruits à fort niveau. Les Systèmes Tactiques de Communication et Protection (TCAPS) sont des protections auditives récentes qui présentent une atténuation qui varie en fonction du niveau sonore extérieur. Ils sont ainsi particulièrement adaptés à la protection contre les bruits impulsionnels, et préservent les sons à bas niveau et l'intelligibilité de la parole. Par conséquent, notre étude se concentre principalement sur les TCAPS. Le premier chapitre de ce mémoire présente plus en détail les enjeux de la lutte contre le bruit, ainsi que les caractéristiques des protections auditives conventionnelles et des TCAPS.

L'étude de la préservation de la capacité à localiser une source sonore avec les TCAPS est l'objet principal de ce travail de thèse. La localisation sonore constitue en effet un point crucial de l'interprétation de l'environnement sonore. Tout d'abord, on conçoit qu'il est important pour un soldat de pouvoir déterminer la position d'une menace, afin d'y répondre rapidement et en sécurité. Mais cette capacité permet aussi d'améliorer l'intelligibilité. Pour finir, la dégradation de la localisation sonore demeure une raison majeure pour laquelle le personnel présente une tendance à ne pas porter ses protections auditives individuelles. Dans le second chapitre, les mécanismes de la localisation sonore humaine dans le plan horizontal sont développés. La restriction spatiale a été choisie, car être capable de localiser une source sonore dans le plan horizontal semble spécialement crucial pour la sécurité de l'auditeur. Cette capacité se base sur l'exploitation de trois indices acoustiques de localisation, qui peuvent être mesurés sur une tête artificielle. Une analyse qualitative de la modification de ces indices par six TCAPS est alors effectuée dans ce chapitre.

L'étude de l'effet des TCAPS sur la localisation sonore se réalise généralement à travers une expérience subjective, au cours de laquelle on demande à des auditeurs de localiser des sources sonores avec et sans protection. Dans notre étude, une telle expérience a été conduite avec 30 auditeurs volontaires et six TCAPS différents à l'Institut franco-allemand de recherches de Saint-Louis. Le protocole expérimental et les résultats sont présentés dans le troisième chapitre de ce mémoire. Nous analysons non seulement l'impact des TCAPS sur le nombre de bonnes réponses, mais aussi sur les différents types d'erreurs de localisation et sur le temps de réponse. Pour finir, l'influence de l'âge et du genre de l'auditeur et son ressenti quant à la difficulté du test sont examinés.

La réalisation et l'analyse des résultats d'une expérience comportementale sont toutefois très chronophages. Dans le cadre du développement d'un prototype de TCAPS préservant la localisation sonore, il ne semble pas faisable de l'évaluer de cette façon. L'objectif principal de l'étude présentée dans ce mémoire est donc le développement d'une méthode d'évaluation de l'effet des TCAPS sur la localisation sonore, basée sur des mesures électroacoustiques. Dans le quatrième chapitre, deux modèles de localisation sont décrits. Le premier, appelé modèle d'appariement, est basé sur l'hypothèse que le système auditif compare les indices de localisation propres à la source sonore à une base de données d'indices en condition d'écoute normale. Le second modèle repose sur un réseau de neurones à trois couches , qui exploite en entrée les indices de localisation pour déterminer en sortie la position de la source. Avec ces deux modèles, nous obtenons une prédiction des performances de localisation sonore en fonction de la condition d'écoute. Ils sont donc appliqués à l'évaluation des TCAPS du point de vue de la localisation. Les résultats de prédiction obtenus sont alors comparés aux observations de l'expérience subjective précédente. La dernière partie de ce chapitre se concentre sur des tentatives d'amélioration des deux modèles.

Le fonctionnement du modèle d'appariement et du réseau de neurones repose sur des mesures sur une tête artificielle, qui représente un auditeur type. La prédiction des performances de localisation avec des TCAPS doit donc être indépendante de la tête artificielle utilisée. En conséquence, la première partie du cinquième chapitre est consacrée à l'examen des différences entre les prédictions du chapitre précédent et celles obtenues sur une deuxième tête artificielle au Conservatoire National des Arts et Métiers. Pour terminer, dans la seconde partie de ce cinquième chapitre sont abordés des exemples d'application des modèles de localisation sonore. Tout d'abord, nous proposons une tentative triviale d'amélioration d'un casque en lui ajoutant un pavillon d'oreille fictif. Une évaluation de bouchons d'oreilles à atténuation uniforme, principalement développés pour les musiciens, est ensuite effectuée grâce aux modèles. Enfin, nous nous intéressons aux effets de deux prothèses auditives sur les performances de localisation sonore.

## CHAPITRE I

### Contexte de l'étude

Ce premier chapitre est consacré à la description des protections auditives. Dans un premier temps, après une rapide introduction sur la réglementation du bruit dans les domaines civil et militaire, les caractéristiques des protections auditives conventionnelles sont détaillées. De plus, nous décrivons les différentes méthodes de mesure de l'atténuation apportée par ces protections. Dans un second temps, la notion de connaissance de la situation, ou *situational awareness*, est introduite et mène à la présentation des systèmes tactiques de communication et protection utilisés dans cette étude.

#### 1 Lutte contre le bruit dans les domaines civil et militaire

#### 1.1 Réglementation

La conservation de l'audition des personnes est capitale. Les personnes atteintes de pertes auditives ne sont plus capables de communiquer correctement avec leurs proches et se sentent isolées de leur environnement sonore. De plus, la surdité professionnelle a un coût très élevé pour la société, comme indiqué dans une récente étude du Conseil National du Bruit (CNB, 2016). En effet, en 2014, la Caisse Primaire d'Assurance Maladie a estimé à 85 millions d'euros le coût de ces maladies, coût que l'on peut ramener à 103 000  $\in$  par personne prise en charge. Le coût répercuté sur les entreprises est estimé en moyenne à 94 000  $\in$ , selon l'Institut National de Recherche et de Sécurité. Ces facteurs devraient inciter à la fois les employeurs et les employés à améliorer la sécurité en environnement bruyant.

Dans le cas où le bruit au travail ne peut pas être contrôlé (réduction du bruit des machines, travail dans des cabines isolées, ...), les personnes soumises à ce bruit doivent porter des Protections Individuelles Contre le Bruit (PICB), de type bouchon d'oreille ou casque antibruit. Le Code du travail<sup>1</sup> prévoit trois actions de prévention différentes en

<sup>1.</sup> Article R4431-2, Créé par Décret n°2008-244 du 7 mars 2008 - art. $\left(\mathrm{V}\right)$ 

fonction des niveaux d'exposition en bruit continu ou impulsionnel. L'exposition au bruit continu est estimée par la moyenne sur une journée de huit heures du niveau sonore  $L_{(eq,8h)}$ avec la pondération fréquentielle A. Le niveau sonore instantané maximal  $L_{peak}$  mesuré avec la pondération C représente l'exposition au bruit impulsionnel<sup>2</sup>. Les trois actions de prévention et leur limite d'application sont les suivantes :

- 1. Exposition moyenne  $L_{(eq,8h)}$  supérieure à 80 dB(A) ou Niveau de crête  $L_{peak}$  supérieur à 135 dB(C)
  - Mise à disposition des PICB
  - Information et formation aux risques
  - Examen audiométrique proposé
- 2.  $L_{(eq,8h)} \ge 85 \text{ dB}(A) \text{ ou } L_{peak} \ge 137 \text{ dB}(C)$ 
  - Mesures de réduction d'exposition au bruit
  - Signalisation des endroits concernés et limitation d'accès
  - Contrôle de l'utilisation des PICB
  - Audiométrie obligatoire
- 3.  $L_{(eq,8h)} \ge 87 \text{ dB}(A)$  ou  $L_{peak} \ge 140 \text{ dB}(C)$ 
  - Valeur limite d'exposition, à ne pas dépasser
  - Mesure prenant en compte l'atténuation des protections auditives

Ces différentes limites s'appliquent aussi dans le domaine militaire. Les véhicules utilisés y constituent les principales sources d'exposition au bruit continu. Par exemple, des niveaux sonores de 106 dB(A) peuvent être observés à l'intérieur d'hélicoptères et des niveaux pouvant atteindre 117 dB(A) ont été mesurés dans des chars (Humes et al., 2006, chap. 3). Ces valeurs étant similaires à celles rencontrées en milieu industriel, la principale différence réside alors dans les sources de bruit impulsionnel.

Alors que dans une forge, par exemple, des bruits impulsionnels maximaux d'environ 144 dB(C) ont été mesurés (Terroir, 2016), des niveaux jusqu'à 190 dB(peak) (amplitude du pic de pression) mesurés à la position du tireur peuvent être observés sur le champ de bataille pour des armes de type anti-blindage et anti-personnel (Humes et al., 2006, Chapitre 3) ou des tirs de mortier (Talcott, 2011). La valeur limite d'exposition au bruit impulsionnel est donc largement dépassée, même lors de l'utilisation d'armes de petit calibre, type revolver 9 mm, dont le bruit peut atteindre 157 dB(peak) au niveau des oreilles du tireur.

L'utilisation et l'optimisation des protections individuelles contre le bruit sont donc cruciales, particulièrement dans le cas du bruit impulsionnel en milieu militaire. Dans la suite de cette section, nous décrivons les PICB dites « conventionnelles », ainsi que les méthodes permettant de les caractériser.

<sup>2.</sup> Article R4431-1 du Code du travail

#### **1.2** Protections auditives conventionnelles

Les protections individuelles contre le bruit peuvent avoir deux formes différentes. D'un côté, les casques (Figure I.1a) sont composés de deux coquilles montées sur un serre-tête. Chacune des coquilles englobe une oreille de l'auditeur et s'ajuste hermétiquement sur son crâne. De manière générale, elles se composent d'une coque en plastique contenant un absorbant acoustique, sous forme de mousse. De l'autre, les bouchons d'oreilles sont, comme leur nom l'indique, à insérer dans le conduit auditif. On les trouve sous forme de mousse à mémoire de forme (Figure I.1b), sous forme de bouchon en polymère avec un design à double ou triple ailette (Figure I.1c), ainsi que sous forme de bouchons d'oreilles et les casques peuvent être portés simultanément pour former une double protection auditive.

Les bouchons d'oreilles ont l'avantage d'être plus discrets que les casques. En particulier, ils s'adaptent bien à l'utilisation de protections individuelles différentes, comme des lunettes. En effet, ces dernières pourraient provoquer un défaut d'étanchéité du casque et diminuer l'atténuation du bruit. Toutefois, les systèmes de protection plus récents prennent en général ces problèmes en compte. D'un autre côté, le principal défaut des bouchons d'oreilles est la difficulté de leur mise en place, excepté pour les bouchons personnalisés. Les auditeurs ont en général tendance à ne pas assez enfoncer les bouchons en mousse dans leur conduit auditif. Si les bouchons à triple ailette sont plus faciles à mettre, le problème réside dans le choix de la taille du protecteur. Une mauvaise pose ou le choix de la mauvaise taille peuvent donc fortement réduire leur efficacité. Il devient donc nécessaire de former les personnels à l'utilisation de ces PICB.

Les différentes PICB présentées sur la Figure I.1 sont dites passives, car ils permettent l'atténuation du bruit sans apport d'énergie extérieur. Il existe aussi des systèmes actifs pour lesquels un apport extérieur d'énergie, principalement électrique, est nécessaire à l'atténuation. Les plus connus sont les systèmes à réduction active du bruit, ou *Active Noise Reduction* (ANR). Leur principe repose sur la production d'interférences destructrices : ils produisent un contre-bruit de même amplitude, mais en opposition de phase par rapport au bruit à atténuer. Ce type de protection auditive est particulièrement efficace à basse fréquence. Il ne fait toutefois pas partie de l'objet de notre étude.

Les protections auditives conventionnelles sont caractérisées par l'atténuation du bruit qu'elles apportent. Cette atténuation s'exprime en décibels et dépend de la fréquence. Toutefois, l'atténuation est souvent réduite à un chiffre unique qui doit être indiqué par le fabricant du protecteur. Par exemple, la norme américaine (ANSI S3.19-1974, 1974) décrit le calcul du *Noise Reduction Rating* (NRR) et la norme internationale (ISO 4869-2 :1994, 1994) celui du *Single Number Rating* (SNR). Cette dernière norme introduit aussi les cotes *High, Middle, Low* (HML) qui donnent des valeurs d'atténuation séparées pour les fréquences graves, médiums et aigües. Le détail du calcul de ces différents indices peut être trouvé dans (NIOSH, 2016). À titre d'exemple, le bouchon d'oreille en mousse de la



(a) Casque 3M<sup>TM</sup> PELTOR<sup>TM</sup> X5A



(b) Bouchons 3M<sup>TM</sup> E-A-R<sup>TM</sup> Classic<sup>TM</sup>



(c) Bouchons 3M<sup>TM</sup> E-A-R<sup>TM</sup> Ultrafit<sup>TM</sup> X (d) Exemple de bouchons personnalisés

FIG. I.1 – Photographies de protections individuelles contre le bruit conventionnelles.

Figure I.1b possède, d'après le fabricant, un NRR de 29 dB, un SNR de 28 dB et des cotes HML respectivement égales à 30, 24 et 22 dB.

Dans le cadre de notre étude, les NRR, SNR ou autres HML ne sont donnés qu'à titre indicatif. Nous nous intéressons plutôt à l'évolution de l'atténuation des protections auditives en fonction de la fréquence. Trois méthodes différentes de mesure de cette atténuation sont décrites dans la section suivante.

#### 1.3 Mesure(s) de l'atténuation des protections auditives

Lorsque l'on parle d'atténuation d'une protection auditive, il faut distinguer deux quantités acoustiques différentes (Berger, 1986, 2005) : la réduction de bruit, notée *Noise Reduction* (NR), et la perte d'insertion, notée *Insertion Loss* (IL). La différence entre les deux est expliquée par la Figure I.2. La NR est égale à la différence entre les niveaux mesurés à l'entrée et à la sortie de la PICB. À l'inverse, pour calculer l'IL, on mesure tout d'abord le niveau reçu au tympan sans protection, que l'on soustrait par le niveau obtenu avec protection. La nuance entre ces mesures réside dans la prise en compte de la



FIG. I.2 – Principe de la mesure de la réduction de bruit et de la perte d'insertion : NR = B' - A' et IL = A - A'. Figure tirée de (Berger, 1986).

fonction de transfert de l'oreille ouverte ou Transfer Function of the Open Ear (TFOE). En reprenant les notations de la Figure I.2, on a TFOE = A - B et IL = NR + TFOE. D'après Berger (2005), la perte d'insertion représente mieux l'efficacité de la protection auditive, le principal avantage de la NR étant la possibilité d'être mesurée pour des sources impulsionnelles. Nous nous intéressons donc à la mesure de la perte d'insertion.

Il existe deux méthodes majeures de mesure de la perte d'insertion d'une protection auditive. La première se base sur l'évaluation des seuils d'audition d'un sujet sans et avec la PICB, et s'appelle la méthode d'atténuation réelle au seuil d'audition, ou *Real-Ear Attenuation at Threshold* (REAT). Elle est normalisée dans (ANSI S12.6-2016, 2016) et (ISO 4869-1 :1990, 1990). L'avantage de cette méthode est d'estimer les performances réelles de la protection sur l'auditeur : elle prend en compte l'atténuation du protecteur, la morphologie de l'auditeur et la mise en place du protecteur. Elle permet donc la prise en compte de variabilités interindividuelles. C'est généralement cette méthode qui est utilisée par les fabricants pour pouvoir ensuite calculer les NRR, SNR ou HML et les indiquer sur leur produit. La méthode REAT générale nécessite l'installation d'un système d'audiométrie en champ diffus dans une cabine insonorisée. Toutefois, pour mesurer la perte d'insertion avec des bouchons d'oreilles, il « suffit » d'un simple audiomètre avec un casque audio assez profond pour pouvoir porter les bouchons en dessous.

La seconde méthode s'appuie aussi sur une mesure sur sujet, mais ne repose pas sur une audiométrie ou une quelconque réponse du sujet à un test. Pour mesurer la perte d'insertion, une sonde est placée entre la PICB et le tympan de l'auditeur. Il s'agit de la méthode à microphone dans l'oreille réelle, ou *Microphone In Real Ear* (MIRE), détaillée dans la norme (ANSI S12.42-2010, 2010). Contrairement à la méthode REAT, elle ne dépend plus de la manière dont l'auditeur répond au test d'audiométrie, et permet donc une mesure plus fiable de l'IL. De plus, elle peut être appliquée pour différents niveaux sonores, alors que la méthode REAT est fixée au seuil d'audition. En revanche, le principal défaut de la méthode MIRE demeure le placement de la sonde entre la protection et le tympan de l'auditeur. Pour un casque ou pour des bouchons d'oreilles, cette sonde peut en effet provoquer une fuite acoustique qui conduirait à une mesure erronée de l'atténuation. Certains fabricants utilisent alors des versions instrumentées de leur PICB et une correction de la fuite pour mesurer l'atténuation avec la méthode MIRE.

Pour finir, la mesure de la perte d'insertion d'une PICB, lorsque l'auditeur est soumis à des bruits continus ou impulsionnels à très forts niveaux, n'est pas possible avec la méthode MIRE standard. On ne peut effectivement pas réaliser la mesure sans protection. Dans ce cas, c'est la réduction de bruit NR qui est mesurée et corrigée par une TFOE. Une autre méthode, dérivée de la méthode MIRE, repose sur l'utilisation d'une tête artificielle à la place d'un auditeur réel. Elle est décrite dans les normes (ANSI S12.42-2010, 2010) et (ISO 4869-3 :2007, 2007). Cette méthode est très attirante puisqu'elle ne nécessite plus d'expérimentation avec des sujets humains. Mais c'est justement le confort de ces sujets, leur morphologie, l'effet de la conduction osseuse et la mise en place de la protection qui influencent substantiellement les performances des PICB (Berger, 2005). De plus, les têtes artificielles n'offrent qu'une simulation du système auditif. Malgré ces limitations considérables, cette méthode permet au moins d'assurer la répétabilité des mesures.

Les dispositifs expérimentaux disponibles pour notre étude ne permettant pas l'application des méthodes REAT ou MIRE standards, l'atténuation de protecteurs auditifs sera évaluée en utilisant une tête artificielle. Celle-ci a été développée par l'Institut francoallemand de recherches de Saint-Louis (ISL) pour répondre aux exigences des normes américaines et internationales pour ce type de mesure en champ impulsionnel (Buck & Parmentier, 1999; Parmentier et al., 2000). Dans notre cas, la mesure de perte d'insertion est réalisée en champ diffus stationnaire avec une source sonore à large bande fréquentielle [40 Hz – 10 kHz]. Pour cela, le stimulus sonore utilisé est un bruit rose (spectre plat en tiers d'octave) et la tête artificielle est placée dans une chambre réverbérante. Ce dispositif est présenté sur la Figure I.3. Le dispositif de sonorisation de la chambre offre la possibilité de mesurer la perte d'insertion pour un niveau sonore extérieur variant entre 70 et 120 dB(SPL).

#### **1.4** Perte d'insertion de protecteurs conventionnels

La Figure I.4 présente les pertes d'insertion du casque et des bouchons d'oreille  $3M^{TM}$  de la Figure I.1, mesurées sur tête artificielle en chambre réverbérante, avec un niveau extérieur de 100 dB(SPL). Les bouchons personnalisés ne font pas l'objet de cette mesure puisqu'ils ne sont pas adaptés à la tête artificielle. Les bouchons d'oreille sont plus efficaces que le casque à basse fréquence puisqu'ils apportent une atténuation d'environ 30 dB entre 100 et 400 Hz. Cette différence entre bouchons et casques est quasiment toujours observée.



FIG. I.3 – Photographies du dispositif de mesure de la perte d'insertion d'une protection auditive sur tête artificielle en champ diffus.

Elle est due à la cavité formée par le casque autour de l'oreille et à un effet de pompage amplifié suivant la souplesse du coussin (Zannin & Gerges, 2006; Boyer et al., 2014).

À plus haute fréquence, les bouchons restent globalement plus performants que le casque, avec une atténuation comprise entre 40 et 60 dB. Mais le casque offre tout de même une perte d'insertion supérieure à 30 dB. L'atténuation maximale, environ 60 dB, est obtenue avec le bouchon 3M<sup>TM</sup> E-A-R<sup>TM</sup> Classic<sup>TM</sup> pour les fréquences aux alentours de 2 kHz. Ce bouchon en mousse est le plus performant des trois, mais il s'agit aussi de celui qui se dégrade le plus facilement : idéalement, il doit être changé tous les jours, alors que le bouchon en polymère et le casque sont réutilisables.

Il semble que la perte d'insertion s'effondre lorsque la fréquence dépasse 10 kHz, quelle que soit la protection. Cet affaiblissement n'est en général pas observé lorsque la méthode REAT est employée. L'atténuation devrait soit rester à peu près constante, soit augmenter avec la fréquence. L'effet observé ici est en réalité causé par une mauvaise efficacité du système de sonorisation à haute fréquence : le niveau mesuré dans la chambre réverbérante par un microphone de contrôle en fonction de la fréquence est compris entre 85 et 90 dB de 100 Hz à 3 kHz. Il chute ensuite jusqu'à 60 dB à 16 kHz, biaisant l'estimation de la perte d'insertion.

En conclusion, les protections auditives classiques de type bouchons d'oreille ou casque peuvent apporter une atténuation suffisante du bruit auquel l'auditeur peut être confronté, que ce soit dans le domaine industriel ou dans le domaine militaire. Idéalement, les porter simultanément permet de sommer leur perte d'insertion et donc de soumettre l'auditeur à des niveaux sonores plus élevés, sans pour autant que son audition soit endommagée. Le choix de la protection auditive à utiliser doit toutefois être réfléchi. En cas de surprotection, l'auditeur peut se retrouver isolé de son environnement sonore, c'est-à-dire être bien protégé, mais ne plus entendre les sons utiles.


FIG. I.4 – Perte d'insertion, en fonction de la fréquence (100 Hz-16 kHz), de trois protecteurs conventionnels mesurée sur tête artificielle en chambre réverbérante. Niveau extérieur : 100 dB(SPL).

# 2 Protection auditive et situational awareness

Si une protection auditive empêche la dégradation de l'audition par les bruits dangereux, elle empêche aussi l'auditeur d'entendre des sons plus faibles, comme la voix. La perception et l'interprétation de l'environnement sonore sont des capacités résumées sous le terme de connaissance de la situation ou *situational awareness*, et à présent considérées comme cruciales pour l'efficacité et la survie de l'auditeur. Pour ces raisons, les auditeurs peuvent décider de ne pas porter leurs PICB et d'exposer leur audition à des environnements dangereux (Bevis et al., 2014). Le premier élément définissant cette connaissance de la situation est la détection des signaux. Les seuils d'audition étant augmentés par les protections auditives conventionnelles, cette détection est détériorée (Casali et al., 2004; Alali, 2011).

L'élément le plus étudié concernant le *situational awareness* est la compréhension de la voix. Il est effectivement concevable que l'intelligibilité puisse être réduite par les bouchons d'oreille ou les casques, puisque l'audibilité est dégradée. Toutefois, les PICB doivent être portées en environnement bruyant et certaines études montrent que dans ce cas, la compréhension de la parole est conservée, voire améliorée, par les protections (Howell & Martin, 1975; Abel et al., 1982, 1993; Lambourg et al., 2016). En effet, si le rapport signal sur bruit n'est pas affecté par la protection, l'effet de masquage du bruit sur la parole diminue avec le niveau sonore perçu. Il semble tout de même que dans le cadre de conditions d'écoute particulières, l'effet des PICB puisse être inversé (McKinley, 2000; Casali et al., 2004). De plus, pour des auditeurs présentant des pertes auditives, les protecteurs ont plus

tendance à dégrader l'intelligibilité (Abel et al., 1982; Suter, 1989; Abel et al., 1993). Pour finir, on peut considérer que les militaires, plutôt sujets à des sources de bruit impulsionnel, se trouvent « la plupart du temps » en environnement silencieux. Les PICB doivent donc avoir un impact négatif sur leur compréhension de la parole.

Dans le but de préserver la capacité des auditeurs à détecter des bruits faibles et à comprendre la parole, tout en leur assurant une protection suffisante contre les bruits à fort niveau, de nouveaux types de protection auditive ont été développés.

# 2.1 Systèmes Tactiques de Communication et Protection

À la différence des PICB conventionnelles, ces nouveaux protecteurs présentent une perte d'insertion qui dépend du niveau sonore : les sons faibles comme la voix sont peu atténués, tandis que l'atténuation est maximale pour les sons forts (Martin, 1979; Humes, 1983; Wilde & Humes, 1987). Ce type de protection est bien adapté à la protection en environnement impulsionnel. Nous appelons globalement ce type de protecteur auditif les systèmes tactiques de communication et protection ou *Tactical Communication And Protective System* (TCAPS), ils sont souvent associés aux systèmes de radiocommunication. Dans cette partie, nous décrivons trois types de TCAPS différents : les bouchons non linéaires, les bouchons talk-through et les casques talk-through. Pour chaque type, deux systèmes sont présentés plus en détail. Notamment, nous présentons des mesures de perte d'insertion pour différents niveaux sonores extérieurs, obtenues sur tête artificielle en chambre réverbérante (cf. Figure I.3). Ces six TCAPS seront ceux utilisés tout au long de notre étude.

# 2.1.1 Bouchons d'oreille non linéaires

Il est possible d'obtenir une atténuation dépendant du niveau sonore extérieur en utilisant la non-linéarité acoustique d'un orifice (Ingard & Ising, 1967). Lorsqu'un filtre exploitant cette propriété est soumis à une onde sonore d'amplitude élevée, il entraîne l'apparition d'un flux turbulent augmentant l'impédance du protecteur, principalement pour les hautes fréquences. Les premiers protecteurs exploitant cette propriété sont apparus au début du XXe siècle (Mallock-Armstrong, 1915). En général, ces filtres se présentent sous la forme de cylindres qui sont alors adaptés à des bouchons d'oreille conventionnels (il n'existe aucun « casque non linéaire » à notre connaissance). Ce type de protecteur a l'avantage d'être ouvert, puisqu'il doit laisser passer les sons faibles. Il permet donc aussi de réduire l'effet d'occlusion. Dans notre étude, deux modèles différents sont testés : les bouchons  $3M^{TM}$  CombatArms<sup>TM</sup> et SureFire<sup>®</sup> EP4 Sonic Defenders<sup>®</sup> Plus.

**3M<sup>™</sup> CombatArms<sup>™</sup> - Génération 4.1** Ce bouchon d'oreille en polymère, représenté sur la Figure I.5a, est actuellement en dotation dans l'armée française. Il comprend un filtre non linéaire, développé à l'Institut franco-allemand de recherches de Saint-Louis



FIG. I.5 – (a) Photographie du bouchon  $3M^{TM}$  Combat $Arms^{TM}$  - Génération 4.1, (b) Perte d'insertion en champ diffus pour 5 niveaux extérieurs différents

(Dancer et al., 1999; Hamery et al., 2000), permettant d'obtenir une atténuation dépendante du niveau sonore. Ce filtre peut être fermé dans les zones de bruit continu, à l'aide d'un interrupteur à bascule (voir Figure I.5a), afin d'obtenir un bouchon d'oreille conventionnel. La bonne mise en place du bouchon dans l'oreille est assurée par un design à triple ailette et une languette s'insérant dans la conque. Dans la suite, ce bouchon est appelé P1.

D'après la documentation fournie par le constructeur, ce bouchon non linéaire présente un SNR de 16 dB en mode ouvert, pour lequel principalement les hautes fréquences sont atténuées, et de 28 dB en mode fermé. L'atténuation dans ce dernier mode est relativement plate entre les hautes ( $H = 28 \ dB$ ), moyennes ( $M = 25 \ dB$ ) et basses fréquences ( $L = 24 \ dB$ ). Enfin, la mesure de la réduction du bruit impulsionnel révèle la variation d'atténuation en fonction du niveau, allant de 4,8 dB pour un pic sonore de 110 dB, à 30,2 dB pour un pic de 190 dB. La perte d'insertion de ce bouchon d'oreilles en bruit continu a été mesurée en chambre réverbérante pour un niveau sonore extérieur variant de 70 dB(SPL) à 110 dB(SPL). Le résultat est représenté sur la Figure I.5b en fonction du niveau sonore n'est effective qu'à haute fréquence (au-dessus de 5 kHz).

SureFire<sup>®</sup> EP4 Sonic Defenders<sup>®</sup> Plus Le bouchon d'oreille présenté sur la Figure I.6a comprend un filtre non linéaire Hocks-Noise-Braker<sup>®</sup> permettant d'obtenir une atténuation dépendante du niveau sonore. Ce filtre peut être fermé dans les zones de bruit continu, à l'aide d'un bouchon en plastique, afin d'obtenir un bouchon d'oreille « classique ». La bonne mise en place du bouchon dans l'oreille est assurée par un design à triple ailette en polymère et un anneau de maintien EarLock<sup>®</sup> se plaçant dans la conque. Dans la suite, ce bouchon est appelé P2.

D'après la documentation fournie par le constructeur, ce bouchon non linéaire présente un NRR de 12 dB en mode ouvert, pour lequel principalement les hautes fréquences sont atténuées, et de 24 dB en mode fermé. L'atténuation dans ce dernier mode est relativement



FIG. I.6 – (a) Photographie du bouchon SureFire<sup>®</sup> EP4 Sonic Defenders<sup>®</sup> Plus, (b) Perte d'insertion en champ diffus pour 5 niveaux extérieurs différents

plate entre les hautes (37,7 dB à 8 kHz), moyennes (35,6 dB à 2 kHz) et basses fréquences (31,7 dB à 125 Hz). Enfin, la mesure de la réduction du bruit impulsionnel révèle la variation d'atténuation en fonction du niveau, allant de 23,9 dB pour un pic sonore de 132 dB, à 30 dB pour un pic de 168 dB. La perte d'insertion de ce bouchon d'oreilles en bruit continu a été mesurée en chambre réverbérante pour un niveau sonore extérieur variant de 70 dB(SPL) à 110 dB(SPL). Le résultat est représenté sur la Figure I.6b en fonction de la fréquence. On observe très clairement que la variation d'atténuation en fonction du niveau sonore n'est effective qu'à haute fréquence (au-dessus de 4 kHz).

# 2.1.2 Bouchons d'oreille talk-through

Les bouchons non linéaires sont des systèmes passifs, car ils ne nécessitent aucune électronique. À l'inverse, les systèmes talk-through obtiennent une variation d'atténuation avec le niveau extérieur par traitement du signal reçu par un microphone placé à l'entrée du protecteur. Il s'agit donc de systèmes actifs, qu'il est important de ne pas confondre avec les systèmes de réduction active du bruit. Dans le cas des bouchons talk-through, l'oreille est totalement protégée, et on trouve généralement un microphone au niveau de l'oreille de l'auditeur et un haut-parleur, permettant de bloquer les bruits dangereux et de restituer les bruits faibles. Les bouchons talk-through sont le plus souvent compatibles avec les systèmes de radiocommunication. De plus, avec une bande passante assez large, ils permettent d'obtenir des qualités sonores très correctes (Killion et al., 2011). Dans notre étude, les deux bouchons talk-through utilisés sont les bouchons Nacre QuietPro<sup>®</sup> et Silynx Clarus<sup>TM</sup>.

**Nacre QuietPro**<sup>®</sup> Ce système électronique de protection auditive et de communication, représenté sur la Figure I.7a, propose une atténuation dépendante du niveau extérieur et obtenue par compression du signal reçu au microphone placé à l'entrée de l'oreille et restitué par un haut-parleur (talk-through). La faible latence du système permet de protéger



FIG. I.7 – (a) Photographie du système Nacre Quiet Pro $^{\circ},$  (b) Perte d'insertion en champ diffus pour 5 niveaux extérieurs différents

correctement l'utilisateur contre les bruits impulsionnels de fort niveau. Les oreilles de l'auditeur sont bouchées par des embouts en mousse présentant un NRR de 27 dB (données d'atténuation fournies par le constructeur). Ils ne laissent passer que la sortie du hautparleur, pour le bon fonctionnement des systèmes talk-through et de radiocommunication.

Un réglage d'amplification est possible et permet à l'utilisateur d'augmenter ses capacités auditives, par exemple pour mieux entendre une autre personne parler ou détecter des bruits faibles. Dans toutes nos études, nous ne considérons pas l'effet de cette amplification et l'avons réglée au minimum. Enfin, l'allumage du système talk-through ne peut se faire qu'après l'insertion des bouchons d'oreilles, car l'appareil réalise une mesure d'étanchéité pour assurer leur bonne mise en place. Cette mesure consiste principalement en l'estimation du niveau sonore à l'intérieur du conduit auditif pendant que le haut-parleur envoie un signal composé de deux sinus à environ 15 et 580 Hz. Lorsqu'un bouchon est mal positionné, c.-à-d. lorsqu'une fuite est présente, le niveau dans le conduit ne provient que de la composante à 580 Hz et se trouve donc amoindri. Le système QuietPro<sup>®</sup> renvoie donc une alarme à l'utilisateur pour lui indiquer ce défaut de placement. Dans la suite, ce bouchon est appelé P3.

La variation de l'atténuation en fonction du niveau extérieur pour les bruits impulsionnels n'étant pas donnée, la perte d'insertion en bruit continu a tout de même été mesurée pour des niveaux extérieurs allant de 70 dB à 110 dB. Ces résultats sont représentés sur la Figure I.7b. On remarque tout d'abord que ce protecteur est très efficace en basse fréquence avec une perte d'insertion d'environ 20 dB à faible niveau et de près de 50 dB à fort niveau. L'atténuation commence à augmenter pour des niveaux supérieurs à 80 dB, et varie de manière significative avec le niveau : à 1 kHz, elle vaut 6 dB à 70 dB, 24 dB à 90 dB et 60 dB à 110 dB. À partir de 100 dB, la perte d'insertion semble avoir atteint



FIG. I.8 – (a) Photographie du système Silynx Clarus<sup>TM</sup>, (b) Perte d'insertion en champ diffus pour 5 niveaux extérieurs différents

un maximum pour les basses fréquences, mais le protecteur continue de compenser l'augmentation du niveau extérieur pour les fréquences supérieures à 400 Hz. On remarque que le système talk-through permet d'obtenir une atténuation presque plate entre 500 Hz et 4 kHz. Toutefois, à partir de 110 dB, elles sont fortement atténuées : on retrouve en fait la perte d'insertion engendrée par le bouchon seul, lorsque le système talk-through est éteint.

Silynx Clarus<sup>™</sup> Le système présenté en Figure I.8a propose une protection contre les bruits continus et impulsionnels, combinée à un système talk-through en environnement calme ou bruyant. À l'instar du système précédent, la variation d'atténuation est réalisée par compression du signal sonore capté par un microphone situé à l'entrée de l'oreille et restitué par un haut-parleur. La mise en place dans l'oreille est assurée par un embout en mousse présentant un NRR de 25 dB.

Un réglage d'amplification est possible et permet à l'utilisateur d'augmenter ses capacités auditives, par exemple pour mieux entendre une autre personne parler ou détecter des bruits faibles. Dans toutes nos études, nous ne considérons pas l'effet de cette amplification et l'avons réglée au minimum. Dans la suite, ce bouchon est appelé P4.

La variation de l'atténuation en fonction du niveau extérieur pour les bruits impulsionnels n'étant pas donnée, la perte d'insertion en bruit continu a tout de même été mesurée pour des niveaux extérieurs allant de 70 dB à 110 dB. Ces résultats sont représentés sur la Figure I.8b. Il apparaît que la variation d'atténuation en fonction du niveau extérieur ne débute qu'entre 90 et 100 dB et existe sur toute la gamme de fréquences étudiée. L'augmentation de l'atténuation varie entre 10 et 20 dB. De façon similaire au bouchon talk-through précédent, ce système semble favoriser le passage des fréquences utiles comprises entre 1 et 4 kHz.



FIG. I.9 – (a) Photographie du casque  $3M^{TM}$  Peltor<sup>TM</sup> ComTac<sup>TM</sup> XP, (b) Perte d'insertion en champ diffus pour 5 niveaux extérieurs différents

### 2.1.3 Casques talk-through

Comme les systèmes talk-through sont adaptés à des bouchons d'oreille, il existe aussi des casques talk-through. Un microphone est placé sur chaque coquille, qui contient aussi un haut-parleur. Les casques talk-through sont parfois appelés casques acoustiquement transparents, mais ce terme ne sera pas utilisé ici. Les casques  $3M^{TM}$  Peltor<sup>TM</sup> ComTac<sup>TM</sup> XP et MSA Supreme<sup>®</sup> Pro sont testés dans notre étude.

**3M**<sup>TM</sup> **Peltor**<sup>TM</sup> **ComTac**<sup>TM</sup> **XP** Le casque présenté sur la Figure I.9a comprend un système talk-through permettant de réduire l'exposition de l'auditeur aux bruits de fort niveau tout en préservant la communication. Lorsque le système électronique n'est pas utilisé, le casque possède un SNR de 28 dB. D'après les données du fabricant, son atténuation est surtout importante dans les hautes fréquences ( $H = 31 \ dB, M = 25 \ dB, L = 16 \ dB$ ). Ces caractéristiques sont plutôt faibles comparées à un casque de chantier « classique ». Cela est dû au fait que le Peltor<sup>TM</sup> ComTac<sup>TM</sup> est relativement fin pour pouvoir être porté sous un casque militaire. Dans la suite, ce casque est appelé P5.

De la même manière que les bouchons talk-through, ce casque talk-through permet un réglage d'amplification que nous avons laissée au minimum pour toutes nos expériences. La Figure I.9b représente la perte d'insertion en bruit continu pour ce casque talk-through, mesurée pour des niveaux extérieurs allant de 70 dB à 110 dB. L'augmentation de l'atténuation avec le niveau est présente jusqu'à 100 dB et principalement pour les fréquences comprises entre 1 kHz et 5 kHz. À basse fréquence (autour de 100 Hz), on observe une amplification induite par le système talk-through (jusqu'à 6 dB) pour les bas niveaux (70 dB) et pour les forts niveaux (100 et 110 dB).

**MSA Supreme<sup>®</sup> Pro** Le casque présenté sur la Figure I.10a est un deuxième casque talk-through permettant de réduire l'exposition de l'auditeur aux bruits de fort niveau tout en préservant la communication. D'après le fabricant, lorsque le système électronique



FIG. I.10 – (a) Photographie du casque MSA Supreme<br/> $^{\circ}$ Pro, (b) Perte d'insertion en champ diffus pour 5 niveaux extérieurs différents

n'est pas utilisé, le casque possède un NRR de 19 dB(A). Comme le casque précédent, un réglage d'amplification est proposé à l'utilisateur, amplification que nous laisserons toujours au minimum par la suite. Dans la suite, ce casque est appelé P6.

Sur la Figure I.10b représentant la perte d'insertion du casque MSA Supreme<sup>®</sup> Pro, on observe une augmentation de l'atténuation pour des niveaux supérieurs ou égaux à 80 dB et pour des fréquences supérieures à 600 Hz. Une nouvelle augmentation de la perte d'insertion se produit pour les niveaux supérieurs à 90 dB, mais seulement en haute fréquence (au-delà de 6 kHz). Pour finir, on remarque peu de différences entre les courbes associées aux niveaux extérieurs de 80 et 90 dB, ainsi qu'entre les courbes associées aux niveaux de 100 et 110 dB. La variation de la perte d'insertion en fonction du niveau extérieur ne semble donc pas continue, mais paraît avoir une forme de courbe en escalier.

Face aux protections auditives conventionnelles, le développement des TCAPS a permis de grandement améliorer la détection des signaux et l'intelligibilité tout en conservant un niveau de protection acceptable (Bockstael et al., 2011; Norin et al., 2011; Bernstein et al., 2014). Dans l'objectif de progresser dans la préservation du *situational awareness* quand l'auditeur porte des protections auditives, le thème que nous traitons dans cette étude est la conservation de la capacité du système auditif à localiser les sources sonores.

# 2.2 Performances de localisation sonore

En condition d'écoute normale, nous pouvons connaître la position d'une source sonore dans l'espace à trois dimensions avec une excellente précision et sans aide visuelle. Cette aptitude est étudiée depuis le début du XXe siècle (Lord Rayleigh, 1907, 1909; Hartley & Fry, 1921; Wallach, 1939). Dans le cas où plusieurs sources sont présentes, cette capacité permet de les distinguer et de ne se concentrer que sur une seule : c'est l'effet *cocktail party*, qui permet de suivre une conversation dans le bruit. Puisque l'auditeur doit être en mesure de localiser une menace éventuelle et que la localisation sonore améliore l'intelligibilité dans le bruit, on comprend que cette aptitude est importante pour le *situational awareness*. En général, son étude se fait dans un repère sphérique. Un stimulus est présenté à des auditeurs, et on leur demande d'indiquer la direction dans laquelle il est perçu. On étudie les performances de localisation d'auditeurs en fonction de l'azimut et de l'élévation, et plus rarement en fonction de la distance. Les principales mesures de ces performances se font par l'estimation des angles d'erreur moyens pour l'azimut et l'élévation, et des taux de confusions avant-arrière et gauche-droite. De nombreuses études présentent les performances de localisation en condition d'oreille libre et en environnement calme (Kuhl & Plantz, 1972; Blauert, 1983; Middlebrooks & Green, 1991; Abel et al., 1999; Bronkhorst, 1995) ou bruyant (Hawley et al., 1999; Stern et al., 2005). À titre d'exemple, prenons une étude récente menée sur une sphère composée de près de 280 sources sonores (Gallagher et al., 2014) : seulement 1,5% d'erreurs supérieures à 45° et 5,6% d'inversions avant-arrière sont observées en condition d'oreille ouverte. Intuitivement, on peut penser qu'utiliser un bouchon ou placer un casque sur les oreilles devrait détériorer cette aptitude.

La dégradation des performances de localisation sonore induite par les protections individuelles contre le bruit a été rapportée dès les années 70 (Atherley & Noble, 1970; Atherley & Else, 1971; Noble & Russell, 1972) et a depuis été observée dans de nombreuses études. Les erreurs de localisation avec des bouchons d'oreille et des casques à atténuation fixe ont été caractérisées en utilisant des bruits à large bande fréquentielle (McKinley, 2000; Brungart et al., 2003b), des signaux de parole (Brungart et al., 2004) ou encore des enregistrements de rechargement d'armes (Vause & Grantham, 1999). Une baisse des performances de localisation a été observée quel que soit le stimulus sonore et a principalement été attribuée à des erreurs d'élévation et des confusions avant/arrière. La dégradation est plus importante pour les casques antibruit que pour les bouchons d'oreilles, et l'est d'autant plus lorsque les deux types de protection sont portés en même temps (Brungart et al., 2003b, 2004; Simpson et al., 2005). De plus, le port de protections auditives aurait un effet substantiel sur le temps de réponse des auditeurs (Bolia & McKinley, 2000).

Des erreurs de localisation sont toujours présentes quand des systèmes tactiques de communication et protection sont utilisés (Brungart & Hobbs, 2007; Zimpfer & Sarafian, 2014). Cependant, leur atténuation, dépendant du niveau sonore, semble engendrer une amélioration de la précision de localisation comparée à l'atténuation fixe des protecteurs auditifs classiques. En effet, les performances de localisation sonore avec des bouchons et casques talk-through (Abel et al., 2007) sont meilleures quand le système électronique est en marche (dans le cas contraire, le protecteur est en fait utilisé comme un protecteur classique). Il en est de même dans le cas des bouchons non linéaires qui peuvent être fermés : dans (Gallagher et al., 2014), par exemple, le pourcentage d'inversions avant-arrière passe de 21,5%, pour un bouchon non linéaire fermé, à 16,9% lorsqu'il est ouvert. Comme avec les PICB conventionnelles, on observe aussi un avantage des bouchons devant les casques, pour lesquels le pourcentage d'inversions avant-arrière peut atteindre 40%

(Brungart & Hobbs, 2007). En revanche, en fonction des études, on peut trouver des différences entre les bouchons non linéaires et les bouchons talk-through (Zimpfer et al., 2012; Zimpfer & Sarafian, 2014) ou non (Talcott et al., 2012). Toutefois, aucun TCAPS ne permet à l'auditeur de retrouver ses performances de localisation sonore en condition d'oreille ouverte.

Nous retiendrons que par rapport aux PICB conventionnelles, les TCAPS apportent une amélioration de l'intelligibilité et des performances de localisation sonore. Ces observations nous ont menés à restreindre notre étude des performances de localisation sonore aux TCAPS. De plus, seule la localisation de source sonore à large bande fréquentielle est étudiée puisque la plupart des bruits que l'on peut trouver sur un champ de bataille sont à large bande. Pour finir, la détection de tirs d'armes à feu, étudiée dans (Talcott et al., 2012), n'est pas abordée, car nous considérons que l'amélioration des performances de localisation sonore avec des TCAPS est plus importante dans un domaine de niveaux sonores dans lequel ces protecteurs devraient produire une atténuation nulle, c'est-à-dire pour la parole et des sons naturels.

# CHAPITRE II

# Effet des TCAPS sur les indices de localisation sonore

Les systèmes tactiques de communication et protection (TCAPS) assurent une protection correcte contre le bruit et préservent l'intelligibilité. Toutefois, elles semblent tout de même dégrader la capacité des auditeurs à localiser les sources sonores. Dans l'objectif de trouver la cause de cette dégradation, ce chapitre décrit tout d'abord les mécanismes de la localisation sonore, principalement basés sur l'utilisation de différents indices. Ensuite, nous détaillons deux protocoles de mesure des fonctions de transfert contenant les indices de localisation, ainsi que les méthodes d'extraction de ces derniers. Pour finir, l'influence des TCAPS sur la localisation sonore est étudiée en comparant les indices obtenus avec et sans les protecteurs.

# 1 Mécanismes de la localisation sonore

Dans (Lord Rayleigh, 1907), une théorie sur les mécanismes de localisation sonore est proposée, la théorie duplex. Celle-ci, toujours utilisée aujourd'hui, indique que la position d'une source sonore peut être estimée à l'aide de deux indices binauraux : les différences interaurales de temps et d'intensité. Ces indices se basent sur la comparaison des signaux reçus aux oreilles de l'auditeur (d'où le terme d'indice binaural). Dans la suite, nous distinguerons donc ces signaux en les nommant  $H_R(\theta, \phi, f)$  et  $H_G(\theta, \phi, f)$ , respectivement pour l'oreille droite et l'oreille gauche, avec  $\theta$  l'azimut de la source,  $\phi$  son élévation et f sa fréquence. La localisation sonore dans l'espace à trois dimensions n'étant pas le sujet principal de ce mémoire, la localisation sur le plan horizontal est principalement développée. Les signaux droite et gauche sont alors respectivement notés  $H_R(\theta, f)$  et  $H_G(\theta, f)$ . Pour finir, l'effet de la distance entre l'auditeur et la source n'est pas considéré.

Le système de coordonnées sphériques utilisé tout au long de cet ouvrage est celui



(a) Coordonnées sphériques verticales (b) Coordonnées sphériques interaurales

FIG. II.1 – Systèmes de coordonnées sphériques utilisés dans cette étude. (a) Le premier définit les angles  $\theta$  et  $\phi$ , respectivement l'azimut et l'élévation. (b) Le second définit l'angle latéral  $\alpha$  et l'angle polaire  $\beta$ .

décrit sur la Figure II.1a : un azimut  $\theta$  de 0° correspondant à la position directement face à l'auditeur et de 90° à la position de l'oreille droite. Le plan horizontal passant par l'axe interaural (axe y sur la figure) correspond à une élévation  $\phi$  égale à 90°. Toutefois, des travaux précédents utilisent un système de coordonnées se basant sur l'axe interaural et définissant un angle latéral  $\alpha$  et un angle polaire  $\beta$ . Le passage d'un repère à l'autre est explicité en Annexe A. Dans la suite, quand nous ferons référence à ces angles à la place de l'azimut et de l'élévation, le système défini en Figure II.1b sera utilisé.

# 1.1 Différence interaurale d'intensité

Afin d'expliquer les mécanismes de la localisation sonore sur le plan horizontal, Lord Rayleigh a commencé par étudier la différence interaurale d'intensité, ou *Interaural Level Difference* (ILD), de sons purs. Elle représente la différence d'amplitude des signaux sonores arrivant à chacune des deux oreilles et s'exprime en décibels (dB). Cette différence est due à un effet d'ombre acoustique créé par la tête de l'auditeur et dépend donc de l'angle d'arrivée de l'onde. Il est observé dans (Lord Rayleigh, 1907) que la localisation d'un son pur à partir de l'ILD est de moins en moins adéquate quand la fréquence du son diminue. En effet, la différence de niveau entre les deux oreilles provient de la diffraction de l'onde acoustique par la tête. Cette diffraction devient importante lorsque la longueur d'onde est inférieure au diamètre de la tête (Stern et al., 2005), c'est-à-dire lorsque la fréquence augmente. Ainsi, l'ILD est considérée comme un indice de localisation prépondérant à haute fréquence : à partir de 500 Hz selon Hartmann (1999), 1 kHz selon Middlebrooks & Green (1991) et 1,5 kHz d'après Stern et al. (2005).

La diffraction de l'onde acoustique par la tête de l'auditeur dépend donc non seulement de la position de la source sonore, mais aussi de sa fréquence. La Figure II.2 représente



FIG. II.2 – Différence interaurale d'intensité mesurée en fonction de la fréquence pour une source sonore située à  $10^{\circ}$ ,  $45^{\circ}$  et  $90^{\circ}$  sur le plan horizontal.

l'ILD mesurée sur une tête artificielle en fonction de la fréquence (entre 100 Hz et 5 kHz) pour une source sonore située à 10°, 45° et 90° sur le plan horizontal. On remarque que l'évolution de l'ILD avec la position de la source dépend fortement de la fréquence. Par exemple, à 300 Hz, elle est plus élevée à 90° qu'à 45°, tandis que l'inverse est observé à 1 kHz. L'estimation de la différence interaurale d'intensité pour des sources sonores à large bande fréquentielle se fait donc par le calcul de la différence des spectres d'énergie droite et gauche intégrés sur la plage de fréquence considérée  $[f_1; f_2]$  (Larcher, 2001) :

$$ILD(\theta) = 10 \log_{10} \left( \frac{\int_{f_1}^{f_2} |H_D(\theta, f)|^2 \,\mathrm{d}f}{\int_{f_1}^{f_2} |H_G(\theta, f)|^2 \,\mathrm{d}f} \right),\tag{II.1}$$

En considérant comme dans l'équation précédente que l'ILD est la différence d'énergie entre l'oreille droite (90°) et l'oreille gauche (270°), on suppose généralement qu'elle sera maximale pour une source sonore placée face à l'oreille droite, et minimale pour une source face à l'oreille gauche. Cependant, la diffraction de l'onde acoustique sur la tête de l'auditeur à haute fréquence introduit une interférence constructive à l'opposé de la position de la source sonore. Cet effet, appelé *bright spot*, se remarque d'un point de vue théorique dans l'étude de la diffraction d'une onde sonore par une sphère rigide (Wiener, 1947) et reste présent lors de l'application à une tête réelle. Il peut être à l'origine d'une baisse des performances de localisation sonore (Macaulay & Hartmann, 2010). La Figure II.3 donne un exemple d'ILD calculée en fonction de l'azimut de la source sonore sur le plan horizontal. On remarque effectivement l'influence du *bright spot* qui se caractérise par la diminution (en valeur absolue) de l'ILD à l'approche de l'axe interaural.



FIG. II.3 – Exemple de différence interaurale d'intensité mesurée en décibels en fonction de l'azimut pour une source sonore sur le plan horizontal avec une tête artificielle.

Pour finir, lorsque Lord Rayleigh a étudié la localisation de sons purs, il a remarqué que nous étions toujours capables de localiser des sources sonores à basse fréquence, alors que la différence interaurale d'intensité est trop faible. Il a donc conjecturé que la différence de phase entre les signaux reçus à chaque oreille constitue le nouvel indice de localisation prépondérant. Cette différence interaurale de phase pour des sons purs correspond à une différence interaurale de temps dans le cas de sources sonores plus complexes.

# 1.2 Différence interaurale de temps

La différence interaurale de temps, ou *Interaural Time Difference* (ITD), représente la différence de temps d'arrivée de l'onde acoustique aux deux oreilles. En effet, si la source est sur le côté de la tête, l'onde sonore arrive à l'oreille contralatérale avec une différence de marche par rapport à l'oreille ipsilatérale. L'ITD permet donc de percevoir l'azimut de la source.

# 1.2.1 Exemple d'ITD

La Figure II.4 représente un exemple d'ITD mesurée sur une tête artificielle en fonction de l'azimut sur le plan horizontal, à partir de mesures réalisées sur une tête artificielle. La valeur maximale de l'ITD sur ce graphique est de 750 µs. Elle est plus grande que celle de 660 µs avancée par Stern et al. (2005) ou 700 µs de Middlebrooks & Green (1991). Cette valeur reste en revanche cohérente avec le calcul de Hartmann (1999) qui estime une ITD maximale de 763 µs pour une tête de rayon 8,75 cm. L'azimut pour lequel l'ITD atteint son maximum n'est pas exactement égal à 90° puisque cela supposerait que la tête artificielle



FIG. II.4 – Exemple de différence interaurale de temps mesurée en millisecondes en fonction de l'azimut pour une source sonore sur le plan horizontal avec une tête artificielle.

soit sphérique et que ses oreilles soient diamétralement opposées.

# 1.2.2 ITD en fonction de la fréquence

On conçoit que la différence interaurale de temps est un indice de localisation prépondérant à basse fréquence en reprenant l'analyse de Lord Rayleigh (1907). Pour des sons purs, l'exploitation de la différence interaurale de phase pour la localisation n'est possible que si la longueur d'onde est supérieure au diamètre de la tête. En posant c la célérité du son dans l'air égale à 340 m.s<sup>-1</sup> et en choisissant le rayon de tête moyen a, utilisé dans (Hartley & Fry, 1921), égal à 8,75 cm, la fréquence limite d'exploitation de la différence interaurale de phase est estimée à 1943 Hz. Cette limite est en général réduite à 1,5 kHz (Larcher, 2001; Busson, 2006; Stern et al., 2005). Pour une fréquence inférieure f, on peut déduire une différence interaurale de temps à partir de la différence de phase (Kuhn, 1977) :

$$ITD(\theta, f) = \frac{\phi_D(\theta, f) - \phi_G(\theta, f)}{2\pi f},$$
 (II.2)

avec  $\phi_D$  et  $\phi_G$  les phases des signaux  $H_D$  et  $H_G$ , respectivement. Cette formulation prend en compte la dépendance de l'ITD à la fréquence. Toutefois, Kuhn (1977) a montré qu'elle en est indépendante pour une fréquence inférieure à 500 Hz.

Alors que la différence interaurale de phase n'est plus exploitable à haute fréquence, une ITD peut être estimée, dans le cas d'une source sonore plus complexe qu'un son pur, comme le retard d'enveloppe entre les signaux droite et gauche (Middlebrooks & Green,



FIG. II.5 – Différence interaurale de temps calculée en millisecondes en fonction de l'azimut pour un son pur à 250 Hz, à 1 kHz, pour une source complexe 2-4 kHz et pour le modèle de Woodworth.

1990):

$$ITD(\theta, f) = \frac{1}{2\pi} \frac{\mathrm{d}(\phi_D(\theta, f) - \phi_G(\theta, f))}{\mathrm{d}f},\tag{II.3}$$

dont la formulation est équivalente à celle d'un retard de groupe.

La Figure II.5 représente l'ITD calculée en fonction de l'azimut dans quatre conditions différentes :

- pour un son pur à 250 Hz;
- pour un son pur à 1 kHz;
- pour une source haute fréquence dont les caractéristiques spectrales principales se trouvent entre 2 et 4 kHz;
- pour le modèle de tête sphérique rigide de rayon 10 cm de Woodworth & Schlosberg (1962) (voir Équation II.5 plus loin).

Les résultats obtenus correspondent à ceux de Carterette & Friedman (1978, chap. 10, sec. 2). On remarque que l'ITD estimée pour des sources haute fréquence est en général plus faible qu'en basse fréquence (Benichoux et al., 2016). En effet, les expériences de Kuhn (1977) montrent que l'ITD peut être approchée à basse fréquence par la formule :

$$ITD(\theta) \approx 3\frac{a}{c}\sin(\theta),$$
 (II.4)

tandis que pour les hautes fréquences, la Figure II.5 montre que l'ITD est proche de l'estimation de l'ITD pour une sphère rigide. On utilise donc la formule de Woodworth & Schlosberg (1962) :

$$ITD(\theta) = \frac{a}{c}(\theta + \sin(\theta)), \qquad (II.5)$$

Le rapport entre basses et hautes fréquences est donc :  $\frac{3\sin(\theta)}{\theta+\sin(\theta)} \ge 1$ . Le maximum de ce rapport, à savoir  $\frac{3}{2}$  lorsque  $\theta = 0$ , se retrouve aussi si on utilise l'approximation à haute fréquence obtenue dans (Kuhn, 1977), et valable pour  $\frac{\pi}{12} \le \theta \le \frac{\pi}{3}$ :

$$ITD \approx 2\frac{a}{c}\sin(\theta),$$
 (II.6)

On peut tout de même considérer que l'ITD ne varie que très peu avec la fréquence (Benichoux et al., 2016). La différence interaurale de temps sera donc considérée dans la suite comme un indice indépendant de la fréquence.

# 1.2.3 Utilisation de l'ITD et de l'ILD

Selon Blauert (1983), les différences interaurales de temps calculées comme un retard de phase sont exploitables jusqu'à 1,6 kHz et les retards d'enveloppe interviennent à haute fréquence. Toujours selon Blauert, l'ILD présente un effet sur toute la gamme de fréquences audibles. Cependant, le choix de l'utilisation de l'ITD ou de l'ILD par le système auditif dépend effectivement du spectre de la source.

L'ILD devenant faible à basse fréquence, elle est considérée comme l'indice de localisation prépondérant à haute fréquence pour les sons purs. La différence interaurale de phase, ambiguë à haute fréquence, sera prépondérante pour les basses (Blauert, 1983, chap. 2, sec. 4). Pour des sources sonores complexes, l'ITD est l'indice le plus important pour les sources principalement basses et l'ILD pour les sources hautes fréquences. Dans le cas de sons à large bande fréquentielle, les deux différences interaurales sont utilisées, mais l'ITD reste toutefois prépondérante (Macpherson & Middlebrooks, 2002). De manière générale, le système auditif utilise ces deux indices et localise la source sonore à la position la plus « réaliste » selon leur valeur (Wightman & Kistler, 1997a). Le terme de « combinaison naturelle des paramètres interauraux » introduit par Gaik (1993) résume bien le fonctionnement du système auditif pour l'utilisation de l'ITD et de l'ILD.

Ces principes généraux s'appliquent dans le cas de la localisation d'une unique source sonore en champ libre. En effet, si deux sources sonores émettent des signaux cohérents depuis des positions différentes, les indices interauraux sont modifiés et il est possible de ne localiser qu'une unique source à une position erronée. Ce principe de sommation de la localisation s'applique aussi pour un nombre de sources supérieur ou dans le cas d'un champ réverbérant (Blauert, 1983, chap. 3, sec. 3). L'utilisation du principe de sources-images permet effectivement d'assimiler ce dernier cas à la localisation multi-sources. En champ réverbérant, l'ITD va donc être fortement affectée par les réflexions, car elle dépend de la cohérence entre les signaux arrivant aux deux oreilles. L'ILD ne va que très peu varier puisque l'absorption des parois augmente avec la fréquence. Ainsi, en champ réverbérant, l'auditeur va inconsciemment délaisser les ITD en faveur des ILD (Hartmann, 1999). Cependant, il existe tout de même un principe de précédence, ou *precedence effect* en anglais (Wallach et al., 1949), selon lequel le jugement pour la localisation est basé sur l'onde « arrivée le plus tôt » pour éviter la prise en compte d'informations inutiles (et incorrectes) provenant des réflexions (Brown et al., 2015).

# 1.3 Limitation de la théorie duplex

Nous avons montré que les différences interaurales de temps et d'intensité permettent de localiser la position d'une source sonore sur le plan horizontal. Il existe cependant des limitations à la *théorie duplex* de Lord Rayleigh. En considérant le modèle de tête sphérique, on remarque que les ITD et les ILD estimées à des positions symétriques par rapport à l'axe interaural sont identiques. Ceci peut conduire à des confusions avantarrière, effectivement observées en laboratoire (Wallach, 1939). Les ensembles de points pour lesquels l'ITD et l'ILD sont constantes forment des hyperboles sur le plan horizontal (Blauert, 1983, chap. 2, sec. 5). Un modèle de localisation sonore uniquement basé sur les différences interaurales ne permettrait donc pas de percevoir la distance de la source.

Dans l'espace à trois dimensions, les hyperboles formées sur le plan horizontal par les ensembles des points iso-ITD (ou iso-ILD) deviennent des hyperboloïdes. Elles peuvent être approchées en champ lointain par des cônes (Blauert, 1983, chap. 2, sec. 5). La Figure II.6 représente deux cônes de confusion différents, que l'on peut caractériser par leur azimut sur le plan horizontal  $\theta_R$  et  $\theta_B$ . En plus des confusions avant-arrière déjà remarquées sur le plan horizontal, la *théorie duplex* ne permet donc pas non plus de percevoir l'élévation de la source sonore. C'est pourquoi les hyperboloïdes engendrant ces erreurs de localisation sont appelées « cônes de confusion » (Shinn-Cunningham et al., 2000). L'utilisation d'un système de coordonnées sphériques interaurales (cf. Figure II.1b) prend ici son sens : fixé l'angle latéral  $\alpha$  revient à se placer sur un cône de confusion.

Le fait que les surfaces iso-ITD et iso-ILD soient des hyperboloïdes provient du modèle de tête sphérique. Ces surfaces existent tout de même pour une tête réelle et sont traditionnellement toujours appelées cônes de confusion. En effet, la différence interaurale de temps varie avec l'élévation. Cette variation a été étudiée par Larcher & Jot (1997) et Savioja et al. (1999) qui proposent respectivement les modèles suivants :

$$ITD(\theta,\phi) = \frac{a}{c}[\arcsin(\sin\phi\sin\theta) + \sin\phi\sin\theta] ,$$
  

$$ITD(\theta,\phi) = \frac{a}{c}(\sin\theta + \theta)\sin\phi ,$$
(II.7)

Dans l'Équation II.7, les dépendances à l'élévation en  $\cos \phi$  des deux publications précédemment citées ont été transformées en  $\sin \phi$  afin de respecter notre définition du système de coordonnées sphériques (voir Figure II.1a). Dans les deux cas, on remarque que la valeur maximale de l'ITD dans un plan d'élévation constante diminue lorsque  $\phi$  tend vers 0 ou  $\pi$ .

Concernant les différences interaurales d'intensité, il a été démontré qu'en champ proche, l'ILD varie significativement avec la distance. Les surfaces iso-ILD auraient donc plutôt la forme d'un « tore de confusion » pour des sources sonores situées à moins d'un mètre de l'auditeur (Shinn-Cunningham et al., 2000).



FIG. II.6 – Représentation de deux cônes de confusion différents et caractérisés par les angles  $\theta_R$  pour le rouge, et  $\theta_B$  pour le bleu.

Ces différentes modifications des cônes de confusion pour une tête réelle ne résolvent toutefois pas les confusions avant-arrière engendrées par la *théorie Duplex* sur le plan horizontal et qui nous intéressent particulièrement dans ce mémoire. De plus, pour de faibles valeurs de différences interaurales (c.-à-d. quand  $\theta$  tend vers 0), les cônes de confusion tendent vers le plan médian (Blauert, 1983, chap. 2, sec. 5). Dans ce plan particulier, il est tout à fait possible de percevoir l'élévation de la source ainsi que de savoir si elle se trouve devant ou derrière, alors que les différences interaurales sont nulles. Dans le cas où l'auditeur est autorisé à réaliser de légers mouvements de la tête, les confusions avantarrière peuvent être évitées (Wallach, 1939; Burger, 1958; Wightman & Kistler, 1999). On considère généralement que la localisation sonore sur le plan médian et sur les cônes de confusion est plutôt permise par les effets de l'oreille externe, de la tête et du torse sur le spectre du signal reçu au tympan (Shaw, 1974; Musicant & Butler, 1984; Middlebrooks et al., 1989; Carlile et al., 2005).

# 1.4 Indices spectraux

# 1.4.1 Origines

Dans un premier temps, il a été conjecturé que le rôle du pavillon de l'oreille dans la localisation était d'ajouter une série de retards au signal arrivant au tympan en fonction de la position de la source (Batteau, 1967). C'est toutefois une théorie plutôt basée sur l'apport de caractéristiques spectrales à ce signal qui est la plus utilisée aujourd'hui.

La forme particulière de l'oreille externe apporte effectivement une coloration spectrale



FIG. II.7 – Distributions de la pression aux fréquences M1 à M5 dans la conque et dans le conduit auditif. Figure tirée de (Shaw & Teranishi, 1968).

dépendante de la position de la source au signal sonore arrivant au tympan. À partir de 3 kHz, le pavillon de l'oreille agit comme un ensemble de résonateurs. Cinq modes acoustiques différents de la conque ont d'ailleurs été mis en évidence dans (Shaw & Teranishi, 1968) entre 2,9 kHz et 12,8 kHz. La Figure II.7 représente la distribution des ventres et nœuds de pression à l'intérieur de la conque et du conduit auditif pour ces cinq fréquences. Lorsque le conduit auditif est bouché, six modes de la conque et du pavillon de l'oreille peuvent être trouvés de 4,2 kHz à 16,7 kHz (Shaw, 1997). Ces modes sont représentés sur la Figure II.8.

La tête et le torse de l'auditeur ont aussi des effets substantiels et dépendants de la direction de la source sur le signal sonore arrivant à l'oreille. L'effet principal de la tête est montré par le calcul analytique de la diffraction du champ de pression par une sphère rigide de Rabinowitz et al. (1993), en considérant que les oreilles sont diamétralement opposées. Ce modèle simple permet de comprendre que suivant la fréquence de la source, l'amplitude du signal mesuré à une oreille varie en fonction de la position de la source. La Figure II.9 représente, en fonction de la fréquence, l'amplification du signal mesuré au niveau d'une oreille et due à la diffraction de l'onde sonore sur une sphère rigide de rayon 10 cm. La source sonore est placée à 1,3 m du centre de la sphère et est déplacée de 0° (face à l'oreille) à 180° (côté opposé à l'oreille) par pas de 45°. Les effets de la tête sont prépondérants pour une fréquence supérieure à 1 kHz. Quand la source se trouve face à l'oreille, on observe une augmentation jusqu'à 6 dB de la pression acoustique. Cet effet diminue au fur et à mesure de l'éloignement de la source de la position frontale. En revanche, on observe une réamplification lorsque la source se trouve du côté opposé à l'oreille : c'est l'effet bright spot (Wiener, 1947; Duda & Martens, 1998) déjà introduit dans la présentation de l'ILD en Section 1.1. Des modèles de tête plus poussés, comme une tête ellipsoïdale (Duda et al., 1999), permettraient d'obtenir des résultats plus proches de la réalité.

L'effet du torse sur le signal sonore arrivant au tympan dépend quant à lui principale-



FIG. II.8 – Distributions de la pression pour les modes de la conque et du pavillon lorsque le conduit auditif est bouché. Figure tirée de (Shaw, 1997).



FIG. II.9 – Amplification du signal reçu à l'oreille par diffraction du champ acoustique sur la sphère rigide en fonction de la fréquence. Le calcul est réalisé pour cinq positions  $\theta$  de source : 0° correspond à la position face à l'oreille et 180° à la position opposée.



FIG. II.10 – Répartition des bandes directionnelles de Blauert (1969) en fonction de la fréquence.

ment de l'élévation de la source. Il est étudié à travers un modèle de double sphère rigide dans (Algazi et al., 2002). Le torse introduit d'importantes réflexions quand l'élévation de la source est grande et un effet d'ombre acoustique conséquent à basse élévation.

Pour conclure, en fonction de la position d'une source sonore, le signal reçu au tympan est significativement modifié par la tête et le torse de l'auditeur et filtré suivant l'excitation ou non des modes de l'oreille externe. Le spectre obtenu dépend donc des caractéristiques spatiales de la source sonore. Ces indices spectraux sont les indices de localisation qui permettent à l'auditeur à la fois de distinguer une source venant de devant d'une autre venant de derrière, et de percevoir l'élévation de la source sonore.

#### 1.4.2 Théorie des bandes directionnelles

À partir de l'étude de ces indices spectraux sur le plan médian, une théorie de bandes directionnelles est proposée dans (Blauert, 1969) : le fait que la source soit devant ou derrière serait directement traduit dans le spectre du signal reçu au tympan par un niveau d'énergie important dans certaines bandes de fréquence. Blauert a ainsi mis en évidence qu'un son présentant d'importantes caractéristiques entre 250 et 500 Hz et entre 4 et 5 kHz sera perçu comme venant de devant, tandis que les bandes distinguant les sons venant de derrière se trouvent entre 1 et 2 kHz et entre 10 et 12 kHz. Il a aussi montré qu'un signal à bande étroite autour de 8 kHz est perçu comme venant d'au-dessus. Ces résultats sont résumés sur la Figure II.10. Ces bandes directionnelles correspondent de plus aux bandes d'amplification (*boost bands*) avant et arrière appliquées à un signal large bande par la tête et le pavillon de l'oreille (Blauert, 1969).

Les bandes directionnelles pour les positions avant, arrière et au-dessus sont aussi mesurées dans (Hebrank & Wright, 1974b). Il est tout d'abord montré que les indices spectraux de localisation dans le plan médian sont compris entre 3,8 et 16 kHz. Les indices avant sont compris entre 3,9 et 8 kHz. L'augmentation de la fréquence entre ces valeurs est liée à la perception d'une augmentation de l'élévation. D'ailleurs, les sons présentant des pics entre 8,1 et 9,1 kHz sont perçus sur le dessus. Enfin, les indices arrière se trouvent entre 10 et 12 kHz. Par rapport à Blauert, Hebrank n'a pas retrouvé d'indices spectraux à des fréquences inférieures à 4 kHz, mais le reste de leurs observations reste en accord.



FIG. II.11 – Exemple de spectres reçus aux oreilles et présentant un creux qui se décale en fréquence en fonction de l'azimut. Le nombre à gauche des courbes représente l'azimut en degrés. Les différents spectres sont séparés verticalement de 10 dB. Cette figure est tirée de (Middlebrooks, 1997)

En revanche, dans (Hebrank & Wright, 1974b; Butler & Belendiuk, 1977; Asano et al., 1990; Middlebrooks, 1997) est introduit le fait que des creux (*notch* en anglais) dans le spectre, qui se déplacent en fonction de l'angle d'incidence, représentent aussi des indices de localisation. Un exemple de creux se déplaçant en fréquence avec l'azimut de la source est donné en Figure II.11.

Les différences que l'on peut trouver entre ces études sont dues au fait qu'ils proviennent d'expériences subjectives. Tous les auditeurs n'ont pas forcément la même façon de répondre (surtout lorsqu'ils font une erreur de localisation), ce qui peut influer sur leur résultat. Enfin, les positions des bandes directionnelles varient aussi entre les auditeurs en fonction de leur morphologie (Middlebrooks, 1999). Des études plus récentes montrent toutefois que les bandes directionnelles resteraient stables en fonction des auditeurs (Itoh et al., 2007; Wallis & Lee, 2015).

Pour introduire cette théorie des bandes directionnelles, les indices spectraux ont principalement été étudiés sur le plan médian puisque les différences interaurales de temps et d'intensité y sont nulles. Toutefois, des études incluant l'azimut de la source sonore montrent que pour une fréquence donnée, on peut trouver une, voire deux, zone de l'espace produisant un maximum de pression acoustique au niveau du tympan (Middlebrooks et al., 1989). L'importance de la largeur de bande fréquentielle, pour distinguer les sources avant et arrière, est aussi montrée sur le plan horizontal (Musicant & Butler, 1984). On y retrouve l'idée de déplacement d'un creux, dans le spectre mesuré à l'oreille, qui se déplace en fonction de l'azimut de la source (Middlebrooks, 1997). Finalement, le manque de perception de l'élévation et les ambiguïtés avant-arrière émis par la *théorie duplex* sont résolus par l'utilisation des caractéristiques spectrales du signal sonore reçu au tympan. En fonction des bandes fréquentielles dans lesquelles ce spectre présente un niveau important d'énergie, la source sonore sera perçue comme venant de derrière, de devant ou du dessus. Un effet remarquable de cette théorie des bandes directionnelles est que, pour une source sonore à bande étroite, la direction perçue dépend principalement de sa fréquence et non de sa position réelle (Blauert, 1969; Butler & Planert, 1976). Des indices spectraux sont aussi disponibles pour distinguer l'azimut d'une source sonore. Toutefois, si les différences interaurales de temps et d'intensité sont disponibles, on leur accordera une plus grande importance dans la perception de cette dimension (Middlebrooks, 1997). La principale précision à apporter à cette théorie est l'étude de l'utilisation des indices spectraux par le système auditif comme des indices monauraux ou des différences interaurales.

# 1.4.3 Indices spectraux monauraux ou binauraux?

La théorie des bandes directionnelles présente un côté paradoxal. En effet, comment savoir si une caractéristique spectrale de premier plan du signal reçu au tympan provient du spectre de la source sonore ou d'un filtrage dépendant de la direction de la source (Hartmann, 1999)? Des signaux à bande étroite sont alors interprétés comme provenant d'une position différente de la réalité. L'importance de la connaissance a priori du contenu spectral de la source apparaît dans ce cas. D'ailleurs, elle est aussi présente dans le cas de la localisation multi-sources de signaux de parole (Stern et al., 2005) : les énoncés clairement identifiés sont correctement localisés.

D'un autre côté, comment expliquer les performances de localisation sonore sur le plan médian ? En supposant que la tête soit symétrique par rapport au plan médian, les signaux reçus aux tympans gauche et droite à partir d'une source sonore située sur ce plan sont égaux et aucune information binaurale n'est donc disponible. L'hypothèse de symétrie peut toutefois être contestée et la présence de « disparité binaurale » étudiée (Searle et al., 1975).

Le traitement des indices spectraux comme monauraux ou binauraux, c'est-à-dire comme des différences interaurales spectrales ou *Interaural Spectral Difference* (ISD), est donc controversé (Carlile et al., 2005), car les indices monauraux dépendraient fortement du spectre de la source. Sur le plan médian, des ISD haute fréquence permettant de distinguer l'élévation de la source ont été mesurées dans (Middlebrooks et al., 1989), s'accordant avec les travaux de Searle. D'autres études, montrant que seule une variation très importante du spectre reçu aux oreilles peut significativement perturber la localisation (Macpherson & Middlebrooks, 2003; Wightman & Kistler, 1997b), sont aussi en faveur de l'exploitation des ISD. Cependant, l'idée de disparité binaurale sur le plan médian est aussi contestée (Hebrank, 1976). Les performances de localisation sonore restent robustes lorsque les ISD sont indisponibles et que seuls des indices monauraux peuvent être exploités (Hebrank & Wright, 1974a).

Dans le but de déterminer la façon dont le système auditif se sert des indices spectraux, des expériences ont été menées en utilisant des techniques de réalité virtuelle (Jin et al., 2004; Zhang & Hartmann, 2010). Les différents auditeurs ont dû localiser des sources sonores alors que leurs indices spectraux étaient modifiés des deux façons suivantes :

— en ne conservant les indices spectraux monauraux que sur une seule oreille;

— en « aplatissant » les indices spectraux d'une oreille, mais en conservant les ISD.

Les résultats montrent qu'aucune de ces deux conditions n'est suffisante pour correctement localiser les sources sonores, et particulièrement pour empêcher les confusions avant-arrière (Zhang & Hartmann, 2010, Exp. 4 et 5). Jin, en particulier, en a conclu que la résolution des confusions avant-arrière et de la perception de l'élévation dans un cône de confusion nécessite à la fois des informations spectrales monaurales et binaurales. Lorsque l'on s'éloigne du plan médian, les indices spectraux de l'oreille ipsilatérale sont quand même prépondérants face à ceux de l'oreille contralatérale (Morimoto, 2001; Carlile et al., 2005). On peut considérer que l'oreille la plus éloignée de la source sonore ne contribue plus à la perception de l'élévation ou à la résolution des confusions avant-arrière dès que la source est séparée du plan médian de plus de 60° (Morimoto, 2001).

Finalement, dans la suite de ce mémoire, nous considérerons les indices spectraux comme étant principalement monauraux, à condition que l'information soit toujours disponible aux deux oreilles. Ce choix est fait, pour une première raison, afin d'éviter la modification des performances de localisation due au manque d'information de l'oreille contralatérale pour des positions de source proche du plan médian. Pour une seconde raison, l'utilisation exclusive des ISD ne paraît pas non plus adaptée à la localisation.

# 2 Head-Related Transfer Functions

Les indices de localisation binauraux et monauraux (ITD, ILD et indices spectraux) sont contenus dans les fonctions de transfert traduisant les phénomènes physiques qui résultent du parcours de l'onde acoustique depuis la source sonore jusqu'aux tympans de l'auditeur. Elles prennent en compte les différentes modifications apportées au signal sonore par le torse, la tête et l'oreille externe de l'auditeur. Ces fonctions de transfert sont communément appelées « Fonctions de Transfert Relatives à la Tête », ou *Head-Related Transfer Function* (HRTF) (Morimoto & Ando, 1980; Wightman & Kistler, 1989a).

Supposons qu'une scène sonore réelle comprend une source de spectre S à la position  $(r, \theta, \phi)$ . Alors, en notant  $H_D(r, \theta, \phi)$  la HRTF entre la source et l'oreille droite et  $H_G(r, \theta, \phi)$  la HRTF gauche, les spectres reçus aux oreilles droite et gauche sont respectivement

$$X_D = H_D(r, \theta, \phi)S$$
  
et  
$$X_G = H_G(r, \theta, \phi)S$$
 (II.8)

Connaissant le spectre de la source, les HRTF peuvent donc être utilisées dans les systèmes de restitution sonore spatialisée. Au casque (Wightman & Kistler, 1989a,b) ou avec des haut-parleurs externes (Morimoto & Ando, 1980), le principe de ces systèmes est la reconstruction exacte des signaux différents reçus aux deux oreilles.

Les HRTF sont mesurées en évaluant les fonctions de transfert entre un haut-parleur placé à différentes positions autour de la tête de l'auditeur et un microphone placé dans le conduit auditif de celui-ci. Les HRTF sont donc des fonctions de transfert propres à chaque auditeur. Dans le cadre de la restitution sonore spatialisée, les HRTF de l'auditeur sont indispensables à une reproduction haute fidélité. Toutefois, des caractéristiques communes à plusieurs auditeurs peuvent être trouvées dans les HRTF. Plusieurs études cherchent donc à individualiser des HRTF tirées d'une base de données (Busson, 2006; Guillon, 2009; Emerit et al., 2016). Dans le cadre de notre étude, nous cherchons à caractériser l'effet de protections auditives sur les performances de localisation. L'effet en question, bien que dépendant de l'auditeur, est tout de même supposé présenter un aspect global. C'est pourquoi nous nous contenterons de l'utilisation d'HRTF mesurées sur tête artificielle. De plus, les performances de localisation peuvent aussi dépendre du spectre de la source. Pour nous affranchir de cette dépendance, nous ne considérons que la localisation d'un son théorique à spectre unitaire, soit une impulsion de Dirac. Cette hypothèse donne S = 1dans l'équation précédente : le spectre du signal recu aux oreilles se réduit aux HRTF correspondantes.

Dans cette partie, nous allons voir comment ces HRTF peuvent être mesurées avec et sans TCAPS ainsi que la manière dont elles sont utilisées pour estimer les différents indices de localisation sonore.

# 2.1 Protocole de mesure

Dans le cadre de notre étude, la mesure d'HRTF a été réalisée sur une tête artificielle ISL (Parmentier et al., 2000) avec le dispositif de localisation sonore du groupe APC. Dans cette section, nous allons tout d'abord décrire le protocole de mesure, puis les différents traitements à apporter aux mesures pour obtenir les HRTF seront développés.

#### 2.1.1 Dispositif expérimental de l'ISL

À l'intérieur d'une cabine d'audiométrie dont les murs sont recouverts d'une mousse absorbante, la tête artificielle ISL est placée au centre d'un arrangement circulaire de 8 haut-parleurs espacés de 45°, de 2,60 m de diamètre sur le plan horizontal (voir Figure II.12). Elle fait face au premier haut-parleur placé à la position 0° et porte, ou non, des protections auditives. Pour avoir des mesures sur 16 positions de source sonore, la tête



FIG. II.12 – Photographie du dispositif de mesure d'HRTF de l'ISL.

pourra être tournée de 22,5° dans le sens trigonométrique. Ce dispositif est schématisé sur la Figure II.13. La position des haut-parleurs lorsque la tête est tournée est représentée en pointillés.

Le schéma de principe de la Figure II.14 présente la mesure effectuée pour obtenir les HRTF. La fonction de transfert est mesurée entre la sortie et l'entrée d'un analyseur de spectre Stanford Research SR780. Celui-ci génère un sinus glissant de 216 points allant de 90 Hz à 17,8 kHz avec un pas logarithmique. Cette sortie de l'analyseur est transmise à un amplificateur de puissance Dayton Audio<sup>®</sup> MA1260 à 12 canaux. 8 de ses sorties sont chacune connectées à un haut-parleur Visaton FR16WP. Un programme informatique en C++ pilotant une carte Arduino<sup>TM</sup>, qui pilote elle-même deux multiplexeurs, permet de choisir la source sonore à utiliser pour la mesure. Le signal émis par la source sonore est reçu à l'oreille de la tête artificielle et envoyé à un amplificateur de mesure Brüel & Kjær Type 2636 puis à l'analyseur de spectre. Pour chacune des 216 fréquences testées, ce dernier calcule la fonction de transfert entre son entrée et sa sortie. Ces fonctions de transfert sont mesurées pour les deux oreilles de la tête artificielle et pour 16 positions de source allant de 0° à 337,5° avec un pas de 22,5°. Pour obtenir des HRTF à partir de ces mesures de fonction de transfert, certaines étapes d'égalisation sont nécessaires.

# 2.1.2 Égalisation des fonctions de transfert

Tout d'abord, avant la mesure de fonction de transfert, il est évidemment important de vérifier que les niveaux sonores délivrés par les 8 sources sont similaires. En effet, les protections auditives avec lesquelles nous allons mesurer les fonctions de transfert ont une atténuation dépendant du niveau sonore. Le niveau global reçu au centre de la tête, en l'absence de celle-ci, est fixé à environ 70 dB(SPL) quand l'analyseur de spectre génère



FIG. II.13 – Schéma du dispositif expérimental. La tête est placée au centre du cercle constitué de 16 sources sonores.



FIG. II.14 – Schéma de principe de la mesure de fonction de transfert.



FIG. II. 15 – Spectre des haut-parleurs, identifiés par leur position, après égalisation du niveau global.

un bruit blanc à la même tension que le sinus glissant utilisé pour la mesure. La réponse en fréquence des 8 haut-parleurs est mesurée pour vérifier que l'égalisation des sources sonores est suffisante. La Figure II.15 présente l'amplitude en décibels des 8 réponses. Les différences entre les haut-parleurs se situent entre 100 Hz et 700 Hz et peuvent atteindre 13 dB. Elles sont sans doute dues à des réflexions sur les parois de la cabine, dont la mousse n'empêche que les réflexions à haute fréquence. Toutefois, le niveau global ne varie que de plus ou moins 0,5 dB par rapport à la moyenne de 67,8 dB(SPL), ce qui reste de l'ordre d'une erreur de mesure et demeure donc acceptable.

Les fonctions de transfert obtenues précédemment ne correspondent pas exactement à des HRTF puisqu'elles contiennent les contributions de la chaîne de mesure, par exemple la réponse en fréquence des haut-parleurs (Busson, 2006, chap. 1, sec. 4). Il est important de les enlever puisque, comme nous venons de voir, les différentes sources sonores ont des réponses différentes en basse fréquence. Afin d'éliminer ces contributions, les fonctions de transfert sont divisées par une mesure de référence réalisée pour chaque source sonore. La mesure de référence est toujours la fonction de transfert entre l'entrée et la sortie de l'analyseur de spectre, mais cette fois en utilisant un microphone Brüel& Kjær Type 4134 placé au centre de la tête, en l'absence de celle-ci. Cette égalisation correspond à la mesure de fonction de transfert en « champ libre » de Blauert (1983, chap. 2, sec. 2).

Les HRTF ainsi obtenues contiennent toujours des caractéristiques qui ne dépendent pas de la direction de la source sonore. On pourra les éliminer en réalisant une égalisation « champ diffus » (Larcher, 2001, chap. 1, sec. 5) : les HRTF sont divisées par leur moyenne sur toutes les positions étudiées. Les fonctions de transfert résultantes sont appelées fonctions de transfert directionnelles, ou *Directional Transfer Function* (DTF), et ne présentent plus ni la résonance du conduit auditif, ni les réponses du microphone de référence et de l'amplificateur de mesure (Middlebrooks & Green, 1990). Dans notre cas, la moyenne des HRTF sur toutes les positions est définie comme un filtre à phase minimale construit à partir de la valeur RMS des 16 amplitudes des HRTF dans le domaine de pression linéaire. Si l'on note  $H_m(f)$  cette moyenne avec la fréquence f et  $H(\theta_i, f)$  la HRTF mesurée à la position  $\theta_i$ , l'amplitude de la moyenne s'écrit :

$$|H_m(f)| = \sqrt{\frac{1}{16} \sum_{i=1}^{16} |H(\theta_i, f)|^2},$$
(II.9)

et, pour que ce soit un filtre à phase minimale (Oppenheim et al., 1999), sa phase se construit comme suit :

$$\arg(H_m(f)) = \Im\{\mathcal{H}[-\log(|H_m(f)|)]\},\tag{II.10}$$

avec  $\Im$  la partie imaginaire et  $\mathcal{H}$  la transformée de Hilbert. Ce choix de la moyenne pour le calcul des DTF permettra dans une prochaine section d'obtenir une estimation plus précise de l'ITD.

Pour terminer, dans le but d'éviter des artefacts dus à un mauvais positionnement angulaire de la tête artificielle, nous faisons l'hypothèse que celle-ci et les protections auditives sont symétriques par rapport au plan médian. Ceci nous permet de moyenner les HRTF (ou DTF) droite et gauche pour chaque fréquence et chaque position :

$$\forall f, \forall \theta, \qquad H(\theta, f) = \frac{1}{2} (H_D(\theta, f) + H_G(2\pi - \theta, f)). \tag{II.11}$$

Dans le cas particulier de mesure d'HRTF avec le casque  $3M^{TM}$  Peltor<sup>TM</sup> ComTac<sup>TM</sup> XP, qui comprend un microphone de radiocommunication sur le côté gauche, l'hypothèse de symétrie est maintenue. Nous supposons en effet que les différences entre les HRTF gauche et droite engendrées par ce microphone n'existent que pour des fréquences supérieures au domaine fréquentiel de mesure.

Une fois les HRTF et DTF ainsi calculées, elles peuvent être utilisées pour estimer les indices de localisation sonore, à savoir les différences interaurales de temps et d'intensité ainsi que les indices spectraux.

# 2.2 Estimation des différences interaurales de temps et d'intensité

Pour estimer les ITD et les ILD, nous utilisons les DTF mesurées à chaque position étudiée en fonction de la fréquence. Nous noterons les DTF droite et gauche  $H_D(\theta, f)$  et  $H_G(\theta, f)$ , respectivement, avec  $\theta$  l'azimut de la source et f la fréquence.

# 2.2.1 Différence interaurale d'intensité

Nous avons vu que les différences interaurales d'intensité sont des indices de localisation efficace à haute fréquence, car elles sont dues à l'effet d'ombre de la tête de l'auditeur. L'ILD

est donc définie comme la différence (en décibels) d'énergie entre les DTF droite et gauche, calculée sur le domaine fréquentiel [ $f_1 = 1,5$  kHz;  $f_2 = 16$  kHz]. D'après l'Équation II.1 du premier chapitre, on a donc :

$$ILD(\theta) = 10 \log_{10} \left( \frac{\int_{f_1}^{f_2} |H_D(\theta, f)|^2 \,\mathrm{d}f}{\int_{f_1}^{f_2} |H_G(\theta, f)|^2 \,\mathrm{d}f} \right),$$
(II.12)

dont l'intégrale se transforme en somme dans notre cas puisque les DTF ont été obtenues pour un nombre fini de fréquences, soit :

$$ILD(\theta) = 10 \log_{10} \left( \frac{\sum_{f_i} |H_D(\theta, f_i)|^2}{\sum_{f_i} |H_G(\theta, f_i)|^2} \right),$$
 (II.13)

avec 1,5 kHz  $\leq f_i \leq$  16 kHz. L'estimation de l'ILD se fait donc bien comme une différence d'énergie entre les DTF droite et gauche, et non comme une moyenne des différences d'amplitude prise fréquence par fréquence entre les DTF droite et gauche.

# 2.2.2 Différence interaurale de temps

Les ITD peuvent être estimées à partir des équivalents des HRTF dans le domaine temporel : les « Réponses Impulsionnelles Relatives à la Tête », ou *Head-Related Impulse Response* (HRIR). Dans notre cas, elles le sont soit avec les HRIR, soit avec les réponses impulsionnelles directionnelles, ou *Directional Impulse Response* (DIR), obtenues par transformée de Fourier inverse des DTF.

**Hypothèse de linéarité de phase** Les HRIR sont des filtres à réponse impulsionnelle finie (RIF) causaux et stables. Elles sont souvent modélisées par le système { $\tau \oplus HRIR_{min}$ }, c'est-à-dire une mise en cascade d'un retard pur  $\tau$  et d'un filtre à phase minimale. Si l'on note  $HRIR_{mixte}(\theta, t)$  un filtre RIF causal et stable et  $HRTF_{mixte}(\theta, f)$  sa transformée de Fourier, alors :

$$HRTF_{mixte}(\theta, f) = HRTF_{min}(\theta, f) HRTF_{excess}(\theta, f), \qquad (II.14)$$

avec  $HRTF_{min}$  un filtre à phase minimale et  $HRTF_{excess}$  un filtre passe-tout. En posant  $HRTF_{mixte} = H(\theta, f)e^{i\phi(\theta, f)}$ , les propriétés d'un filtre à phase minimale (Oppenheim et al., 1999, chap. 5, sec. 6) donnent :

$$|HRTF_{min}(\theta, f)| = H(\theta, f)$$
  

$$\phi_{min}(\theta, f) = \Im \{\mathcal{H}[-\log(H(\theta, f))]\}$$
  

$$|HRTF_{excess}(\theta, f)| = 1$$
  

$$\phi_{excess}(\theta, f) = \phi(\theta, f) - \phi_{min}(\theta, f)$$
  
(II.15)

On fait l'hypothèse que la phase de la composante passe-tout  $\phi_{excess}(\theta, f)$  est linéaire. Cette hypothèse est justifiée dans le cadre des HRTF tant que la source sonore n'est pas trop proche de l'axe interaural (Busson, 2006, chap. 2, sec. 1). On a donc :

$$\phi_{excess}(\theta, f) = 2\pi f \tau \,. \tag{II.16}$$

 $\tau$  est finalement la pente de la phase de la composante passe-tout. En déterminant cette pente pour les HRTF droite et gauche, on trouve la différence interaurale de temps :

$$ITD(\theta) = \tau_G - \tau_D \,. \tag{II.17}$$

La différence se fait dans cet ordre pour que l'ITD soit du même signe que l'ILD à chaque position. Il est donc possible d'estimer l'ITD par régression linéaire de la phase de la composante passe-tout des HRTF droite et gauche. Cette méthode d'estimation de l'ITD n'est cependant pas assez robuste, particulièrement pour une source proche de l'axe interaural. Elle repose de plus sur des calculs plutôt lourds, alors que des méthodes moins compliquées sont plus efficaces.

Estimation de seuil Les HRIR sont des réponses impulsionnelles et l'ITD représente leur différence de temps d'arrivée à l'oreille. Pour une HRIR, le temps d'arrivée à l'oreille peut être estimé comme le temps auquel la réponse impulsionnelle dépasse un certain seuil, en général choisi comme un pourcentage de sa valeur maximale. La valeur du seuil est arbitraire, mais choisir 50 % de la valeur maximale semble être un bon compromis (Busson, 2006, chap. 2, sec. 1). Cette méthode est représentée sur la Figure II.16, dans laquelle des HRIR droite et gauche mesurées à la position  $\theta = 90^{\circ}$  sont tracées en fonction du temps. Les barres verticales représentent les seuils à 50 % de la valeur maximale de ces réponses impulsionnelles.

Cette méthode d'estimation de l'ITD est la plus simple à mettre en œuvre. Il reste tout de même l'inconvénient que, pour des positions contralatérales, le niveau de seuil choisi peut être noyé dans le bruit de mesure et un deuxième pic peut apparaître.

**Maximum d'intercorrélation** La différence interaurale de temps étant une différence de temps d'arrivée entre deux signaux, on peut l'évaluer comme le retard qui maximise l'intercorrélation entre ces deux signaux. L'équation suivante formalise cette méthode :

$$ITD(\theta) = \arg\max_{\tau} \left( \int_{-\infty}^{+\infty} h_G(t) h_D(t-\tau) dt \right), \qquad (II.18)$$

en notant  $h_G(t)$  et  $h_D(t)$  les réponses impulsionnelles gauche et droite<sup>1</sup>, respectivement. Dans le cas des HRIR, l'intégrale se réduit à un domaine temporel fini puisque ce sont des filtres à réponse impulsionnelle finie. Enfin, à la place des HRIR droite et gauche, nous estimons l'ITD par maximum d'intercorrélation à partir des DIR droite et gauche. Ceci permet d'éviter certains artefacts lors de l'estimation de l'ITD : la Figure II.17 présente l'ITD estimée à partir des HRIR et DIR du mannequin KEMAR mesurées avec une haute définition spatiale. On remarque effectivement la présence d'un décrochement près de l'axe interaural pour les estimations à partir des HRIR.

<sup>1.</sup> À noter que pour des signaux complexes, la définition de l'intercorrélation utiliserait le conjugué  $h_D^*(t)$ 



FIG. II.16 – Estimation de l'ITD par étude du seuil des HRIR droite et gauche (position 90°) :  $ITD = \tau_G - \tau_D = 690 \ \mu s.$ 



FIG. II.17 – ITD estimée par la méthode du maximum d'intercorrélation, à partir des HRIR et DIR du mannequin KEMAR mesurées dans (Gardner & Martin, 1994).

Cette méthode d'estimation de l'ITD est la plus robuste des 3 méthodes présentées dans cette section, et sera celle que nous utiliserons dans la suite. Pour terminer, nous avons jugé plus pratique de calculer l'intercorrélation des DIR à partir des DTF, comme la transformée de Fourier inverse de la densité spectrale d'interaction<sup>2</sup>.

# 2.3 Modèle auditif de lissage des HRTF

Les HRTF obtenues en suivant le protocole de l'ISL possèdent une résolution fréquentielle importante. Elles contiennent en effet 216 points répartis de manière logarithmique de 90 Hz à 17,8 kHz. Cette résolution n'est pas nécessaire puisque la cochlée, dans l'oreille interne, réalise un filtrage du signal sonore par un banc de filtres passe-bandes. La gamme de fréquences audibles est alors subdivisée en bandes fréquentielles, appelées bandes critiques, dont la largeur augmente de manière logarithmique avec leur fréquence centrale (Zwicker, 1961). Ces bandes peuvent être approchées par des bandes de tiers d'octave pour des fréquences supérieures à 300 Hz, et par des bandes d'octaves à basse fréquence (Fastl & Zwicker, 2007, Ch. 8, Sec. 7).

Dans le but de simuler plus précisément la tonotopie cochléaire, les DTF sont passées dans un banc de filtres auditifs conduisant à une représentation multicanal du mouvement de la membrane basilaire engendré par l'onde acoustique (Patterson & Holdsworth, 1996). Le banc de filtres est composé de 28 filtres gammatones d'ordre 4 (Holdsworth et al., 1988; Hohmann, 2002). Un tel filtre est défini par sa réponse impulsionnelle g(t) construite par multiplication d'une distribution Gamma et d'un signal sinusoïdal :

$$\forall t \ge 0, \qquad g(t) = t^{n-1} e^{-2\pi f_b t} \cos(2\pi f_c t),$$
 (II.19)

avec n l'ordre du filtre (ici n = 4),  $f_c$  sa fréquence centrale et  $f_b$  sa largeur de bande. Les 28 fréquences centrales du banc de filtres sont réparties de manière logarithmique entre 100 Hz et 16 kHz en appliquant l'espacement de canaux de (Slaney, 1993). Les largeurs de bande correspondantes sont déduites des fréquences centrales en calculant les largeurs de bande rectangulaire équivalente, ou ERB (*Equivalent Rectangular Bandwidths* en anglais), de Glasberg et Moore :

$$ERB = \frac{f_c}{EarQ} + minBW, \tag{II.20}$$

avec EarQ = 9,26449 le facteur de qualité et minBW = 24,7 la largeur minimale (Glasberg & Moore, 1990). La largeur de bande du filtre est ensuite estimée en utilisant la correction de Patterson  $f_b = 1,019ERB$  (Patterson & Holdsworth, 1996).

En pratique, nous réalisons le filtrage des DTF par chaque gammatone dans le domaine fréquentiel en la multipliant par sa fonction de transfert (cf. Annexe B) :

$$G(f) = \frac{\Gamma(n)}{2} \left[ (2\pi f_b + 2i\pi (f - f_c))^{-n} + (2\pi f_b + 2i\pi (f + f_c))^{-n} \right], \qquad (\text{II.21})$$

<sup>2.</sup> Voir Cross-Correlation Theorem sur le site Internet MathWorld



FIG. II.18 – Fonctions de transfert des 28 gammatones du banc de filtres utilisé pour simuler la tonotopie cochléaire.

avec  $\Gamma$  la fonction Gamma. L'amplitude normalisée des fonctions de transfert des 28 gammatones du banc de filtre que nous utilisons est représentée sur la Figure II.18. Pour finir, le filtrage des DTF se fait de manière à préserver leur énergie totale. La Figure II.19 donne un exemple de DTF lissée avec cette méthode. La courbe finale ne suit pas exactement la courbe originale, principalement à basse fréquence, en raison de la condition de conservation de l'énergie totale.

Nous nous arrêtons là pour la simulation du traitement du signal reçu par les oreilles. Nous ne traitons que la réponse du système de localisation sonore à un signal de spectre unitaire, et laissons de côté les aspects spectro-temporels. Il est tout de même possible de modéliser le fonctionnement des cellules ciliées et la réponse du nerf auditif (Meddis, 1986; McEwan & Van Schaik, 2000; Sumner et al., 2002).

# 2.4 Indices de localisation sonore avec TCAPS

Les méthodes d'estimation des différences interaurales de temps et d'intensité, et le traitement des indices spectraux précédemment développés sont appliqués pour étudier la dégradation des indices de localisation induite par les systèmes tactiques de communication et protection. Les résultats présentés dans cette partie proviennent de mesures d'HRTF effectuées avec le protocole de l'ISL (voir Section 2.1.1) et en utilisant les six TCAPS présentés dans la Section 2.1, Chapitre I. Pour plus de clarté, rappelons que P1 désigne le bouchon non linéaire 3M<sup>TM</sup> CombatArms<sup>TM</sup>, P2 le bouchon non linéaire SureFire<sup>®</sup> EP4 Sonic Defenders<sup>®</sup> Plus, P3 le bouchon talk-through Nacre QuietPro<sup>®</sup>, P4 le bouchon talk-through Silynx Clarus<sup>TM</sup>, P5 le casque talk-through 3M<sup>TM</sup> Peltor<sup>TM</sup> ComTac<sup>TM</sup> XP et P6 le casque talk-through MSA Supreme<sup>®</sup> Pro.


FIG. II.19 – Exemple de lissage de DTF par la banque de filtres auditifs gammatones.

Cette section décrit les modifications engendrées par les TCAPS sur l'ITD, puis sur l'ILD, et enfin sur les indices spectraux. Un exemple de métrique quantifiant la modification des indices spectraux est aussi développé. Ces résultats sont similaires à ceux présentés dans (Joubaud et al., 2015) pour tous les TCAPS sauf P4 et en ne mesurant les HRTF que tous les 45°, de 22,5° à 337,5°.

#### 2.4.1 Effet des TCAPS sur l'ITD

La Figure II.20 représente les différences interaurales de temps estimées pour chaque TCAPS en fonction de l'azimut. Les protecteurs auditifs ont été regroupés par type (bouchons non linéaires, bouchons talk-through, casques talk-through). Quand les oreilles de la tête artificielle sont laissées libres (cf. Figure II.20a), l'ITD est maximale sur l'axe interaural et atteint environ 730 µs, valeur qui concorde avec les observations faites en Section 1.2.1, Chapitre I.

Sur les Figures II.20b et II.20c, on observe que l'ITD est très peu affectée par les bouchons non linéaires et le bouchon talk-through P3. Son augmentation (en valeur absolue) maximale, obtenue avec P3, est inférieure à 70 µs et est atteinte aux positions voisines de l'axe interaural, soit 67,5° et 292,5°. L'augmentation de l'ITD est encore davantage accentuée par le bouchon P4 et les casques P5 et P6, et atteint 280 µs pour P4 et P5 et 220 µs pour P6, toujours autour de l'axe interaural. Cette augmentation de l'ITD peut être interprétée comme une augmentation de la distance entre les oreilles : en supposant que le chemin acoustique entre le microphone du protecteur (ou l'entrée du filtre pour P1 et P2) et le tympan est exactement le même pour les deux oreilles, l'ITD est effectivement égale à la différence de temps d'arrivée de l'onde sonore entre les deux microphones (ou



FIG. II.20 – ITD estimée en fonction de l'azimut à partir des mesures d'HRTF avec et sans TCAPS.

les deux entrées de filtre pour P1 et P2).

L'augmentation de l'ITD pourrait avoir comme effet de concentrer la localisation de source sonore autour de l'axe interaural. Toutefois, ces modifications de l'ITD restent de l'ordre des différences juste perceptibles, *Just-Noticeable Difference* (JND), pour ces positions (Busson, 2006, chap. 2, sec. 4). Leur effet sur la localisation est donc plutôt considéré comme négligeable. On retiendra finalement que les TCAPS ne semblent pas dégrader considérablement la différence interaurale de temps. Ceci est en accord avec les résultats d'expériences subjectives antérieures qui montrent que les TCAPS dégradent principalement la localisation avant-arrière, permise par les indices spectraux. Dans (Brungart & Hobbs, 2007), la conservation de l'ITD a aussi été observée, même si elle n'a pas empêché l'observation d'erreurs droite/gauche lors de l'expérience subjective.

#### 2.4.2 Effet des TCAPS sur l'ILD

Les différences interaurales d'intensité estimées avec TCAPS et regroupées par type sont données en Figure II.21. En condition d'écoute normale (cf. Figure II.21a), l'ILD est maximale pour les positions proches de l'axe interaural et atteint presque 15 dB. On observe de plus une réduction de près de 5 dB de l'ILD, aux positions 90° et 270°, due au phénomène de *bright spot* mis en lumière dans la Section 1.1 du Chapitre I.

Les bouchons non linéaires P1 et P2, sur la Figure II.21b, ne semblent pas engendrer une dégradation importante de l'ILD. Effectivement, la variation maximale observée pour ses bouchons est inférieure à 3 dB. En revanche, avec les bouchons talk-through et particulièrement P3, l'ILD est substantiellement dégradée. Avec P3, elle varie entre 1 et 4 dB par rapport à la condition d'écoute normale, et pour P4, la variation est comprise entre 1 et 2,6 dB.

La plus importante détérioration de l'ILD apparaît avec las casques talk-through P5 et P6 sur la Figure II.21d. Elle résulte en un aplatissement de la différence interaurale d'intensité, qui ne dépasse pas 5 dB en valeur absolue. La compression de l'ILD pour ses protecteurs auditifs varie de 5 dB pour les positions proches du plan médian à 13 dB au niveau de l'axe interaural. Puisque cette modification de l'ILD est nettement supérieure aux JND (Francart & Wouters, 2007), elle pourrait engendrer une concentration significative de la localisation sonore vers le plan médian. En revanche, nous n'observons pas le changement de signe trouvé dans (Brungart & Hobbs, 2007) et qui pourrait conduire à un accroissement de la confusion gauche-droite.

La diminution de l'ILD, comme l'augmentation de l'ITD, peut encore une fois s'interpréter comme une augmentation de la distance entre les oreilles. La Figure II.22 témoigne de cet effet : l'ILD a été estimée sur un modèle de tête sphérique dont le point de mesure s'éloigne de l'oreille jusqu'à une distance de 4 cm. La compression de l'ILD observée ici est toutefois encore loin de celle obtenue avec les casques talk-through. La diffraction de l'onde sonore sur leurs coques doit donc aussi contribuer à la compression de l'ILD.



FIG. II. 21 – ILD estimée en fonction de l'azimut à partir des mesures d'HRTF avec et sans TCAPS.



FIG. II.22 – ILD estimée à partir d'un modèle de tête sphérique pour lequel la distance entre le point de mesure et l'oreille varie de 0 cm à 4 cm.

La dégradation des indices de localisation entraînant une augmentation de la confusion avant-arrière n'a pas été mise en évidence sur les indices binauraux que sont l'ITD et l'ILD. Ce n'est pas surprenant, car ce sont justement ces indices de localisation qui engendrent une ambiguïté avant-arrière et les indices spectraux qui la résolvent. Même si les différences interaurales sont détériorées par les TCAPS, surtout l'ILD, nous nous attendons plus particulièrement à une importante dégradation des indices spectraux.

#### 2.4.3 Effet des TCAPS sur les indices spectraux

Nous nous intéressons à la dégradation des indices spectraux par les TCAPS qui, d'après les études précédentes, empêche la résolution des ambiguïtés avant-arrière. Pour représenter cette dégradation, les DTF mesurées avec et sans TCAPS sont moyennées par quadrant (les mesures aux positions 0°, 90°, 180° et 270° sont mises à l'écart), résultant en quatre spectres représentant les indices spectraux avant et arrière, ipsilatéraux et contralatéraux. Pour chacun de ces côtés, nous étudions la différence des indices spectraux avant et arrière. Ainsi une valeur positive (resp. négative) correspond à un indice mettant en évidence une source frontale (resp. arrière). Les indices spectraux avant-arrière ipsilatéraux et contralatéraux sont donnés pour chaque TCAPS sur la Figure II.23.

En condition d'écoute normale (Figure II.23a), on observe que la caractéristique spectrale principale permettant la discrimination des sources avant et arrière se situe vers 4 kHz pour l'oreille ipsilatérale et correspond à une bande directionnelle de Blauert (1969). La présence de composantes spectrales importantes pour des fréquences supérieures à 10 kHz correspond à une source arrière pour l'oreille ipsilatérale. Il s'agit aussi d'une bande direc-



FIG. II.23 – Indices spectraux avant-arrière ipsilatéraux et contralatéraux en fonction de la fréquence pour chaque TCAPS.

tionnelle mise en évidence sur la Figure II.10. Pour ces fréquences, l'oreille contralatérale joue le rôle inverse. À 1,5 kHz, les deux oreilles présentent aussi des indices spectraux opposés, mais moins importants que les indices précédents.

Les indices spectraux précédemment identifiés sont globalement conservés par les bouchons non linéaires P1 et P2 (cf. Figures II.23b et II.23c). En effet, les différences entre P2 et la condition d'écoute normale peuvent atteindre 5 dB, mais ne dépassent pas 2 dB pour les fréquences inférieures à 8 kHz quelle que soit l'oreille étudiée. Les principales différences se trouvent donc à haute fréquence. Concernant P1, les mêmes tendances sont observées, mais les différences avec la condition d'oreille libre deviennent substantielles pour le côté ipsilatéral à partir de 6 kHz, fréquence pour laquelle une faille est observée. Pour le côté contralatéral, les différences principales sont observées à partir de 8 kHz.

Avec les bouchons talk-through P3 et P4, les indices spectraux avant-arrière ipsilatéraux et contralatéraux sont fortement dégradés (plus de 3 dB à partir de 3 kHz). En étudiant la Figure II.23d, on remarque que P3 a tendance à amplifier les indices spectraux identifiés en oreille libre à 1,5 kHz et au-dessus de 10 kHz pour l'oreille ipsilatérale. Le principal indice à 4 kHz est en revanche fortement détérioré et les indices spectraux de l'oreille contralatérale chutent entre 1,5 et 5 kHz. Au maximum, les différences entre les indices spectraux avant-arrière de P3 et de la condition d'oreille libre peuvent atteindre 11 dB. Pour P4, les différences avec la condition d'écoute normale sont moins importantes (inférieures à 6 dB). Par contre, il est plus difficile de retrouver les caractéristiques de discrimination avant-arrière de la condition d'écoute normale, alors que de nouveaux pics et de nouvelles failles sont introduits, par exemple à 150 Hz et 7 kHz (cf. Figure II.23e).

Comme pour l'ILD, les dégradations les plus importantes des indices spectraux s'observent avec les casques talk-through sur les Figures II.23f et II.23g. On observe des modifications des indices spectraux avant-arrière supérieures à 3 dB à partir de 700 Hz avec P5 et 400 Hz avec P6. Avec des différences maximales de 9 dB pour P5 et de plus de 13 dB pour P6, il devient très difficile de retrouver les indices spectraux de la condition d'écoute normale :

- le pic à 4 kHz pour l'oreille ipsilatérale a disparu et est presque inversé avec les deux casques;
- au-dessus de 10 kHz, les indices spectraux ipsilatéraux sont inversés;
- les indices à 1,5 kHz des deux oreilles sont amplifiés et de même signe (inversion des indices contralatéraux) pour P5;
- les indices ipsilatéraux à 1,5 kHz sont inversés pour P6.

Par l'étude des indices spectraux avant-arrière ipsilatéraux et contralatéraux, nous sommes finalement capables de prédire un classement des TCAPS basé sur leur dégradation de ces indices, et donc sur l'augmentation éventuelle des erreurs avant-arrière. Ce classement réalise un tri des TCAPS par type. La détérioration la plus importante des indices spectraux est engendrée par les casques talk-through. Les bouchons talk-through

TAB. II.1 – ISSD (en  $\mathrm{dB}^2)$  estimée pour chacun des six systèmes tactiques de communication et protection.

TCAPS	P1	P2	$\mathbf{P3}$	P4	P5	P6
ISSD	$3,\!9$	$^{3,3}$	$9,\!6$	5,1	10,5	14,7

conservent bien certaines caractéristiques spectrales, mais moins efficacement que les bouchons non linéaires. La classification des TCAPS selon la dégradation des indices spectraux respecte l'ordre déjà établi par des expériences subjectives (Brungart & Hobbs, 2007; Zimpfer et al., 2012; Zimpfer & Sarafian, 2014).

Afin de quantifier nos observations, la moyenne sur toutes les directions de la variance de la différence entre les DTF mesurées avec TCAPS et celles mesurées en oreille libre est calculée et s'exprime en «  $dB^2$  ». Cette métrique est similaire à l'ISSD, introduite dans (Middlebrooks, 1999), qui compare les DTF de différents sujets. Il s'agit donc d'une mesure inter-sujet que nous adaptons ici à une mesure intra-sujet (différence de DTF pour une même tête). C'est pourquoi nous l'appellerons toujours ISSD. Une ISSD proche de zéro reflète une grande similarité spectrale. Le résultat de cette métrique pour nos six TCAPS et moyenné sur les 16 positions de mesure est donné dans le Tableau II.1. Ces données permettent une nouvelle fois de conclure sur la classification des TCAPS par type. Au sein d'un même type, il semblerait toutefois que les performances de localisation sonore avec P2 soient meilleures qu'avec P1. De même, l'ISSD de P4 est inférieure à celle de P3 et l'ISSD de P5 est inférieure à celle de P6. Cette métrique prédit donc le classement : P2, P1, P4, P3, P5, P6.

#### 2.5 Conclusion sur les HRTF mesurées avec les TCAPS

Pour conclure, à partir de l'observation de la modification des indices de localisation, nous avons la possibilité de retrouver des résultats qualitatifs de dégradation des performances de localisation engendrée par les TCAPS. Si aucune protection ne change substantiellement l'ITD, l'ILD est compressée par les casques talk-through, et à moindre mesure par les bouchons talk-through. Les dégradations les plus considérables sont observées pour les indices spectraux, ce qui explique que la principale cause de la baisse des performances de localisation lors d'une expérience subjective sur le plan horizontal est l'augmentation des confusions avant-arrière.

Nous supposons donc qu'il est possible de prédire les performances de localisation sonore avec TCAPS grâce à des mesures de DTF sur une tête artificielle. Dans la suite, deux modèles de localisation sonore sont développés pour ensuite être appliqués aux systèmes tactiques de communication et protection : un modèle d'appariement et un réseau de neurones.

# CHAPITRE III

# Expérience subjective de localisation sonore

Dans ce chapitre, l'expérience subjective de localisation sonore, décrivant l'influence des TCAPS sur les performances, est décrite. Tout d'abord, le protocole expérimental est détaillé, puis les résultats sont présentés. Comme cette étude servira de base à l'évaluation des performances obtenues par des modèles de prédiction, beaucoup de facteurs sont pris en compte. Nous étudions notamment le nombre de bonnes réponses, le type d'erreurs commises, la distribution des erreurs sur le plan horizontal ou l'influence des TCAPS sur le temps de réponse.

# 1 Protocole expérimental

#### 1.1 Auditeurs

30 auditeurs volontaires, 4 femmes et 28 hommes, tous membres du personnel de l'Institut franco-allemand de recherches de Saint-Louis, ont participé à cette étude. Au moment de l'expérience, ils étaient âgés de 20 à 54 ans. La moyenne d'âge était d'environ 35 ans, et la médiane de 32 ans. Une partie de ces sujets avaient déjà participé à une expérience subjective de localisation sonore similaire, mais pas moins que quatre mois plus tôt.

Avant l'expérience, nous avons fait passer un test d'audition à chaque auditeur. Il s'agit d'un test de Békésy à fréquence fixe, pour les octaves de 125 Hz à 8 kHz. Lorsque l'auditeur entend un son, il appuie sur un bouton-poussoir et continue d'appuyer tant qu'il l'entend. L'auditeur relâche le bouton dès et tant qu'il n'entend plus le stimulus. Pour chaque fréquence, la mesure se poursuit jusqu'à ce que le sujet ait appuyé et relâché trois fois le bouton. Cette méthode permet de limiter à la fois l'influence du temps de réaction de l'auditeur lorsque le niveau augmente, et l'effet de persistance lorsque le niveau diminue.

La courbe d'audition de chaque auditeur est comparée avec les seuils d'audition donnés





FIG. III.1 – Photographies du dispositif expérimental pour les expériences subjectives de localisation sonore.

en fonction de l'âge dans la norme (ISO 7029 :2000, 2000). En conclusion, nous avons trouvé que tous les auditeurs avaient une audition normale pour leur âge.

#### 1.2 Dispositif expérimental

Pour cette expérience subjective, nous avons utilisé le dispositif de localisation sonore de l'ISL (cf. Figure III.1). Dans une cabine d'audiométrie dont les murs ont été recouverts de mousse absorbante, le sujet est assis, sa tête placée au centre d'un cercle de 2,60 m de diamètre et composé de 8 haut-parleurs espacés de 45°. Ces derniers sont cachés par un rideau opaque afin de supprimer tout indice visuel de localisation des sources sonores. L'auditeur doit se positionner face à une croix jaune située à la position d'azimut 0° sur le rideau. Une technique de stéréophonie panoramique permet de produire des sources sonores virtuelles entre les haut-parleurs (Blanck et al., 2015). Ainsi le dispositif permet d'avoir à disposition 16 sources sonores différentes, schématisées sur la Figure III.2, et espacées de 22,5°, de 0° à 337,5°.

Le stimulus émis par les sources sonores réelles et virtuelles est un burst de bruit blanc de 280 ms à un niveau de 70 dB(SPL). Nous avons choisi un bruit à large bande fréquentielle afin que tous les indices de localisation soient disponibles. En effet, nous avons vu qu'un bruit à bande étroite peut être localisé par des sujets à une position dépendante de sa fréquence centrale et indépendante de sa position réelle (Blauert, 1969; Butler & Planert, 1976). Le niveau de 70 dB(SPL) est choisi, car la conservation des performances de localisation en portant les TCAPS paraît particulièrement cruciale pour des sons à bas niveau, comme la parole. Enfin, la courte durée du stimulus sonore empêche l'auditeur de réaliser des mouvements de tête pour faciliter la localisation de la source. Le bruit de fond de la cabine, lorsque tous les appareils sont en marche, est de 30 dB(A).

L'expérience est contrôlée par un ordinateur à écran tactile, représenté sur le schéma de la Figure III.3. Un programme en langage C choisi l'identifiant de la source à utiliser (1, ..., 16) et envoie l'information à une carte programmable Arduino<sup>TM</sup> MEGA 2560 via



FIG. III.2 – Disposition des sources sonores : l'auditeur est placé au centre du cercle constitué de 16 sources sonores. Les sources sonores virtuelles sont représentées en pointillés.



FIG. III.3 – Schéma du dispositif expérimental.

un port série. La carte pilote deux multiplexeurs Analog Devices ADG1606 assurant le routage du signal audio vers les entrées d'un amplificateur de puissance Dayton Audio<sup>®</sup> MA1260 connecté à huit haut-parleurs Visaton FR16WP. Le signal audio correspond au burst de bruit blanc et est transmis par la sortie audio de l'ordinateur.

#### 1.3 Procédure

Lorsque le stimulus sonore est émis, l'auditeur indique la direction dans laquelle il l'a perçu à l'aide de l'écran tactile représentant l'arrangement circulaire des 16 sources sonores. La position réelle de la source, la réponse de l'auditeur et son temps de réponse (temps entre la fin de l'émission du son et la réponse) sont enregistrés. Le choix dans la position perçue est forcé, c'est-à-dire que le sujet ne peut pas proposer une autre direction qu'une des 16 possibles. L'ordre de passage des sources sonores est aléatoire, à la seule condition que chacune d'entre elles passe exactement 5 fois. Une série de localisations est donc composée de 80 sons.

À la fin de chaque série, le sujet quantifie son ressenti quant à la difficulté du test sur une échelle continue. Il doit de plus estimer son nombre de bonnes réponses pour la série.

Afin de faciliter la prise en main du dispositif de localisation sonore, une séquence d'entraînement a été introduite. L'auditeur réalise tout d'abord un « tour d'horizon » pour se familiariser avec le stimulus sonore et l'écran tactile. Il s'agit d'une série de 32 sons dont la provenance est contrôlée par l'auditeur lui-même. La séquence se poursuit avec une série d'entraînement à la localisation de 32 sons (2 par position), semblable à une série en condition d'écoute normale, mais avec un feedback visuel : après la réponse de l'auditeur, le dispositif indique la position réelle de la source sur l'écran tactile.

En raison du temps requis par l'expérience et pour éviter l'émergence d'effets de fatigue, chaque auditeur a participé à trois séances sur trois jours différents. Lors de la première séance, le test d'audiométrie est effectué, suivi d'une séquence d'entraînement. Ensuite, l'auditeur accomplit une série de localisations en condition d'écoute normale (aussi appelée condition d'oreille libre ou condition d'oreille nue), puis 4 séries en condition d'écoute avec TCAPS. La durée totale de cette première séance est d'environ 50 minutes. Le déroulement des deuxième et troisième séances est similaire au précédent, excepté que l'audiométrie n'est pas refaite. Ces deux séances durent donc moins longtemps, à savoir environ 35 minutes.

En conclusion, chaque auditeur effectue au total 15 séries de localisation, 3 en condition d'écoute normale et 2 avec chacun des 6 TCAPS. L'ordre de passage des protecteurs est aléatoire, à la condition qu'aucun d'entre eux ne passe deux fois dans la même séance et que tous aient été utilisés avant de les tester une seconde fois.

#### 1.4 Interprétation des erreurs de localisation

La réponse d'un auditeur au test de localisation n'est pas considérée uniquement comme bonne ou mauvaise. Nous distinguons six types de réponses différents, comme indiqué sur la Figure III.4 dans le cas où l'azimut réel de la source est égal à  $22,5^{\circ}$ . Par rapport à la localisation exacte, nous considérons qu'une réponse à plus ou moins  $22,5^{\circ}$  dans le bon quadrant reste acceptable. Ce type de réponse est appelé localisation voisine. Si la différence entre la position réelle de la source et la réponse excède  $22,5^{\circ}$ , mais que la réponse demeure dans le bon quadrant, nous rentrons dans le cadre d'une erreur grossière. Dans tous les autres cas, la position perçue par l'auditeur n'est pas dans le même quadrant que la position réelle de la source. Ceci correspond soit à une erreur avant-arrière, soit à une erreur gauche-droite, soit à une erreur mixte (gauche-droite et avant-arrière). Il est important de noter que les positions  $0^{\circ}$ ,  $90^{\circ}$ ,  $180^{\circ}$  et  $270^{\circ}$  appartiennent chacune à deux quadrants. Ainsi, aucune erreur avant-arrière n'est possible aux positions d'azimut égal à  $90^{\circ}$  ou  $270^{\circ}$ , de même qu'aucune erreur gauche-droite n'est possible à  $0^{\circ}$  ou  $180^{\circ}$ .



FIG. III.4 – Interprétation des erreurs de localisation dans le cas où la source est à la position d'azimut  $22,5^\circ$ 

#### 1.5 Analyse des données

Les données des tests de localisation sonore de tous les auditeurs sont analysées dans le but de légitimer l'existence d'effets de différents facteurs sur les performances de localisation et le temps de réponse. Les TCAPS utilisés représentent le principal facteur étudié, mais nous pourrons par exemple aussi nous intéresser à l'effet de l'âge des auditeurs. L'influence de ces facteurs est déterminée par une analyse de variance (ANOVA), en général à mesures répétées. Le résultat d'une ANOVA est donné sous la forme de la statistique F calculée, de la p-valeur associée et des degrés de liberté de l'analyse. Si un effet significatif d'un facteur est révélé par l'ANOVA, une analyse post-hoc est réalisée pour comparer les niveaux de ce facteur deux à deux. Dans ce cas, un test HSD de Tukey, ou test de la différence franchement significative, est effectué.

Cependant, l'analyse de variance requiert le respect de certaines conditions d'applications, comme la normalité et l'homoscédasticité (homogénéité des variances) des échantillons. La condition de normalité est étudiée par un test de Kolmogorov-Smirnov. Si les échantillons sont normaux, la condition d'homoscédasticité est étudiée par un test de Levene. Sinon, le test de Brown-Forsythe est appliqué.

En cas de violation de l'hypothèse d'homoscédasticité uniquement, les degrés de liberté de l'analyse sont corrigés en suivant la procédure de Greenhouse-Geisser (Méot, 2014). Les différentes corrections de cette procédure sont appelées corrections de borne inférieure, de Huynh-Feldt et de Greenhouse-Geisser. En revanche, si la normalité des échantillons n'est pas respectée non plus, une méthode non paramétrique, le test de Kruskal-Wallis, est utilisée pour évaluer la signification statistique des différences entre les échantillons. Le test de Dunn permet la comparaison entre les échantillons pris deux à deux.

TCAPS	Sans	P1	P2	P3	P4	P5	P6
Type de réponse							
Localisations exactes	47,9	37	$37,\!3$	32,7	29,3	26,1	$22,\!9$
Localisations voisines	76,3	67,7	$67,\!3$	65,3	60,2	$53,\!6$	51,2
Erreurs grossières	$^{1,4}$	$^{3,4}$	$^{3,8}$	$^{4,4}$	6	$^{6,8}$	$^{6,9}$
Erreurs avant-arrière	$^{2,2}$	$^{8,5}$	$^{8,5}$	$^{9,8}$	12,7	18,7	20,5
Erreurs gauche-droite	$^{0,1}$	$_{0,3}$	$_{0,3}$	$^{0,3}$	$^{0,8}$	$_{0,5}$	$0,\!9$
Erreurs mixtes	0	$^{0,1}$	$^{0,1}$	$_{0,2}$	$0,\!3$	$^{0,4}$	$^{0,5}$

TAB. III.1 – Nombres moyens de localisations exactes, de localisations voisines et d'erreurs pour les 7 conditions d'écoute (sur 80)

# 2 Analyse des résultats

### 2.1 Performances de localisation sonore

Pour chaque TCAPS et pour la condition d'écoute normale, les résultats des auditeurs sur les différentes séances sont moyennés. Le Tableau III.1 répertorie les moyennes sur tous les auditeurs des nombres de localisations exactes et d'erreurs pour chaque condition d'écoute. Quand aucune protection auditive n'est portée, les performances de localisation sonore sont excellentes : environ 60% de localisations exactes et 95% de localisations voisines. Les quelques erreurs subsistantes sont essentiellement des erreurs grossières et avantarrière. En revanche, une augmentation du nombre d'erreurs avant-arrière et, à moindre mesure, du nombre d'erreurs grossières est constatée dès qu'un TCAPS est utilisé. Elle engendre une chute du nombre de localisations voisines et une détérioration encore plus importante du nombre de localisations exactes. Cela se traduit par une réduction de la précision, définie comme le rapport du nombre de localisations exactes sur le nombre de localisations voisines, qui décroît de 63% sans TCAPS à 45% pour P6. Les erreurs gauchedroite et mixtes connaissent elles aussi une augmentation lorsqu'un TCAPS est utilisé, mais elles restent négligeables par rapport aux autres types d'erreurs.

Pour finir, ces résultats montrent aussi les différences existantes entre les différents types de TCAPS. La multiplication des erreurs de localisation est manifestement supérieure dans le cas des casques talk-through qu'avec les bouchons d'oreilles. Les différences entre bouchons non linéaires et talk-through sont toutefois moins marquées, particulièrement entre P1, P2 et P3. Les différentes analyses suivantes vont permettre de confirmer ces observations.

#### 2.1.1 Étude du nombre de localisations exactes

La Figure III.5a représente la variabilité interindividuelle du nombre de localisations exactes pour chaque condition d'écoute sous la forme de « boîtes à moustaches ». La médiane est indiquée en rouge, tandis que les limites haute et basse de la boîte bleue représentent respectivement les troisième et premier quartiles, notés Q3 et Q1. Les « moustaches » en pointillés montrent la répartition globale des données sous Q1 et au-dessus de



FIG. III.5 – (a) Nombre de localisations exactes pour chaque condition d'écoute, représenté par des boîtes à moustaches. (b) Répartition des nombres moyens de localisations exactes en fonction de l'azimut pour chaque type de TCAPS utilisé dans l'expérience subjective.

Q3, à condition que la différence entre ces données et le quartile le plus proche ne dépasse pas 2 fois l'espace interquartile. Dans le cas contraire, les données sont représentées par une croix rouge<sup>1</sup>.

Cette première figure permet une étude graphique des différences entre les conditions d'écoute. Aucun TCAPS ne semble permettre aux auditeurs d'obtenir des performances aussi bonnes qu'en condition d'écoute normale. Le classement des TCAPS par type établi précédemment se retrouve ici, même s'il ne semble pas y avoir de différence significative entre les résultats obtenus avec le bouchon talk-through P4 et le casque talk-through P5. De plus, il est intéressant de remarquer que la baisse conséquente du nombre de localisations exactes induite par les TCAPS n'est pas accompagnée d'une augmentation de la variabilité interindividuelle. Au contraire, le plus petit écart interquartile (Q3 - Q1) est obtenu pour le casque P6.

Une analyse de variance à mesures répétées montre que l'effet de la condition d'écoute sur le nombre de localisations exactes est significatif  $\{F_{(6;174)} = 55, 15; p < 0,001\}$ . L'analyse post-hoc révèle qu'une différence significative existe entre la condition d'écoute normale et les conditions avec TCAPS, entre le groupe (P1, P2) et le groupe (P4, P5, P6), entre (P3, P4) et P6, et finalement entre P3 et P5. Ainsi, concernant le nombre de localisations exactes, le classement des TCAPS par type est valide, mais la frontière entre chaque type est mince : P3 engendre un nombre de localisations exactes proche de celui des bouchons non linéaires, et le casque talk-through P5 ne diffère pas significativement du bouchon talk-through P4.

La répartition du nombre moyen de localisations exactes en fonction de l'azimut est

<sup>1.</sup> Cette méthode peut permettre de détecter les valeurs aberrantes dans un échantillon statistique. Pour notre étude, nous avons choisi de garder toutes les données, puisqu'une valeur aberrante correspond simplement à une personne plus ou moins performante pour la tâche de localisation.

présentée sur la Figure III.5b. La valeur maximale est 5, puisque chaque position de source est testée 5 fois dans chaque série de localisations (cf. Section 1.3). Pour ne pas surcharger le diagramme polaire, les données sont regroupées par type de TCAPS. Le diagramme correspondant à la légende P1-P2, par exemple, représente ainsi la répartition moyenne obtenue avec des bouchons non linéaires.

Dans la condition d'oreille libre, le maximum de localisations exactes est atteint pour la position de source juste en face de l'auditeur. Les performances sont globalement plus mauvaises pour les sources arrière, mais atteignent un nouveau pic pour les positions situées sur l'axe interaural ( $\theta = 90^\circ$  et  $\theta = 270^\circ$ ). De plus, les résultats ne sont pas non plus symétriques par rapport au plan médian, en raison d'une réduction des performances dans le quadrant arrière droit. Enfin, on remarque aussi que les auditeurs semblent obtenir un plus grand nombre de localisations exactes pour les positions correspondant à des sources sonores réelles. En effet, pour ce type de source, ils réalisent en moyenne presque 1 bonne réponse de plus qu'avec une source virtuelle. Ceci peut se comprendre dans le cas des sources virtuelles proches de l'axe interaural, puisque la précision de localisation y est plus faible. En revanche, la baisse de performance pour les sources virtuelles proches du plan médian peut être attribuée, justement, à l'utilisation de ce type de source.

Avec les bouchons non linéaires P1 et P2, on observe sur la Figure III.5b une réduction du nombre de localisations exactes dans toutes les directions, excepté aux positions 112,5° et 135°. Ces dernières correspondent aux positions du quadrant arrière droit pour lesquelles une baisse des performances était déjà présente en condition d'oreille nue. La dégradation due aux bouchons non linéaires se traduit en moyenne par 0,7 localisation exacte de moins dans chaque direction. Une détérioration similaire des performances de localisation est induite par les bouchons talk-through P3 et P4 dans l'hémisphère avant. À l'arrière, le nombre de localisations exactes est encore plus réduit pour atteindre 1,1 bonne réponse de moins en moyenne par direction. Pour finir, comme il a été remarqué dans les analyses précédentes, les casques talk-through engendrent la plus forte réduction du nombre de localisations exactes. La Figure III.5b montre que cette chute est comparable à celle engendrée par les bouchons non linéaires et talk-through pour les positions proches des « pôles » (0°, 90°, 180° et 270°), mais se révèle encore plus conséquente aux autres positions. La dégradation maximale par rapport à la condition d'oreille libre est obtenue pour un azimut de 315° et vaut 2,5 bonnes réponses.

Les différentes observations de la réduction du nombre de localisations exactes engendrée par les TCAPS ont permis de les classer par type et de montrer leur effet en fonction de la position réelle de la source. Dans la suite, nous étudions le nombre de localisations voisines obtenu avec et sans TCAPS. L'étude du Tableau III.1 a déjà révélé que l'effet des TCAPS sur ce nombre semble être moindre, démontrant une probable baisse de la précision de localisation.



FIG. III.6 – (a) Nombre de localisations voisines pour chaque condition d'écoute, représenté par des boîtes à moustaches. (b) Répartition des nombres moyens de localisations voisines en fonction de l'azimut pour l'expérience subjective.

#### 2.1.2 Étude du nombre de localisations voisines

Nous avons considéré qu'une erreur de localisation inférieure ou égale à 22,5° demeurait acceptable et nommée localisation voisine. Bien entendu, le nombre de localisations exactes est compris dans le nombre de localisations voisines. Pour chaque TCAPS et pour la condition d'écoute normale, les données concernant ce nombre et obtenues dans l'expérience subjective de localisation sont représentées sous la forme de boîtes à moustaches sur la Figure III.6a.

On retrouve la classification des TCAPS par type obtenue précédemment. Néanmoins, les performances obtenues avec le bouchon talk-through P3 ne semblent pas différées significativement de celles obtenues avec les bouchons non linéaires. Par contre, les nombres de localisations voisines correspondant aux casques P5 et P6 sont clairement inférieurs : leur médiane passe sous la barre des 60 localisations voisines pour une série de 80 sons. Encore une fois, les performances obtenues en condition d'écoute normale sont remarquablement meilleures qu'avec les TCAPS. De plus, la dispersion des données sur l'ensemble des auditeurs est plus réduite. Ces observations peuvent faire douter de la normalité de l'échantillon en condition d'écoute normale, mais le test de Kolmogorov-Smirnov nous en assure. En revanche, l'homogénéité des variances des différents échantillons n'est pas respectée.

Une ANOVA à mesure répétée montre que la condition d'écoute a un effet significatif sur le nombre de localisations voisines  $\{F_{(6,174)} = 78,6; p < 0,001\}$ . Ce résultat reste vrai quand la borne inférieure des degrés de liberté est utilisée (Méot, 2014). La signification statistique des différences entre les conditions d'écoute prises deux à deux est évaluée par le test HSD de Tukey. Tout d'abord, le nombre de localisations voisines obtenu en condition d'oreille nue est significativement plus élevé que celui de chaque TCAPS. Ensuite, des différences significatives sont trouvées entre le groupe (P1, P2, P3) et le groupe (P4, P5, P6), ainsi qu'entre le bouchon talk-through P4 et les casques. La classification des TCAPS par type qui est ressortie de l'étude du nombre de localisations exactes devient ici une séparation entre bouchons (non linéaires et talk-through) et casques. Le nombre de localisations voisines obtenu avec le bouchon talk-through P4 reste toutefois significativement différent de ceux obtenus avec les autres bouchons.

La répartition des localisations voisines en fonction de la position réelle de la source est moyennée sur l'ensemble des auditeurs et représentée par un diagramme polaire sur la Figure III.6b. En condition d'écoute normale, la moyenne sur toutes les directions est de 4,8 localisations voisines. On peut donc dire qu'à 22,5° près, les performances de localisation des auditeurs sont quasi parfaites dans cette condition. La mise en commun de ces observations avec celles réalisées sur le nombre de localisations exactes montre que la précision de localisation, que l'on définit ici comme le rapport du nombre de localisations exactes sur celui de localisations voisines, est plus haute pour des sources sonores situées devant l'auditeur. La précision est tout de même faible pour les positions proches de l'axe interaural.

Avec les bouchons non linéaires, la distribution des localisations voisines reste à peu près circulaire, avec un décrochement notable à 180°. La baisse par rapport à la condition d'oreille nue est de 0,55 localisation voisine en moyenne. Elle se réduit à 0,85 pour les bouchons talk-through et à 1,5 pour les casques talk-through. Le décrochement à l'arrière se fait plus insistant pour les systèmes talk-through. Avec les casques, on retrouve une détérioration moins importante au niveau des pôles et maximale vers le centre des quatre quadrants. Le nombre minimal de localisations voisines est obtenu à la position d'azimut 315° et vaut 2,16.

Pour finir, nous pouvons étudions les nombres de localisations exactes et voisines conjointement. La Figure III.7 représente les valeurs moyennes sous forme de croix. La dispersion sur l'ensemble des auditeurs est illustrée sous forme d'ellipse. Ses axes sont calculés à partir de la matrice de covariance des observations faites pour les deux variables étudiées. Les valeurs propres de cette matrice de covariance définissent la taille des axes de l'ellipse et ses vecteurs propres leur direction. Cette représentation nous permet d'observer à nouveau l'isolement des performances en condition d'oreille libre, la similarité de celles obtenues avec les deux bouchons non linéaires, ainsi que les différences entre les bouchons talk-through P3 et P4. De plus, la rotation des ellipses, ou ce que l'on pourrait appeler leur « pente », apporte une information sur la précision de localisation : plus la direction du grand axe est tournée vers l'axe vertical (axe des localisations exactes), plus la précision est élevée. La perte de précision introduite par les différents TCAPS s'observe donc par la rotation du grand axe des ellipses vers l'axe horizontal.

#### 2.1.3 Remarque sur l'influence du type de source sonore

La stéréophonie panoramique, permettant de créer des sources virtuelles, peut avoir une influence sur les performances de localisation. En effet, il a précédemment été remarqué



FIG. III.7 – Représentation des valeurs moyennes (croix) des nombres de localisations exactes et voisines dans chaque condition d'écoute. Les ellipses représentent les matrices de covariance des différents échantillons.

que le nombre de localisations exactes est plus important pour les sources réelles. Le type de source ne semble au contraire pas affecter le nombre de localisations voisines, quelle que soit la condition d'écoute. Pour justifier de l'utilisation de cette métrique plutôt que du nombre de localisations exactes, nous l'étudions en fonction de la condition d'écoute et du type de source (réelle ou virtuelle).

Pour chaque auditeur, le nombre de localisations voisines est considéré séparément suivant le type de source. Ces données sont analysées par une ANOVA à mesures répétées à deux facteurs intra-sujets, la condition d'écoute et le type de source. Comme dans la sous-section précédente, elle révèle une influence significative de la condition d'écoute sur le nombre de localisations voisines  $\{F_{(6,174)} = 78,6; p < 0,001\}$ . L'analyse post-hoc montre une différence significative entre la condition d'écoute normale et les conditions avec TCAPS, entre (P1, P2) et (P4, P5, P6), et entre (P3, P4) et (P5, P6). Ces résultats sont très similaires à ceux obtenus sans prendre en compte la stéréophonie panoramique.

L'ANOVA à deux facteurs permet néanmoins de négliger l'influence globale du type de source  $\{F_{(1,29)} = 0,04; p = 0,85\}$ . Si nous avions étudié le nombre de localisations exactes, cette influence aurait été très significative. Toutefois, l'interaction de la condition d'écoute et du type de source est significative sur le nombre de localisations voisines  $\{F_{(6,174)} = 7,76; p < 0,001\}$ . Comme de nombreuses interactions ne sont pas pertinentes, l'analyse post-hoc de l'interaction est réalisée en deux parties.

Tout d'abord, nous analysons, à l'aide du test HSD de Tukey, l'effet du type de source sachant la condition d'écoute. Dans ce cas, on obtient une différence significative entre les sources réelles et virtuelles uniquement dans le cas du casque talk-through P5. Avec ce TCAPS, le nombre moyen de localisations voisines est significativement plus élevé pour des sources virtuelles (28,2 contre 25,4 avec des sources réelles). Ensuite, l'effet de la condition d'écoute est étudié séparément pour chaque type de source. Dans les deux cas, on trouve une différence significative entre la condition d'écoute normale et les conditions avec TCAPS, entre (P1, P2, P3) et (P4, P5, P6), et entre P4 et (P5, P6). La seule différence notable entre les sources réelles et virtuelles est, pour ces dernières, une différence significative entre les casques P5 et P6.

Pour conclure, par rapport à l'étude précédente du nombre de localisations voisines, on remarque que la différence significative entre P3 et P4 est causée par un effet d'interaction entre la condition d'écoute et le type de source. En s'affranchissant de ce dernier facteur, une classification des TCAPS par type est bien obtenue.

#### 2.1.4 Conclusion sur l'étude des performances de localisation

L'étude des nombres de localisations exactes et voisines obtenus dans l'expérience subjective de localisation sonore montre qu'aucun TCAPS ne permet de conserver les performances de localisation en condition d'écoute normale. Toutefois, les différents types de TCAPS induisent des performances différentes. Les bouchons non linéaires dégradent moins la localisation que les bouchons talk-through, mais la limite est mince : statistiquement parlant, les performances obtenues avec P3 ne sont pas différentes de celles obtenues avec P1 et P2. En revanche, il apparaît clairement que les casques talk-through entraînent la pire détérioration des performances de localisation. Globalement, ces résultats sont cohérents avec différentes études réalisées avec des TCAPS (Abel et al., 2007; Brungart & Hobbs, 2007; Carmichel et al., 2007; Zimpfer et al., 2012; Zimpfer & Sarafian, 2014).

Une classification des TCAPS par type a été observée et peut être confirmée en moyennant, pour chaque auditeur, les nombres de localisations voisines obtenus avec chaque type de TCAPS. Une ANOVA à mesures répétées sur ces données montre un effet significatif du type de TCAPS  $\{F_{(2,58)} = 133,4; p < 0,001\}$ . Le test HSD de Tukey révèle bien une différence significative entre chaque type de TCAPS.

Dans la suite, nous étudions plus en détail l'effet des TCAPS sur les différents types d'erreurs : erreur grossière, avant-arrière, gauche-droite ou mixte. Une étude sur le temps de réponse est aussi réalisée.

#### 2.2 Étude des erreurs

La définition des différents types d'erreurs a été donnée précédemment et est récapitulée sur la Figure III.4. D'après l'étude du Tableau III.1, la chute des performances de localisation engendrée par les TCAPS se traduit principalement par une augmentation des erreurs avant-arrière, et à moindre mesure des erreurs grossières. Ces deux types d'erreurs sont donc tout d'abord analysés dans cette partie, et la faible évolution des erreurs gauche-droite et mixtes est développée ensuite.

#### 2.2.1 Erreurs avant-arrière

La dispersion sur l'ensemble des sujets et la répartition moyenne en fonction de l'azimut sont données pour le nombre d'erreurs avant-arrière sur les Figures III.8a et III.8b. Comme pour le nombre de localisations voisines, les principales différences attendues sont entre la condition d'oreille nue et les TCAPS, ainsi qu'entre les bouchons et les casques. Encore une fois, la faible dispersion des résultats pour la condition d'écoute normale empêche la validation de l'hypothèse d'homoscédasticité. Avec la correction de borne inférieure, l'effet de la condition d'écoute sur le nombre d'erreurs avant-arrière est significatif selon l'ANOVA  $\{F_{(6,174)} = 68,8; p < 0,001\}$ . La comparaison des échantillons pris deux à deux conduit presque aux mêmes observations que pour le nombre de localisations voisines : la condition d'écoute normale est significativement différente des conditions avec TCAPS, et tous les bouchons conduisent à un nombre d'erreurs avant-arrière significativement inférieur à celui des casques. Ceci renforce l'affirmation que la confusion avant-arrière est la principale source d'erreurs de localisation dans cette expérience subjective. Les seules différences significatives entre bouchons sont observées entre (P1, P2) et P4.

Si l'on observe la répartition moyenne des erreurs avant-arrière en fonction de l'azimut (cf. Figure III.8b), on remarque que les quelques erreurs avant-arrière présentes en condition d'oreille libre apparaissent pour des sources sonores proches de l'axe interaural. Lorsque l'auditeur porte des bouchons non linéaires, non seulement les erreurs avant-arrière augmentent pour ces positions, mais de nouvelles confusions sont présentes, principalement juste derrière l'auditeur. Avec les bouchons talk-through, le pic d'erreurs avant-arrière trouvé pour les bouchons non linéaires est élargi, tandis qu'à l'avant, le nombre d'erreurs reste à peu près constant.

En revanche, dans les deux hémisphères avant et arrière, le nombre d'erreurs avantarrière est considérablement augmenté avec les casques talk-through P5 et P6. Nous rappelons toutefois que pour éviter de surcharger les diagrammes polaires, nous avons moyenné les résultats obtenus pour les TCAPS de même type. Dans le cas de la confusion avantarrière, les deux casques talk-through ont des effets opposés. Lorsque la source sonore est placée devant l'auditeur, P5 engendre un nombre moyen d'erreurs avant-arrière égal à environ 1,08 alors que ce nombre atteint 2,08 avec P6. À l'inverse, on observe 1,62 confusion avant-arrière avec P5 et 0,85 avec P6 lorsque la source sonore est dans l'hémisphère arrière. Finalement, pour simplifier, avec le casque talk-through P5, une source située à l'arrière aura tendance à être perçue à l'avant alors qu'une source à l'avant sera plutôt perçue du bon côté. Cet effet a déjà été observé avec le casque P5 lors d'une expérience subjective préliminaire (Joubaud & Zimpfer, 2014), mais les haut-parleurs n'étaient pas cachés par le rideau opaque. Nous en avions conclu que les auditeurs pouvaient se référer à des indices visuels dans le cas de confusions avant-arrière acoustiques. Cette nouvelle étude montre



FIG. III.8 – Représentations des nombres d'erreurs avant-arrière suivant la condition d'écoute.

qu'il s'agit en fait d'un effet particulier du casque P5 puisque P6 engendre quant à lui l'effet opposé.

La dégradation des performances de localisation induite par les TCAPS et dévoilée par cette expérience subjective est donc principalement due à une augmentation substantielle des erreurs avant-arrière. Cette hausse s'observe surtout juste derrière l'auditeur pour les bouchons non linéaires et plus globalement dans l'hémisphère arrière pour les bouchons talk-through. L'augmentation se généralise devant et derrière l'auditeur avec les casques talk-through, même si un hémisphère est privilégié selon le modèle utilisé.

#### 2.2.2 Erreurs grossières

La localisation dans le bon quadrant, mais avec une erreur supérieure ou égale à 45°, appelée erreur grossière, constitue la deuxième source de dégradation de la capacité des auditeurs à localiser les sources sonores (cf. Tableau III.1). Elle est toutefois moins importante, comme l'atteste la Figure III.9a, mais respecte elle aussi la classification des différents protecteurs par type : le nombre moyen d'erreurs grossières ne dépasse pas 1,4 en condition d'écoute normale, 3,8 avec les bouchons non linéaires, et peut atteindre 6 et 6,9 avec les bouchons et les casques talk-through, respectivement.

Statistiquement parlant, la condition d'écoute a encore une fois un effet significatif sur le nombre d'erreurs grossières selon l'ANOVA  $\{F_{(6,174)} = 22,8; p < 0,001\}$  (avec la correction de borne inférieure). Pour le nombre d'erreurs grossières, le test HSD de Tukey révèle que la condition d'oreille libre est significativement différente des conditions avec TCAPS. On observe de plus une différence significative entre (P1, P2) et (P4, P5, P6), et entre P3 et (P5, P6). Cette fois, la distinction entre les différentes conditions d'écoute est ténue. On peut toutefois en tirer une nouvelle séparation entre bouchons non linéaires et casques talk-through.

La répartition des erreurs grossières en fonction de l'azimut est donnée en Figure III.9b.



FIG. III.9 – Représentations des nombres d'erreurs grossières suivant la condition d'écoute.

Il est important de noter que l'échelle a été modifiée par rapport à la Figure III.8b, afin que l'augmentation de ces erreurs liée à l'utilisation des TCAPS, détectée sur le Tableau III.1 et confirmée par l'ANOVA, puisse être observée. Les erreurs grossières émergent sur les positions avant, arrière, gauche et droite pour les bouchons non linéaires et sont amplifiées pour les bouchons talk-through. Dans le cas des casques talk-through, les erreurs grossières sont concentrées en majorité sur les positions multiples de 45°, surtout à l'arrière, résultant en un diagramme polaire en forme d'étoile. Ce résultat peut être surprenant puisqu'il correspond aux positions des sources réelles. Toutefois, il doit plutôt provenir d'une plus grande difficulté à localiser pour ces positions. En particulier, les positions correspondant aux « pôles » sont celles pour lesquelles le plus grand nombre d'erreurs grossières est possible, puisqu'elles appartiennent à deux quadrants successifs.

En conclusion, notons que le nombre d'erreurs grossières réalisé par les auditeurs montre globalement une différence entre les bouchons et les casques, mais que son importance reste toutefois bien inférieure à celle des erreurs avant-arrière.

#### 2.2.3 Erreurs gauche-droite et mixtes

Les nombres moyens d'erreurs gauche-droite et mixtes ont été étudiés dans le Tableau III.1 et trouvés négligeables par rapport aux confusions avant-arrière et aux erreurs grossières. Pour cette raison, et parce qu'ils représentent une erreur de localisation grave, nous considérons ces deux types d'erreurs conjointement dans cette partie. Les représentations sous forme de boîtes à moustache et de diagramme polaire ne sont pas données, car elles fournissent trop peu d'informations.

Le faible nombre d'erreurs gauche-droite et mixte dans la plupart des conditions d'écoute nous empêche de réaliser une ANOVA. En effet, les conditions de normalité et d'homoscédasticité ne sont pas respectées par les échantillons. La statistique du test de Kruskal-Wallis est toutefois significative  $\{H_{(6)} = 34,05; p < 0,001\}$ , et le test de Dunn

montre principalement que les données de la condition d'écoute normale sont significativement inférieures de celles obtenues avec les systèmes talk-through, bouchons et casques. On peut aussi noter que le casque P6 engendre un nombre d'erreurs gauche-droite et mixtes significativement supérieur à ceux obtenus avec les bouchons non linéaires P1 et P2.

Les erreurs gauche-droite et mixtes se produisent essentiellement lorsque la source sonore est proche du plan médian (azimut égal à 22,5°, 157,5°, 202,5° ou 337,5°). Mais cet effet des TCAPS reste éphémère. En effet, même pour les casques talk-through, on ne dénombre en moyenne que 0,42 erreur gauche-droite et mixte sur l'ensemble de ces positions. Lorsque la source sonore est proche du plan médian, les différences interaurales de temps et d'intensité sont quasiment nulles. Mais c'est aussi pour ces positions que la localisation est la plus précise : les auditeurs sont capables de détecter une variation d'ITD correspondant à une modification de 1° de la position de la source (Hartmann, 1999). Le faible nombre d'erreurs gauche-droite et mixtes observé dans cette expérience subjective montre que les TCAPS ne doivent pas substantiellement dégrader l'ITD et l'ITD.

#### 2.2.4 Conclusion sur l'analyse des erreurs

L'expérience subjective de localisation sonore avec les TCAPS a permis de montrer que la dégradation des performances engendrée par les protecteurs auditifs conduit principalement à des confusions avant-arrière. Le nombre d'erreurs grossières augmente aussi de manière notable, tandis que les erreurs gauche-droite et mixtes restent négligeables. Il semble donc que les TCAPS apportent surtout une détérioration des indices spectraux, alors que les différences interaurales de temps et d'intensité seraient préservées.

De manière générale, les faibles nombres d'erreurs obtenus en condition d'écoute normale ne sont retrouvés avec aucun TCAPS. Ensuite, une séparation peut être faite entre les bouchons et les casques, ces derniers engendrant les plus grands nombres d'erreurs. Cependant, cette démarcation est mince puisque le bouchon talk-through P4 entraîne un nombre d'erreurs grossières similaire à celui des casques. Pour finir, même si les performances obtenues avec le bouchon talk-through P3 sont proches de celles des bouchons non linéaires, les TCAPS peuvent globalement être classifiés par type.

Dans la suite, nous étudions l'influence des TCAPS sur le temps de réponse des auditeurs.

#### 2.3 Temps de réponse

Dans la Section 1.3 décrivant le protocole de cette expérience de localisation, il est indiqué que le temps de réponse de l'auditeur est enregistré. À chaque fois que l'auditeur doit localiser une source sonore, avec ou sans TCAPS, nous avons mesuré le temps entre la fin de l'émission du stimulus et sa réponse. La présente section présente l'étude du temps de réponse en fonction des performances de localisation sonore, ainsi que sa distribution sur le plan horizontal.

#### 2.3.1 Temps de réponse en fonction des performances

Le temps de réponse peut être étudié en fonction de la justesse de la réponse de l'auditeur. Les Figures III.10a, III.10b et III.10c représentent la répartition des temps de réponse moyens de l'ensemble des auditeurs pour chaque condition d'écoute, respectivement dans les cas où l'auditeur réalise une localisation exacte, une localisation voisine ou une quelconque erreur de localisation. Globalement, le temps de réponse est compris entre 1 seconde et 2,5 secondes dans le cas des localisations exactes et voisines. Lorsqu'une erreur de localisation est produite, il peut atteindre plus de 3 s.

Sur chacune des figures, on remarque ensuite le fait surprenant que le temps de réponse semble généralement plus long en condition d'écoute normale qu'en condition avec TCAPS. Effectivement, le temps moyen de réponse en condition d'oreille libre, que la localisation soit bonne ou mauvaise, est de 1,66 s, alors qu'il reste compris entre 1,44 s et 1,47 s avec les TCAPS. Par contre, aucune différence substantielle de temps de réponse ne semble se dégager entre les protections auditives. Enfin, il apparaît que le temps de réponse dans le cas d'une erreur de localisation sur la Figure III.10c est plus long que dans le cas d'une localisation exacte (Figure III.10a) ou voisine (Figure III.10b). Si l'on rassemble toutes les conditions d'écoute, le temps moyen pour une localisation exacte atteint 1,42 s, 1,44 s pour une localisation voisine et 1,6 s pour une erreur.

Les observations précédentes sont confirmées par une analyse de variance à deux facteurs intra-sujets (tous les auditeurs ont testé les sept conditions d'écoute et ont réalisé des localisations exactes, voisines et des erreurs dans chaque condition). Les conditions d'application sont respectées pour chacun des deux facteurs. L'ANOVA affirme l'effet significatif de la condition d'écoute  $\{F_{(6;174)} = 8,92; p < 0,01\}$ . L'analyse post-hoc montre qu'en condition d'écoute normale, le temps de réponse est significativement plus élevé qu'avec les TCAPS. L'effet du second facteur (justesse de la réponse) est aussi significatif  $\{F_{(2;58)} = 68,68; p < 0,001\}$ . On retrouve par le test de Tukey que le temps de réponse est significativement plus long lorsque les auditeurs font une erreur de localisation. En revanche, on ne trouve aucune différence de temps de réponse entre les cas où des localisations exactes et voisines sont produites.

#### 2.3.2 Distribution du temps de réponse sur le plan horizontal

Pour aller plus loin, nous nous intéressons à la distribution du temps de réponse moyen en fonction de l'azimut. Les cas des localisations exactes et voisines rendant des résultats similaires, seul le second est étudié ici. De plus, les nombres d'erreurs gauche-droite et mixtes étant relativement faibles, on ne considère que les erreurs avant-arrière et grossières. Les Figures III.11a, III.11b et III.11c représentent la distribution du temps de réponse pour chaque type de TCAPS, respectivement pour les localisations voisines, les erreurs avant-arrière et les erreurs grossières. Le temps de réponse est fixé à 0 aux positions pour lesquelles aucune donnée n'est disponible.



FIG. III.10 – Temps de réponse moyens (en secondes) des auditeurs pour chaque condition d'écoute, représentés sous forme de boîtes à moustaches, (a) dans le cas de localisations exactes, (b) pour les localisations voisines, et (c) pour les erreurs de localisation.

Le temps de localisation voisine (Figure III.11a) est presque réparti uniformément avec tous les types de TCAPS. Il semble minimum lorsque la source sonore est placée juste devant l'auditeur. En condition d'oreille nue, le temps de réponse est systématiquement plus élevé que dans les conditions avec TCAPS. De plus, il augmente en moyenne de 90 ms lorsque la source est placée à l'arrière. Cet effet existe aussi avec les TCAPS, mais n'est que de 34 ms pour les bouchons non linéaires, 54 ms pour les bouchons talk-through et 2 ms pour les casques talk-through.

La Figure III.11b représente le temps de réponse dans le cas d'une erreur avant-arrière. On remarque une accentuation des différences entre la condition d'écoute normale et les conditions avec TCAPS, ce qui pourrait être une explication de la hausse globale du temps de réponse en condition d'oreille libre. Cependant, cette accentuation s'observe principalement dans l'hémisphère droit. D'après la Figure III.8b, cela correspond à des positions de source pour lesquelles un nombre négligeable de confusions avant-arrière est observé.

Concernant les TCAPS, on peut observer que le temps d'erreur avant-arrière est légèrement plus élevé lorsque la source sonore est derrière l'auditeur. Dans ce cas, le temps de réponse augmente en moyenne de 270 ms pour les bouchons non linéaires, de 250 ms pour les bouchons talk-through et de 135 ms pour les casques talk-through.

Le temps de réponse pour des erreurs grossières en condition d'écoute normale (cf. Figure III.11c) souffre du même problème que pour les erreurs avant-arrière : il semble qu'il y ait des positions privilégiées pour lesquelles le temps de réponse augmente, mais en réalité ce sont des positions pour lesquelles très peu d'erreurs grossières sont commises. De même, pour les bouchons talk-through P3 et P4, le temps d'erreur grossière est maximal aux azimuts 67,5° et 292,5° mais peu d'erreurs grossières sont produites à ces positions d'après la Figure III.9b. Finalement, on peut tout de même regarder les différences entre les hémisphères avant et arrière, et gauche et droit pour les TCAPS. Lorsque l'auditeur commet une erreur grossière, son temps de réponse est en moyenne augmenté de 106 ms avec P1 et P2 (la valeur nulle à 292,5° est exclue du calcul) et de seulement 47 ms avec P5 et P6 si la source est derrière lui. En revanche, ce temps est plus long de 200 ms lorsque la source est devant pour les bouchons talk-through. Ceci est probablement dû aux maximums évoqués précédemment. Enfin, des différences notables sont aussi présentes entre les hémisphères gauche et droit. Quel que soit le TCAPS utilisé, le temps d'erreur grossière est plus élevé dans l'hémisphère gauche. Cette différence est maximale pour les casques talk-through et atteint 120 ms.

#### 2.3.3 Conclusion sur l'étude du temps de réponse

L'influence des TCAPS sur le temps de réponse des auditeurs pendant un test de localisation sonore a été développée. L'hypothèse la plus évidente était que les TCAPS augmentent significativement le temps de réponse. Mais au contraire, nous avons observé



FIG. III.11 – Distribution du temps de réponse moyen (en secondes) pour chaque type de TCAPS en fonction de l'azimut de la source, (a) pour les localisations voisines, (b) pour les erreurs avant-arrière, et (c) pour les erreurs grossières.

un temps de réponse significativement plus élevé en condition d'écoute normale. Les auditeurs mettent-ils plus de temps à répondre lorsque leurs performances de localisation sont les meilleures ? Il paraît plus probable que le protocole expérimental soit la cause de cet effet : la condition d'écoute normale est systématiquement testée en première à chaque session. Il aurait fallu l'inclure dans la randomisation des conditions d'écoute pour pouvoir en tirer des conclusions comportementales. De plus, ce temps de réponse plus élevé peut résulter d'une séquence d'entraînement à la localisation trop courte.

En fait, la modification du temps de réponse due aux TCAPS est indirecte. Dans les parties précédentes, nous avons vu que les protections auditives entraînent une augmentation importante des erreurs de localisation. Or, il apparaît que le temps de réponse est significativement plus long lorsqu'une erreur est commise. Cette observation fait ressortir un point critique pour la survivance de l'auditeur : en portant des TCAPS, non seulement sa capacité à localiser une source sonore est dégradée, mais son temps de réaction est aussi rallongé quand cette capacité lui fait défaut. Par contre, notre protocole expérimental ne nous a pas permis de déceler un autre effet significatif des TCAPS : l'augmentation du temps de réponse à un test d'identification d'une source audiovisuelle. Ceci a été montré dans (Bolia & McKinley, 2000) et (McKinley, 2000) pour des bouchons ou casques à atténuation fixe. Dans (Simpson et al., 2005), il est même montré que ce temps de réponse, lorsque l'auditeur porte des bouchons et un casque comme double protection, est quasiment égal à celui obtenu lorsqu'il doit chercher une source visuelle sans aide auditive. Le temps de réponse à ce type de test est tout de même amélioré avec des TCAPS, que ce soit des bouchons non linéaires (Gallagher et al., 2014) ou des casques talk-through (Carmichel et al., 2007). Quelle que soit la condition d'écoute, le temps de réponse en cas d'erreur grossière est plus élevé dans l'hémisphère gauche. Cet effet pourrait être lié à la latéralité des auditeurs, même s'il n'a jamais, à notre connaissance, été trouvé de différences de temps de réponse entre droitiers et gauchers. En revanche, il a déjà été observé que les droitiers présentent de meilleures performances de localisation sonore (nombre de bonnes réponses) dans l'hémisphère droit, tandis que l'inverse apparaît pour les gauchers (Ocklenburg et al., 2010).

En conclusion, le temps de réponse des auditeurs est intimement lié à leurs performances de localisation. Il semble qu'améliorer les TCAPS au sens de la localisation sonore permettrait aussi de limiter la perte de temps de réaction. On peut même supposer qu'une telle amélioration pourrait améliorer le temps de recherche d'une source audiovisuelle, puisque l'utilisation de TCAPS par rapport aux protections à atténuation fixe réduit ce temps. Dans la suite, nous analysons les effets de l'âge et du genre, à la fois sur le temps de réponse et sur les performances de localisation.



FIG. III.12 – Nombre de localisations exactes et temps de réponse (en secondes) moyens pour chaque condition d'écoute et en fonction de l'âge. Les barres d'erreur représentent les erreurs types calculées sur l'ensemble des auditeurs.

#### 2.4 Effets de l'âge et du genre

Dans une étude préliminaire portant sur une expérience subjective de localisation avec le casque P5 (Joubaud & Zimpfer, 2014), nous avons montré qu'il existait un effet significatif de l'âge sur le temps de réponse en condition d'oreille libre. Dans le cadre de la présente étude, nous étudions l'effet de l'âge sur le nombre de localisations exactes ainsi que sur le temps de réponse global. L'analyse de cet effet pour les différents types d'erreurs n'est pas effectuée, car aucun résultat intéressant n'en ressort.

Pour étudier l'influence de l'âge, les auditeurs ont été divisés en deux groupes : ceux ayant moins de 32 ans (médiane des âges) et ceux étant plus âgés. Les Figures III.12a et III.12b représentent respectivement le nombre de localisations exactes et le temps de réponse moyens pour chaque condition d'écoute et pour chacun de ces deux groupes. Sur la première, on remarque que le nombre moyen de localisations exactes est systématiquement inférieur pour les sujets plus âgés. Toutefois, d'après les erreurs types, il ne semble pas que cette différence soit représentative de l'échantillon total. Cette hypothèse est vérifiée par une analyse de variance à un facteur intra-sujet (la condition d'écoute) et un facteur intersujet (l'âge), qui montre que l'âge n'a pas d'effet significatif  $\{F_{(1;28)} = 1,72; p = 0,20\}$ . En revanche, l'ANOVA révèle un effet « presque significatif » de l'interaction entre l'âge et la condition d'écoute  $\{F_{(6;168)} = 2,07; p = 0,06\}$ . Il pourrait donc y avoir un effet particulier de l'âge selon la condition d'écoute, par exemple pour le bouchon P1 et le casque P5 d'après la Figure III.12a. Ces résultats sont en accord avec la littérature, dans laquelle on trouve qu'en général, l'impact de l'âge est plus marqué pour des stimuli à bande fréquentielle étroite (Dobreva et al., 2011). Avec des protecteurs auditifs, on peut trouver un effet minime de l'âge, mais seulement en condition d'écoute normale (Abel & Hay, 1996).

Concernant l'impact de l'âge sur le temps de réponse, la même remarque peut être faite : en moyenne, les auditeurs les plus jeunes semblent répondre plus vite, quelle que



FIG. III.13 – Nombre de localisations exactes et temps de réponse (en secondes) moyens pour chaque condition d'écoute et en fonction du genre. Les barres d'erreur représentent les erreurs types calculées sur l'ensemble des auditeurs.

soit la condition d'écoute. Mais une fois de plus, l'analyse de variance montre que l'influence de l'âge n'est pas significative  $\{F_{(1;28)} = 0,56; p = 0,46\}$ , et il en est cette fois de même concernant l'interaction entre l'âge et la condition d'écoute  $\{F_{(6;168)} = 0,36; p = 0,90\}$ . Finalement, contrairement aux résultats obtenus dans (Joubaud & Zimpfer, 2014), nous n'observons pas de conséquence remarquable de l'âge sur le temps de réponse au test de localisation sonore.

D'un autre côté, nous pouvons aussi étudier les différences de performance de localisation qui peuvent exister entre les hommes et les femmes. Les résultats de cette expérience ne peuvent cependant pas être généralisés puisque la population d'auditeurs comprend seulement quatre femmes. Les Figures III.13a et III.13b représentent respectivement le nombre de localisations exactes et le temps de réponse moyens pour chaque condition d'écoute et en fonction du genre de l'auditeur. Il apparaît que, globalement, les femmes semblent obtenir un nombre plus élevé de localisations exactes pour toutes les conditions d'écoute, excepté P3. Cette différence de performance de localisation n'est toutefois pas significative  $\{F_{(1;28)} = 0.53; p = 0.47\}$  selon une ANOVA à un facteur intra-sujet (la condition d'écoute) et un facteur inter-sujet (le genre de l'auditeur). La Figure III.13a amène à penser que le genre des auditeurs peut avoir un impact substantiel sur le nombre de localisations exactes dans des conditions d'écoute particulières comme avec les TCAPS P1, P2 ou P4. Mais d'après l'ANOVA, cette interaction n'est pas significative  $\{F_{(6;168)} = 0,58; p = 0,74\}$ . Les résultats précédents se généralisent aux localisations voisines ainsi qu'aux différentes erreurs de localisation. Dans une configuration expérimentale différente (Zimpfer et al., 2012), il est souligné que les femmes semblent commettre plus d'erreurs gauche-droite. Mais nos résultats sont plus en accord avec ceux de Zündorf et al. (2011) qui ne trouvent aucune différence significative entre les performances des hommes et des femmes pour localiser une unique source. En revanche, il est montré que les hommes ont un avantage pour localiser une source sonore parmi plusieurs.

Sur la Figure III.13b, l'étude des erreurs types montre que les temps de réponse des hommes et des femmes ne sont pas si distincts. En moyenne, les femmes sont tout de même plus rapides dans toutes les conditions d'écoute sauf avec les bouchons talk-through. De plus, le genre des auditeurs ne paraît pas avoir d'effet sur une condition d'écoute en particulier. Ces observations sont confirmées par une analyse de variance : le genre n'a globalement aucun impact sur le temps de réponse  $\{F_{(1;28)} = 0,03; p = 0,86\}$ , et l'effet d'interaction entre le genre et la condition d'écoute n'est pas non plus significatif  $\{F_{(6;168)} = 0,86; p = 0,53\}$ . Finalement, en considérant toujours que le nombre de femmes était bien inférieur dans cette étude, on ne trouve aucune influence sérieuse du genre sur les performances de localisation, avec ou sans TCAPS.

#### 2.5 Étude du ressenti des auditeurs

Après chaque série du test de localisation, les auditeurs ont répondu à un court questionnaire (cf. Section 1.3). Nous leur avons demandé d'estimer leur nombre de bonnes réponses, c'est-à-dire de localisations exactes, et de quantifier leur ressenti quant à la difficulté du test. On peut observer l'estimation moyenne du nombre de localisations exactes en fonction de leur nombre réel pour chaque condition d'écoute sur la Figure III.14. Les ellipses représentent les matrices de covariance des données (de manière similaire à la Figure III.7). En comparant les valeurs moyennes à la droite d'équation y = x, on remarque que l'estimation moyenne des auditeurs est plutôt proche de la réalité. Cependant, l'observation des ellipses montre que pour les conditions d'écoute sans TCAPS et avec P1, P2 et P3, les auditeurs ont plutôt eu tendance à sous-estimer leur performance.

De plus, il apparaît sur la Figure III.14 que les nombres de localisations exactes estimés par les auditeurs permettent de classifier les TCAPS. Une analyse de variance affirme que la condition d'écoute à un impact significatif sur le nombre de localisations exactes estimé  $\{F_{(6;174)} = 39,3; p < 0,001\}$ . Le test HSD de Tukey développe ce résultat : des différences significatives sont révélées entre la condition d'oreille nue et les conditions avec TCAPS, ainsi qu'entre les bouchons (non linéaires et talk-through) et les casques. Globalement, les auditeurs ont donc ressenti la réduction de leurs performances de localisation induite par les protections auditives, et plus particulièrement par les casques.

La Figure III.15 représente la répartition de la difficulté ressentie (en pourcentage) en fonction des nombres de localisations exactes estimés. D'un côté, on observe un regroupement des valeurs moyennes calculées avec les bouchons, et de l'autre un regroupement des casques. La difficulté ressentie semble minimale pour la condition d'écoute normale, et reste très liée au nombre de localisations exactes estimé. Si ce nombre, donné ici sur une série de 80 sons, est calculé en pourcentage, alors la droite en pointillé sur la figure est la droite d'équation y = 100 - x. On remarque donc qu'en moyenne, les auditeurs ont donné une difficulté de test inversement proportionnelle à leur estimation de leur



FIG. III.14 – Représentation des valeurs moyennes (croix) des nombres de localisations exactes réels et estimés dans chaque condition d'écoute. Les ellipses représentent les matrices de covariance des différents échantillons. La courbe en pointillés est la droite d'équation y = x.

nombre de bonnes réponses. Ceci est confirmé par le fait qu'une ANOVA réalisée sur la difficulté  $\{F_{(6;174)} = 21,6; p < 0,001\}$  retourne presque les mêmes résultats que l'ANOVA précédente : différence significative entre la condition d'oreille nue et les conditions avec TCAPS, et entre les bouchons et les casques, à l'exception de P4 dont la difficulté est ressentie de façon analogue à P5.

L'hypothèse suivante a été émise avant l'étude des réponses au questionnaire : dans la mesure où les TCAPS dégradent la capacité à localiser une source sonore, l'auditeur ne devrait pas « s'en rendre compte ». Sa réponse doit bien être la position à laquelle il a perçu le son. Nous pouvions donc nous attendre à ce que les auditeurs estiment un nombre de bonnes réponses quasiment constant en fonction de la condition d'écoute, et qu'ils ressentent tout de même une augmentation de la difficulté à la tâche de localisation. En réalité, les auditeurs semblent avoir directement associé leur estimation du nombre de bonnes réponses à la difficulté, c'est-à-dire que d'une certaine manière, ils savent quand ils commettent une erreur de localisation. Ceci explique l'hésitation illustrée par l'augmentation du temps de réponse en cas d'erreur. Finalement, le questionnaire aurait pu être réduit à une seule des deux questions. D'ailleurs, si on note D la difficulté en pourcentage, et E le nombre estimé de bonnes réponses transformé en pourcentage, une ANOVA à deux facteurs intra-sujets ne montre aucune différence significative entre E et 100 - D  $\{F_{(1;29)} = 1,29; p = 0,26\}$ .



FIG. III.15 – Représentation des valeurs moyennes (croix) des nombres de localisations exactes estimés et de la difficulté ressentie dans chaque condition d'écoute. Les ellipses représentent les matrices de covariance des différents échantillons. La courbe en pointillés est la droite d'équation  $y = 100 - \frac{100}{80}x$ .

# 3 Conclusion

L'expérience comportementale décrite dans ce chapitre nous a permis d'étudier la dégradation de la capacité à localiser les sources sonores induite par les systèmes tactiques de communication et protection. Les performances de localisation en condition d'écoute normale ne sont pas retrouvées, même avec les bouchons non linéaires. Cependant, on peut noter que les auditeurs ont réalisé des séquences d'entraînement en condition d'oreille libre, et que cette condition est toujours la première testée pendant chaque session. Ceci pourrait augmenter les différences de performances de localisation que nous trouvons entre la condition d'écoute normale et les conditions avec TCAPS.

De plus, l'utilisation de sources virtuelles réalisées par stéréophonie panoramique peut être un facteur de dégradation de la localisation sonore. Les indices spectraux générés par ce type de source ne sont clairement pas identiques à ceux de la source réelle correspondante, mais sont égaux à la somme des indices spectraux des deux sources utilisées. Il a aussi déjà été montré que les auditeurs ont des comportements différents lorsqu'ils sont confrontés à des sources résultant de la stéréophonie panoramique (Baumgartner & Majdak, 2015). En utilisant uniquement des sources réelles, les différences interindividuelles pourraient être réduites, menant à des différences plus significatives entre les conditions d'écoute. Néanmoins, l'influence du type de source (réelle ou virtuelle) sur les performances de localisation sonore des auditeurs a été jugée négligeable par rapport à l'effet des TCAPS.

Finalement, l'observation des performances de localisation avec les systèmes tactiques de communication et protection a permis de classer ces protecteurs auditifs par type : les bouchons non linéaires conduisent à de plus mauvaises performances que les bouchons talk-through, et les casques talk-through engendrent la pire dégradation de la capacité à localiser les sources sonores. La principale cause de ces dégradations est l'augmentation des erreurs avant-arrière, qui peut être directement reliée au degré de détérioration des indices spectraux dans le chapitre précédent. Dans la suite, cette relation est exploitée pour tenter de prédire la dégradation des performances de localisation sonore à partir de mesures d'HRTF sur tête artificielle.
# CHAPITRE IV

# Prédiction des performances de localisation

Dans les chapitres précédents, nous avons montré que les TCAPS dégradent principalement les indices spectraux apportés à l'auditeur, ce qui se traduit par une augmentation importante des confusions avant-arrière. L'objectif de ce chapitre est de réussir à prédire la dégradation des performances de localisation sonore engendrée par les TCAPS. Cette prédiction est réalisée à partir des indices de localisation extraits des HRTF mesurées, avec et sans protection auditive, sur la tête artificielle de l'ISL (cf. Chapitre II Section 2.1.1). Deux modèles de localisation différents sont alors développés au cours du chapitre : un modèle d'appariement et un réseau de neurones à trois couches. Le premier compare les indices de localisation dégradés à une base de données d'indices en condition d'écoute normale, et est décrit dans la première partie de ce chapitre. Ensuite, le second modèle est développé. Il estime la position d'une source sonore par l'extraction des informations contenues dans les indices de localisation correspondants. Pour une source sonore et une condition d'écoute (par exemple avec un casque), la sortie espérée dès deux modèles est une distribution de probabilités de localisation en fonction de l'azimut.

Après la description des deux modèles de localisation sonore, les prédictions de performance qu'ils procurent avec les TCAPS sont comparées aux résultats de l'expérience subjective précédente. Pour chacune des 16 positions de source testées dans cette expérience, la sortie des modèles est évaluée. Il en résulte des matrices de probabilités de localisation, de dimension  $(16 \times 16)$ , dans chaque condition d'écoute. Dans la troisième partie de ce chapitre, soit ces matrices sont directement analysées, soit une étude de métriques est réalisée (probabilité de localisation voisine, d'erreurs, ...). Un examen de la vraisemblance entre les distributions de probabilités prédites et les observations subjectives est aussi effectué.

Les dernières parties de ce chapitre sont consacrées à des tentatives d'amélioration du modèle d'appariement et du réseau de neurones. Pour le premier, nous nous concentrons principalement sur le respect strict du rôle des différents indices de localisation dans la perception des dimensions latérale et polaire. L'amélioration du réseau de neurones repose sur l'optimisation de son algorithme d'apprentissage.

Ces deux modèles de localisation sonore et la comparaison de leurs prédictions avec l'expérience subjective ont fait l'objet d'une publication dans le *Journal of the Acoustical Society of America* (Joubaud et al., 2017).

# 1 Modèle d'appariement

# 1.1 Principe

Le premier modèle de localisation sonore basé sur les DTF (cf. Chapitre II, Section 2.1) que nous allons présenter est appelé modèle d'appariement et se base sur un modèle de traitement des indices spectraux par le système auditif.

L'hypothèse principale faite dans le cadre de ce modèle est que le système auditif compare le stimulus sonore à une base de données de DTF dans le but de localiser la source. Le stimulus sonore en question correspond au spectre de la source filtré par une DTF particulière. On l'appelle le stimulus proximal (Middlebrooks, 1992). Dans le cadre de notre étude, nous avons supposé que le spectre de la source est unitaire. Le stimulus proximal correspond donc à la DTF correspondant à la position de la source. Cette méthode de comparaison avec une base de données suppose que le système auditif établit une relation directe entre les DTF de la base et la position d'une source sonore dans l'espace (Middlebrooks, 1992; Zakarauskas & Cynader, 1993; Langendijk & Bronkhorst, 2002; Baumgartner et al., 2013, 2014).

La principale difficulté de cette méthode réside dans la définition du processus de comparaison entre le stimulus proximal et les DTF de la base de données. On peut le considérer comme une mesure de similarité spectrale.

# 1.2 Mesure de similarité spectrale

Dans (Middlebrooks, 1992), la similarité spectrale est calculée comme la corrélation croisée entre le stimulus et la DTF de la base de données. Alternativement, pour comparer des DTF mesurées sur différents auditeurs, Middlebrooks (1999) a aussi proposé l'ISSD, définie précédemment comme la variance de la différence entre les amplitudes des deux spectres exprimés en décibels. Dans (Langendijk & Bronkhorst, 2002), les performances de modèles de localisation utilisant la méthode de corrélation croisée et une méthode utilisant l'écart-type (c.-à-d. la racine carrée de la variance) de la différence des spectres sont étudiées. Les résultats obtenus à partir du calcul de vraisemblances montrent que la méthode de l'écart-type conduit à des performances plus proches de celles obtenues par expérience subjective. C'est pourquoi cette méthode est utilisée dans des modèles de localisation (Baumgartner et al., 2013).

Cependant, la variation des performances de localisation sur les plans sagittaux en fonction de l'angle latéral semble mieux représentée par la différence entre les gradients positifs du stimulus et de la DTF de la base de données (Baumgartner et al., 2014). Cette observation rejoint aussi les résultats de Zakarauskas & Cynader (1993). De plus, grâce au calcul de gradient, cette méthode ne prend pas en compte la différence d'amplitude globale entre les DTF et représente bien une mesure de similarité spectrale.

Nous avons donc choisi d'utiliser la méthode décrite dans (Baumgartner et al., 2014). Elle nécessite toutefois une adaptation au plan horizontal, puisque cette étude se concentre sur les performances de localisation en fonction de l'angle polaire, pour différents angles latéraux. Cette adaptation se traduit principalement par l'utilisation des différences interaurales de temps et d'intensité.

# 1.3 Développement du modèle d'appariement

Le schéma de principe du modèle d'appariement est donné en Figure IV.1. Nous allons tout d'abord décrire le traitement des indices spectraux, puis celui des différences interaurales, pour finir par la mise en commun des indices de similarité résultants.

# 1.3.1 Entrée du modèle et base de données

Nous avons déjà vu que la base de données est constituée des DTF mesurées en condition d'écoute normale et que l'entrée du modèle est le stimulus proximal. Ce dernier est la DTF correspondant à la position réelle de la source sonore, car le stimulus est supposé unitaire. Cependant, nous allons devoir comparer la sortie du modèle d'appariement avec les performances moyennes des auditeurs de l'expérience subjective. Or celle-ci exploite des sources sonores virtuelles pour disposer de 16 positions. La modélisation doit donc respecter le même paradigme.

La stéréophonie panoramique ne permet pas de reproduire les HRTF (ou les DTF) naturelles venant d'une source réelle placée à la position simulée. Elle engendre plutôt une sommation des HRTF venant des deux sources utilisées. Pour l'entrée du modèle d'appariement, nous simulons alors la stéréophonie panoramique en utilisant, comme stimulus proximal, la moyenne des deux DTF mesurées aux positions réelles les plus proches. La base de données demeure constituée des DTF mesurées en condition d'écoute normale avec 16 sources réelles, puisqu'elles représentent les indices de localisation stockés par le système auditif, indépendamment du type de source sonore.

# 1.3.2 Filtrage cochléaire

L'étape de filtrage cochléaire correspond à l'application de la banque de filtres gammatones, définie dans le Chapitre II Section 2.3, aux DTF d'entrée et stockées. Leur comparaison ne se fait donc que sur les composantes des 28 fréquences centrales des filtres.



Probabilité

FIG. IV.1 – Schéma de principe du modèle d'appariement.

# 1.3.3 Gradient positif

Avec  $f_k$  la k-ième fréquence centrale des 28 gammatones utilisées pour le lissage, soit  $D_{sp}^{\varsigma}(\theta_i, f_k)$  la DTF représentant le stimulus proximal à la position de source  $\theta_i$ . L'indice i est compris entre 1 et  $N_p$ , le nombre total de positions de mesure. La base de données contenant les DTF mesurées en condition d'écoute normale à toutes les positions se note  $\{D_{bd}^{\varsigma}(\theta_j, f_k), 1 \leq j \leq N_p\}$ . L'exposant  $\varsigma$  est égal à g ou d, représentant respectivement les DTF gauche ou droite. Pour chaque  $k = 2, \ldots, 28$ , les gradients positifs du stimulus proximal  $\tilde{D}_{sp}^{\varsigma}$  et de la base de données  $\tilde{D}_{bd}^{\varsigma}$  (pour tout  $j = 1, \ldots, N_p$ ) sont calculés :

$$\begin{cases} \tilde{D}_{sp}^{\varsigma}(\theta_i, f_k) = \max\left(D_{sp}^{\varsigma}(\theta_i, f_k) - D_{sp}^{\varsigma}(\theta_i, f_{k-1}), 0\right) \\ \tilde{D}_{bd}^{\varsigma}(\theta_j, f_k) = \max\left(D_{bd}^{\varsigma}(\theta_j, f_k) - D_{bd}^{\varsigma}(\theta_j, f_{k-1}), 0\right) \end{cases}$$
(IV.1)

#### 1.3.4 Comparaison et similarité

La comparaison entre les gradients de l'entrée et les DTF stockées par le système auditif est estimée comme une différence absolue moyenne d'après Baumgartner et al. (2016). Cette métrique est calculée pour chaque position des DTF de la base de données :

$$\forall j = 1, \dots, N_p \qquad \delta^{\varsigma}(\theta_i, \theta_j) = \frac{1}{27} \sum_{k=2}^{28} \left| \tilde{D}_{sp}^{\varsigma}(\theta_i, f_k) - \tilde{D}_{bd}^{\varsigma}(\theta_j, f_k) \right| \,. \tag{IV.2}$$

Le résultat de l'Équation IV.2 est compris entre 0 et  $+\infty$ , de petites valeurs correspondant à une grande similarité entre les DTF. Dans le but d'obtenir un nombre compris entre 0 et 1 (1 signifiant la plus grande similarité possible), la métrique  $\delta^{\varsigma}$  est normalisée par une fonction gaussienne pour donner la similarité :

$$P_{\delta}^{\varsigma}(\theta_i, \theta_j) = e^{-\frac{1}{2}(\frac{\delta^{\varsigma}(\theta_i, \theta_j)}{\sigma_{\delta}})^2} .$$
(IV.3)

Cette normalisation introduit un nouveau paramètre  $\sigma_{\delta}$  caractérisant la largeur de la gaussienne. Dans (Baumgartner et al., 2014), la similarité est obtenue avec une fonction sigmoïde à deux paramètres : l'un décrivant le degré de sélectivité et l'autre caractérisant la sensibilité de l'auditeur. Le paramètre  $\sigma_{\delta}$  dans l'Équation IV.3 est analogue au degré de sélectivité. Nous ignorons les paramètres propres à un auditeur, car notre objectif est de prédire les performances globales de localisation à l'aide d'une tête artificielle. Un algorithme d'optimisation, décrit dans une prochaine section, fixe  $\sigma_{\delta}$  pour obtenir cette prédiction.

#### **1.3.5** Pondération binaurale

La similarité spectrale est calculée séparément pour les deux oreilles (paramètre  $\varsigma$ ). On considère que le système auditif utilise les deux oreilles pour localiser, mais exclusivement les indices spectraux ipsilatéraux pour des sources sonores à un angle latéral supérieur



FIG. IV.2 – Pondérations appliquées aux indices de similarité gauche et droit en fonction de l'azimut.

à 60° (Morimoto, 2001). Les indices de similarité gauche et droit sont donc pondérés en utilisant la même méthode que Baumgartner et al. (2014), qui retrouvent les résultats de Morimoto (2001) et Macpherson & Sabin (2007). La pondération appliquée à la similarité de droite est définie pour  $\theta_i = -90^\circ, \ldots, 270^\circ$  par :

$$W^{d}(\theta_{i}) = \begin{cases} (1 + e^{-\theta_{i}/\Theta})^{-1}, & -90^{\circ} \le \theta_{i} \le 90^{\circ} \\ (1 + e^{-(180 - \theta_{i})/\Theta})^{-1}, & 90^{\circ} \le \theta_{i} \le 270^{\circ} \end{cases}$$
(IV.4)

La pondération de gauche est simplement calculée comme  $W^g(\theta_i) = 1 - W^d(\theta_i)$ . Dans l'équation précédente, le paramètre  $\Theta$  est fixé à 13° d'après Baumgartner et al. (2014). La Figure IV.2 représente ces deux pondérations en fonction de l'azimut (de 0° à 355°). Pour finir, l'indice de similarité globale résultant est :

$$P_{\delta}(\theta_i, \theta_j) = W^d P^d_{\delta}(\theta_i, \theta_j) + W^g P^g_{\delta}(\theta_i, \theta_j).$$
(IV.5)

# 1.3.6 Utilisation des différences interaurales de temps et d'intensité

Si le modèle d'appariement n'exploitait que les indices spectraux, comme décrit précédemment, nous n'aurions pas correctement modélisé la localisation sonore sur le plan horizontal. Dans cette dimension, l'ITD et l'ILD sont les indices de localisation prépondérants. Leur exploitation se doit donc d'être incluse dans le modèle.

Dans (Middlebrooks, 1992), l'ILD est utilisée depuis la valeur absolue de la différence des ILD correspondant aux DTF d'entrée et de la base de données. L'opposé de cette valeur absolue est ensuite transformé en score standard. La distribution ainsi obtenue a une moyenne nulle et un écart-type unitaire. Une valeur positive indique alors des ILD proches.

Un nombre compris entre 0 et 1 est tout de même plus approprié à la mesure de similarité. Nous avons choisi de normaliser les différences entre les ITD et ILD de l'entrée et de la base de données de la même manière que la mesure de similarité spectrale, en utilisant une gaussienne. Soient  $ITD_{sp}(\theta_i)$  et  $ILD_{sp}(\theta_i)$  respectivement l'ITD et l'ILD du stimulus proximal à la *i*-ème position, et soient  $ITD_{bd}(\theta_j)$  et  $ILD_{bd}(\theta_j)$  respectivement l'ITD et l'ILD du stimulus proximal à la *i*-ème position de la base de données. Les différences d'ITD et d'ILD normalisées sont calculées comme suit :

$$P_{ITD}(\theta_i, \theta_j) = e^{-\frac{1}{2} \left(\frac{ITD_{sp}(\theta_i) - ITD_{bd}(\theta_j)}{\sigma_{ITD}}\right)^2},$$
(IV.6)

$$P_{ILD}(\theta_i, \theta_j) = e^{-\frac{1}{2} \left(\frac{ILD_{sp}(\theta_i) - ILD_{bd}(\theta_j)}{\sigma_{ILD}}\right)^2}.$$
 (IV.7)

Deux paramètres, fixant la largeur des gaussiennes, sont encore introduits :  $\sigma_{ITD}$  et  $\sigma_{ILD}$ . Ils reflètent les biais d'ITD et d'ILD en termes d'erreurs de localisation et sont donc étroitement liés aux différences juste perceptibles (JND) des ITD et ILD. Dans (Busson, 2006, chap. 2, sec. 1), il est montré que la JND de l'ITD est comprise entre 10  $\mu$ s pour les sources frontales et 150  $\mu$ s pour les sources proches de l'axe interaural. De façon similaire, la différence juste perceptible de l'ILD se trouve entre 1,4 et 2,6 dB selon la bande fréquentielle considérée (Francart & Wouters, 2007). Comme le paramètre  $\sigma_{\delta}$  de la gaussienne donnant la similarité spectrale, les paramètres  $\sigma_{ITD}$  et  $\sigma_{ILD}$  seront fixés par un algorithme d'optimisation.

#### 1.3.7 Pondération des indices de localisation

Les sorties du modèle d'appariement sont finalement  $P_{\delta}$ ,  $P_{ITD}$  et  $P_{ILD}$ . Ces indices de similarité étant indépendants, la première idée était de définir leur produit comme un indice de similarité globale. Toutefois, Macpherson & Middlebrooks (2002) ont démontré que les différents indices de localisation ne sont pas pondérés de façon égale par le système auditif. Dans le cas d'une source sonore à large bande fréquentielle, un poids plus fort est donné à l'ITD par rapport à l'ILD. Ils ont aussi confirmé le fait que les indices spectraux monauraux ont un rôle mineur dans la perception de l'angle latéral. Dans notre cas, leur rôle est de discriminer les sources avant et arrière.

Pour respecter ces observations, les poids  $w_{\delta}$ ,  $w_{ITD}$  et  $w_{ILD}$  sont attribués respectivement aux indices de similarité obtenus avec les indices spectraux, l'ITD et l'ILD. Une similarité globale est calculée comme la moyenne géométrique pondérée :

$$P(\theta_i, \theta_j) = (P_{\delta}(\theta_i, \theta_j)^{w_{\delta}} P_{ITD}(\theta_i, \theta_j)^{w_{ITD}} P_{ILD}(\theta_i, \theta_j)^{w_{ILD}})^{1/(w_{\delta} + w_{ITD} + w_{ILD})} .$$
(IV.8)

Le calcul est réalisé pour chaque position  $\theta_j$  des DTF de la base de données. Finalement,  $P(\theta_i, \theta_j)$  est normalisé de façon à ce que sa somme sur toutes les positions  $\theta_j$  soit égale à 1. L'estimation de P pour chaque position de source  $\theta_i$  nous permet d'obtenir une matrice de probabilité de localisation (de taille  $16 \times 16$  dans notre cas), dont les colonnes correspondent aux positions de source et les lignes aux positions de mesure de la base de données. Ces probabilités devraient refléter les performances globales de localisation sonore obtenues par une expérience subjective. Mais dans le but d'obtenir cette équivalence, les différents paramètres du modèle d'appariement doivent être optimisés.

# 1.4 Optimisation des paramètres

Dans le modèle d'appariement développé dans cette section, six paramètres inconnus ont été introduits : les degrés de sélectivité  $\sigma_{\delta}$ ,  $\sigma_{ITD}$  et  $\sigma_{ILD}$ , et les pondérations  $w_{\delta}$ ,  $w_{ITD}$ et  $w_{ILD}$ , attribués à chaque indice de localisation. L'objectif du processus d'optimisation de ces paramètres est de faire en sorte que le modèle d'appariement retourne des performances de localisation proches de celles obtenues dans l'expérience subjective en condition d'oreille libre. Nous pourrons ensuite étudier le comportement du modèle lorsqu'on lui applique des DTF mesurées avec des TCAPS. Finalement, c'est l'étape d'optimisation des paramètres du modèle d'appariement qui nécessite la conservation du paradigme de l'expérience subjective et l'utilisation de DTF simulant les sources virtuelles.

Pour procéder à l'optimisation, les réponses de l'ensemble des auditeurs en condition d'écoute normale sont rassemblées en une matrice de confusion, dont les colonnes représentent les positions cibles et les lignes les réponses des auditeurs. On transforme cette matrice de confusion en matrice de probabilité de localisation, que l'on nommera matrice subjective, afin de pouvoir la comparer avec la sortie du modèle  $P(\theta_i, \theta_j)$ . Dans le processus d'optimisation, nous utiliserons aussi une matrice représentant les probabilités de localisation à partir d'une réponse au hasard, appelée la matrice-chance. Pour une source à la position  $\theta_i$ , la probabilité de localisation au hasard est constante, quel que soit  $\theta_j$ . Dans notre cas, tous les éléments de cette matrice sont donc égaux à  $\frac{1}{16}$ .

Les différences interaurales permettent la perception de la dimension latérale. On s'attend donc à ce que  $P_{ITD}$  et  $P_{ILD}$  retournent une forte similarité à la position réelle ainsi qu'à la position symétrique par rapport à l'axe interaural. Le rôle de  $P_{\delta}$  représentant les indices spectraux est ensuite d'effectuer la discrimination avant-arrière. Afin d'assurer que le rôle de chaque indice de localisation est préservé dans l'application de l'Équation IV.8, le processus d'optimisation est divisé en deux parties : on cherche tout d'abord les paramètres relatifs aux ITD et ILD, puis les paramètres relatifs aux indices spectraux sont optimisés.

# 1.4.1 Paramètres relatifs aux différences interaurales

Pour simuler la détection de l'angle latéral à l'aide de l'ITD et de l'ILD, le modèle d'appariement est appliqué en négligeant les indices spectraux. Nous fixons donc  $w_{\delta} = 0$ dans l'Équation IV.8. La sortie du modèle d'appariement en résultant ne peut toutefois pas



FIG. IV.3 – Schéma de la projection sur la dimension latérale depuis la dimension azimutale. Les éléments des positions arrière sont ajoutés aux éléments avant. Les flèches rouges ne représentent que le cas des sources réelles.

être comparée aux résultats de l'étude subjective, puisque celle-ci n'empêche pas l'utilisation des indices spectraux. Nous faisons alors l'hypothèse que les différences interaurales de temps et d'intensité sont parfaitement symétriques par rapport à l'axe interaural. Cela nous permet de ne prendre en compte que les positions frontales, dont l'azimut est compris entre -90° et 90°, et donc de correctement simuler la perception de l'angle latéral grâce aux ITD et ILD. Concrètement, pour cette étape d'optimisation, les similarités  $P_{ITD}$  et  $P_{ILD}$  des Équations IV.6 et IV.7 ne sont calculées que pour  $-90^\circ \leq \theta_i, \theta_j \leq 90^\circ$ . La sortie du modèle  $P(\theta_i, \theta_j)$  est donc finalement de dimensions (9 × 9).

L'hypothèse de symétrie par rapport à l'axe interaural est aussi appliquée à la matrice subjective et à la matrice-chance de la façon suivante : leurs éléments correspondant aux positions arrière (90°  $< \theta < 270°$ ) sont ajoutés aux éléments avant. Il s'agit en fait d'une projection sur la dimension latérale depuis la dimension azimutale. La Figure IV.3 schématise cette transformation. Dans notre configuration, il en résulte des matrices de dimensions (9 × 9), qui sont ensuite de nouveau normalisées pour obtenir des probabilités de localisation pour chaque position cible (c.-à-d. pour chaque colonne). Ces matrices peuvent alors être comparées à la sortie du modèle d'appariement.

La fonction-coût utilisée pour l'optimisation est inspirée de celle de Baumgartner et al. (2014). Il s'agit de l'erreur quadratique entre les probabilités estimée P et subjective  $P_s$ , normalisée par la probabilité de localisation au hasard  $P_h$ . Elle est calculée pour toutes les positions  $\theta_i$  et  $\theta_j$  situées devant l'auditeur :

$$E = \sqrt{\frac{1}{81} \sum_{-90^{\circ} \le \theta_i, \theta_j \le 90^{\circ}} \left(\frac{P(\theta_i, \theta_j) - P_s(\theta_i, \theta_j)}{P_h(\theta_i, \theta_j)}\right)^2}.$$
 (IV.9)

E dépend des paramètres liés aux différences interaurales de temps et d'intensité  $\sigma_{ITD}$ ,  $\sigma_{ILD}$ ,  $w_{ITD}$  et  $w_{ILD}$ . Nous cherchons donc à trouver le jeu de paramètres la minimisant.

Nous pouvons nous attendre à des écarts-types des gaussiennes (Équations IV.6 et

IV.7) compris entre les valeurs minimale et maximale de la JND de l'ITD et de l'ILD, soit respectivement 10  $\mu$ s et 150  $\mu$ s, et 1,4 et 2,6 dB. Ces valeurs proviennent toutefois d'études comprenant un dispositif expérimental différent (Busson, 2006; Francart & Wouters, 2007). Pour ne pas restreindre excessivement le domaine d'étude, E sera calculée pour 5  $\mu s \leq \sigma_{ITD} \leq 500 \ \mu s$  avec un pas de 1  $\mu$ s et pour 0,1 dB  $\leq \sigma_{ILD} \leq 5$  dB avec un pas de 0.1 dB.

Les pondérations des deux indices binauraux peuvent en théorie être quelconques. Mais puisque  $w_{\delta} = 0$ , on remarque dans l'Équation IV.8 qu'elles sont divisées par leur somme. On peut donc mesurer l'erreur E pour la pondération  $w_{ILD}$  comprise entre 0 et 1 et fixer  $w_{ILD} = 1 - w_{ITD}$ . Ces estimations sont réalisées avec un pas de 0,01.

La fonction-coût E est évaluée suivant les trois dimensions  $\sigma_{ITD}$ ,  $\sigma_{ILD}$  et  $w_{ITD}$ . Les paramètres optimaux sont ceux qui minimisent cette fonction-coût. En réalisant cette optimisation à partir de la matrice subjective obtenue en condition d'oreille libre dans le chapitre précédent, nous obtenons les valeurs suivantes :

- écart-type de l'ITD  $\sigma_{ITD} = 101 \ \mu s$ ;
- écart-type de l'ILD  $\sigma_{ILD} = 4,2 \text{ dB};$
- pondération de l'ITD  $w_{ITD} = 0,68$ ;
- pondération de l'ILD  $w_{ILD} = 1 w_{ITD} = 0.32$ .

La Figure IV.4 représente E en fonction de chacun des paramètres d'optimisation, lorsque les deux autres sont fixés à leur valeur optimale. Par exemple, la Figure IV.4a donne la courbe de E en fonction de  $\sigma_{ITD}$  quand  $\sigma_{ILD} = 4,2$  dB et  $w_{ITD} = 0,68$ . Cette courbe montre que pour l'écart-type de l'ITD, le minimum de la fonction-coût est clairement défini. Concernant la pondération de l'ITD, le minimum est plus difficile à percevoir, mais il reste unique et global sur le domaine étudié. Le tracé de l'erreur en fonction de l'écart-type de l'ILD montre que le minimum est moins marqué et que ce paramètre aurait plus tendance à atteindre une asymptote. Ce qui a été confirmé par l'étude de E pour des valeurs de  $\sigma_{ILD}$  allant jusqu'à 20 dB. Nous avons aussi observé que des jeux de paramètres différents pouvaient donner une valeur de E très proche du minimum global. Cependant, ils engendrent une prépondérance de l'ILD sur l'ITD qui ne devrait pas être observée pour un stimulus large bande.

Par rapport aux valeurs de JND données précédemment, les écarts-types optimaux de l'ITD et de l'ILD sont relativement élevés. Cela suggère qu'en moyenne, l'estimation de l'angle latéral au cours de l'expérience subjective de localisation ne requiert pas une grande sensibilité aux variations des indices binauraux. De plus, la grande valeur de  $\sigma_{ILD}$ par rapport aux résultats de Francart & Wouters (2007) est certainement causée par le fait que notre étude se concentre sur la localisation de source à large bande fréquentielle. Ensuite, nous observons que l'ITD demeure l'indice prépondérant dans la perception de l'angle latéral. Ceci est aussi dû au fait que l'on considère la localisation d'une source sonore à large bande fréquentielle. Dans cette condition, Macpherson & Middlebrooks (2002) ont trouvé des pondérations de l'ITD et de l'ILD respectivement égales à 0,82



FIG. IV.4 – Optimisation des paramètres liés aux ITD et ILD pour le modèle d'appariement : (a)  $\sigma_{ITD}$ , (b)  $\sigma_{ILD}$  et (c)  $w_{ITD}$ .

et 0,52 qui, une fois normalisées, deviennent 0,61 et 0,39. Nous retrouvons bien cette domination de l'ITD avec de plus des valeurs de pondérations très proches de celles de Macpherson & Middlebrooks. Il faut toutefois noter que la prépondérance de l'ITD serait légèrement réduite si les entrées du modèle étaient uniquement les DTF mesurées avec des sources réelles : en appliquant la méthode d'optimisation dans ce cas, nous trouvons que  $w_{ITD} = 0,60$ . L'utilisation de sources virtuelles semble donc entraîner une augmentation de l'importance de l'ITD par rapport à l'ILD.

# 1.4.2 Paramètres relatifs aux indices spectraux

Les écarts-types et les pondérations des différences interaurales de temps et d'intensité trouvés dans la première phase de l'optimisation sont utilisés pour réaliser la seconde phase. Celle-ci permet de trouver les paramètres relatifs aux indices spectraux  $\sigma_{\delta}$  et  $w_{\delta}$ . De façon similaire à l'optimisation précédente, la sortie complète du modèle d'appariement, sous forme de matrice de dimensions (16 × 16), est comparée à la matrice subjective pour obtenir une fonction-coût à minimiser :

$$E = \sqrt{\frac{1}{256} \sum_{\theta_i, \theta_j} \left( \frac{P(\theta_i, \theta_j) - P_s(\theta_i, \theta_j)}{P_h(\theta_i, \theta_j)} \right)^2},$$
 (IV.10)

où P,  $P_s$  et  $P_h$  représentent toujours respectivement les probabilités de localisation estimée, subjective et au hasard. Mais cette fois, le calcul de la valeur RMS se fait sur l'ensemble des couples de positions  $\theta_i$  et  $\theta_j$ , soit 256 couples. La fonction-coût E dépend à présent uniquement des paramètres  $\sigma_{\delta}$  et  $w_{\delta}$ .

Peu d'informations sont disponibles pour définir le domaine d'étude de  $\sigma_{\delta}$ . Un paramètre similaire est introduit dans (Langendijk & Bronkhorst, 2002) et est fixé à 2 dB, mais dans le contexte de la localisation sonore sur des plans sagittaux et sans utiliser le gradient positif des DTF. On peut tout de même prévoir le comportement de la fonction-coût lorsque  $\sigma_{\delta}$  tend vers l'infini. D'après l'Équation IV.3,  $P_{\delta}^{\varsigma}$  va tendre vers 1 quelles que soient les positions  $\theta_i$  et  $\theta_j$ . La fonction-coût va donc avoir un comportement asymptotique. Son minimum doit être inférieur à cette valeur limite puisqu'il paraît inimaginable que  $P_{\delta}^{\varsigma} = 1$ .



FIG. IV.5 – Optimisation des paramètres liés aux indices spectraux pour le modèle d'appariement : (a)  $\sigma_{\delta}$  et (b)  $w_{\delta}$ .

Le paramètre  $\sigma_{\delta}$  doit finalement être testé sur un domaine suffisamment large pour pouvoir observer le début du comportement asymptotique. Celui que nous avons choisi s'étend de 0,01 à 20 dB, avec un pas de 0,01 dB.

La pondération des indices spectraux de l'Équation IV.8,  $w_{\delta}$ , peut aussi être quelconque. Mais nous avons vu que le rôle de ces indices de localisation est principalement de permettre la discrimination avant-arrière. Les différences interaurales de temps et d'intensité restent les indices prépondérants pour la localisation sonore sur le plan horizontal. Comme la somme de leur pondération est égale à 1, nous supposons que la pondération optimale des indices spectraux doit être inférieure à 1. Elle est donc testée de 0 à 1 avec un pas de 0,01.

Le minimum global de la fonction-coût est obtenu pour les valeurs de paramètres suivantes :

- écart-type des indices spectraux  $\sigma_{\delta} = 0.1 \text{ dB}$ ;
- pondération des indices spectraux  $w_{\delta} = 0.06$ .

La Figure IV.5 présente la fonction-coût en fonction de chacun des deux paramètres, lorsque l'autre est fixé à sa valeur optimale. Sur chaque courbe, le minimum de la fonctioncoût est clairement défini. Pour le paramètre  $\sigma_{\delta}$ , ce minimum est bien situé avant le début d'une asymptote : en augmentant  $\sigma_{\delta}$  jusqu'à 100 dB, nous avons observé que E tend vers environ 1,85. La valeur optimale de 0,1 dB pour  $\sigma_{\delta}$  peut être indirectement comparée à la valeur de 6 dB<sup>-1</sup> du degré de sélectivité  $\tau$  obtenu dans (Baumgartner et al., 2014). Les deux paramètres caractérisent la sélectivité de leur fonction respective, c'est-à-dire sa largeur. La largeur à mi-hauteur est donc choisie comme mesure de comparaison entre les fonctions. Ainsi, la fonction sigmoïde de Baumgartner correspond approximativement à une fonction gaussienne d'écart-type égal à 0,16 (cf. Figure IV.6). La sélectivité spectrale obtenue dans notre étude est donc finalement plus petite, mais assez similaire à celle de Baumgartner et al. (2014).

L'optimisation de la pondération  $w_{\delta}$  face à celle des différences interaurales de temps et



FIG. IV.6 – Comparaison de la fonction sigmoïde de Baumgartner et al. (2014), de la fonction gaussienne correspondante ( $\sigma_{\delta} = 0,16$ ) et de la fonction gaussienne obtenue au cours du processus d'optimisation ( $\sigma_{\delta} = 0,1$ ).

d'intensité confirme le fait que les indices binauraux dominent la localisation sonore sur le plan horizontal. Dans (Macpherson & Middlebrooks, 2002), pour la localisation de l'angle latéral, le poids des indices spectraux est systématiquement trouvé inférieur à 0,1 et le plus souvent non significativement différent de 0. Dans notre étude, même l'utilisation des indices spectraux pour la discrimination avant-arrière n'entraîne pas une augmentation substantielle de  $w_{\delta}$  puisqu'elle est fixée à la valeur de 0,06. On s'attend en revanche à ce que  $w_{\delta}$  devienne plus importante que les pondérations  $w_{ITD}$  et  $w_{ILD}$  dans le cadre de la perception de l'élévation.

# 1.5 Conclusion sur le modèle d'appariement

Avec les paramètres obtenus au cours des deux phases d'optimisation précédentes, le modèle d'appariement est complet et permet de prédire les performances de localisation sonore en condition d'oreille libre les plus proches de celles observées au cours de l'expérience subjective. En remplaçant les DTF d'entrée du modèle par celles mesurées avec des systèmes tactiques de communication et protection, nous devrions pouvoir prédire la dégradation de la capacité à localiser les sources sonores qu'ils entraînent. Ces résultats sont exploités dans une partie ultérieure, car la prochaine décrit une seconde modélisation des mécanismes de localisation basée sur un réseau de neurones.



FIG. IV.7 – Schéma d'un perceptron à m entrées.

# 2 Réseau de neurones

La seconde approche a pour objectif l'implémentation d'un réseau de neurones modélisant le processus de localisation du système auditif (Neti et al., 1992b; Chung et al., 2000; Jin et al., 2000). L'idée est « d'entraîner » le réseau de neurones à identifier la position des sources sonores à partir de DTF mesurées en condition d'écoute normale, puis de lui appliquer en entrée les DTF dégradées par les TCAPS. Pour ce modèle, la mise en relation des DTF en condition d'oreille libre avec les positions dans l'espace est plus indirecte que pour le modèle d'appariement : à la place de stocker ces DTF, le réseau de neurones va trouver et mémoriser leurs caractéristiques permettant la localisation d'une source sonore. Ce type de modélisation a tout d'abord été étudié à partir de DTF de chat (Middlebrooks & Knudsen, 1984; Neti et al., 1992b), puis sur des DTF humaines (Chung et al., 2000; Jin et al., 2000).

Dans cette partie, nous présentons la modélisation du traitement de l'information spatiale par un réseau de neurones à trois couches. Le type de neurone utilisé est tout d'abord décrit, puis mis en réseau avec d'autres neurones. Ensuite, la méthode d'apprentissage du réseau par rétropropagation du gradient de l'erreur est développée. Pour finir, l'application du réseau dans le cadre de la localisation sonore avec TCAPS est détaillée.

# 2.1 Modélisation

# 2.1.1 Modèle de neurone

Un modèle basique de neurone est le modèle de McCulloch-Pitts. Ce neurone réalise une somme pondérée de ses entrées, qu'il compare à un biais. Cette somme est appelée l'activité du neurone et ses pondérations les poids synaptiques. Si l'activité est supérieure à la valeur seuil, le neurone renvoie 1 et 0 sinon. C'est l'activation du neurone. Cette dernière étape apporte un caractère non linéaire au fonctionnement du neurone. La Figure IV.7 donne un exemple de neurone de McCulloch-Pitts à m entrées, de biais b et de fonction d'activation f. Pour trouver la sortie s du neurone, on calcule l'activité v :

$$v = \sum_{i=1}^{m} w_i u_i + b,$$
 (IV.11)

et on lui applique la fonction d'activation f. Ainsi, on obtient :

$$s = f(v) = \begin{cases} 1 & si \, v \ge 0 \\ 0 & si \, v < 0 \end{cases}$$
(IV.12)

L'évaluation binaire de l'activité par une fonction seuil peut être remplacée par l'application d'une autre fonction d'activation. Dans (Haykin, 1998, chap. 1, sec. 3), deux autres exemples sont donnés. La fonction linéaire par morceaux :

$$f(v) = \begin{cases} 1, & v \ge a \\ \frac{v+a}{2a}, & -a < v < a \\ 0, & v \le -a \end{cases},$$
(IV.13)

avec *a* le facteur d'amplification, peut être considérée comme une approximation d'un amplificateur non linéaire. Elle se réduit à une fonction seuil lorsque le facteur d'amplification est infiniment grand. Le second exemple de fonction d'activation est la fonction sigmoïde, représentée en Figure IV.8, et définie par :

$$f(v,a) = \frac{1}{1 + e^{-av}},$$
 (IV.14)

avec a sa pente. Nous noterons à présent f(v) la fonction sigmoïde de pente unitaire f(v, 1). La sigmoïde a l'avantage d'être infiniment différentiable. En utilisant la fonction sigmoïde comme fonction d'activation du neurone de McCulllogh-Pitts, on obtient un modèle classique de neurone appelé le *perceptron*.

En pratique, le biais du neurone est traité comme un poids synaptique de la somme pondérée, appliqué à une entrée constante égale à 1. Finalement, on peut écrire, à partir des Équations IV.11 et IV.14, l'équation du perceptron :

$$\begin{cases} s = f(\sum_{i=0}^{m} w_i u_i) \\ f(v) = \frac{1}{1 + e^{-v}} \end{cases},$$
 (IV.15)

avec  $u_0 = 1$  et  $w_0 = b$ .

# 2.1.2 Mise en réseau

Les perceptrons ne permettent de résoudre que des problèmes relativement simples (problèmes linéairement séparables). La modélisation de la localisation sonore se fait par une mise en réseau de perceptrons, sous la forme de plusieurs couches de neurones. Ce réseau se nomme perceptron multicouche quand les neurones du réseau (ou nœuds) sont des perceptrons avec une fonction d'activation de type sigmoïde. Il s'agit d'un type de réseau de neurones non bouclé, qui « réalise une (ou plusieurs) fonction(s) de ses entrées par composition des fonctions réalisées par chacun des neurones » (Dreyfus et al., 2008).



FIG. IV.8 – Graphe de la fonction sigmoïde pour différentes pentes a.



FIG. IV.9 – Schéma d'un réseau de neurones à une couche couchée. Il comprendmentrées, nnœuds dans sa couche cachée etpnœuds de sortie.

La Figure IV.9 représente un exemple de réseau de neurones à 3 couches, totalement connecté. Par convention, la première couche du réseau, ou couche d'entrée, n'est pas constituée de neurones et ne réalise aucune opération sur les entrées. La seconde couche est appelée couche cachée, et la dernière couche de sortie. Le réseau est considéré comme totalement connecté lorsque chaque neurone d'une couche prend comme entrées toutes les sorties de la couche précédente. Notre exemple de réseau comprend m entrées, n nœuds dans sa couche cachée et p sorties. Notons  $w_{ij}^c$  le poids synaptique du j-ème nœud de la couche cachée correspondant à la i-ème entrée. En reprenant ce qui a été dit à la fin de la section précédente, le biais de chaque nœud s'écrit  $w_{0j}^c$ , avec  $x_0 = 1$ . Ainsi, les sorties de la couche cachée se calculent de la façon suivante :

$$\forall j \in \{1, \dots, n\}: \qquad y_j = f(\sum_{i=0}^m w_{ij}^c x_i),$$
 (IV.16)

Le calcul est similaire pour trouver les sorties de la dernière couche de neurones (et du réseau). En notant  $w_j k^s$  la pondération du k-ième nœud de la couche de sortie appliquée à la sortie de la couche cachée  $y_j$  et  $w_{0k}^s$  le biais de chaque nœud avec  $y_0 = 1$ , on obtient :

$$\forall k \in \{1, \dots, p\}: \qquad z_k = f(\sum_{j=0}^n w_{jk}^s y_i).$$
 (IV.17)

# 2.2 Apprentissage du réseau

### 2.2.1 Rétropropagation du gradient

Les inconnues d'un réseau de neurones sont les différents poids synaptiques de ses couches. Pour les déterminer, il faut utiliser des exemples. Supposons que nous disposons d'entrées et de sorties théoriques connues, que nous notons  $(x^t, z^t)$ . Pour chaque couple, on définit l'erreur du réseau comme la différence quadratique entre la sortie théorique et la sortie réelle du réseau :

$$E = \frac{1}{2} \sum_{k=1}^{p} (z_k^t - z_k)^2.$$
 (IV.18)

L'algorithme d'apprentissage par rétropropagation du gradient est un processus itératif qui consiste en la modification des poids synaptiques dans la direction opposée au gradient de l'erreur  $\overrightarrow{\nabla} E$ . Si l'on note **W** le vecteur des pondérations du réseau et r le niveau d'itération de l'algorithme d'apprentissage, on peut écrire :

$$\mathbf{W}(r+1) = \mathbf{W}(r) - \epsilon \overrightarrow{\nabla} E, \qquad \epsilon > 0.$$
 (IV.19)

Le paramètre  $\epsilon$  est fixe et représente un taux de correction. Il est en général inférieur à 1. Comme indiqué précédemment, les inconnues du réseau sont les poids synaptiques. Le gradient de l'erreur quadratique correspond donc au vecteur des dérivées partielles de l'erreur par rapport à ces poids et se note :

$$\vec{\nabla}E = \frac{\partial E}{\partial \mathbf{W}} = \left\{ \begin{array}{ll} \frac{\partial E}{\partial w_{ij}^c}, & 0 \le i \le m, \quad 1 \le j \le n\\ \frac{\partial E}{\partial w_{jk}^s}, & 0 \le j \le n, \quad 1 \le k \le p \end{array} \right\},$$
(IV.20)

en conservant les notations du réseau de neurones de la Figure IV.9. Le processus d'apprentissage nécessite donc le calcul de chacune de ces dérivées partielles. En utilisant la propriété de la dérivée de la fonction sigmoïde, à savoir f'(v) = f(v)(1 - f(v)), on obtient :

$$\forall j \in [0,n], \forall k \in [1,p], \qquad \frac{\partial E}{\partial w_{jk}^s} = -y_j z_k (z_k^t - z_k)(1 - z_k), \qquad (\text{IV.21})$$

 $\operatorname{et}$ 

$$\forall i \in [0, m], \, \forall j \in [1, n], \qquad \frac{\partial E}{\partial w_{ij}^c} = -x_i^t y_j (1 - y_j) \sum_{k=1}^p w_{jk}^s z_k (1 - z_k) (z_k^t - z_k), \quad (\text{IV.22})$$

comme indiqué dans (Haykin, 1998, chap. 4, sec. 3), mais avec la pente a de la fonction sigmoïde égale à 1 dans notre cas.

# 2.2.2 Algorithme d'apprentissage

Pour commencer l'algorithme d'apprentissage, on initialise aléatoirement les poids synaptiques et on choisit  $\epsilon$ . À chaque itération r de l'apprentissage, on calcule les dérivées partielles de l'erreur quadratique, puis on modifie les poids synaptiques de la couche de sortie,  $\forall j \in [0, n]$  et  $\forall k \in [1, p]$ ,

$$w_{jk}^{s}(r+1) = w_{jk}^{s}(r) + \epsilon y_{j} z_{k}(z_{k}^{t} - z_{k})(1 - z_{k}), \qquad (\text{IV.23})$$

et ceux de la couche cachée,  $\forall i \in [0, m]$  et  $\forall j \in [1, n]$ ,

$$w_{ij}^c(r+1) = w_{ij}^c(r) + \epsilon x_i^t y_j (1-y_j) \sum_{k=1}^p w_{jk}^s(r) z_k (1-z_k) (z_k^t - z_k) .$$
(IV.24)

L'opération est répétée tant que l'erreur quadratique reste supérieure à un seuil choisi.

# 2.3 Application à la localisation sonore

# 2.3.1 Initialisation

Pour revenir au cadre de la localisation sonore avec et sans TCAPS, nous utilisons en entrée du réseau l'amplitude en décibels des DTF droite et gauche mesurées sur tête artificielle. Elle est néanmoins normalisée en fixant à 1 dB sa valeur maximum et à 0 dB les valeurs inférieures à -40 dB :

$$x_{i} = \begin{cases} \frac{D(f_{i})+40}{\max(D)+40} & si D(f_{i}) > -40 \text{ dB} \\ 0, & sinon \end{cases}$$
(IV.25)

Cette normalisation est réalisée dans (Neti et al., 1992b) avec un seuil de -30 dB, mais cette étude est conduite sur des HRTF de chats. Nous ne pouvons pas vérifier si cette valeur est applicable à notre cas, c'est pourquoi nous avons arbitrairement choisi le minimum de l'amplitude des DTF mesurées en condition d'oreille libre.

À l'instar du modèle d'appariement, nous souhaitons comparer la sortie du réseau de neurones avec les performances de localisation de l'expérience subjective. Il est donc nécessaire de simuler l'utilisation des sources virtuelles par stéréophonie panoramique. Cette simulation est encore une fois réalisée en prenant la moyenne des deux DTF des positions réelles encadrant la position virtuelle.

Les DTF droite et gauche utilisées en entrée ont, avant la normalisation, été lissées par la banque de 28 filtres gammatones. Le réseau comprend donc 56 entrées. Cependant, comme il a été remarqué dans le développement du modèle d'appariement, le fait de ne pas utiliser les différences interaurales de temps et d'intensité pourrait conduire à l'apparition d'erreurs gauche-droite. Elles ne sont pas voulues, car absentes des performances humaines de localisation sonore. Pour chaque position, l'ITD et l'ITD ont donc été ajoutées au réseau de neurones, portant son nombre d'entrées à 58. Comme les DTF, les différences interaurales sont normalisées, cette fois entre -1 et 1. Pour cela, l'ITD est divisée par 1 ms, et l'ILD par 20 dB (l'ITD et l'ILD ne dépassent jamais ces valeurs d'après nos mesures).

Le nombre de neurones dans la couche cachée a été fixé à 20. Ce chiffre pourra être modifié ultérieurement, mais il a dans un premier temps semblé présenter un bon compromis entre temps de calcul et précision. Le réseau est totalement connecté donc chaque neurone de la couche cachée possède 59 poids synaptiques (58 entrées + un biais). Pour terminer, les sorties du réseau sont au nombre de 16, une pour chaque position de mesure des DTF. Chaque neurone de la couche de sortie comprend 21 poids synaptiques. Au total, le nombre de coefficients inconnus dans notre réseau de neurones s'élève à 1180 pour la couche cachée et 336 pour la couche de sortie. En préparation de l'apprentissage du réseau de neurones, chacun de ces coefficients sont initialisés aléatoirement entre -0.5 et 0.5.

# 2.3.2 Entrées et sorties théoriques

L'apprentissage du réseau de neurones est réalisé suivant la méthode décrite dans la section précédente. Les DTF et les différences interaurales de temps et d'intensité calculées en condition d'oreille libre sont définies comme les entrées théoriques du réseau. Pour chacune d'entre elles, la première intuition pour la sortie théorique était de fixer à 1 la sortie correspondant à la position réelle de la source et à 0 les 15 autres sorties. Mais il est finalement préférable d'exploiter les performances de localisation de l'expérience subjective en condition d'écoute normale.

À chaque position de source  $\theta_i$  est ainsi attribuée une distribution de von Mises de paramètres  $\mu_i$  et  $\kappa$ , dont la densité de probabilité s'écrit en fonction de la position de sortie du réseau  $\theta_k$ :

$$f(\theta_k) = \frac{e^{\kappa \cos(\theta_k - \mu_i)}}{2\pi I_0(\kappa)}, \qquad (\text{IV.26})$$

avec  $I_0$  la fonction de Bessel modifiée d'ordre 0. Le paramètre  $\mu_i$  est supposé égal à la position de la source  $\theta_i$ , puisqu'il est équivalent à l'espérance d'une loi normale. Le paramètre de concentration  $\kappa$  a une fonction inverse à celle de la variance d'une loi normale. Nous l'avons estimé pour chaque position de source à partir des observations de l'expérience subjective et en utilisant la méthode décrite dans (Sra, 2012) et développée dans l'Annexe C. Nous ne conservons finalement que la moyenne de ces estimations, égale à 13,2. Pour donner un ordre d'idée, cette valeur est équivalente à un écart-type de 15,8° pour une loi normale classique.

Pour conclure, la sortie théorique du réseau est la densité de probabilité précédente, normalisée afin qu'elle soit égale à 1 quand  $\theta_k$  est égal à  $\theta_i$ . Quand l'entrée théorique du réseau représente les indices de localisation d'une source à la position  $\theta_i$ , la sortie s'écrit :

$$\forall k \in \{1, \dots, p\} \qquad z_k^t = e^{\kappa \cos(\theta_k - \theta_i) - \kappa}, \qquad (\text{IV.27})$$

avec  $\kappa = 13,2.$ 

# 2.3.3 Apprentissage

Le seuil de l'erreur quadratique doit être choisi pour permettre d'obtenir des résultats relativement précis sans pour autant provoquer une surspécialisation du réseau. Le réseau ne doit pas seulement être capable de prédire des performances de localisation à partir des entrées théoriques, mais doit aussi pouvoir généraliser ses mécanismes à de nouvelles entrées. Nous avons alors trouvé que fixer le seuil de l'erreur quadratique à 0,1 représentait un bon compromis. De même, le taux de correction  $\epsilon$  représente le degré de correction que l'on apporte aux poids synaptiques à chaque itération. S'il est trop petit, la correction apportée à chaque itération sera faible, donc l'apprentissage sera trop long. En revanche, si  $\epsilon$  est trop grand, nous risquons d'empêcher l'erreur quadratique d'atteindre le minimum recherché. Encore une fois par rapport au temps de calcul engendré, nous avons fixé  $\epsilon$  à la valeur de 0,2.

Chaque itération de l'algorithme d'apprentissage consiste en la correction de tous les coefficients du réseau selon les Équations IV.23 et IV.24, à partir des entrées et sorties théoriques à chaque position  $\theta_i$ . L'ordre dans lequel les données de ces 16 positions sont utilisées est modifié aléatoirement à chaque itération. La fonction-coût est recalculée uniquement quand les corrections ont été appliquées pour l'ensemble des positions  $\theta_i$ . Avec le taux de correction et les entrées et sorties théoriques que nous avons définis, l'optimisation des poids synaptiques du réseau de neurones, telle que l'erreur quadratique soit inférieure au seuil voulu, est réalisée en environ 2300 itérations par un programme MATLAB<sup>®</sup>.

Nous avons vu dans la première étape de l'apprentissage que les coefficients du réseau sont initialisés aléatoirement. Ceci entraîne forcément la production de réseaux de neurones uniques. Pour pouvoir étudier les différences générées entre plusieurs réseaux à partir des mêmes entrées et sorties théoriques, l'apprentissage a été réalisé 200 fois, chacune avec une initialisation aléatoire. Les éventuelles différences de performance de localisation dues aux différences de réseaux seront représentées sous forme de barres d'erreur.

# 2.3.4 Utilisation du réseau

Après l'étape d'apprentissage, nous disposons de 200 réseaux de neurones différents. Le principe de notre étude repose ensuite sur l'évaluation de leur sortie lorsqu'on leur applique des indices de localisation mesurés avec des TCAPS en entrée. En général, nous étudierons la sortie moyennée sur les 200 réseaux. La légitimité de cette utilisation de l'ensemble des réseaux est montrée en étudiant succinctement leur sortie en condition d'oreille libre.

Les réseaux ont été séparés en deux groupes de 100. La sortie de chaque réseau en condition d'écoute normale est représentée par une matrice de dimensions  $16 \times 16$  dont les colonnes correspondent aux positions réelles de la source sonore et les lignes aux 16 positions testées. Les matrices de probabilité sont moyennées dans les deux groupes de réseaux. La différence maximale de probabilité observée entre les deux groupes est seulement de 3‰ et l'erreur quadratique moyenne est quant à elle inférieure à 1‰.

La sortie du réseau de neurones représente des probabilités de localisation et est a fortiori comprise entre 1 et 0. Dans la suite, nous ne nous intéresserons qu'à des différences de probabilité supérieures à 1%. Nous pouvons conclure que les différences entre réseaux, engendrées par l'initialisation aléatoire des poids synaptiques, sont négligeables et que l'exploitation de la moyenne de leur sortie est représentative de l'ensemble des réseaux.

# 2.4 Conclusion sur le réseau de neurones

L'implémentation d'un réseau de neurones à trois couches a été décrite dans cette partie. Il réalise une prédiction de la position d'une source à partir des indices de localisation sonore utilisés par le système auditif. Au total, 200 réseaux de neurones différents ont été entraînés avec les ITD, les ILD et les indices spectraux mesurés en condition d'oreille libre dans le but de reproduire les performances globales obtenues au cours de l'expérience subjective dans cette condition. Dans la suite, les indices de localisation dégradés par les TCAPS sont appliqués aux réseaux, et leur sortie est normalisée pour représenter une matrice de probabilité de localisation. Elle est alors comparée aux performances des auditeurs de l'expérience subjective, ainsi qu'à la sortie du modèle d'appariement développé précédemment.

# 3 Performance des modèles

Pour chaque position cible (les 16 positions de mesure de DTF), les modèles précédemment définis renvoient une probabilité de localisation pour chacune des 16 positions tests. Dans cette partie, nous étudions la capacité des modèles à prédire des performances de localisation sonore proches de celles que nous avons observées au cours de l'expérience subjective. De manière analogue à la phase d'optimisation du modèle d'appariement, nous allons utiliser les résultats de l'expérience subjective sous forme de matrices de confusion normalisées pour obtenir des matrices de probabilité de localisation. Dans un premier temps, les matrices de dimensions ( $16 \times 16$ ) calculées pour chaque modèle et pour l'expérience subjective, dans chacune des sept conditions d'écoute, sont décrites.

Dans un second temps, les définitions de localisation exacte, de localisation voisine et des différents types d'erreurs données dans la Section 1.4 du Chapitre III sont à nouveau utilisées. Cependant, dans le chapitre en question, nous avons étudié les *nombres* de localisations et d'erreurs. À présent, nous ne nous intéressons qu'à des probabilités et de nouvelles métriques doivent donc être introduites.

D'après les définitions, on remarque que pour une position de source donnée, la réponse ou la prédiction est forcément soit une localisation voisine, soit une erreur grossière, avant-arrière, gauche-droite ou mixte. Pour chaque position cible (c'est-à-dire dans chaque colonne des matrices de probabilité), les nombres de localisations voisines et des différents types d'erreurs sont donc divisés par leur somme. On obtient alors des probabilités de localisation et d'erreur. Le nombre de localisations exactes est quant à lui divisé par le nombre de localisations voisines, créant alors une mesure de la précision de localisation.

Au total, pour chaque modèle et l'expérience subjective et pour chaque condition d'écoute, nous présenterons cinq métriques différentes :

- la probabilité de localisation voisine  $P_{loc}$ ;
- la probabilité d'erreur avant-arrière  $P_{fb}$ ;
- la probabilité d'erreur grossière  $P_{sev}$ ;
- la probabilité d'erreur gauche-droite ou mixte  $P_{lr}$ ;
- la précision de localisation  $P_{acc}$ .

Pour finir, une mesure globale de comparaison des prédictions faites par les modèles et des résultats de l'expérience subjective est présentée. Elle est basée sur un calcul de vraisemblance.

# 3.1 Étude des matrices de probabilité

Les matrices de probabilité de localisation, calculées pour l'expérience subjective, le modèle d'appariement et le réseau de neurones, sont montrées respectivement sur les Figures IV.10, IV.11 et IV.12 dans toutes les conditions d'écoute. Les probabilités sont codées sur une échelle de couleur allant du noir (0) au blanc (1) et représentées pour chaque couple de positions source (abscisse) et test (ordonnée). Les résultats en condition d'oreille libre sont tout d'abord étudiés.



FIG. IV.10 – Probabilités de localisation de l'expérience subjective pour chaque couple de positions cible et test, dans les sept conditions d'écoute.



FIG. IV.11 – Sorties du modèle d'appariement pour chaque couple de positions cible et test, dans les sept conditions d'écoute.



FIG. IV.12 – Sorties du réseau de neurones pour chaque couple de positions cible et test, dans les sept conditions d'écoute.

# 3.1.1 En condition d'oreille libre

Nous nous intéressons aux Figures IV.10a, IV.11a et IV.12a. Une localisation exacte correspond à la diagonale allant de la gauche en haut à la droite en bas. Pour l'expérience subjective, on retrouve les résultats décrits dans le Chapitre III : la probabilité de localisation voisine est globalement élevée et la précision est plus faible pour les positions proches de l'axe interaural. De plus, la probabilité de localisation exacte est en général plus importante pour les positions correspondant à des sources réelles que pour celles des sources virtuelles.

Le réseau de neurones (Figure IV.12a) conduit sans surprise à une localisation presque parfaite de la source, puisqu'il a été entraîné à le faire en condition d'écoute normale. La largeur de la diagonale dépend uniquement du paramètre de concentration  $\kappa$  de la sortie théorique utilisée lors de la phase d'apprentissage. Le principal défaut de cette modélisation semble être le fait de ne pas prendre en compte la variation de précision en fonction de l'azimut de la source.

D'un autre côté, des confusions avant-arrière persistent pour le modèle d'appariement sur la Figure IV.11a. Elles sont illustrées par des probabilités non nulles dans les deux sousdiagonales. Cependant, une probabilité plus importante est accordée à la position réelle de la source, démontrant la résolution des ambiguïtés avant-arrière par les indices spectraux. À l'inverse du réseau de neurones, la perte de précision autour de l'axe interaural, observée dans l'expérience subjective, est modélisée par le modèle d'appariement. On retrouve aussi approximativement les différences de probabilité de localisation exacte entre sources réelles et virtuelles.

Si les tendances de localisation prédites par le modèle d'appariement et par le réseau de neurones sont finalement plutôt correctes, les divergences qu'ils montrent avec les résultats de l'expérience subjective ne sont pas négligeables. Avec le modèle d'appariement, les principales erreurs de prédiction se situent sur la diagonale de localisation voisine et sur les sous-diagonales représentant les erreurs avant-arrière. Elles atteignent des proportions particulièrement grandes, jusqu'à 0,34, lorsque la source est juste devant l'axe interaural (soit aux positions d'azimut  $67,5^{\circ}$  et  $292,5^{\circ}$ ). Avec le réseau de neurones, les erreurs de prédiction sont réparties essentiellement sur la diagonale de localisation exacte. L'erreur maximale est atteinte juste devant l'auditeur et vaut environ 0,37. Ceci s'explique par le fait que la précision prédite par le réseau de neurones en condition d'oreille libre ne dépend pas de la position de la source, alors que l'expérience subjective a montré que la précision était maximale pour une source à  $0^{\circ}$ .

Les paramètres du modèle d'appariement ont été optimisés par rapport aux résultats de l'expérience subjective en condition d'écoute normale et l'apprentissage des réseaux de neurones s'est fait à partir de ces mêmes résultats. L'application, aux deux modèles, des indices de localisation mesurés avec les TCAPS devrait donc encore accentuer les différences avec les résultats de l'expérience subjective.

# 3.1.2 Avec les TCAPS

Nous avons vu dans les chapitres précédents comment les systèmes tactiques de communication et protection dégradent les indices de la localisation et comment cela se traduit au cours d'une expérience subjective de localisation. Les conséquences de ces dégradations en termes de probabilité de localisation pendant l'expérience subjective sont dépeintes sur les Figures IV.10b à IV.10g. Les conclusions que l'on peut en tirer sont bien sûr les mêmes que dans l'analyse de l'expérience subjective (cf. Chapitre III). Une augmentation substantielle de la probabilité d'erreur avant-arrière est observée. Elle dépend fortement du type de TCAPS utilisé : les bouchons non linéaires P1 et P2 ont un impact moins important que les bouchons talk-through P3 et P4, et les casques P5 et P6 entraînent la plus forte probabilité d'erreur avant-arrière. On constate de plus un étalement de la diagonale principale, caractérisant une perte en précision de localisation.

Les probabilités prédites par le modèle d'appariement avec les TCAPS sont représentées sur les Figures IV.11b à IV.11g. Nous avons vu qu'en condition d'oreille libre, ce modèle reflétait bien l'impact des indices spectraux sur la discrimination des sources avant et arrière. Avec les quatre bouchons d'oreille, la détérioration des indices spectraux entraîne une réduction de la différence entre la probabilité de localisation voisine (diagonale principale) et la probabilité d'erreur avant-arrière (sous-diagonales). Ceci résulte bien d'une incapacité des indices spectraux à permettre la différenciation avant-arrière. De manière plus détaillée, le bouchon non linéaire P1 conserve une probabilité de localisation exacte supérieure à 0,5 pour des sources arrière. Pour le bouchon non linéaire P2 et le bouchon talk-through P4 (cf. Figures IV.11c et IV.11e, respectivement), le modèle d'appariement prédit une conservation des performances de localisation pour les sources réelles, qui n'est que faiblement observée avec les matrices subjectives de P1 et P2. Cette conservation existe uniquement pour les sources réelles arrière avec le bouchon talk-through P3. Pour finir, quel que soit le bouchon, le modèle ne présente pas d'étalement de la diagonale principale comparable à l'expérience subjective. Il demeure donc globalement trop précis.

Les Figures IV.11f et IV.11g montrent que la probabilité de localisation n'est préservée que sur l'axe interaural avec les casques talk-through. Pour les autres positions la probabilité de localisation est équivalente à la probabilité d'erreur avant-arrière, voire inférieure derrière l'auditeur pour P5 et aux positions proches de l'axe interaural pour P6. Comme avec les bouchons, le modèle d'appariement prédit bien l'augmentation substantielle de la confusion avant-arrière avec les casques talk-through, mais pas la réduction de précision observée dans l'expérience subjective.

Selon la Figure IV.12c, le réseau de neurones prédit une probabilité de localisation très peu détériorée par le bouchon P2. Pour la majorité des positions de source, la probabilité de localisation exacte est même supérieure à celle de la condition d'oreille libre. Si des probabilités d'erreur avant-arrière notables sont tout de même introduites pour P1 (cf. Figure IV.12b), la même remarque peut être faite. On retiendra que le réseau de neurones,

comme le modèle d'appariement, a tendance à rester trop précis avec les bouchons non linéaires. L'étalement de la diagonale principale est mieux modélisé pour les bouchons talkthrough sur les Figures IV.12d et IV.12e, de même que l'augmentation de la probabilité de localisation dans les deux sous-diagonales. Néanmoins, avec P4, des valeurs importantes de probabilité persistent dans la diagonale principale à l'avant, mais aussi à l'arrière (à l'exception de la position d'azimut 180°).

Les prédictions des performances de localisation sonore avec les casques talk-through réalisée par le réseau de neurones sont présentées sur les Figures IV.12f et IV.12g. Pour l'expérience subjective, la diagonale principale et les deux sous-diagonales sont toujours visibles, même pour les casques talk-through. Ce n'est plus le cas avec le réseau de neurones. À l'exception de celles se trouvant sur l'axe interaural, le réseau prédit systématiquement que les sources sonores sont soit juste devant l'auditeur, soit juste derrière. Généralement, une plus grande probabilité est donnée aux positions avant.

Pour conclure sur l'observation des différentes densités de probabilité, nous retiendrons que les deux modèles de localisation sonore prédisent une trop grande précision. Toutefois, il semble que la dégradation des performances de localisation sonore engendrée par les TCAPS soit relativement bien modélisée, en particulier l'augmentation des erreurs avantarrière. Enfin, les prédictions de probabilité de localisation obtenues avec le réseau de neurones apparaissent plus proches des observations de l'expérience subjective que les prédictions du modèle d'appariement, sauf dans le cas des casques talk-through.

# 3.2 Analyse des métriques

Les probabilités de localisation voisine, d'erreur avant-arrière, d'erreur grossière d'erreur gauche-droite ou mixte, et la précision de localisation sont présentées dans cette section. Nous les étudions dans chaque condition d'écoute et dans le cadre de l'expérience subjective, du modèle d'appariement et du réseau de neurones. Ces différentes métriques sont présentées en Figure IV.13. Les valeurs données pour l'expérience subjective sont des moyennes calculées sur tous les sujets et toutes les sessions, et les barres d'erreur représentent l'écart-type des différents échantillons. De même, pour le réseau de neurones, les valeurs moyennes et les écarts-types sont calculés à partir des 200 réseaux générés.

# 3.2.1 Probabilité de localisation voisine

La Figure IV.13a présente la probabilité de localisation voisine, soit la probabilité de localisation à plus ou moins 22,5°, pour toutes les conditions d'écoute. On peut observer que le modèle d'appariement sous-estime systématiquement la probabilité de localisation voisine. La différence par rapport aux probabilités subjectives est de 0,15 environ, sauf pour P4 et P6, pour lesquels la différence est respectivement de 0,04 et 0,05. À l'inverse, les réseaux de neurones prédisent des probabilités de localisation voisine analogues aux performances humaines, en condition d'écoute normale et avec les bouchons d'oreille. Les



FIG. IV.13 – Probabilités de localisation voisine, d'erreurs avant-arrière, grossières, gauche-droite, mixte et précision de localisation estimées pour l'expérience subjective, le modèle d'appariement et le réseau de neurones, dans chaque condition d'écoute. Les données pour le réseau de neurones sont volontairement décalées vers la droite pour plus de visibilité.

différences entre les valeurs correspondantes et celles de l'expérience subjective sont effectivement inférieures aux écarts-types calculés sur l'ensemble des auditeurs. De plus, le réseau de neurones prédit de meilleures performances avec P2 qu'avec les autres bouchons. Ses prédictions concernant les casques sont en revanche plus mauvaises que celles du premier modèle (les différences avec les valeurs subjectives atteignent 0,3). Finalement, à l'exception de la hausse prédite pour P2 par le réseau de neurones, les deux modèles prédisent des probabilités de localisation voisine presque constantes pour les bouchons d'oreille et une dégradation substantielle avec les casques. Cette tendance est semblable aux résultats du test de Dunn pour l'expérience subjective, dans le Chapitre III.

# 3.2.2 Probabilité d'erreur avant-arrière

La probabilité de localisation voisine erronée prédite par le modèle d'appariement se reflète dans la probabilité d'erreur avant-arrière présentée en Figure IV.13b. Dans la partie précédente, nous avons vu que ce modèle prédit des erreurs avant-arrière non négligeables, même en condition d'oreille libre. Cette observation est confirmée par cette nouvelle figure montrant que le modèle d'appariement surestime la probabilité d'erreur avant-arrière dans toutes les conditions d'écoute. D'un autre côté, le réseau de neurones conduit encore une fois à des prédictions plus réalistes pour la condition d'écoute normale et les bouchons, mais pas pour les casques. Une différence notable peut encore être observée entre P1 et P2 (environ 0,16).

# 3.2.3 Probabilité d'erreur grossière

Les erreurs grossières ont été identifiées comme la seconde cause de dégradation des performances de localisation au cours de l'expérience subjective. Comme indiqué sur la Figure IV.13c, le modèle d'appariement fournit des probabilités d'erreur grossière proche de 0 pour la condition d'oreille libre et les bouchons. Même si ces valeurs appartiennent aux déviations de l'expérience subjective, les probabilités d'erreur grossière données par le réseau de neurones sont plus semblables aux valeurs subjectives. Pour les casques talkthrough, les deux modèles prédisent une augmentation des erreurs grossières aussi observée dans l'expérience subjective, bien que les valeurs du réseau de neurones soient surestimées.

# 3.2.4 Probabilité d'erreur gauche-droite ou mixte

Les probabilités d'erreurs gauche-droite et mixte sont présentées dans chaque condition d'écoute sur la Figure IV.13d. On observe des nombres insignifiants de ce type d'erreurs dans l'expérience subjective et ceci est plutôt bien modélisé par le modèle d'appariement. En effet, la différence maximale avec les moyennes subjectives n'est que de 0,03 et est obtenue avec P5. Les probabilités d'erreurs gauche-droite et mixte prédites par le réseau de neurones sont à l'inverse surestimées, spécialement avec les casques P5 et P6.

# 3.2.5 Précision de localisation

Concernant la précision de localisation sonore, le modèle d'appariement la surestime en condition d'écoute normale et avec les bouchons alors qu'elle est correcte avec les casques. Cela nous amène à penser que si ce modèle prédisait correctement la précision avec les bouchons d'oreille, il la sous-estimerait avec les casques. Le réseau de neurones prédit correctement la précision en condition d'oreille libre et avec les casques, mais la surestime avec les bouchons. Les valeurs obtenues sont quand même à la limite des écarts-types de l'expérience subjective. Cette dernière montre que la précision de localisation est plus grande avec les bouchons non linéaires qu'avec les autres TCAPS. Les modèles prédisent le même genre de tendance, sauf qu'un pic de précision est introduit par le réseau de neurones avec le bouchon talk-through P4.

# 3.3 Vraisemblance des modèles

Dans le but de comparer les sorties des deux modèles de localisation à l'expérience subjective, la concordance entre ces sorties et les performances humaines est quantifiée en calculant la log-vraisemblance des observations subjectives avec les deux modèles. La logvraisemblance  $\mathcal{L}$  est estimée pour chaque condition d'écoute selon l'Équation IV.28, avec Nle nombre total d'observations (7200 en condition d'oreille libre, 4800 avec les TCAPS) et  $P_i$  la probabilité de la *i*-ème observation donnée par un modèle dans la condition d'écoute correspondante.  $\mathcal{L}$  est divisée par N afin de prendre en compte les nombres d'observations différents entre la condition d'écoute normale et les autres. Ainsi, on a :

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} \log(P_i) \,. \tag{IV.28}$$

Plus la log-vraisemblance est élevée, plus la distribution de probabilités prédites par le modèle correspond à la répartition des réponses des auditeurs à l'expérience de localisation. Pour un modèle « parfait », respectant exactement les résultats de l'expérience subjective,  $\mathcal{L}$  ne tend pas vers 0, mais vers un maximum correspondant au calcul de la vraisemblance à partir des distributions de probabilités subjectives. Dans la suite, ce maximum de vraisemblance est toujours indiqué pour l'étude des performances des deux modèles de localisation.

La Figure IV.14 représente la log-vraisemblance calculée pour le modèle d'appariement, le réseau de neurones et l'expérience subjective dans les sept conditions d'écoute. Elle confirme le fait que le réseau de neurones prédit des performances de localisation sonore plus proches des performances humaines que le modèle d'appariement, et ce dans toutes les conditions d'écoute sauf avec les casques. Cependant, on peut noter que les vraisemblances des deux modèles présentent des tendances très similaires. La vraisemblance maximale est obtenue en condition d'oreille libre pour les deux modèles. Cet effet provient de l'optimisation des paramètres du modèle d'appariement et de l'apprentissage du réseau



FIG. IV.14 – Log-vraisemblance des modèles de localisation calculée à partir des observations de l'expérience subjective, dans chaque condition d'écoute. Les données de l'expérience subjective permettent le calcul de la vraisemblance maximale.

de neurones à partir de cette condition d'écoute. Dans les conditions avec TCAPS, la logvraisemblance du réseau de neurones chute d'approximativement -2 pour les bouchons à -3,1 et -4,1 avec les casques P5 et P6, respectivement. Pour le modèle d'appariement, elle baisse de -2 pour les bouchons non linéaires à -2,7 pour les bouchons talk-through et -3avec les casques talk-through. En bref, plus les performances réelles de localisation sont mauvaises, par exemple avec les casques, plus les modèles ont des difficultés à correctement modéliser ces performances.

# 3.4 Conclusion sur les performances des modèles

Dans cette partie, nous avons montré que les prédictions du réseau de neurones sont plus similaires aux performances obtenues dans l'expérience subjective que les prédictions du modèle d'appariement. Le réseau de neurones exagère néanmoins l'impact des casques talk-through. Nous avons vu qu'il prédit notamment une augmentation substantielle des erreurs gauche-droite et mixte, alors que ces types d'erreurs sont jugées négligeables dans l'expérience subjective. Le modèle d'appariement est moins réaliste, mais prédit tout de même des tendances correctes en fonction de la condition d'écoute. Toutefois, avec le bouchon talk-through P3, il sous-estime les performances de manière conséquente.

Dans la suite de ce chapitre, des tentatives d'amélioration des modèles de localisation sonore sont décrites et évaluées. Le modèle d'appariement est revu pour mieux simuler la séparation entre la perception de l'angle latéral par les différences interaurales de temps et d'intensité et la discrimination avant-arrière par les indices spectraux. Pour le réseau de neurones, la sortie théorique utilisée dans la phase d'apprentissage est modifiée pour prendre en compte la variation de précision en fonction de la position de la source sonore.

# 4 Correction du modèle d'appariement

Les divergences constatées précédemment entre les performances subjectives de localisation et les prédictions du modèle d'appariement sont certainement reliées à l'optimisation des paramètres du modèle. Ils sont effectivement indépendants de la position réelle de la source, alors que la précision des auditeurs est plus élevée pour une source frontale que pour une source latérale. Dans (Baumgartner et al., 2014), une cartographie sensori-motrice est implémentée pour simuler cette dépendance dans la dimension polaire. Dans le contexte de notre étude, il peut être résolu par la correction des paramètres du modèle d'appariement. Les écarts-types  $\sigma_{\delta}$ ,  $\sigma_{ILD}$  et  $\sigma_{ILD}$  relatifs aux indices spectraux, aux ITD et aux ITD doivent être optimisés de façon à obtenir une variation graduelle de sensibilité en fonction de la position de la source sonore.

Dans cette partie, nous allons donc modifier l'optimisation des écarts-types relatifs aux différences interaurales de temps et d'intensité et aux indices spectraux. Mais tout d'abord, le principe de fonctionnement du modèle d'appariement est revu.

# 4.1 Révision du principe du modèle

Dans la Section 1.4, la procédure d'optimisation du modèle d'appariement est décrite. Afin de séparer les calculs des paramètres liés aux ITD et aux ILD des paramètres relatifs aux indices spectraux, cette procédure est divisée en deux étapes : la première consiste à projeter la sortie du modèle dans la dimension latérale pour obtenir  $\sigma_{ITD}$ ,  $\sigma_{ILD}$ ,  $w_{ITD}$  et  $w_{ILD}$ . À la suite de cette étape, l'optimisation de  $\sigma_{\delta}$  et  $w_{\delta}$  est censée simuler la résolution des ambiguïtés avant-arrière par les indices spectraux.

La projection dans la dimension latérale et l'hypothèse sur le rôle des indices spectraux peuvent être insérées directement dans la définition du modèle d'appariement. Cette modification repose sur la décomposition d'une distribution de probabilités dans la dimension azimutale en un produit d'une composante latérale et d'une composante polaire, décrite en Annexe D. Pour une position de source  $\theta_i$ , si  $P(\theta_i, \theta_j)$  est une distribution de probabilités dans la dimension  $\theta_j$ , on peut écrire :

$$P(\theta_i, \theta_j) = P(\theta_i, \alpha, \beta) = \begin{cases} P(\theta_i, \alpha) P(\theta_i, \beta), & \text{si } \alpha \neq \pm 90^\circ, \\ P(\theta_i, \alpha), & \text{si } \alpha = \pm 90^\circ, \end{cases}$$
(IV.29)

avec  $P(\theta_i, \alpha)$  la composante latérale,  $P(\theta_i, \beta)$  la composante polaire,  $\alpha$  l'angle latéral et  $\beta$ l'angle polaire (cf. Annexe A). Dans notre cas, l'angle latéral  $\alpha$  est compris entre  $-90^{\circ}$ et 90°, avec un pas de 22,5°, et l'angle polaire  $\beta$  est égal à 0° (hémisphère avant) ou 180° (hémisphère arrière). Une telle décomposition suppose que la perception de l'angle polaire

Condition d'écoute	Sans	P1	P2	P3	P4	P5	P6
Erreur de reconstruction ( $\times 10^{-2}$ )	$0,\!14$	$0,\!5$	0,48	0,38	0,7	$0,\!63$	$0,\!67$

TAB. IV.1 – Erreur de décomposition/reconstruction des probabilités de l'expérience subjective dans chaque condition d'écoute.

grâce aux indices spectraux est indépendante de la perception de l'angle latéral par les ITD et ILD.

D'après Carlile et al. (2005), l'hypothèse d'indépendance est bien supportée par la littérature. Notamment, Morimoto et al. (2003) ont montré que les indices spectraux utilisés pour déterminer l'angle polaire sont indépendants de l'angle latéral. Toutefois, il a aussi été observé qu'une bonne localisation dans cette dimension nécessite la connaissance de l'angle latéral, par exemple à l'aide des ITD (Martin et al., 2004). Avant de détailler les modifications du modèle d'appariement, il est donc important de vérifier cette hypothèse.

## 4.1.1 Vérification de l'hypothèse d'indépendance

Les distributions de probabilité de localisation moyennes obtenues à partir des observations de l'expérience subjective sont décomposées et reconstruites selon la méthode de l'Annexe D. Pour chaque condition d'écoute, l'erreur de reconstruction est calculée comme la moyenne de la valeur absolue de la différence entre les probabilités initiale et reconstruite. Les résultats sont regroupés dans le Tableau IV.1. L'erreur de reconstruction moyenne est inférieure à 0,01 dans toutes les conditions d'écoute et est maximale avec P4 (0,007). Les probabilités qui nous intéressent sont en général de l'ordre de 0,1. De plus, l'erreur de reconstruction est encore plus faible en condition d'écoute normale. Nous en concluons que l'hypothèse d'indépendance des composantes latérale et polaire est valide dans le cadre de la localisation sonore sur le plan horizontal.

À partir de cette hypothèse, nous pouvons redéfinir le modèle d'appariement : sa sortie, une distribution de probabilité dans la dimension azimutale, se décompose comme le produit d'une composante latérale et d'une composante polaire. Le schéma de principe de cette modélisation est représenté sur la Figure IV.15.

## 4.1.2 Composante latérale

Cette composante est obtenue uniquement à partir des différences d'ITD et d'ILD entre la source et la base de données. Les probabilités  $P_{ITD}$  et  $P_{ILD}$  sont calculées pour chaque couple  $(\theta_i, \theta_j)$ , en normalisant les résultats des Équations IV.6 et IV.7, respectivement. Elles introduisent les écarts-types  $\sigma_{ITD}$  et  $\sigma_{ILD}$ . À présent, seule leur composante latérale est conservée, donnant  $P_{ITD}(\theta_i, \alpha)$  et  $P_{ILD}(\theta_i, \alpha)$ . Pour l'ITD, le calcul de cette composante se fait comme suit :

$$P_{ITD}(\theta_i, \alpha) = P_{ITD}(\theta_i, \alpha, \beta = 0^\circ) + P_{ITD}(\theta_i, \alpha, \beta = 180^\circ).$$
(IV.30)



Probabilité

FIG. IV.15 – Schéma de principe du modèle d'appariement modifié.
Sur l'axe interaural ( $\alpha = \pm 90^{\circ}$ ), la probabilité est déjà indépendante de l'angle polaire, donc  $P_{ITD}(\theta_i, \alpha) = P_{ITD}(\theta_i, \theta_j)$ . La composante latérale de  $P_{ILD}$  se calcule de manière analogue. Pour finir, une probabilité globale de localisation dans la dimension latérale est calculée comme la moyenne géométrique pondérée suivante :

$$P_{lat}(\theta_i, \alpha) = P_{ITD}(\theta_i, \alpha)^{w_{lat}} P_{ILD}(\theta_i, \alpha)^{1 - w_{lat}}.$$
 (IV.31)

On introduit ici une unique pondération  $w_{lat}$ , comprise entre 0 et 1. Pour finir  $P_{lat}$  est normalisée pour pouvoir être considérée, pour chaque position de source  $\theta_i$ , comme une distribution de probabilités dans la dimension latérale. Le calcul de la composante latérale du modèle d'appariement est résumé dans la partie droite de la Figure IV.15.

## 4.1.3 Composante polaire

Nous avons supposé que, sur le plan horizontal, les indices spectraux servent exclusivement à discriminer les sources avant et arrière. La similarité spectrale  $P_{\delta}(\theta_i, \theta_j)$  est toujours calculée grâce aux Équations IV.3 et IV.5, et transformée en probabilité dans la dimension azimutale. Enfin, comme il est décrit dans la partie gauche de la Figure IV.15, seule sa composante polaire, notée  $P_{pol}(\theta_i, \beta)$  est conservée. On la calcule suivant l'équation :

$$P_{pol}(\theta_i, \beta) = \sum_{-90^\circ < \alpha < 90^\circ} P_{\delta}(\theta_i, \alpha, \beta).$$
(IV.32)

Ce résultat dépend toujours du paramètre  $\sigma_{\delta}$ , mais aucune pondération n'est attribuée aux indices spectraux puisqu'ils apportent l'information polaire indépendamment des ITD et ILD.

Pour finir, la sortie du modèle d'appariement est obtenue par reconstruction d'une distribution dans la dimension azimutale, à partir des composantes latérale  $P_{lat}(\theta_i, \alpha)$  et polaire  $P_{pol}(\theta_i, \beta)$ . De façon analogue à l'Équation IV.29, cette sortie s'obtient en calculant :

$$P(\theta_i, \theta_j) = P(\theta_i, \alpha, \beta) = \begin{cases} P_{lat}(\theta_i, \alpha) P_{pol}(\theta_i, \beta), & \text{si } \alpha \neq \pm 90^\circ, \\ P_{lat}(\theta_i, \alpha), & \text{si } \alpha = \pm 90^\circ, \end{cases}$$
(IV.33)

Les différents paramètres du modèle sont de nouveau optimisés à partir des observations de l'expérience subjective en condition d'oreille libre.

#### 4.1.4 Optimisation

Les composantes latérale et polaire du modèle d'appariement sont respectivement comparées aux composantes latérale et polaire des probabilités moyennes de localisation en condition d'écoute normale résultant de l'expérience subjective. Comme dans la Section 1.4, le critère d'erreur est une valeur RMS de la différence entre le modèle et les observations, normalisée par la probabilité de localiser au hasard.

L'optimisation de la composante latérale permet de fixer les paramètres  $\sigma_{ITD}$ ,  $\sigma_{ILD}$ et  $w_{lat}$ . Le critère d'erreur est minimisé pour un écart-type de l'ITD  $\sigma_{ITD} = 0,103$  ms



FIG. IV.16 – Optimisation des paramètres du modèle d'appariement. Composante latérale : (a)  $\sigma_{ITD}$ , (b)  $w_{lat}$ . Composante polaire : (c)  $\sigma_{\delta}$ .

(cf. Figure IV.16a). Cette valeur est très proche de celle obtenue avec la définition précédente du modèle d'appariement. Des différences sont toutefois constatées pour les autres paramètres : la pondération des différences interaurales  $w_{lat}$  est égale à 1 d'après la Figure IV.16b. L'optimisation de la composante latérale est donc indépendante de la sensibilité à l'ILD. Si l'optimisation est réalisée uniquement pour l'ILD (en imposant  $w_{lat} = 0$ ), on obtient tout de même  $\sigma_{ILD} = 2$  dB, qui est compris entre les valeurs minimale et maximale de la JND de l'ILD selon Francart & Wouters (2007). Mais dans ce cas, le minimum de l'erreur, normalisée par rapport au hasard, est supérieur à 1.

Cette nouvelle version du modèle d'appariement n'a « pas besoin » de l'ILD pour déterminer l'angle latéral de la source sonore. On a déjà vu que l'ITD est prépondérante dans le cas de sources à large bande fréquentielle, mais les données de Macpherson & Middlebrooks (2002) nous permettaient de prévoir une pondération d'environ 0,61. Si cette valeur avait été fixée avant l'optimisation, alors le modèle aurait cherché à minimiser le rôle de l'ILD en augmentant considérablement  $\sigma_{ILD}$ . En effet, d'après nos tests dans ce cas, le minimum de l'erreur n'est pas atteint pour  $\sigma_{ILD} = 100$  dB.

La comparaison des composantes polaires de la sortie du modèle et des observations subjectives permet de fixer la sensibilité des indices spectraux  $\sigma_{\delta} = 0.14$  dB (cf. Figure IV.16c). Cette valeur est très proche de l'écart-type de la gaussienne équivalente à la sigmoïde de Baumgartner et al. (2014) estimée en Section 1.4.

Les performances du modèle d'appariement défini dans cette section sont évaluées et comparées à celles du modèle initial dans une prochaine section. Dans la suite, l'optimisation des écarts-types relatifs aux ITD, aux ILD et aux indices spectraux, en fonction de la position de la source, est étudiée.

## 4.2 Dépendance en azimut des paramètres

# 4.2.1 Écarts-types relatifs aux ITD et ILD

Les différences juste perceptibles (JND) des différences interaurales de temps et d'intensité varient en fonction de l'azimut de la source. La définition d'une valeur fixe pour



FIG. IV.17 – (a) Optimisation de la pondération  $w_{lat}$ . (b) Écart-type de l'ITD  $\sigma_{ITD}$  optimisé en fonction de l'azimut de la source.

les écarts-types  $\sigma_{ITD}$  et  $\sigma_{ILD}$  (cf. Équations IV.6 et IV.7) ne permet pas d'exploiter cette propriété.

L'optimisation de la composante latérale du modèle d'appariement minimise un critère d'erreur global entre cette composante et celle obtenue dans l'expérience subjective. Le critère d'erreur peut être calculé séparément pour chaque position de source, et les écartstypes relatifs aux ITD et aux ILD optimisés en conséquence. Cette méthode nécessite beaucoup plus de temps de calcul puisque, dans notre cas, il faut trouver 16  $\sigma_{ITD}$ , 16  $\sigma_{ILD}$  et une pondération  $w_{lat}$ . Les paramètres liés à l'ITD sont donc calculés séparément de ceux liés à l'ILD, puis la pondération entre les deux indices est déterminée. Pour chaque paramètre à optimiser, la gamme d'étude est la même qu'en Section 1.4.

Encore une fois, le critère d'erreur est minimisé lorsque la pondération  $w_{lat}$  est fixée à 1 (cf. Figure IV.17a). Le modèle d'appariement ne prend que l'ITD en compte, et seul cet indice est donc étudié dans la suite. La Figure IV.17b représente l'écart-type de l'ITD,  $\sigma_{ITD}$ , en fonction de l'azimut. Il est compris entre 70 µs et 150 µs, et reste donc entre les extremums de JND trouvés dans (Busson, 2006, chap. 2, sec. 1). L'écart-type est minimal sur l'axe interaural et est globalement plus élevé à l'arrière. Les variations en « dents de scie » montrent une différence de sensibilité à l'ITD entre les sources réelles et virtuelles. Elles devraient donc permettre au modèle de prédire les différences de performances observées entre les types de source.

Nous nous attendions à obtenir un profil d'écart-type correspondant à l'évolution des JND de l'ITD avec l'azimut : minimal à 0° et 180°, et en augmentation quand la source se rapproche de l'axe interaural. Dans l'hémisphère avant,  $\sigma_{ITD}$  est bien plus élevé aux positions voisines de l'axe interaural qu'au niveau du plan médian. Dans l'hémisphère arrière, en revanche, cette tendance est difficilement observable sur la Figure IV.17b, probablement en raison de l'éffet des sources virtuelles et de la faible résolution spatiale. Pour finir, la minimisation de l'écart-type sur l'axe interaural traduit une bonne aptitude à déterminer l'angle latéral à ces positions dans l'expérience subjective (voir la Figure IV.10a).



FIG. IV.18 – Écart-type relatif aux indices spectraux  $\sigma_{\delta}$  optimisé en fonction de l'azimut de la source.

## 4.2.2 Écart-type relatif aux indices spectraux

À l'instar des écarts-types relatifs aux différences interaurales de temps et d'intensité,  $\sigma_{\delta}$  est aussi optimisé en fonction de l'azimut de la source. La minimisation de l'erreur entre la composante polaire du modèle et celle obtenue dans l'expérience subjective est donc réalisée pour chaque position de source  $\theta_i$ . La gamme d'étude du paramètre s'étend toujours de 0,01 à 20 dB, avec un pas de 0,01 dB.

La Figure IV.18 représente l'écart-type relatif aux indices spectraux optimisé en fonction de l'azimut de la source. Il est compris entre 0,5 et 0,25 dB, et est donc réparti autour de la valeur d'optimisation unique de 0,14 dB trouvée précédemment.  $\sigma_{\delta}$  est le plus faible pour les positions proches de l'axe interaural, c'est-à-dire les positions pour lesquelles des erreurs avant-arrière subsistent même en condition d'écoute normale (cf. Chapitre III, Section 2.2). Ces erreurs proviennent donc du fait que, par rapport aux autres positions, une faible différence spectrale entraînera tout de même une erreur de localisation.

Lorsque la source sonore est plus proche du plan médian, l'écart-type est plus élevé, aux environs de 0,2 dB. Comme dans le cas de l'écart-type de l'ITD, on observe une nette différence entre les azimuts correspondant à des sources réelles et ceux des sources virtuelles. Pour ces dernières,  $\sigma_{\delta}$  est systématiquement plus faible. La composante polaire du modèle d'appariement est donc plus sensible aux différences spectrales pour les sources virtuelles.

Dans la section suivante, les différentes modifications du modèle d'appariement sont évaluées.

# 4.3 Étude des corrections

Pour évaluer les modifications apportées au modèle d'appariement, nous comparons dans un premier temps les composantes latérale et polaire de sa sortie avec celles obtenues à partir de l'expérience subjective en condition d'écoute normale. Ensuite, les nouvelles prédictions des performances de localisation sonore avec les TCAPS sont étudiées. Dans ces deux sections, on se réfère au modèle d'appariement décrit au début de ce chapitre en tant que « modèle original ». Le modèle utilisant la décomposition latérale/polaire est appelé « modèle revu » lorsque ses paramètres sont indépendants de  $\theta_i$ , et « modèle revu  $f(\theta)$  » dans le cas contraire.

## 4.3.1 En condition d'écoute normale

Les nouvelles définitions du modèle d'appariement décrites dans cette partie se basent sur la décomposition des distributions de probabilité dans la dimension azimutale en un produit d'une composante latérale et d'une composante polaire. L'évaluation de la pertinence des modifications apportées au modèle passe donc tout d'abord par l'étude de ces composantes. Nous nous restreignons à la condition d'oreille nue, pour laquelle le modèle est optimisé à partir des observations de l'expérience subjective.

Les composantes latérales des observations et des prédictions sont représentées en fonction de l'angle latéral  $\alpha$  et pour chaque position de source  $\theta_i$  sur la Figure IV.19. Au cours de l'expérience subjective, la détermination de l'angle latéral était plus précise à proximité du plan médian à l'avant. À l'arrière, les performances sont plus mauvaises, excepté pour une source à 180°. Les modèles original et revu (paramètres indépendants de l'azimut) présentent une prédiction de l'angle latéral plus uniforme en fonction de l'azimut de la source. Avec le premier, la probabilité de trouver l'angle latéral est maximale lorsque la source est sur l'axe interaural. À l'inverse, le modèle revu prédit une plus grande confusion quand la source est sur ou juste derrière cet axe. Quand les paramètres du modèle revu dépendent de l'azimut, la composante latérale de la sortie du modèle d'appariement semble plus proche de celle obtenue à partir de l'expérience subjective. L'augmentation de  $\sigma_{ITD}$ à l'arrière observée sur la Figure IV.17b permet notamment au modèle de simuler correctement la baisse de précision dans cet hémisphère. On observe toutefois que la probabilité de trouver l'angle latéral, lorsque la source est juste devant l'axe interaural, est toujours plus élevée que dans le cas de l'expérience subjective.

La Figure IV.20 représente la composante polaire avant, en condition d'oreille libre et en fonction de l'azimut, des observations subjectives et des modèles d'appariement. L'expérience subjective montre que la distinction avant-arrière est très précise au niveau du plan médian et devient moins évidente quand la source se rapproche de l'axe interaural. Le modèle d'appariement original prédit une tendance correcte pour la composante polaire, mais sous-estime la précision au niveau du plan médian. Les composantes polaires prédites par les modèles d'appariement revus sont significativement plus réalistes. Dans le cas où



FIG. IV.19 – Composante la térale des observations subjectives et de la sortie des modèles d'appariement en condition d'écoute normale.



FIG. IV.20 – Composantes polaires avant en fonction de l'azimut de la source pour la condition d'écoute normale.

les paramètres dépendent de la position de la source, la différence de probabilité avec l'expérience subjective est toujours inférieure à 0,07, alors qu'elle atteint 0,35 quand les paramètres sont indépendants de  $\theta_i$ .

La comparaison des composantes latérale et polaire subjectives et de celles prédites par les modèles d'appariement montre que, en condition d'écoute normale, le modèle basé sur la décomposition latérale/polaire et optimisant ses paramètres en fonction de l'azimut permet de prédire des performances de localisation plus proches de la réalité. Dans la suite, les performances des modèles appliqués aux TCAPS sont évaluées.

### 4.3.2 Application aux TCAPS

Les modèles d'appariement, original et revus, sont appliqués aux six TCAPS de notre étude. Dans la Section 3, nous avons vu que le modèle original prédit exclusivement des erreurs avant-arrière et très peu d'erreurs d'autres types. Ces observations restent vraies avec les deux modèles revus. En conséquence, dans cette section, nous n'étudions que la probabilité de localisation voisine et la vraisemblance des observations avec les modèles.

Les modèles d'appariement prédisent des probabilités de localisation voisine qui sont comparées à celle observée dans l'expérience subjective en Figure IV.21a. Par rapport au modèle original, les prédictions des modèles revus sont plus réalistes en condition d'écoute normale, avec les bouchons non linéaires P1 et P2 et le casque talk-through P6. Avec P5, la probabilité de localisation voisine est au contraire largement sous-estimée. Finalement, l'utilisation de la décomposition latérale/polaire semble apporter une amélioration des prédictions, mais pas dans toutes les conditions d'écoute.



FIG. IV.21 – Étude des performances prédites par les différents modèles d'appariement dans chaque condition d'écoute.

L'évaluation des modèles d'appariement passe aussi par l'étude de la vraisemblance des observations subjectives. La Figure IV.21b dévoile que, par rapport au modèle original, les modèles revus entraînent une légère augmentation de la vraisemblance (environ 0,2) en condition d'écoute normale, mais aussi une réduction avec les TCAPS. Si cette baisse ne dépasse pas 0,6 avec les bouchons non linéaires et P6, elle devient substantielle avec les autres systèmes talk-through. En particulier, elle est maximale avec le bouchon P3. La prise en compte de la dépendance des paramètres du modèle à l'azimut permet une légère amélioration de la vraisemblance avec les TCAPS, par rapport au modèle revu dont les paramètres sont indépendants de l'azimut.

Les modèles d'appariement revus entraînent, dans la plupart des conditions d'écoute, à la fois une amélioration de la prédiction de la probabilité de localisation voisine et une réduction de la vraisemblance. Ces observations semblent contradictoires en premier lieu, mais peuvent être expliquées par un manque de prédiction d'erreurs grossières par les modèles. En effet, les observations de l'expérience subjective présentent un nombre conséquent de ce type d'erreur. Si les modèles d'appariement prédisent une probabilité proche de zéro aux positions correspondantes, la vraisemblance en est fortement réduite. S'ils permettent une meilleure prédiction des nombres de localisations voisines et d'erreurs avant-arrière, il semble tout de même que les modèles revus soient globalement moins réalistes. Ceci peut être confirmé par l'étude des matrices de probabilités de la Figure E.1 dans l'Annexe E.

Le modèle d'appariement original montrait déjà une baisse des performances avec P3 par rapport à P4, et à moindre mesure avec P5 par rapport à P6. Avec les modèles revus, les écarts de vraisemblance sont accentués pour les deux TCAPS. Concernant la probabilité de localisation voisine, les modèles revus prédisent surtout une forte réduction avec le casque P5. Ces mauvaises performances sont essentiellement dues à une augmentation des erreurs avant-arrière. Nous avons vérifié que les mêmes résultats étaient obtenus après

Condition d'écoute	Sans	P1	P2	$\mathbf{P3}$	P4	P5	P6
Paramètre							
Écart-type $\sigma_{ITD}$ (µs)	103	151	146	205	159	192	194
Écart-type $\sigma_{\delta}$ (dB)	$0,\!14$	$0,\!23$	$0,\!22$	20	$0,\!41$	20	$0,\!17$

TAB. IV.2 – Écarts-types  $\sigma_{ITD}$  et  $\sigma_{\delta}$  du modèle d'appariement revu optimisés pour chaque condition d'écoute.

avoir remesuré les HRTF avec P3 et P5. Le modèle d'appariement, principalement sous ses formes exploitant la décomposition latérale/polaire, ne paraît pas être adapté à la prédiction des performances de localisation avec les TCAPS P3 et P5.

La dernière supposition est vérifiée par le calcul des écarts-types relatifs à l'ITD et aux indices spectraux du modèle d'appariement revu (sans dépendance à l'azimut) pour chaque TCAPS : pour une condition d'écoute, l'algorithme d'optimisation est utilisé à partir des observations subjectives et des HRTF de cette condition. Les valeurs obtenues pour  $\sigma_{ITD}$ et  $\sigma_{\delta}$  sont indiquées dans le Tableau IV.2. L'écart-type de l'ITD des TCAPS P3 et P5 n'est pas significativement différent de celui des autres conditions d'écoute. En revanche, l'écart-type des indices spectraux est beaucoup plus grand avec P3 et P5. D'ailleurs, la valeur obtenue est égale à la borne supérieure de la gamme d'étude du paramètre. Si celleci est augmentée, par exemple à la valeur aberrante de 200 dB,  $\sigma_{\delta}$  est fixée à 200 dB par l'optimisation pour P3 et P5. En conclusion, le modèle d'appariement n'est pas adapté aux TCAPS P3 et P5, ou en tout cas aux HRTF que nous avons mesurées avec, car la discrimination avant-arrière par les indices spectraux est empêchée.

Pour conclure, la révision du modèle d'appariement, en exploitant la décomposition latérale/polaire, permet d'améliorer la prédiction des probabilités de localisation voisine et d'erreur avant-arrière en sacrifiant la vraisemblance. Cette dernière est légèrement moins dégradée lorsque les paramètres du modèle dépendent de l'azimut de la source. De plus, une étape de vérification de la pertinence du modèle pour un TCAPS particulier semble nécessaire. En prêtant attention à cette nouvelle étape, c'est le modèle d'appariement revu, avec dépendance à l'azimut, qui est retenu.

# 5 Correction du réseau de neurones

Le réseau de neurones est « entraîné » à localiser des sources sonores à partir d'une sortie théorique dont la précision ne dépend pas de la position de la source. De façon analogue au modèle d'appariement original, il peut être modifié pour tenter d'obtenir des performances de localisation plus proches de celles observées dans l'expérience subjective. Trois améliorations possibles du réseau de neurones sont donc détaillées dans la section suivante. Leur validité est ensuite étudiée en décrivant les prédictions en condition d'oreille nue, ainsi qu'en calculant leur vraisemblance vis-à-vis des observations de l'expérience subjective.



FIG. IV.22 – (a) Paramètre de concentration  $\kappa$  obtenu en fonction de l'azimut à partir des résultats de l'expérience subjective en condition d'écoute normale. (b) Sortie théorique du réseau de neurones, constituée des distributions de von Mises correspondantes.

# 5.1 Définition des améliorations

Les principaux paramètres du réseau de neurones pouvant être modifiés pour tenter de l'améliorer sont la concentration  $\kappa$ , et plus généralement la sortie théorique de l'algorithme d'apprentissage, et la limite de l'erreur quadratique qui fixe l'arrêt de cet apprentissage.

## 5.1.1 Dépendance de $\kappa$ vis-à-vis de la position

Dans la Section 2.3, le paramètre de concentration  $\kappa$ , définissant la sortie théorique du réseau de neurones, est une moyenne calculée sur l'ensemble des positions. Pour rappel, elle provient de l'estimation des paramètres d'une distribution de von Mises à partir des résultats de l'expérience subjective en condition d'oreille libre. L'ajustement de cette valeur pour chaque azimut  $\theta$  dans la phase d'apprentissage pourrait mener à des performances de localisation plus réalistes.

Quand on ne moyenne pas les différentes valeurs de  $\kappa$  obtenues pendant l'estimation, on obtient une variation en fonction de la position décrite en Figure IV.22a. On remarque que la concentration est plus élevée pour les positions correspondant aux sources sonores réelles. Ce résultat rejoint les observations faites sur les performances des auditeurs de l'expérience subjective dans la Section 2.1, Chapitre III. De plus, on remarque que la concentration est globalement plus faible à l'arrière. L'évolution de la concentration est donc inversement proportionnelle à celle des paramètres  $\sigma_{ITD}$  et  $\sigma_{\delta}$  du modèle d'appariement optimisés précédemment.

Dans la suite, le réseau de neurones est entraîné avec la sortie théorique constituée de distributions de von Mises dont la concentration dépend de la position de la source. Cette sortie théorique est représentée sur la Figure IV.22b. L'augmentation de la précision de localisation quand la concentration  $\kappa$  augmente est bien observée.

Sortie théorique	Limite de l'erreur quadratique		
	E = 0,1	E = 0,01	
Distribution de von Mises, $\kappa$ fixe	2300	non calculé	
Distribution de von Mises, $\kappa = f(\theta)$	2500	11000	
Observations subjectives	2700	5500	

TAB. IV.3 – Nombre d'itérations nécessaires à l'apprentissage du réseau de neurones en fonction de sa sortie théorique et du seuil d'apprentissage E.

## 5.1.2 Utilisation directe des données subjectives

La création d'une sortie théorique pour le réseau de neurones à partir d'une distribution de von Mises permet dans l'absolu de ne pas avoir besoin des données d'une expérience subjective en choisissant une distribution arbitraire. Dans le cadre de notre étude, une telle expérience a été réalisée et les paramètres de la distribution estimée à partir des résultats. Il semble tout de même évident que les performances de prédiction du réseau de neurones pourraient être améliorées en utilisant directement les données de l'expérience subjective.

Comme pour l'optimisation des paramètres du modèle d'appariement, l'apprentissage du réseau de neurones est aussi effectué à partir des performances de localisation moyennes des auditeurs en condition d'écoute normale et normalisées de manière à former une matrice de probabilités de localisation sonore. La sortie théorique est dans ce cas la matrice de localisation précédemment représentée sur la Figure IV.10a. Les principales différences avec la sortie théorique constituée des distributions de von Mises (cf. Figure IV.22b) sont les probabilités de localisation exacte sur la diagonale principale. Les probabilités de localisation voisine devraient en revanche être similaires.

## 5.1.3 Réduction de la limite de l'erreur quadratique

Dans la Section 2.3, on arrête l'apprentissage du réseau de neurones lorsque l'erreur quadratique entre sa sortie et la sortie théorique devient inférieure à 0,1. Les deux tentatives d'amélioration des performances du réseau énoncées précédemment seront aussi appliquées en réduisant le seuil de l'erreur quadratique à 0,01.

Ce changement de limite dans le processus d'apprentissage a pour conséquence l'augmentation du nombre d'itérations nécessaires pour l'atteindre. Le Tableau IV.3 répertorie le nombre moyen d'itérations, pour chaque type de sortie théorique utilisée dans cette section et pour les seuils d'apprentissage 0,1 et 0,01. L'utilisation des concentrations dépendantes de l'azimut ou des observations subjectives n'entraîne pas d'augmentation substantielle du temps de calcul quand E = 0,1. À l'inverse, quand E = 0,01, le nombre d'itérations est quadruplé pour les distributions de von Mises dépendantes de l'azimut, et doublé pour les observations subjectives.

Dans la section suivante, les performances de prédiction des différents réseaux de neurones sont comparées.

# 5.2 Étude des corrections

Comme pour l'étude des modifications du modèle d'appariement, celle des améliorations du réseau de neurones est réalisée en deux parties : l'étude des performances de prédiction en condition d'oreille libre, et l'évaluation de la probabilité de localisation voisine et de la vraisemblance obtenues avec les systèmes tactiques de communication et protection. Comme le processus d'apprentissage initialise aléatoirement les poids synaptiques, il est nécessaire de générer plusieurs réseaux. Comme pour le réseau original, pour chaque modification nous générons 200 réseaux différents. Les résultats donnés dans cette section sont des valeurs moyennées à partir de ces 200 réseaux.

Dans la suite, on se réfère au réseau de neurones original avec la légende « von Mises,  $\kappa$  fixe, E=0.1 ». L'utilisation de distributions de von Mises dont la concentration dépend de l'azimut est repérée par la légende « von Mises,  $\kappa = f(\theta)$  », l'utilisation des données subjectives est notée « Observations » et le seuil d'apprentissage utilisé est indiqué.

#### 5.2.1 En condition d'écoute normale

La probabilité de localisation en fonction de l'azimut de la source et de l'azimut testé est représentée sur la Figure IV.23 pour l'expérience subjective et pour chaque réseau de neurones. On remarque tout d'abord que l'utilisation d'une concentration dépendante de l'azimut, ou des observations subjectives, pour la sortie théorique permet effectivement au réseau d'introduire une variation de la précision en fonction de la position de la source. Dans le cas où  $\kappa$  est fixe, on n'observe effectivement aucune variation de précision. L'augmentation du seuil de l'erreur d'apprentissage entraîne de plus la réduction du nombre d'erreurs avant-arrière résiduelles, observées dans le cas où E = 0,1.

Quand la sortie théorique est construite à partir des observations subjectives, la probabilité de localisation prédite est, de façon peu surprenante, la plus réaliste. Mais comme le réseau est capable de simuler correctement les probabilités de localisation exacte, de localisation voisine et d'erreur grossière, le seuil de l'erreur quadratique est atteint alors que subsistent des erreurs avant-arrière (voire gauche-droite et mixtes). Dans le cas où la sortie théorique est définie à partir de distributions de von Mises, le nombre d'erreurs avant-arrière prédit est plus réaliste. Toutefois, le seuil de l'erreur est atteint alors que la variation de précision en fonction de l'azimut reste approximative.

# 5.2.2 Application aux TCAPS

Pour chaque tentative d'amélioration du réseau de neurones, la probabilité de localisation voisine  $P_{loc}$  et la vraisemblance des observations subjectives sont calculées dans chacune des sept conditions d'écoute. Les résultats sont reportés sur les Figures IV.24a et IV.24b. Sur la première, la courbe correspondant au réseau original (von Mises,  $\kappa$  fixe, E=0.1) est confondue avec celles correspondant aux légendes « Observations, E=0.01 »



FIG. IV.23 – Étude des performances prédites en condition d'écoute normale par les différents réseaux de neurones. Comparaison avec les performances obtenues dans l'expérience subjective.



FIG. IV.24 – Étude de la probabilité de localisation voisine et de la vraisemblance obtenues avec les différents réseaux de neurones dans chaque condition d'écoute.

et « von Mises,  $\kappa = f(\theta)$ , E=0.1 ». Sur la seconde figure, elle est de nouveau confondue avec « von Mises,  $\kappa = f(\theta)$ , E=0.1 ».

La plus mauvaise prédiction de probabilité de localisation voisine est obtenue pour la sortie théorique basée sur les observations subjectives, avec le seuil E = 0,1. Nous avons précédemment observé que, même en condition d'écoute normale, ce réseau introduit toujours un nombre significatif d'erreurs avant-arrière. Cette observation peut être mise en relation avec les mauvaises performances constatées ici.

La Figure IV.24a montre aussi que les performances prédites par le réseau original sont équivalentes à celles obtenues quand la sortie théorique est composée soit de distributions de von Mises, avec E = 0,1, soit des observations subjectives, avec E = 0,01. En condition d'écoute normale et avec les bouchons, une sortie théorique basée sur des distributions de von Mises et un seuil d'apprentissage égal à 0,01 permettent une meilleure prédiction de la probabilité de localisation voisine. Dans la section précédente, nous avons remarqué que ce réseau produit la probabilité de localisation la plus précise pour atteindre le seuil d'apprentissage, aux dépens de la variation de précision avec l'azimut. Finalement, tous les réseaux de neurones prédisent les mêmes mauvaises performances avec les casques talkthrough. Plus généralement, tous les réseaux présentent la même évolution de  $P_{loc}$  en fonction de la condition d'écoute (notamment son augmentation avec P2).

Sur la Figure IV.24b, on observe dans un premier temps que l'utilisation d'un paramètre de concentration dépendant de la position de la source sonore n'engendre pas d'augmentation de la vraisemblance du réseau de neurones par rapport à l'utilisation d'une concentration moyenne fixe. D'ailleurs, la réduction du seuil de l'erreur quadratique au cours de l'apprentissage entraîne même une chute de la vraisemblance avec les TCAPS.

Ensuite, l'utilisation directe des données en condition d'oreille libre de l'expérience subjective permet d'augmenter la vraisemblance du réseau de neurones dans chaque condition d'écoute. Encore une fois, on observe un effet négatif de la réduction du seuil de l'erreur, principalement pour les casques talk-through P5 et P6. Comme indiqué dans la Section 2.3, le choix d'un seuil d'erreur trop faible entraîne la surspécialisation du réseau de neurones : il n'est plus capable d'utiliser des entrées trop différentes de celles utilisées pendant l'apprentissage pour localiser les sources sonores.

En conclusion, la sortie théorique du réseau de neurones permettant d'obtenir la vraisemblance la plus élevée est l'utilisation directe des données de l'expérience subjective. Néanmoins, pour que la prédiction de la probabilité de localisation voisine soit correcte, il est nécessaire d'augmenter le seuil d'apprentissage. Ceci reste acceptable puisque la vraisemblance n'en est que peu affectée. Quand la sortie théorique est basée sur des distributions de von Mises dépendant de l'azimut de la source, la vraisemblance reste acceptable pour E = 0.1. Les performances prédites dans ce cas sont similaires à celles du réseau original et du réseau généré à partir des observations subjectives (E = 0.01). En revanche, si le seuil d'apprentissage est diminué, la prédiction de  $P_{loc}$  est amélioré, mais en sacrifiant la vraisemblance. Comme il a précédemment été observé pour le modèle d'appariement, ce réseau de neurones doit être trop précis : de faibles probabilités sont prédites pour les erreurs grossières. Comme ces erreurs sont observées dans l'expérience subjective, la vraisemblance est substantiellement réduite. Dans la suite, nous conserverons donc principalement le réseau de neurones dont la sortie théorique est basée sur les observations de l'expérience subjective, avec E = 0.01. Les matrices de probabilités obtenues avec ce réseau dans chaque condition d'écoute sont représentées sur la Figure E.3 dans l'Annexe E. Celui défini à partir des distributions de von Mises dont la concentration dépend de l'azimut, avec E = 0,1, est tout de même conservé puisqu'il produit une alternative à l'utilisation des résultats d'une expérience subjective.

# 6 Discussion

# 6.1 Résultats finals

Après correction des deux modèles de localisation sonore développés dans ce chapitre, les prédictions des performances avec les systèmes tactiques de communication et protection ont été améliorées. Les Figures IV.25a et IV.25b représentent respectivement la probabilité de localisation voisine et la vraisemblance calculées dans chaque condition d'écoute. En condition d'écoute normale, les deux modèles retenus présentent des performances équivalentes, car ils sont tous les deux optimisés à partir des observations subjectives dans cette condition. De manière générale, leurs performances sont toujours correctes avec les bouchons non linéaires et talk-through. On notera tout de même une augmentation avec P2 pour le réseau de neurones et une baisse avec P3 pour le modèle d'appariement. D'après la Figure IV.25b, la vraisemblance des observations est équivalente pour les deux modèles en condition d'oreille libre et avec P6. Le réseau de neurones présente la plus grande vraisemblance dans les autres conditions d'écoute.



(a) Probabilité de localisation voisine  $P_{loc}$ . (b) Vraisemblance des observations.

FIG. IV.25 – Comparaison des performances de localisation et de la vraisemblance du modèle d'appariement et du réseau de neurones retenus, dans chaque condition d'écoute.

Le réseau de neurones ne permet toujours pas de prédire des performances réalistes avec les casques talk-through. Une surestimation du nombre d'erreurs gauche-droite demeure la principale raison de ce problème (cf. Figure E.3, Annexe E). Avec P6, un nombre important d'erreurs grossières est aussi estimé. Concernant le modèle d'appariement, nous avons vu que les mauvaises performances prédites avec P3 et P5 viendraient du fait que la modélisation n'est pas adaptée aux HRTF mesurées avec ces TCAPS. Enfin, la faible vraisemblance obtenue dans presque toutes les conditions d'écoute avec le modèle d'appariement est due à un manque de prédiction d'erreurs grossières, c'est-à-dire à une trop grande précision du modèle dans le bon quadrant (cf. Figure E.1, Annexe E).

# 6.2 Effet des TCAPS sur les performances de localisation sonore

Le réseau de neurones corrobore la classification des TCAPS décrite dans le Chapitre III, ainsi que dans (Zimpfer & Sarafian, 2014) et (Joubaud et al., 2015). Ceci est aussi vrai pour le modèle d'appariement si l'on ne considère pas P3 et P5. Il est en effet clair que la détérioration des indices de localisation induite par les TCAPS peut être divisée en deux groupes, correspondant aux bouchons d'oreille et aux casques. Une explication triviale de cette séparation repose sur la proportion dans laquelle l'oreille est couverte par la protection auditive. Les casques causent ainsi une dégradation des performances de localisation prédites plus grande que celle des bouchons non linéaires ou talk-through.

À l'instar de l'étude statistique menée sur les résultats de l'expérience subjective dans le Chapitre III, on ne trouve que très peu de différences entre les bouchons d'oreille. Le modèle d'appariement prédit des performances similaires entre les bouchons non linéaires et P4, tandis que les résultats du réseau de neurones sont équivalents pour P1 et les bouchons talk-through. Le fait que les bouchons talk-through semblent plus remplir la conque ne joue donc pas substantiellement sur les performances de localisation prédites. En fait, nous nous attendons plutôt à trouver des différences significatives entre les types de bouchons d'oreille dans la détermination de l'élévation. En effet, celui-ci est perçu majoritairement grâce aux indices spectraux à haute fréquence. Or les systèmes talk-through présentent généralement une fréquence de coupure autour de 8 kHz. Cette hypothèse est cohérente avec les résultats comportementaux de Zimpfer & Sarafian (2014).

Nous avons vu que, selon le réseau de neurones, P2 serait tout de même le TCAPS qui aurait le plus faible impact sur la capacité à localiser les sources sonores. La distinction physique principale entre P2 et l'autre bouchon non linéaire P1 réside dans une obstruction moins complète de l'oreille (cf. Figures I.5a et I.6a). De façon similaire, en utilisant le processus d'appariement décrit dans (Langendijk & Bronkhorst, 2002), Durin et al. (2014) ont montré que plus une aide auditive est discrète (c'est-à-dire moins elle obstrue le pavillon de l'oreille), plus les HRTF enregistrées avec sont similaires à celles de la condition d'oreille libre. Néanmoins, une telle différence entre P1 et P2 n'a pas été observée au cours de l'expérience subjective.

# 6.3 Limitations

Nous avons tenté de réduire les divergences constatées entre les performances subjectives de localisation et les prédictions en faisant en sorte que les paramètres des modèles dépendent de la position de la source. Si cette tentative a permis de réduire l'écart avec les performances subjectives, on constate que la probabilité de localisation voisine (et a fortiori la probabilité de localisation exacte) est généralement toujours sous-estimée par les modèles. Une surestimation de la probabilité d'erreurs avant-arrière en est la cause principale. Contrairement aux modèles qui utilisent des HRTF mesurées à une position fixe, les auditeurs de l'expérience subjective ont pu améliorer leur discrimination avant-arrière par de légers mouvements de tête, même « s'ils n'en avaient pas le droit » (Thurlow et al., 1967; Wightman & Kistler, 1999; Blauert, 1983, chap. 2, sec. 5). En effet, McAnally & Martin (2014) ont montré qu'une rotation rapide de seulement 4° peut permettre d'améliorer significativement la discrimination avant-arrière. La prise en compte du mouvement de tête dans les modèles de localisation est possible, mais nécessiterait une mesure d'HRTF beaucoup plus précise (par exemple tous les degrés), ainsi que la définition d'un critère moyennant les différents indices de localisation engendrés.

Le modèle d'appariement revu est basé sur l'hypothèse que les indices spectraux sont utilisés uniquement pour la discrimination avant-arrière. Les résultats de Macpherson & Middlebrooks (2002) confirment que le poids des indices spectraux monauraux est négligeable dans la dimension latérale pour des auditeurs normo-entendants. Cependant, des auditeurs atteints de surdité unilatérale sont capables de repondérer ces indices pour localiser les sources sonores sans ITD ou ILD (Slattery & Middlebrooks, 1994). Le modèle d'appariement ne peut présentement pas simuler cette repondération puisque seule l'information polaire est extraite de la comparaison spectrale.

Concernant le réseau de neurones, son algorithme d'apprentissage par rétropropaga-

tion du gradient de l'erreur est très efficace, mais peut conduire à une surspécialisation du réseau. Nous avons observé cet effet dans le cas où la sortie théorique est basée sur des distributions de von Mises dépendantes de l'azimut, avec un seuil d'apprentissage égal à 0,01. La mauvaise prédiction des performances de localisation avec les casques talk-through en est sa principale conséquence. Le problème de surspécialisation du réseau pourrait éventuellement être résolu par l'implémentation d'un autre type de réseau de neurones, par exemple un réseau maximally fault-tolerant (Neti et al., 1992a). De plus, notre réseau de neurones prédit un nombre conséquent, et irréaliste, d'erreurs gauche-droite. Nous pensons que cette mauvaise prédiction provient d'un manque de pondération des ITD et ILD par rapport aux indices spectraux. Un nouveau réseau de neurones avec plus d'informations a priori sur l'utilisation des différents indices de localisation pourrait être défini. En particulier, nous pourrions appliquer le principe de décomposition latérale/polaire du modèle d'appariement pour y arriver.

L'objectif de notre étude est de retrouver les performances comportementales de localisation sonore à partir de modèles basés sur des HRTF mesurées sur un seul auditeur type, soit une tête artificielle. Le contraste entre les performances subjectives et les prédictions peuvent alors provenir de l'absence de variabilité interindividuelle dans les modèles. Les HRTF sont propres à chaque auditeur, et l'interaction entre les TCAPS et les indices spectraux peuvent l'être aussi. Il est donc d'importance majeure que les prédictions des modèles soient indépendantes de la tête artificielle choisie. De façon alternative, des HRTF mesurées sur les auditeurs ayant participé à l'expérience subjective pourraient être utilisées.

Les idiosyncrasies des auditeurs ne mènent pas seulement à des HRTF différentes, mais aussi à l'individualisation des facteurs non acoustiques. En appliquant le modèle de localisation de Baumgartner et al. (2013), Majdak et al. (2014) ont démontré que l'écart-type de la gaussienne définissant l'indice de similarité était variable en fonction de l'auditeur considéré. Dans notre cas, cela signifie que les paramètres  $\sigma_{\delta}$ ,  $\sigma_{ILD}$  et  $\sigma_{ILD}$  et les poids du réseau de neurones dépendent aussi du sujet. L'amélioration des modèles de localisation pourrait donc passer par l'utilisation d'HRTF individuelles et une optimisation des paramètres pour chaque auditeur. Ceci devrait conduire non seulement à des prédictions plus réalistes de l'impact des TCAPS sur leurs performances, mais aussi à un outil de sélection du protecteur auditif le plus adapté à un auditeur particulier au sens de la localisation sonore. Cependant, l'utilisation d'HRTF individuelles n'est pas envisageable dans le cadre de l'évaluation d'un prototype de TCAPS. Le remplacement d'une expérience subjective de localisation par une campagne de mesure d'HRTF avec et sans protection ne paraît pas raisonnable. La « construction » d'HRTF individuelles avec TCAPS pourrait toutefois être possible en exploitant une banque d'HRTF individuelles en condition d'écoute normale et des fonctions de transfert de TCAPS. Ces dernières seraient obtenues en supprimant les effets de la tête artificielle des HRTF mesurées avec les protecteurs.

Quoi qu'il en soit, les performances de localisation prédites à partir d'une tête artificielle peuvent être exploitées dans le cadre de l'amélioration des systèmes tactiques de communication et protection. À titre d'exemple, une attention particulière peut être portée à la directivité, au placement des microphones et à la fréquence de coupure des systèmes talk-through (Chung et al., 2008; Brungart et al., 2003a), à la forme des coquilles des casques (Rubak & Johansen, 1999), ou bien à la longueur du filtre des bouchons non linéaires.

# 6.4 Conclusion

Deux modèles de localisation sonore basés sur des HRTF mesurées sur le plan horizontal et exploitant l'ITD, l'ILD et les indices spectraux ont été développés. Ils sont optimisés afin que leurs sorties correspondent aux performances humaines en condition d'oreille libre, obtenues par une expérience subjective avec des sources réelles et virtuelles. Appliqué aux TCAPS, le réseau de neurones prédit des performances réalistes, sauf pour les casques talk-through. De plus, aucune erreur gauche-droite ou mixte n'a été observée dans l'expérience subjective, alors que le réseau de neurones en prédit un nombre significatif avec les casques. Le modèle d'appariement surestime principalement la probabilité d'erreur avant-arrière et conduit à des performances irréalistes pour deux TCAPS en particulier. De plus, aucune erreur grossière n'est engendrée par ce modèle. Toutefois, dans les autres conditions d'écoute, le modèle d'appariement prédit des probabilités de localisation voisine et d'erreur avant-arrière plus proches des performances humaines.

Finalement, la classification des TCAPS par type trouvée au terme d'une très longue expérience comportementale ( $\sim 60$  heures) a pu être retrouvée à partir de modèles de localisation basés sur des HRTF ( $\sim 2$  heures). Il paraît donc faisable d'utiliser ces modèles pour évaluer l'impact d'un prototype de TCAPS sur les performances de localisation sonore. Néanmoins, l'indépendance des prédictions vis-à-vis de la tête artificielle utilisée pour les mesures d'HRTF reste à vérifier.

# CHAPITRE V

# Validation de la modélisation et application

Dans le chapitre précédent, deux méthodes permettant de prédire les performances de localisation sonore ont été décrites. Elles utilisent les indices de localisation décrits dans le Chapitre II. Leurs prédictions à partir des DTF mesurées avec les TCAPS sont comparables aux résultats de l'expérience subjective de localisation sonore du Chapitre III. Le présent chapitre vise, dans un premier temps, à valider les deux modèles de prédiction, en étudiant l'indépendance des résultats par rapport à la tête artificielle utilisée. Ensuite, les deux méthodes sont appliquées dans le cadre d'une tentative d'amélioration des performances d'un casque talk-through, de l'étude de bouchons d'oreille à « atténuation uniforme » et de la prédiction de l'effet de prothèses auditives sur la localisation sonore.

# 1 Indépendance vis-à-vis de la tête artificielle

La prédiction des performances de localisation sonore à partir de DTF différentes de celles de l'auditeur peut conduire à des différences importantes par rapport aux performances réelles. L'objectif de notre étude est néanmoins de proposer une méthode s'affranchissant de sujets humains pour obtenir des performances de localisation proches de la réalité. Il n'est en effet pas pertinent de remplacer une expérience subjective avec 30 auditeurs par une série de mesures d'HRTF sur chacun de ces auditeurs, d'autant plus si des mesures avec TCAPS sont nécessaires. La prédiction des performances de localisation à partir d'une tête artificielle a donc été retenue, mais, encore une fois, deux têtes artificielles différentes procurent des HRTF différentes et donc théoriquement des performances de localisation différentes. Dans cette partie, le modèle d'appariement et le réseau de neurones développés dans le chapitre précédent sont appliqués aux HRTF mesurées sur la tête artificielle KEMAR avec le dispositif expérimental du CNAM. L'objectif est de



FIG. V.1 – Photographie du dispositif expérimental du CNAM. Le mannequin KE-MAR est placé au centre d'une sphère de 50 haut-parleurs.

montrer que l'utilisation d'une tête artificielle différente de celle de l'ISL permet d'obtenir des prédictions de performances de localisation sonore similaires.

Dans un premier temps, le dispositif de mesure d'HRTF est présenté. Ensuite, nous étudions le comportement de l'optimisation des deux modèles de localisation avec ces nouveaux indices de localisation. Enfin, les prédictions des performances de localisation sonore obtenues sont comparées à celles du chapitre précédent.

# 1.1 Dispositif expérimental du CNAM

Les mesures ont été réalisées au Conservatoire National des Arts et Métiers (CNAM) dans la sphère de reproduction sonore développée par Lecomte (2016). La tête artificielle utilisée est un mannequin possédant des épaules et un buste : le Knowles Electronics Manikin for Acoustic Research (KEMAR) (Burkhard & Sachs, 1975). À l'intérieur d'une chambre traitée acoustiquement, le mannequin KEMAR est placé de manière à ce que sa tête soit au centre d'une sphère de 2,40 m de diamètre, constituée de 50 haut-parleurs. Une photographie du mannequin ainsi positionné est donnée en Figure V.1. Comme dans le Chapitre II Section 2.1.1, la tête peut porter des protections auditives et fait face au haut-parleur situé à la position d'azimut nul sur le plan horizontal. Ce plan contient 8 haut-parleurs séparés de 45°. La tête sera encore une fois tournée de 22,5° dans le sens trigonométrique pour avoir 16 mesures sur le plan horizontal. Le schéma de la Figure II.13 correspond donc aussi à ce dispositif. Contrairement à la tête artificielle ISL, le mannequin possède deux pavillons de tailles différentes. La mesure sur les deux oreilles conduit donc à deux jeux de fonctions de transfert différents, correspondant chacun à une oreille. Pour chaque jeu, on obtient les fonctions de transfert de l'autre oreille en supposant que le mannequin est parfaitement symétrique. Dans la suite, seules les mesures de l'oreille gauche



FIG. V.2 – Exemple de l'effet du lissage par liftrage cepstral sur l'amplitude d'une fonction de transfert.

sont conservées, car son pavillon correspond à une physionomie moyenne.

Le calcul de fonctions de transfert est réalisé par un programme MATLAB<sup><sup>w</sup></sup> qui génère lui-même un sinus glissant transmis au haut-parleur utilisé. La méthode appliquée dans ce protocole pour obtenir les fonctions de transfert est celle détaillée dans (Farina, 2007) et utilisée dans (Zhang et al., 2009). Les réponses impulsionnelles obtenues sont échantillonnées à 48 kHz et contiennent 52800 points. Ceci entraîne une résolution fréquentielle inférieure à 1 Hz pour les fonctions de transfert. Ces dernières sont donc très bruitées. Dans l'objectif de les lisser, nous utilisons deux techniques. La première consiste à appliquer une fenêtre de Tukey à la réponse impulsionnelle. Elle peut être assimilée à un lobe de cosinus convolué avec une fenêtre rectangulaire (Harris, 1978). Cette méthode lisse la fonction de transfert tout en conservant les propriétés temporelles. Des failles dans le spectre d'amplitude sont toutefois observables. La seconde technique consiste en l'extraction de l'enveloppe spectrale de la réponse impulsionnelle par liftrage cepstral (Childers et al., 1977). Elle renvoie un excellent lissage du spectre d'amplitude, mais dégrade substantiellement la phase (empêchant alors d'évaluer l'ITD). Pour ne conserver que les avantages de ces méthodes, les fonctions de transfert que nous gardons sont construites à partir du module du résultat du liftrage cepstral et de la phase du résultat du fenêtrage. La Figure V.2 représente un exemple d'amplitude de fonction de transfert avant et après ce lissage, en fonction de la fréquence. Pour finir, pour avoir la même gamme de fréquences que les mesures sur la tête artificielle ISL, nous ne conservons que les points correspondant à une fréquence comprise entre 90 Hz et 17,8 kHz.

Comme dans le protocole de l'ISL décrit dans le Chapitre II, les étapes d'égalisation des

Modèle d'appariement	Paramètre	DTF ISL	DTF KEMAR	
Original	$\sigma_{ITD}$ (ms)	0,101	0,106	
	$\sigma_{ILD}$ (dB)	4,2	$^{4,4}$	
	$\sigma_{\delta} (dB)$	$_{0,1}$	$0,\!36$	
	$w_{ITD}/w_{\delta}$	$0,\!68 \neq 0,\!06$	$0,74 \ / \ 0,48$	
Revu	$\sigma_{ITD}$ (ms)	0,103	0,119	
	$\sigma_{ILD}$ (dB)	2	3	
	$w_{lat}$	1	0,96	
	$\sigma_{\delta} (dB)$	$0,\!14$	$0,\!65$	
Revu $f(\theta)$	$\sigma_{ITD}$ (ms)	$0,05 \le \sigma_{ITD} \le 0,15$	$0,08 \le \sigma_{ITD} \le 0,2$	
	$w_{lat}$	1	1	
	$\sigma_{\delta} \ (dB)$	$0,05 \le \sigma_{\delta} \le 0,25$	$0, 1 \le \sigma_{\delta} \le 20$	

TAB. V.1 – Paramètres des trois modèles d'appariement optimisés à partir des DTF mesurées sur la tête artificielle ISL et sur le mannequin KEMAR

fonctions de transfert utilisées pour obtenir les DTF sont appliquées. Les mêmes méthodes permettent aussi d'estimer les différents indices de localisation. Le filtrage par la banque de filtres gammatones est d'ailleurs encore nécessaire puisque la résolution fréquentielle de ces DTF est très importante (environ 1 Hz).

# 1.2 Optimisation du modèle d'appariement avec le KEMAR

Dans le Chapitre IV, trois modèles d'appariement sont comparés. Nous les avons appelés le modèle original, le modèle revu et le modèle revu avec des paramètres dépendants de la position. Nous avons jugé que ce dernier permet une meilleure prédiction des performances de localisation sonore avec et sans TCAPS. Dans cette section, l'optimisation des trois modèles d'appariement à partir des DTF du mannequin KEMAR est étudiée.

Le Tableau V.1 répertorie les paramètres des trois modèles d'appariement obtenus à partir des DTF de la tête artificielle ISL et des DTF du mannequin KEMAR. Pour le modèle original, la principale différence concerne les paramètres relatifs aux indices spectraux : l'écart-type  $\sigma_{\delta}$  et la pondération  $w_{\delta}$  sont largement augmentés par rapport à ceux obtenus précédemment. Cela signifie que des différences plus importantes entre les DTF du KEMAR sont présentes, et que le modèle accorde une plus grande importance aux indices spectraux. Les paramètres relatifs aux différences interaurales de temps et d'intensité sont au contraire très peu affectés par le changement de tête artificielle : les écarts-types sont plutôt grands par rapport aux valeurs de JND, et l'ITD est toujours prépondérante par rapport à l'ILD.

Des observations similaires peuvent être faites pour le modèle d'appariement revu, basé sur la décomposition latérale/polaire. La valeur de  $\sigma_{\delta}$  est presque cinq fois plus grande que celle obtenue avec les DTF ISL. L'écart-type de l'ITD a augmenté de seulement 16 µs et l'ILD n'est quasiment pas prise en compte (pondération de 1 - 0.96 = 0.04). Dans le cas où les écarts-types dépendent de la position de la source, les extremums de  $\sigma_{ITD}$ 



FIG. V.3 – Composantes polaires avant, obtenues avec les DTF du KEMAR, en fonction de l'azimut de la source pour la condition d'écoute normale.

sont similaires à ceux obtenus dans le chapitre précédent. En revanche, le maximum de l'écart-type des indices spectraux atteint la limite du domaine d'optimisation (20 dB). Plus particulièrement, cette limite est atteinte au niveau de l'axe interaural (67,5°, 270° et 292,5°). Pour ces positions, le modèle n'arrive pas à faire de distinction avant-arrière. Sur le reste du plan horizontal,  $\sigma_{\delta}$  reste inférieur à 1 dB, mais il est toujours supérieur à celui obtenu avec les DTF ISL.

D'après la comparaison des paramètres du Tableau V.1, il semble que l'application des modèles d'appariement aux DTF mesurées sur le mannequin KEMAR entraîne principalement des différences dans la dimension polaire, par rapport aux DTF mesurées sur la tête artificielle ISL. Cette supposition est confirmée par la Figure V.3 représentant la composante polaire avant de chaque modèle d'appariement en fonction de la position de la source. Seule la prise en compte de la dépendance de  $\sigma_{\delta}$  à la position de la source permet d'obtenir une composante polaire proche de celle de l'expérience subjective. Avec les DTF ISL, le modèle revu en permettait aussi une prédiction correcte. On remarque aussi que les erreurs de prédiction de la composante polaire sont principalement présentes pour les sources virtuelles.

Pour finir, nous comparons la probabilité de localisation voisine et la vraisemblance prédites par les trois modèles d'appariement, avec et sans TCAPS, aux résultats de l'expérience subjective. D'après la Figure V.4a, le modèle revu n'engendre pas une amélioration de la prédiction de  $P_{loc}$  par rapport au modèle original, et le modèle revu  $f(\theta)$  prédit une probabilité de localisation voisine plus proche de la réalité. Ce dernier modèle montre, de manière similaire à l'expérience subjective, une classification des TCAPS par type, avec peu de différences entre les bouchons non linéaires et P3, et entre P4 et les casques talk-



FIG. V.4 – Probabilités de localisation voisine et vraisemblances des trois modèles d'appariement du mannequin KEMAR pour chaque condition d'écoute.

through. L'amélioration apportée par la prise en compte de la dépendance des paramètres à l'azimut de la source est principalement causée par la meilleure prédiction de la composante polaire. Ainsi, la surestimation des erreurs avant-arrière est réduite. En revanche, comme dans le cas de la tête artificielle ISL, trop peu d'erreurs grossières sont prédites par les trois modèles. Cela conduit à une réduction de la vraisemblance (cf. Figure V.4b). À l'inverse de la probabilité de localisation voisine, la plus mauvaise vraisemblance est produite par le modèle revu  $f(\theta)$ . Comme avec les DTF ISL, une prédiction correcte des probabilités de localisation voisine et d'erreur avant-arrière n'entraîne pas forcément une amélioration de la prédiction globale des distributions de probabilités.

En conclusion, l'utilisation de DTF mesurées sur une autre tête artificielle conduit à des paramètres relatifs aux ITD et ILD équivalents, mais aussi à des différences de paramètres relatifs aux indices spectraux. Dans notre cas, la sensibilité à ces derniers est augmentée de manière conséquente. Pour certaines positions de source, le modèle d'appariement revu  $f(\theta)$  n'est même plus capable de réaliser la distinction avant-arrière. Si ce problème peut aussi être la cause d'une diminution de la vraisemblance, on retrouve tout de même que le modèle revu  $f(\theta)$  conduit à des prédictions des performances de localisation sonore plus proches de celles de l'expérience subjective. Dans une prochaine section, les prédictions avec les TCAPS de ce modèle sont comparées à celles engendrées par les DTF de la tête artificielle ISL. Cette comparaison est aussi faite avec les prédictions du réseau de neurones généré avec les DTF du mannequin KEMAR, dont l'optimisation est discutée dans la section suivante.

## 1.3 Optimisation du réseau de neurones avec le KEMAR

Le réseau de neurones développé dans le Chapitre IV permet de prédire les performances de localisation sonore avec les TCAPS grâce à son apprentissage en condition d'écoute normale. Ce processus est basé sur l'ajustement de la sortie du réseau à une sortie théorique, dans le cas où les entrées du réseau sont les indices de localisation mesurés en condition d'oreille libre. Trois sorties théoriques différentes ont été comparées en utilisant les DTF mesurées sur la tête artificielle ISL :

- une distribution de von Mises à concentration fixe (von Mises,  $\kappa$  fixe);
- une distribution de von Mises dont la concentration dépend de la position de la source (von Mises,  $\kappa = f(\theta)$ );
- les probabilités de localisation dérivées des observations de l'expérience subjective (Observations).

De plus, l'algorithme d'apprentissage est stoppé quand l'erreur quadratique entre la sortie du réseau et la sortie théorique atteint une valeur fixe. Dans le chapitre précédent, nous avons fixé ce seuil d'apprentissage à 0,1 ou 0,01. Les réseaux de neurones générés par ces différents seuils et sorties théoriques à partir des DTF mesurées sur le mannequin KEMAR sont évalués.

En condition d'écoute normale (utilisée pour l'apprentissage), les matrices de probabilités de localisation sonore données par chaque réseau de neurones, à partir des DTF du KEMAR, sont comparées aux résultats de l'expérience subjective sur la Figure V.5. L'apprentissage à partir de distributions de von Mises à concentration fixe ne permet pas de prédire correctement les variations de précision avec l'azimut. Ces dernières sont présentes quand les observations de l'expérience subjective sont directement utilisées, mais des erreurs avant-arrière résiduelles subsistent, même si le seuil d'apprentissage est réduit à 0,01. Des performances correctes sont obtenues à partir de distributions de von Mises dont la concentration dépend de l'azimut. Le nombre d'erreurs avant-arrière est particulièrement réduit en baissant le seuil d'apprentissage. Finalement, les observations faites avec les DTF du mannequin KEMAR sur la Figure V.5 sont les mêmes que celles faites avec les DTF de la tête artificielle ISL dans le Chapitre IV, Section 5.2.

Les DTF mesurées sur le mannequin KEMAR avec les TCAPS sont appliquées aux cinq réseaux de neurones. Les Figures V.6a et V.6b représentent respectivement les probabilités de localisation voisine et les vraisemblances des réseaux dans les sept conditions d'écoute. Le réseau basé sur les observations subjectives avec un seuil d'apprentissage de 0,1 sous-estime largement  $P_{loc}$  dans toutes les conditions d'écoute, probablement à cause des erreurs avant-arrière résiduelles. Les autres réseaux prédisent cette métrique correctement en condition d'écoute normale et avec les bouchons non linéaires. Avec les systèmes talk-through,  $P_{loc}$  est sous-estimé d'environ 0,17 pour les bouchons et d'environ 0,21 pour les casques. Le classement des TCAPS par type est aussi très marqué. Enfin, le réseau de neurones (von Mises,  $\kappa = f(\theta)$ , E = 0,01) produit une surestimation de la probabilité de localisation voisine avec les bouchons non linéaires.

L'étude de la vraisemblance des réseaux de neurones (Figure V.6b) montre que la vraisemblance est particulièrement améliorée quand les observations de l'expérience subjective sont utilisées pour l'apprentissage. La définition de la sortie théorique à l'aide de distribu-



FIG. V.5 – Étude des performances prédites en condition d'écoute normale par les différents réseaux de neurones basés sur les DTF du KEMAR. Comparaison avec les performances obtenues dans l'expérience subjective.



FIG. V.6 – Probabilités de localisation voisine et vraisemblances des cinq réseaux de neurones du mannequin KEMAR pour chaque condition d'écoute.

tions de von Mises réduit la vraisemblance, particulièrement pour les casques talk-through et le bouchon P3. De plus, la diminution du seuil d'apprentissage pour les distributions de von Mises dont la concentration dépend de l'azimut entraîne une réduction supplémentaire de la vraisemblance.

En conclusion, comme avec les DTF ISL, le réseau de neurones généré à partir des données subjectives et des DTF du mannequin KEMAR permet d'obtenir les performances de localisation sonore les plus réalistes, à condition que le seuil d'apprentissage soit suffisamment petit. L'utilisation de distributions de von Mises pour la sortie théorique reste toutefois une option, cette fois, tant que le seuil d'apprentissage n'est pas trop petit. Si la vraisemblance en est particulièrement affectée, la prédiction de la probabilité de localisation voisine est similaire à celle obtenue à partir des observations subjectives. Dans la section suivante, les prédictions des performances de localisation avec les TCAPS, produites à partir des DTF de la tête ISL ou du KEMAR, sont comparées.

#### 1.4 Comparaison des têtes artificielles

Quelle que soit la tête artificielle utilisée, nous avons trouvé que les performances de localisation sonore étaient le mieux estimées d'un côté par le modèle d'appariement revu  $f(\theta)$ , basé sur la décomposition latérale/polaire et dont les paramètres dépendent de la position de la source, et de l'autre par le réseau de neurones optimisé à partir des observations subjectives (E = 0,01). Dans cette partie, nous comparons les résultats de ces deux modèles en fonction de la tête artificielle utilisée : la tête artificielle ISL ou le mannequin KEMAR. Les différentes métriques développées dans le chapitre précédent sont utilisées :

- la probabilité de localisation voisine  $P_{loc}$ ;
- la probabilité d'erreur avant-arrière  $P_{fb}$ ;
- la probabilité d'erreur grossière  $P_{sev}$ ;

- la probabilité d'erreurs gauche-droite et mixte  $P_{lr}$ ;
- la précision de localisation  $P_{acc}$ ;
- la vraisemblance des observations.

#### 1.4.1 Modèle d'appariement

La comparaison des deux têtes artificielles à travers les six métriques pour le modèle d'appariement est regroupée sur la Figure V.7. On remarque tout d'abord que les prédictions de la probabilité de localisation voisine sont similaires, surtout pour P1, P2, P4, et P6 (cf. Figure V.7a). Pour ces TCAPS, la différence entre les deux têtes ne dépasse pas 0,05. En condition d'écoute normale,  $P_{loc}$  est sous-estimée avec le mannequin KEMAR. En revanche, sa réduction substantielle (et non-réaliste) avec P3 et P5, trouvée avec la tête ISL, n'est pas observée avec cette nouvelle tête.

Les observations précédentes se reflètent sur la probabilité d'erreur avant-arrière sur la Figure V.7b. Avec la tête KEMAR,  $P_{fb}$  est légèrement sous-estimée avec P6, mais la différence avec la valeur moyenne subjective est inférieure à l'écart-type. Pour les autres TCAPS, quelle que soit la tête utilisée, la probabilité d'erreur avant-arrière est systématiquement surestimée. Toutefois, les différences avec les valeurs moyennes subjectives restent petites, excepté avec P3 et P5 pour la tête ISL.

La Figure V.7c présente la probabilité d'erreur grossière prédite par le modèle d'appariement. Elle ne dépasse jamais 0,02 sauf dans le cas de P6 avec le mannequin KEMAR où elle atteint 0,15. Les erreurs grossières sont la deuxième source d'erreurs de localisation dans l'expérience subjective du Chapitre III et leur nombre augmente avec les TCAPS, principalement avec les casques talk-through. Le manque de prédiction de ce type d'erreur par le modèle d'appariement peut conduire à une réduction de la vraisemblance.

Dans l'expérience subjective, le nombre d'erreurs gauche-droite ou mixtes commis par les auditeurs est jugé négligeable. S'il semble que le modèle d'appariement en produit pour les systèmes talk-through, la probabilité d'erreur gauche-droite ne dépasse pas 0,05. Enfin, la plus grande différence entre les deux têtes artificielles est observée avec le bouchon talkthrough P4, pour lequel la différence de  $P_{lr}$  entre les deux têtes est de 0,04. Nous pouvons donc considérer que le modèle d'appariement prédit correctement la probabilité d'erreur gauche-droite, indépendamment de la tête artificielle utilisée.

La précision de localisation représente la proportion de localisations exactes parmi les localisations voisines (à plus ou moins 22,5°). La Figure V.7e montre que le modèle d'appariement optimisé à partir du mannequin KEMAR prédit plus précisément la précision  $P_{acc}$ que celui de la tête ISL. En effet, la différence moyenne avec les résultats de l'expérience subjective est inférieure à 0,04 avec le KEMAR, et dépasse 0,08 avec la tête ISL. Toutefois, avec cette dernière, la surestimation de la précision de localisation ne dépasse que très peu l'écart-type de l'expérience subjective.

Finalement, une mesure globale de l'accord entre les distributions de probabilités pré-



FIG. V.7 – Probabilités de localisation voisine, d'erreurs avant-arrière, grossières, gauche-droite, précision de localisation et vraisemblance estimées dans chaque condition d'écoute, pour l'expérience subjective et les modèles d'appariement obtenus avec la tête ISL et le mannequin KEMAR.

dites par le modèle d'appariement et les observations subjectives est donnée par la vraisemblance (ou log-vraisemblance dans notre cas), représentée en Figure V.7f. Les deux têtes artificielles permettent d'obtenir des vraisemblances similaires en condition d'écoute normale et avec les bouchons non linéaires. Avec les systèmes talk-through (excepté P6), la plus grande vraisemblance est observée pour le mannequin KEMAR. Avec P3 et P5, la mauvaise prédiction des performances de localisation pour la tête ISL en est la cause. En revanche, pour P4, la différence de vraisemblance est étonnante, car les deux têtes présentent des probabilités de localisation voisine et d'erreurs proches dans cette condition. Avec le casque talk-through P6, la sous-estimation des erreurs avant-arrière et la surestimation des erreurs grossières par le mannequin KEMAR provoquent une réduction de la vraisemblance par rapport à la tête ISL.

Pour conclure, dans le chapitre précédent, nous avons observé une importante dégradation des prédictions du modèle d'appariement (revu  $f(\theta)$ ) pour les TCAPS P3 et P5. Elle a été attribuée au fait que les DTF mesurées sur la tête artificielle ISL avec ces systèmes n'étaient pas adaptées à la modélisation. On peut alors remarquer qu'excepté pour ces deux TCAPS, le modèle d'appariement prédit des performances de localisation sonore équivalentes avec les deux têtes artificielles. Les matrices de probabilités prédites par le modèle d'appariement dans chaque condition d'écoute sont données en Annexe E pour la tête ISL et le mannequin KEMAR, respectivement sur les Figures E.1 et E.2.

#### 1.4.2 Réseau de neurones

Le réseau de neurones entraîné à partir des résultats de l'expérience subjective en condition d'oreille nue, avec un seuil d'apprentissage faible (E = 0.01), a été appliqué à la tête artificielle ISL et au mannequin KEMAR. La Figure V.8 représente les probabilités de localisation voisine et d'erreurs, la précision et la vraisemblance des réseaux de neurones des deux têtes. Sur le premier graphique, on remarque que  $P_{loc}$  n'est pas beaucoup affectée par l'utilisation d'une tête artificielle différente. La probabilité de localisation voisine avec les bouchons talk-through est plus élevée pour la tête ISL (différence de 0,1 pour P4), mais le contraire est observé pour les casques talk-through. En condition d'oreille libre et avec les bouchons non linéaires, la différence entre les deux têtes ne dépasse pas 0,06. Dans le Chapitre IV, nous avons vu que le réseau de neurones surestime la dégradation des performances de localisation avec les casques talk-through. Ce résultat est toujours obtenu sur la Figure V.8a, que ce soit avec la tête ISL ou le mannequin KEMAR. Avec ce dernier, les différences entre les probabilités de localisation voisine prédites avec P3 et P4 et les valeurs moyennes de l'expérience subjective sont supérieures aux écarts-types subjectifs. Pour finir, la légère surestimation de  $P_{loc}$  par le réseau de neurones avec P2, observée précédemment sur la tête ISL, est aussi présente pour le mannequin KEMAR.

La Figure V.8b représente la probabilité d'erreur avant-arrière prédite par le réseau de neurones. Les deux têtes artificielles conduisent à des résultats très similaires : la différence



FIG. V.8 – Probabilités de localisation voisine, d'erreurs avant-arrière, grossières, gauche-droite, précision de localisation et vraisemblance estimées dans chaque condition d'écoute, pour l'expérience subjective et les réseaux de neurones obtenus avec la tête ISL et le mannequin KEMAR.

maximale de  $P_{fb}$  entre les deux vaut 0,11 et est atteinte avec P4. Comme pour la probabilité de localisation voisine, la probabilité d'erreur avant-arrière est moins bien prédite pour les systèmes talk-through. En particulier, pour le mannequin KEMAR, la différence de  $P_{fb}$  est la moyenne subjective dépasse l'écart-type subjectif avec P3, P4 et P5.

Concernant les probabilités d'erreurs grossières et gauche-droite, respectivement représentée sur les Figures V.8c et V.8d, les mêmes prédictions sont réalisées par le réseau de neurones pour les deux têtes artificielles. La différence entre les deux têtes ne dépasse effectivement pas 0,04 pour  $P_{sev}$  et 0,02 pour  $P_{lr}$ . Les observations faites dans le Chapitre IV, Section 3 sont toujours valables : le réseau de neurones prédit correctement la probabilité d'erreur grossière dans toutes les conditions d'écoute, sauf avec P6, et des erreurs gauche-droite sont présentes pour P3, P5 et P6.

La précision de localisation représentée en Figure V.8e est systématiquement surestimée par le réseau de neurones, mais les différences avec les moyennes subjectives sont inférieures aux écarts-types correspondants. Si pour les casques talk-through P5 et P6, la tête ISL et le mannequin KEMAR donnent des résultats légèrement différents, dans toutes les autres conditions d'écoute, la précision  $P_{acc}$  estimée par le réseau de neurones est indépendante de la tête artificielle utilisée.

Pour finir, le calcul de la log-vraisemblance permet d'évaluer la concordance entre les prédictions du réseau de neurones et les observations de l'expérience subjective. D'après la Figure V.8f, la tête ISL permet d'obtenir une meilleure vraisemblance que le mannequin KEMAR avec les TCAPS P3 et P5. Le contraire est observé pour le casque talk-through P6. Dans les autres conditions d'écoute, les deux têtes entraînent une vraisemblance équivalente. De plus, elles présentent une évolution similaire de la vraisemblance en fonction du type de TCAPS.

Comme dans le cas du modèle d'appariement, nous pouvons conclure que l'utilisation de DTF mesurées sur deux têtes artificielles différentes ne conduit pas à des différences substantielles de prédiction des performances de localisation par le réseau de neurones. Les matrices de probabilités prédites par le réseau de neurones dans chaque condition d'écoute sont données en Annexe E pour la tête ISL et le mannequin KEMAR, respectivement sur les Figures E.3 et E.4.

# 1.5 Conclusion sur l'indépendance des modèles à la tête artificielle

Nous avons étudié les prédictions faites par les deux modèles de localisation sonore à partir de DTF mesurées sur deux têtes artificielles différentes. Pour le modèle d'appariement, les principales différences de performance entre les deux têtes sont observées avec les TCAPS P3 et P5. Nous avions remarqué dans le chapitre précédent que les DTF mesurées sur la tête ISL avec ces protections ne sont pas adaptées à la modélisation. Les prédictions apportées à partir du mannequin KEMAR avec P3 et P5 sont bien meilleures. La plus grande résolution fréquentielle du protocole de mesure du CNAM a dû permettre

l'ajout d'indices spectraux non captés par la résolution logarithmique du protocole ISL. Concernant le réseau de neurones, nous avons jugé que l'utilisation d'une tête artificielle différente ne provoque des modifications notables des prédictions ni en condition d'oreille nue ni avec les TCAPS. De plus, chaque configuration modèle/tête permet de retrouver la classification des TCAPS par type résultant de l'expérience subjective de localisation sonore.

En conclusion, si les HRTF et les performances de localisation sonore dépendent de l'auditeur, nous avons tout de même pu développer des modèles qui prédisent les performances globales obtenues au terme d'une expérience subjective, et qui restent indépendants de la tête artificielle utilisée. D'ailleurs, on remarque que les prédictions avec le mannequin KEMAR sont correctes, alors que les DTF utilisées ont été mesurées dans une cabine différente de celle dans laquelle l'expérience subjective s'est déroulée.

# 2 Applications

Le principal objectif des modèles de prédiction développés dans nos travaux est d'offrir une méthode d'évaluation de prototypes de TCAPS du point de vue de la localisation sonore, sans nécessiter d'expérience subjective. Pour donner un exemple d'une telle application, nous avons tenté d'améliorer les performances de localisation du casque MSA Supreme<sup>®</sup> Pro (P6). Cette amélioration est décrite dans la première section de cette partie. Dans un second temps, nous étudions l'influence des bouchons d'oreille à « atténuation plate », particulièrement adaptés pour les musiciens, sur les performances de localisation. Pour finir, les deux modèles de prédiction sont appliqués à l'étude de prothèses auditives.

# 2.1 Amélioration des performances d'un casque talk-through

Dans les Chapitres III et IV, nous avons montré que, comparé aux autres TCAPS, le port de casques talk-through engendrait la plus sévère dégradation des performances de localisation sonore. La détérioration des indices spectraux en est la principale cause et conduit à une augmentation substantielle des confusions avant-arrière. En effet, l'ITD est conservée et reste prépondérante pour la localisation de sources à large bande par rapport à l'ILD compressée par les casques talk-through (cf. Chapitre II, Section 2.4). L'amélioration des performances de localisation d'un casque passe donc par la récupération d'indices spectraux permettant la distinction avant-arrière.

## 2.1.1 Principe

La perte des indices spectraux utiles engendrée par le port d'un casque est due au fait que le pavillon auditif soit complètement recouvert. Toutefois, par rapport à un casque conventionnel, un casque talk-through a l'avantage de posséder un « point de mesure » de l'onde sonore incidente : son microphone d'entrée. Ce capteur est généralement omnidirectionnel pour pouvoir capter la plus grande majorité des sons environnants. Sur les



FIG. V.9 – Casque talk-through MSA Supreme<sup> $\circ$ </sup> Pro : (a) modèle modifié avec un pavillon fermé, (b) modèle modifié avec un pavillon ouvert.

casques talk-through P5 et P6 étudiés précédemment, le microphone est placé à l'avant de la coque, contrariant légèrement son caractère omnidirectionnel. Toutefois, cet effet ne crée pas d'indices spectraux exploitables par l'auditeur pour distinguer les sources avant et arrière.

Notre tentative d'amélioration d'un casque talk-through est basée sur la création d'indices spectraux avant-arrière, en modifiant la forme de la coque du casque. L'idée proposée par Rubak & Johansen (1999) est de rendre une forme de pavillon auditif au casque. Dans notre cas, nous ajoutons un modèle simplifié de pavillon de tête artificielle au niveau du microphone d'entrée.

# 2.1.2 Matériel

Dans cette étude, nous utilisons le casque talk-through MSA Supreme<sup>®</sup> Pro, noté P6. Il a été choisi, car sa géométrie se prête plus facilement à l'application des modifications que le casque P5. La Figure V.9 présente les modifications apportées au casque : un pavillon simplifié de tête artificielle ISL est imprimé et ajusté sur la coque. Deux pavillons différents sont créés :

- le premier est fermé devant le microphone pour que l'onde atteigne ce dernier par le faux conduit auditif;
- le second est ouvert, il ne cache pas le microphone d'entrée.

Les mesures d'HRTF avec le casque P6 modifié ont été réalisées avec la tête et le dispositif expérimental de l'ISL (Chapitre II, Section 2.1.1). Les performances de localisation sont évaluées grâce au modèle d'appariement et au réseau de neurones.

Modèle	Métrique	Sans	P6 normal	P6 fermé	P6 ouvert
Appariement	$P_{loc}$	0.96	0.6	0.4	0.67
	$P_{sev}$	0	0.02	0.21	0.01
	$P_{fb}$	0.04	0.33	0.34	0.27
	$P_{lr}$	0	0.05	0.05	0.05
	$P_{acc}$	0.64	0.47	0.37	0.54
Réseau de neurones	$P_{loc}$	0.91	0.34	0.24	0.33
	$P_{sev}$	0.03	0.18	0.32	0.18
	$P_{fb}$	0.05	0.37	0.35	0.4
	$\dot{P_{lr}}$	0.01	0.11	0.09	0.09
	$P_{acc}$	0.62	0.51	0.47	0.52

TAB. V.2 – Métriques prédites par le modèle d'appariement et le réseau de neurones dans le cadre de l'amélioration du casque P6.

# 2.1.3 Résultats

Le Tableau V.2 répertorie les cinq métriques  $P_{loc}$ ,  $P_{sev}$ ,  $P_{fb}$ ,  $P_{lr}$  et  $P_{acc}$  calculées en condition d'écoute normale et pour le casque P6, avec et sans les pavillons. Les résultats de la colonne P6 normal ne correspondent pas exactement aux valeurs obtenues dans le chapitre précédent (cf. Figure IV.25a pour  $P_{loc}$ ), car les mesures d'HRTF ont aussi été refaites dans cette condition. Pour le modèle d'appariement, l'effet du pavillon fermé sur les performances de localisation avec le casque se révèle négatif : la probabilité de localisation voisine est réduite de 0,2 à cause d'une augmentation équivalente de la probabilité d'erreur grossière. Cette dernière engendre aussi une baisse de la précision de localisation. Des observations similaires peuvent être faites concernant les prédictions du réseau de neurones.

Les prédictions des deux modèles de localisation diffèrent en revanche pour le pavillon ouvert. Le réseau de neurones prédit des performances équivalentes à celles du casque sans modification. On peut toutefois noter une légère augmentation de la probabilité d'erreur avant-arrière et une réduction de la probabilité d'erreur gauche-droite. Les performances prédites avec le pavillon ouvert par le modèle d'appariement sont meilleures qu'avec le casque non modifié :  $P_{loc}$  est augmenté de 0,07 par rapport à ce cas. Cette amélioration provient de la diminution de  $P_{fb}$  : le pavillon ouvert contribue bien à la création d'indices spectraux avant-arrière. Pour finir, une meilleure précision de localisation  $P_{acc}$  est obtenue avec le pavillon ouvert.

Dans le Chapitre IV, nous avons observé que le réseau de neurones ne permet pas une prédiction correcte des performances de localisation avec les casques. Dans cette section, nous avons tout de même étudié ses prédictions, car l'évolution des dégradations était respectée avec les autres TCAPS. Il semble finalement que la dégradation de la localisation sonore engendrée par un casque talk-through, et prédite par le réseau de neurones, soit trop importante pour pouvoir remarquer une amélioration des indices spectraux. Les résultats du modèle d'appariement sont par conséquent gardés, puisqu'ils ont été jugés réalistes avec les casques talk-through (au moins avec P6 pour la tête artificielle ISL).
L'ajout d'un pavillon à la coquille du casque MSA Supreme<sup>®</sup> Pro a permis d'améliorer la discrimination avant-arrière, mais seulement dans le cas du pavillon ouvert. Le pavillon fermé n'est pas assez évolué pour simuler le conduit auditif de manière effective. De plus, il recouvre le microphone d'entrée de façon trop importante. Le pavillon ouvert ne cache pas le microphone et ne crée aucun conduit auditif. Néanmoins, il entraîne des différences entre les DTF avant et arrière qui ne peuvent pas être obtenues lorsque seule la coquille du casque altère le caractère omnidirectionnel du microphone.

### 2.1.4 Conclusion sur l'amélioration du casque talk-through

L'ajout d'indices spectraux avant-arrière par la mise en place d'un faux pavillon d'oreille ouvert permet d'augmenter les performances de localisation sonore, d'après le modèle d'appariement. Cette évolution devrait toutefois être confirmée par une expérience subjective. En effet, les modèles donnent les résultats pour un auditeur idéal, simulé par une tête artificielle. L'amélioration des performances constatée ici pourrait être « noyée » dans la variabilité interindividuelle, et donc être non significative. Le rôle des méthodes de prédiction des performances de localisation est toutefois renforcé : ils permettent une évaluation rapide des modifications apportées dans le cadre du développement d'un prototype.

L'expérience présentée dans cette section ne sert que d'exemple d'applications des modèles. L'application d'un modèle de pavillon à la coque d'un casque n'est pas adaptée, non seulement d'un point de vue esthétique, mais aussi pratique. Les systèmes talk-through étudiés P5 et P6 ont effectivement une forme qui permet de les porter sous un casque tactique. Cette possibilité serait empêchée par notre solution. L'idée de Rubak & Johansen (1999) semble plus faisable. À la place de ces modifications physiques, l'amélioration des systèmes talk-through, bouchon ou casque, pourrait plutôt être réalisée dans le traitement électronique du signal. Par exemple, une méthode d'optimisation de l'ILD avec les casques est proposée dans (Sällberg et al., 2007). En effet, nous avons montré dans le Chapitre II et dans (Joubaud et al., 2015) que l'ILD est dégradée par ces systèmes. Cette amélioration pourrait augmenter les performances de localisation de sources haute fréquence (ILD prépondérante). Pour finir, nous pouvons imaginer un système de restitution de l'environnement sonore en 3D, lié à un algorithme de détection de la position de la source. L'utilisation d'HRTF individualisées permettrait aussi de meilleures performances.

### 2.2 Étude des performances de bouchons à atténuation uniforme

Dans cette section, une nouvelle application des modèles de prédiction est proposée. Les performances de localisation observées précédemment avec les bouchons non linéaires P1 et P2 sont comparées à celles prédites par les modèles avec des bouchons d'oreille à atténuation uniforme.



(a) Bouchon ETY

(b) Bouchon ER20

FIG. V.10 – Bouchons d'oreille à atténuation uniforme : (a) Etymotic ETY•Plugs<sup>®</sup>, (b) Etymotic ER•20<sup>®</sup> XS.

### 2.2.1 Description des protecteurs

Sur le plan horizontal, la dégradation des performances de localisation engendrée par les protections auditives résulte majoritairement de la détérioration des indices spectraux. En effet, au filtrage directionnel réalisé par l'oreille externe s'ajoute l'effet du protecteur sur l'onde incidente. Les indices spectraux pourraient théoriquement être conservés si la protection auditive utilisée avait une atténuation uniforme, c'est-à-dire constante avec la fréquence. Ainsi, l'unique effet du protecteur serait d'atténuer le niveau reçu au tympan, sans altérer les indices de localisation.

L'atténuation uniforme d'une protection auditive peut être obtenue par traitement du signal dans le cas des systèmes actifs. Certains modèles récents de protection auditive offrent cette option. Les premiers protecteurs avec une telle propriété sont toutefois passifs : un filtre combinant diaphragme et résonateur acoustiques reproduit la fonction de transfert de l'oreille externe, en particulier la fréquence de résonance du conduit auditif à environ 2,7 kHz (Killion et al., 1988; Niquette, 2009). Ce type de protection auditive a été développé pour répondre au besoin des musiciens de conserver la qualité du son d'un point de vue fréquentiel, tout en étant protégé (Patel, 2008).

Les bouchons d'oreille à atténuation uniforme utilisés dans cette section sont des bouchons prémoulés à triple ailette, disponibles dans le commerce :

— le bouchon Etymotic ETY•Plugs<sup>®</sup>, simplement noté ETY dans la suite;

— le bouchon Etymotic  $\text{ER} \bullet 20^{\circ}$  XS, simplement noté ER20 dans la suite.

Une photographie de ces bouchons est donnée en Figure V.10. D'après le fabricant, la principale différence entre ces bouchons réside dans l'amélioration de la clarté à haute fréquence pour le bouchon ER20.

Les HRTF sur le plan horizontal ont été mesurées, avec les deux bouchons à atténuation uniforme, grâce au dispositif expérimental du CNAM, et donc avec le mannequin KEMAR. Le modèle d'appariement et le réseau de neurones, optimisés à partir des mesures en condition d'oreille libre sur la même tête, sont appliqués pour évaluer l'effet de

Modèle	Métrique	Sans	P1	$\mathbf{P2}$	ETY	<b>ER20</b>
Appariement	$P_{loc}$	0.88	0.76	0.78	0.86	0.79
	$P_{sev}$	0	0.02	0.02	0.01	0.02
	$P_{fb}$	0.12	0.2	0.18	0.13	0.17
	$P_{lr}$	0	0.02	0.02	0	0.02
	$P_{acc}$	0.62	0.61	0.59	0.62	0.61
Réseau de neurones	$P_{loc}$	0.91	0.8	0.85	0.82	0.84
	$P_{sev}$	0.03	0.04	0.05	0.04	0.04
	$P_{fb}$	0.05	0.14	0.08	0.13	0.11
	$P_{lr}$	0.01	0.02	0.02	0.01	0.01
	$P_{acc}$	0.63	0.62	0.61	0.62	0.66

TAB. V.3 – Comparaison des métriques prédites par le modèle d'appariement et le réseau de neurones avec différents bouchons d'oreille passifs.

ces protections auditives sur les performances de localisation sonore.

### 2.2.2 Résultats

Le Tableau V.3 regroupe les prédictions des deux modèles de localisation sans protection et avec les bouchons P1, P2, ETY et ER20. Les six métriques  $P_{loc}$ ,  $P_{sev}$ ,  $P_{fb}$ ,  $P_{lr}$  et  $P_{acc}$  ont été calculées. Pour le modèle d'appariement, la probabilité de localisation voisine est particulièrement améliorée avec le bouchon ETY, puisque la  $P_{loc}$  correspondante est très proche de celle de la condition d'écoute normale. Avec le bouchon ER20, on observe seulement une très légère augmentation de la métrique par rapport aux bouchons non linéaires. L'optimisation des performances par les bouchons à atténuation uniforme résulte majoritairement de la réduction des erreurs avant-arrière. En effet, les métriques  $P_{sev}$  et  $P_{lr}$  sont très faibles dans toutes les conditions d'écoute considérées. La précision de localisation avec les bouchons reste aussi très stable et équivalente à celle observée en condition d'oreille nue.

Les résultats obtenus avec le réseau de neurones sont différents. Si les bouchons ETY et ER20 entraînent une amélioration des performances par rapport à P1, le bouchon non linéaire P2 reste le plus proche des performances sans protection. L'évolution de la probabilité d'erreur avant-arrière reflète celle de la probabilité de localisation voisine. Encore une fois, les probabilités d'erreur grossière et gauche-droite sont faibles et plutôt constantes en fonction de la condition d'écoute. Il en est de même pour  $P_{acc}$ , excepté qu'une légère augmentation est observée avec le bouchon ER20.

Dans le Chapitre IV, nous avons observé que le réseau de neurones prédit une différence notable de performance de localisation entre les deux bouchons non linéaires P1 et P2. L'expérience subjective n'a toutefois pas permis de retrouver cette différence en prenant la moyenne sur tous les auditeurs. Cet effet propre au réseau de neurones pourrait empêcher l'observation de l'amélioration apportée par une atténuation uniforme.

#### 2.2.3 Conclusion sur les bouchons à atténuation uniforme

D'après les deux modèles de prédiction, les bouchons ETY et ER20 entraînent des performances de localisation légèrement meilleures que celles du bouchon non linéaire P1. Le modèle d'appariement donne la même conclusion avec P2, mais pas le réseau de neurones. Seule une expérience subjective pourrait confirmer lequel des deux modèles prédit la bonne tendance. Dans (Gallagher et al., 2014), le bouchon ETY est comparé à d'autres bouchons d'oreilles, dont des modèles équivalents à P2, dans une expérience de localisation sonore dans une sphère. La supériorité du bouchon ETY pour la localisation est nettement observable dans cette étude. Il semble donc que, dans notre cas, la prédiction la plus réaliste provient du modèle d'appariement. Cependant, les différences entre P1, P2, ETY et ER20 ne sont pas très importantes. L'amélioration des performances de localisation grâce aux bouchons à atténuation uniforme est donc plus marquée lorsque l'élévation est prise en compte, comme dans (Gallagher et al., 2014).

# 2.3 Prédiction de la dégradation des performances de localisation créée par des aides auditives

La majeure partie de nos travaux se focalise sur les protections auditives. Le cas des aides auditives, ou prothèses, peut aussi être étudié de façon similaire. En effet, on peut grossièrement considérer ces aides comme des bouchons d'oreille dont le rôle est plus d'amplifier le signal incident que de l'atténuer. La conservation de la capacité à localiser les sources sonores lors du port de prothèses est importante, puisqu'elle est non seulement liée au confort du patient appareillé, mais aussi à sa sécurité. Dans cette section, les HRTF mesurées avec deux systèmes différents d'aide auditive sont étudiées, puis les modèles de prédiction sont appliqués aux indices de localisation correspondants.

### 2.3.1 Contexte et description des prothèses

La capacité d'un auditeur à localiser une source sonore dans le plan horizontal dépend fortement de sa possibilité d'exploiter les ITD et ILD. Même si Slattery & Middlebrooks (1994) ont démontré que certains patients atteints de surdité unilatérale pouvaient utiliser les indices spectraux monauraux pour percevoir l'angle latéral, nous nous plaçons dans le cas d'une surdité bilatérale. De plus, puisque des mesures sur tête artificielle sont étudiées, la surdité est supposée parfaitement symétrique. Ces hypothèses permettent de se concentrer uniquement sur l'évaluation de l'effet des aides auditives sur les performances de localisation sonore.

Les prothèses auditives sont souvent classées en fonction de leur forme et de la position du microphone. On distingue cinq catégories :

1. Invisible dans le canal (*invisible-in-the-canal*, IIC) : placée profondément dans le conduit auditif, quasiment invisible à l'oeil nu;

- 2. Complètement dans le canal (*completely-in-the-canal*, CIC) : placée dans le conduit auditif;
- 3. Dans le canal (*in-the-canal*, ITC) : placée à l'entrée du conduit auditif;
- 4. Dans l'oreille (*in-the-ear*, ITE) : placée dans la conque;
- 5. Derrière l'oreille (*behind-the-ear*, BTE) : placée derrière l'oreille, le microphone est en général au-dessus du pavillon.

L'influence de la forme de ces divers systèmes sur les performances de localisation est étudiée dans (Durin et al., 2014) sans y inclure l'effet du traitement du signal par le dispositif. Les aides auditives de types IIC et CIC dégradent très peu les performances, car le filtrage directionnel de l'oreille externe est conservé. Celui-ci est légèrement dégradé pour les types ITD et ITE, mais devient absent pour les systèmes BTE. En effet, le point de mesure se situe au-dessus du pavillon, supprimant ainsi les indices spectraux. Lorsque des auditeurs appareillés sont testés, le même type de performances de localisation est obtenu (Best et al., 2010; Van den Bogaert et al., 2011).

Dans notre étude, deux prothèses de type BTE, que nous notons A1 et A2, sont évaluées. Pour pallier le défaut de positionnement du microphone, des traitements du signal reçu sont disponibles :

- la communication binaurale : adaptation binaurale de la compression appliquée au signal, dans le but de préserver l'ILD;
- le mode directionnel : reproduction du filtrage directionnelle réalisé par l'oreille externe.

Les mesures d'HRTF avec ces prothèses ont été réalisées au CNAM dans le cadre d'un stage audioprothésiste. Les systèmes d'amplification de A1 et A2 ont été ajustés pour produire un gain unitaire (le mannequin KEMAR n'a pas de pertes auditives). Avec les deux prothèses, les HRTF sont d'abord mesurées quand la communication binaurale et le mode directionnel fonctionnent. Ensuite, pour A2, les deux fonctions sont désactivées. Concernant A1, seule la communication binaurale peut être éteinte.

Les mesures sont conduites sur le plan horizontal, pour des sources sonores placées tous les 45°, de 0° à 315°. La fréquence de mesure maximale a été fixée à environ 8 kHz. Dans les sections suivantes, nous analysons les indices de localisation produits par les aides auditives, puis les deux modèles de prédiction sont appliqués aux HRTF mesurées.

### 2.3.2 Différence interaurale de temps

Les Figures V.11a et V.11b représentent l'ITD estimée en condition d'écoute normale et avec les prothèses A1 et A2, respectivement. Sur la première, on remarque que le système A1 n'influence pas l'ITD. La différence maximale par rapport à l'ITD en condition d'oreille nue ne vaut que 60 µs, et est atteinte sur l'axe interaural. En revanche, sur la Figure V.11b, l'ITD est diminuée lorsque le système A2 est éteint, principalement pour les positions proches de l'axe interaural. Sur ce dernier, la différence d'ITD avec la condition d'écoute



FIG. V.11 – Différence interaurale de temps estimée avec et sans audioprothèses.

normale atteint 180 µs. Cette dégradation de l'ITD est néanmoins résolue par l'allumage de l'appareil.

Les prothèses auditives de type BTE semblent très peu altérer la différence interaurale de temps, malgré le placement du microphone au-dessus du pavillon de l'oreille. Toutefois, la réduction de l'ITD observée dans le cas « A2 Off » révèle un effet négatif de l'extinction du mode directionnel. Cette option, principalement chargée de reproduire les indices spectraux, serait donc aussi bénéfique pour la conservation de l'ITD.

#### 2.3.3 Différence interaurale d'intensité

Les différences interaurales d'intensité estimées avec et sans les audioprothèses sont représentées sur la Figure V.12. Comparée à l'ITD, l'ILD est beaucoup plus affectée par les aides auditives. Avec la prothèse A2, on observe des différences par rapport à la condition d'oreille libre qui atteignent environ 7 dB. La communication binaurale semble améliorer l'ILD, mais on constate une différence entre l'avant et l'arrière. Cet effet est aussi observé en condition d'oreille nue, mais il pourrait être amplifié par le mode directionnel.

Avec la prothèse A1, l'extinction de la communication binaurale entraîne une dégradation de l'ILD plus importante que pour A2. La différence avec la condition d'écoute normale vaut 8,5 dB sur l'axe interaural. L'impact de la prothèse sur l'ILD est néanmoins largement plus important quand la communication binaurale est réactivée. Dans ce cas, la différence interaurale d'intensité est quasiment annulée, et même inversée (elle vaut -1,3 dB à 90°). L'adaptation de la compression pour la conservation de l'ILD ne fonctionne donc pas comme espéré. Il semble plutôt que le système réalise une moyenne des signaux arrivant à chaque prothèse, annulant ainsi l'ILD. La légère inversion observée peut être causée par une augmentation de la compression du signal du côté ipsilatéral.

En conclusion, on peut noter que les prothèses auditives de type BTE altèrent l'ILD de manière significative. La compression du signal a déjà été authentifiée comme la cause de cette détérioration : les signaux les plus forts sont moins amplifiés par le système,



FIG. V.12 – Différence interaurale d'intensité estimée avec et sans audioprothèses.

donc le signal ipsilatéral est atténué par rapport au signal contralatéral. Les fabricants répondent à ce problème par l'introduction d'une communication entre les deux appareils, afin de restituer la différence interaurale d'intensité. Si cette adaptation binaurale de la compression semble efficace avec la prothèse A2, elle ne remplit pas correctement cette fonction pour la prothèse A1.

### 2.3.4 Indices spectraux

Les indices spectraux monauraux permettent de résoudre les ambiguïtés avant/arrière introduites par les différences interaurales. Ils sont estimés comme la différence entre les HRTF avant moyennes et les HRTF arrière moyennes (cf. Chapitre II, Section 2.4.3), et sont donnés sur la Figure V.13 pour les oreilles ipsi- et contralatérales. Pour s'affranchir de la forte précision de mesure (résolution de 1 Hz), les indices spectraux sont filtrés par une banque de 20 gammatones dont la fréquence centrale est répartie de manière logarithmique de 100 Hz à 8 kHz.

Tout d'abord, on remarque que les indices spectraux sont détériorés par la prothèse A1, quel que soit le réglage utilisé. Du côté ipsilatéral, le système directionnel introduit tout de même un pic à 5 kHz qui est aussi observé en condition d'écoute normale. Ce pic est renforcé par l'utilisation de la communication binaurale. Enfin, l'augmentation des indices spectraux ipsilatéraux à partir de 2 kHz est présente, mais moins marquée. Du côté contralatéral, les pics à 800 Hz et 3 kHz sont conservés, et un nouveau émerge à 5 kHz. De plus, les indices spectraux contralatéraux sont amplifiés à basse fréquence.

Toujours d'après la Figure V.13, l'audioprothèse A2 engendre une dégradation substantielle des indices spectraux dans le cas « A2 Off ». Celui-ci correspond en particulier à



FIG. V.13 – Indices spectraux avant-arrière ipsi- et contralatéraux obtenus à partir d'HRTF mesurées avec et sans audioprothèses.

une extinction du mode directionnel et donc à une captation du son sans effet du pavillon de l'oreille. Cela conduit à des différences très peu marquées entre les indices spectraux ipsi- et contralatéraux. Quand le mode directionnel est activé (avec la communication binaurale), A1 conserve correctement les indices ipsilatéraux. En revanche, ceux de l'autre oreille ne sont pas retrouvés. On constate notamment une inversion du pic à 3 kHz.

La simulation du filtrage directionnel opéré par le pavillon auditif permet une meilleure conservation des indices spectraux avant-arrière. Les prothèses de type BTE proposant cette option devraient donc empêcher les confusions entre les sources avant et arrière. En particulier, le système A2 paraît prometteur sur ce point. Dans la suite, les modèles de prédiction développés dans le chapitre précédent, dans le contexte de la protection auditive, sont appliqués aux DTF mesurées avec les prothèses A1 et A2.

### 2.3.5 Application des modèles de localisation

Les DTF mesurées avec les aides auditives ont été réalisées pour huit positions sur le plan horizontal. Pour pouvoir optimiser les modèles de localisation sonore à partir des HRTF en condition d'écoute normale, des données subjectives ou théoriques sont nécessaires. Nous avons choisi d'utiliser les données de l'expérience subjective du Chapitre III, et de recréer le paradigme d'utilisation des sources virtuelles. Les HRTF aux positions correspondantes sont définies comme la moyenne des HRTF aux positions réelles. Les paramètres du modèle d'appariement obtenus dans cette étude sont similaires à ceux trouvés avec le mannequin KEMAR dans la Section 1.2. Cependant, la forte prépondérance de l'ITD par rapport à l'ILD n'est pas retrouvée. En effet, la pondération  $w_{lat}$  est égale à 0,53: l'ITD et l'ILD sont utilisées de manière équilibrée par le modèle d'appariement.

Les modèles de prédiction sont appliqués dans chaque condition d'écoute :

- sans aide auditive;
- avec A1 en mode directionnel et avec communication binaurale : « A1 On » ;
- avec A1 en mode directionnel seul : « A1 Off »;
- avec A2 en mode directionnel et avec communication binaurale : « A2 On » ;
- avec A2 sans option : « A2 Off ».

La Figure V.14 représente les métriques  $P_{loc}$ ,  $P_{sev}$ ,  $P_{fb}$ ,  $P_{lr}$  et  $P_{acc}$  obtenues dans ces conditions avec les deux modèles. Tout d'abord, on remarque que le modèle d'appariement et le réseau de neurones prédisent des performances analogues dans toutes les conditions d'écoute. La probabilité de localisation voisine dévoile une très importante diminution des performances avec la prothèse A1, ainsi que dans le cas « A2 Off ». Avec la prothèse A2 en mode directionnel et avec la communication binaurale,  $P_{loc}$  paraît mieux conservée, malgré une différence de 0,2 par rapport à la condition d'écoute normale. Le modèle d'appariement et le réseau de neurones prédisent des évolutions semblables de  $P_{loc}$ . Notamment, ils trouvent que les deux prothèses en mode Off engendrent des performances analogues. En revanche, la différence de  $P_{loc}$  entre les cas « A1 On » et « A1 Off » diffère entre les



FIG. V.14 – Métriques  $P_{loc}$ ,  $P_{sev}$ ,  $P_{fb}$ ,  $P_{lr}$  et  $P_{acc}$  calculées avec et sans prothèses auditives, et obtenues par le modèle d'appariement et le réseau de neurones.

modèles. Elle ne dépasse pas 0,1 selon le réseau de neurones, et vaut 0,29 pour le modèle d'appariement.

L'évolution de la probabilité de localisation voisine est liée à celle de la probabilité d'erreur avant-arrière. La valeur de  $P_{fb}$  la plus proche de la condition d'oreille libre est obtenue avec « A2 On ». La forte différence entre cette condition et le cas où les options de A2 sont éteintes montre l'efficacité de son mode directionnel. Pour A1, la probabilité d'erreur avant-arrière est peu modifiée par l'utilisation ou non de la communication binaurale, particulièrement selon le réseau de neurones. Enfin, les erreurs avant-arrière seules ne permettent pas d'expliquer l'impact négatif des prothèses sur les performances de localisation sonore.

La probabilité d'erreur grossière prédite par les modèles, quelle que soit la prothèse utilisée, est supérieure à celle trouvée précédemment pour les casques talk-through. On remarque que  $P_{sev}$  est plus importante avec la prothèse A1 qu'avec A2. L'utilisation de la communication binaurale ou du mode directionnel n'a pas d'effet sur la probabilité d'erreur grossière. Concernant les erreurs gauche-droite, le modèle d'appariement prédit une  $P_{lr}$  de 0,15 pour « A1 On » et de moins de 0,02 pour les autres conditions. Avec le réseau de neurones, le maximum de  $P_{lr}$  est atteint dans la même condition, et elle reste toujours supérieure à 0,04 avec les prothèses. La valeur la plus proche de la condition d'oreille libre est obtenue pour « A2 On ». Ces résultats reflètent la mauvaise communication binaurale de la prothèse A1, qui annule l'ILD.

La précision de localisation en condition d'écoute normale vaut 0,63 pour les deux modèles. Elle représente la proportion de localisations exactes parmi les localisations voisines. Avec les aides auditives, le modèle d'appariement prédit une  $P_{acc}$  d'environ 0,4 pour chaque prothèse. Le réseau de neurones mène à une précision proche de la condition d'écoute normale pour « A1 Off » et « A2 On », mais à des diminutions de  $P_{acc}$  de 0,18 et 0,14 pour « A1 On » et « A2 Off », respectivement.

#### 2.3.6 Conclusion sur les prothèses auditives

La prédiction des performances de localisation sonore avec A1 et A2, prédite par le modèle d'appariement et le réseau de neurones, montre la conséquence du positionnement du microphone au-dessus du pavillon pour les prothèses de type BTE. La disparition des indices spectraux engendre une augmentation substantielle des confusions avant-arrière. L'application d'un mode directionnel, censé reproduire ces indices, permet de réduire considérablement ces confusions pour la prothèse A2. Avec le système A1, ce mode est toujours en marche, mais il est moins efficace. L'observation des indices spectraux avant-arrière prévoyait déjà cette conclusion. Les valeurs de  $P_{fb}$  observées sont similaires à celles trouvées précédemment avec des casques talk-through, ce qui reste tout de même surprenant. Le manque de mesure à haute fréquence (pas de données au-delà de 8 kHz) pourrait donc être critiqué. Les audioprothèses récentes présentent effectivement des fréquences de coupure étendues (environ 12 kHz). Les modèles de localisation devraient prédire une meilleure résolution des ambiguïtés avant-arrière à partir de mesures à plus hautes fréquences.

Dans le but de préserver l'ILD, les prothèses de type BTE proposent aussi un système de communication binaurale qui adapte la compression du signal entre les deux oreilles. Cette option ne fonctionne pas avec A1. Au contraire, la différence interaurale d'intensité est annulée. Ceci conduit à une augmentation des erreurs gauche-droite. Un défaut de réglage pourrait être la cause de cette dégradation des performances, car peu d'erreurs gauche-droite ont été observées au cours de l'expérience subjective avec aide menée dans le cadre du stage audioprothésiste. Pour finir, la dégradation de l'ILD, qui reste conséquente dans toutes les conditions d'écoute, conduit aussi à la prédiction d'importantes probabilités d'erreur grossière. La principale cause de cette dégradation réside dans la pondération équilibrée de l'ITD et de l'ILD utilisée par les deux modèles. Dans le cas de sources sonores à large bande fréquentielle, la différence interaurale de temps devrait être prépondérante.

L'application du modèle d'appariement et du réseau de neurones montre l'effet des aides auditives sur les performances de localisation sonore. La dégradation observée est obtenue par comparaison avec une condition d'écoute normale, c'est-à-dire sans perte auditive. Dans le cas réel d'un patient atteint de surdité et appareillé, la prothèse auditive peut engendrer une amélioration de sa capacité à localiser les sources sonores. L'expérience subjective, conduite au cours du stage audioprothésiste avec les prothèses étudiées ici, tend vers cette conclusion. Toutefois, il a déjà été observé que d'autres aides auditives peuvent avoir un impact négatif sur les performances de l'auditeur, comparé à la condition non appareillée (Van den Bogaert et al., 2006).

## 2.4 Conclusion sur la validation des modèles de prédiction

Dans ce chapitre, nous avons montré que la prédiction des performances de localisation sonore avec le modèle d'appariement et le réseau de neurones ne dépend pas de la tête artificielle utilisée pour les mesures d'HRTF. Le choix de la prédiction par des modèles de localisation est donc validé par rapport à la réalisation d'une expérience subjective très chronophage. De plus, l'étude de l'impact global des protections auditives sur la capacité à localiser les sources sonores ne nécessite pas de connaître les HRTF individuelles des auditeurs.

Afin de consolider la légitimité des modèles de prédiction, trois applications différentes ont été proposées. Malgré quelques légères différences, nous avons tout de même pu constater que le modèle d'appariement et le réseau de neurones prédisent des performances de localisation semblables. Tout d'abord, nous avons montré que l'apport physique d'indices spectraux avant-arrière, par le placement d'un pavillon d'oreille factice, pouvait améliorer les performances de localisation avec les casques talk-through. Ensuite, le modèle d'appariement et le réseau de neurones n'ont pas permis de retrouver le gain apporté du point de vue de la localisation, d'après des études précédentes, par les bouchons à atténuation uniforme destinés aux musiciens. De plus, ces protecteurs ne seraient pas adaptés aux militaires, car ils ne possèdent pas une atténuation dépendante du niveau sonore. Or celle-ci est particulièrement efficace pour protéger contre les sources de bruit impulsionnel et conserver la communication. Pour finir, les modèles ont été utilisés pour évaluer des prothèses auditives, dont le microphone est placé au-dessus de l'oreille. Nous avons pu montrer qu'au moins un système, qui reproduit le filtrage directionnel du pavillon et qui conserve l'ILD par communication binaurale, engendre une amélioration significative des performances de localisation sonore. Il serait intéressant d'adapter ce type de système aux casques et aux bouchons talk-through, dans le but d'éviter les confusions avant-arrière.

# Conclusion

# 1 Résultats

L'impact des systèmes tactiques de communication et protection sur les performances de localisation sonore sur le plan horizontal a été abordé. De manière classique tout d'abord, les réponses de 30 auditeurs à une expérience de localisation avec TCAPS ont été examinées. Nous avons pu constater une dégradation conséquente des nombres de localisations exactes et voisines, engendrée par les TCAPS. L'augmentation des erreurs avant-arrière constitue sa principale cause. Aucun TCAPS n'entraîne de performances de localisation analogues à la condition d'écoute normale. Cette expérience a tout de même permis de classifier les TCAPS par type : les bouchons non linéaires affectent moins les performances de localisation sonore que les bouchons talk-through, et la pire détérioration est observée avec les casques talk-through. À cause de la fréquence de coupure des systèmes actifs, nous pouvons nous attendre à de plus amples différences entre les protections passives et actives en étudiant la localisation dans la dimension de l'élévation.

Les résultats de l'expérience subjective ont pu être qualitativement prévus grâce à l'observation des indices de localisation estimés avec et sans protection. Les différences interaurales de temps et d'intensité sont très peu affectées par les bouchons non linéaires et talk-through. Si les casques talk-through ne dégradent pas non plus l'ITD, les ILD observées sont par contre compressées. Nous en avons conclu que la perception de la dimension gauche-droite, c'est-à-dire de l'angle latéral, ne devait pas être substantiellement dégradée par les TCAPS dans le cas de sources à large bande fréquentielle. À l'inverse, l'observation des indices spectraux permettant de résoudre les ambiguïtés avant-arrière révèle un impact conséquent des protecteurs. L'analyse de cette dégradation permet de retrouver la classification des TCAPS par type. Il s'avère donc possible de prédire l'effet des TCAPS sur les performances de localisation sonore à partir de mesures d'HRTF sur tête artificielle.

Deux modèles de localisation sonore exploitant les HRTF mesurées avec les TCAPS ont alors été développés dans cette étude : un modèle d'appariement et un réseau de neurones à trois couches. Le premier est basé sur une décomposition de la dimension azimutale en une composante latérale, déterminée par l'ITD et l'ILD, et une composante polaire, obtenue grâce aux indices spectraux. Les deux modèles sont optimisés à partir des résultats de l'expérience subjective précédente en condition d'oreille libre. Les prédictions obtenues avec les TCAPS sont globalement réalistes. En particulier, le réseau de neurones prédit des performances de localisation sonore très proches des observations subjectives avec les bouchons d'oreilles. En revanche, avec les casques talk-through, il surestime les erreurs avant-arrière et grossières, voire les erreurs gauche-droite. Le modèle d'appariement sousestime les erreurs grossières : il est trop précis dans le bon quadrant. Ses probabilités de localisation voisine et d'erreur avant-arrière sont néanmoins semblables aux résultats de l'expérience subjective.

L'adaptation des deux modèles de localisation à des HRTF mesurées sur une tête artificielle différente, et même avec un dispositif expérimental différent, engendre encore une fois des prédictions analogues aux observations subjectives. Ces résultats montrent qu'il est possible d'évaluer des TCAPS, du point de vue de la localisation sonore, à partir des HRTF d'un auditeur type, en l'occurrence une tête artificielle. Les modèles de localisation nécessitent toujours les données d'une expérience subjective en condition d'écoute normale pour leur optimisation, mais l'objectif d'étude des TCAPS sans avoir recours à des auditeurs est presque atteint.

Pour finir, trois exemples d'application des modèles de prédiction ont été présentés. Nous avons vu que la création d'indices spectraux avant-arrière sur un casque talk-through, en ajoutant un pavillon d'oreille factice, permet de réduire les confusions correspondantes. Ensuite, les bouchons d'oreilles à atténuation uniforme, c'est-à-dire constante en fonction de la fréquence, n'engendre pas d'amélioration significative des performances de localisation dans le plan horizontal par rapport aux bouchons non linéaires. Enfin, le modèle d'appariement et le réseau de neurones ont aussi pu être appliqués à des prothèses auditives de type BTE. Elles présentent une réduction des performances de localisation sonore entraînée par le placement du microphone au-dessus du pavillon de l'oreille. Pour un système en particulier, cette réduction est corrigée grâce à la communication binaurale qui conserve l'ILD, et à une adaptation de la directivité du microphone pour reconstruire les indices spectraux. Au total, les modèles de localisation ont été utilisés dans quatre applications. Pour chacune d'entre elles, nous avons constaté que les performances prédites par le modèle d'appariement et le réseau de neurones sont similaires.

L'expérience comportementale réalisée dans cette étude, ainsi que les modèles de prédiction qui s'ensuivent, se focalise sur les performances de localisation sonore d'auditeurs « naïfs », c'est-à-dire de personnes n'étant pas habituées à porter les protections auditives étudiées. Or la plasticité fonctionnelle du cerveau humain permet au système auditif de s'entraîner à localiser avec une audition perturbée (Carlile, 2014). En effet, en bouchant les cavités des pavillons auditifs pour dégrader les indices spectraux, Hofman et al. (1998) ont montré que des performances de localisation correctes pouvaient être retrouvées à partir de quatre à six semaines. Cette durée d'apprentissage est réduite à environ onze jours si une seule oreille est modifiée (Van Wanrooij & Van Opstal, 2005). En cas de dégradation de la perception latérale seulement (par modification de l'ITD), le temps de récupération des performances de localisation est inférieur à dix jours (Trapeau & Schönwiesner, 2011, 2015). Le système auditif peut donc s'adapter à de nouvelles HRTF, tout en restant capable d'exploiter les HRTF normales. De meilleures performances de localisation avec les TCAPS peuvent donc être attendues pour des auditeurs les utilisant quotidiennement. De plus, l'apprentissage pourrait être accéléré par un entraînement approprié.

Carlile et al. (2014) ont montré que des séances d'entraînement à la localisation avec feedback permettent de retrouver des performances correctes au bout de 10 jours alors que les auditeurs portent des bouchons remplissant 40% de l'oreille externe. De même, dans (Casali & Robinette, 2015), une séance quotidienne d'une heure d'entraînement à la localisation sur le plan horizontal avec des bouchons ou un casque talk-through permet aux auditeurs d'atteindre des performances proches de celles obtenues en condition d'écoute normale après douze jours. Finalement, en plus de l'amélioration des systèmes de protection auditive du point de vue de la localisation sonore, un protocole d'entraînement à la localisation avec ces systèmes pourrait être mis au point pour faciliter la prise en main des TCAPS par le personnel.

# 2 Perspectives

Nous avons précédemment pointé du doigt le besoin de conduire une expérience subjective en condition d'écoute normale, afin de pouvoir optimiser le modèle d'appariement et le réseau de neurones. Les prédictions de ce dernier à partir de distributions de probabilités théoriques (distributions de von Mises) semblent néanmoins aussi acceptables. La définition d'une distribution de probabilités théorique qui entraîne des résultats corrects avec le modèle d'appariement permettrait de s'affranchir complètement des expériences comportementales. Concernant le réseau de neurones, nous avons conclu que la prédiction d'erreurs gauche-droite irréalistes provient d'un manque de pondération des indices binauraux, qui pourrait être résolu par l'utilisation d'une décomposition latérale/polaire analogue à celle du modèle d'appariement. Nous pensons que cette amélioration faciliterait aussi l'évaluation des performances de localisation sonore dans la dimension de l'élévation.

Le rôle des indices de localisation sonore sur une sphère est mieux représenté dans le système de coordonnées polaires interaurales. Les différences interaurales de temps et d'intensité permettent la perception de l'angle latéral. Pour un angle latéral fixe, l'angle polaire définit un cône de confusion dont la résolution est assurée par les indices spectraux. L'application de la décomposition latérale/polaire pour les deux modèles de prédiction paraît adaptée à l'étude de l'effet des TCAPS sur une sphère. La mesure d'HRTF pour les 50 sources sonores de la sphère du CNAM a déjà été réalisée avec les six TCAPS utilisés dans cette étude. Ces HRTF n'ont cependant pas été exploitées au-delà du plan horizontal. L'ajout d'une nouvelle dimension à la prédiction de l'impact des TCAPS sur la localisation sonore constitue donc un futur axe de recherche pour l'ISL.

Le modèle d'appariement et le réseau de neurones déjà définis sur le plan horizontal procurent un outil bien plus rapide qu'une expérience subjective pour évaluer l'effet d'une protection auditive sur la localisation sonore. La suite de l'étude présentée dans ce mémoire consiste donc principalement en l'analyse des caractéristiques des TCAPS propres à la localisation. Concernant les bouchons non linéaires, il semble que l'idéal serait un filtre à atténuation uniforme, comme les bouchons pour musiciens, mais qui reste non linéaire. Des bouchons plus discrets, c'est-à-dire qui obstruent moins la conque de l'auditeur, constituent aussi une première solution. Ensuite, nous avons remarqué que la forme physique des casques talk-through pourrait être améliorée pour la localisation. Cependant, l'évolution des systèmes talk-through devrait plutôt s'effectuer du point de vue traitement du signal. Tout d'abord, une augmentation de la fréquence de coupure s'impose pour que la totalité des indices spectraux soit disponible. Ensuite, l'ILD peut être conservée grâce à un traitement binaural du signal. Enfin, il est important d'introduire des indices spectraux exploitables par l'auditeur et dépendants de la position de la source sonore. Ce point peut être résolu en travaillant sur la directivité des systèmes talk-through. Finalement, nous pouvons conclure que l'amélioration des protections auditives du point de vue de la localisation sonore pourrait se baser sur les systèmes qui existent déjà dans le domaine de l'audioprothèse.

# Bibliographie

- Abel, S. M., Alberti, P. W., Haythornthwaite, C., & Riko, K. (1982). Speech intelligibility in noise : effects of fluency and hearing protector type. *Journal of the Acoustical Society* of America, **71**(3), 708–715.
- Abel, S. M., Armstrong, N. M., & Giguère, C. (1993). Auditory perception with leveldependent hearing protectors : the effects of age and hearing loss. *Scandinavian Audio*logy, 22(2), 71–85.
- Abel, S. M., Giguère, C., Consoli, A., & Papsin, B. C. (1999). Front/back mirror image reversal errors and left/right asymmetry in sound localization. Acta Acustica united with Acustica, 85(3), 378–389.
- Abel, S. M. & Hay, V. H. (1996). Sound localization : the interaction of aging, hearing loss and hearing protection. *Scandinavian Audiology*, 25(1), 3–12.
- Abel, S. M., Tsang, S., & Boyne, S. (2007). Sound localization with communications headsets : Comparison of passive and active systems. Noise & Health, 9(37), 101–107.
- Alali, K. A. (2011). Azimuthal localization and detection of vehicular backup alarms under electronic and non-electronic hearing protection devices in noisy and quiet environments.
  PhD thesis, Faculty of the Virginia Polytechnic Institute and State University. 208 pages.
- Algazi, V. R., Duda, R. O., Duraiswami, R., Gumerov, N. A., & Tang, Z. (2002). Approximating the head-related transfer function using simple geometric models of the head and torso. *Journal of the Acoustical Society of America*, **112**(5), 2053–2064.
- ANSI S12.42-2010 (2010). American National Standards Institute : Methods for the measurement of insertion loss of hearing protection devices in continuous or impulsive noise using microphone-in-real-ear or acoustical test fixture procedures. Acoustical Society of America.
- ANSI S12.6-2016 (2016). American National Standards Institute : Methods for measuring the real-ear attenuation of hearing protectors. Acoustical Society of America.

- ANSI S3.19-1974 (1974). American National Standards Institute : Measurement of Real-Ear Hearing Protector Attenuation and Physical Attenuation of Earmuffs.
- Asano, F., Suzuki, Y., & Sone, T. (1990). Role of spectral cues in median plane localization. Journal of the Acoustical Society of America, 88(1), 159–168.
- Atherley, G. & Else, D. (1971). Effect of ear-muffs on the localization of sound under reverberant conditions. In Proc. Roy. Soc. Med., volume 64 (pp. 203–205).
- Atherley, G. & Noble, W. (1970). Effect of ear-defenders (ear-muffs) on the localization of sound. British Journal of Industrial Medicine, 27, 260–265.
- Banerjee, A., Dhillon, I. S., Ghosh, J., & Sra, S. (2005). Clustering on the unit hypersphere using von Mises-Fisher distributions. *Journal of Machine Learning Research*, 6, 1345– 1382.
- Batteau, D. W. (1967). The role of the pinna in human localization. In *Proc. R. Soc. Lond.*, volume **B 168** (pp. 158–180).
- Baumgartner, R. & Majdak, P. (2015). Modeling localization of amplitude-panned virtual sources in sagittal planes. *Journal of the Audio Engineering Society*, **63**(7), 562–569.
- Baumgartner, R., Majdak, P., & Laback, B. (2013). Assessment of sagittal-plane sound localization performance in spatial-audio applications. In *The Technology of Binaural Listening* chapter 4. Springer.
- Baumgartner, R., Majdak, P., & Laback, B. (2014). Modeling sound-source localization in sagittal planes for human listeners. *Journal of the Acoustical Society of America*, 136(2), 791–802.
- Baumgartner, R., Majdak, P., & Laback, B. (2016). Erratum : Modeling sound-source localization in sagittal planes for human listeners [J. Acoust. Soc. Am. 136(2), 791-802 (2014)]. Journal of the Acoustical Society of America, 140(4), 2456.
- Benichoux, V., Rébillat, M., & Brette, R. (2016). On the variation of interaural time differences with frequency. *Journal of the Acoustical Society of America*, **139**.
- Berger, E. H. (1986). Methods of measuring the attenuation of hearing protection devices. Journal of the Acoustical Society of America, **79**(6), 1655–1687.
- Berger, E. H. (2005). Preferred methods for measuring hearing protector attenuation. In *Internoise, Rio de Janeiro, Brazil.*
- Bernstein, E. R., Brammer, A. J., & Yu, G. (2014). Augmented warning sound detection for hearing protectors. *Journal of the Acoustical Society of America*, **135**(1), EL29–EL34.

- Best, V., Kalluri, S., McLachlan, S., Valentine, S., Edwards, B., & Carlile, S. (2010). A comparison of CIC and BTE hearing aids for three-dimensional localization of speech. *International Journal of Audiology*, 49, 723–732.
- Bevis, Z. L., Semeraro, H. D., van Besouw, R. M., Rowan, D., Lineton, B., & Allsopp, A. J. (2014). Fit for the frontline? A focus group exploration of auditory tasks carried out by infantry and combat support personnel. *Noise & Health*, 16(69), 127–135.
- Blanck, G., Zimpfer, V., & Joubaud, T. (2015). Création de sources sonores virtuelles pour des tests subjectifs de localisation sonore. Technical Report ISL-R 124/2015. 71 pages.
- Blauert, J. (1969). Sound localization in the median plane. Acustica, 22(4), 205–213.
- Blauert, J. (1983). Spatial Hearing : The Psychophysics of Human Sound Localization. MIT Press. 427 pages.
- Bockstael, A., De Coensel, B., Botteldooren, D., D'Heanens, W., Keppler, H., Maes, L., Philips, B., Swinnen, F., & Bart, V. (2011). Speech recognition in noise with active and passive hearing protectors : a comparative study. *Journal of the Acoustical Society of America*, **129**(6), 3702–3715.
- Bolia, R. S. & McKinley, R. L. (2000). The effects of hearing protectors on auditory localization : Evidence from audio-visual target acquisition. *International Journal of Occupational Safety and Ergonomics*, 6(3), 309–319.
- Boyer, S., Doutres, O., Sgard, F., Laville, F., & Boutin, J. (2014). Objective assessment of the sound paths through earmuff components. *Applied Acoustics*, 83, 76–85.
- Bronkhorst, A. W. (1995). Localization of real and virtual sound sources. Journal of the Acoustical Society of America, 98(5), 2542–2553.
- Brown, A. D., Stecker, G. C., & Tollin, D. J. (2015). The precedence effect in sound localization. Journal of the Association for Research in Otolaryngology, 16, 1–28.
- Brungart, D. S., Eades, C. S., Simpson, B. D., & Kordik, A. J. (2003a). The effect of microphone placement on localization accuracy with electronic pass-through earplugs. In *IEEE WASPAA*, New Paltz, NY, USA (pp. 149–152).
- Brungart, D. S. & Hobbs, B. W. (2007). A comparison of acoustic and psychoacoustic measurements of pass-through hearing protection devices. In *IEEE WASPAA*, New Paltz, NY, USA (pp. 70–73).
- Brungart, D. S., Kordik, A. J., & Simpson, B. D. (2004). The effects of single and double hearing protection on the localization and segregation of spatially-separated speech signals. *Journal of the Acoustical Society of America*, **116**(4), 1897–1900.

- Brungart, D. S., Kordik, A. J., Simpson, B. D., & McKinley, R. L. (2003b). Auditory localization in the horizontal plane with single and double hearing protection. Aviation, Space and Environmental Medicine, 74(9), 937–946.
- Buck, K. & Parmentier, G. (1999). Artificial heads for high-level impulse sound measurement. In 6th International Congress on Sound and Vibration, Copenhagen, Denmark.
- Burger, J. (1958). Front-back discrimination of the hearing system. Acustica, 8(5), 301–302.
- Burkhard, M. D. & Sachs, R. M. (1975). Anthropometric manikin for acoustic research. Journal of the Acoustical Society of America, 58(1), 214–222.
- Busson, S. (2006). Individualisation d'indices acoustiques pour la synthèse binaurale (Acoustic cues individualization for binaural synthesis). PhD thesis, Université de la Méditerranée Aix-Marseille II. 313 pages.
- Butler, R. A. & Belendiuk, K. (1977). Spectral cues utilized in the localization of sound in the median saggital plane. Journal of the Acoustical Society of America, 61(5), 1264– 1269.
- Butler, R. A. & Planert, N. (1976). The influence of stimulus bandwidth on localization of sound in space. *Perception & Psychophysics*, **19**(1), 103–108.
- Carlile, S. (2014). The plastic ear and perceptual relearning in auditory spatial perception. Frontiers in Neuroscience, 8.
- Carlile, S., Balachandar, K., & Kelly, H. (2014). Accommodating to new ears : The effects of sensory and sensory-motor feedback. *Journal of the Acoustical Society of America*, 135(4), 2002–2011.
- Carlile, S., Martin, R., & McAnally, K. (2005). Spectral information in sound localization. International Review of Neurobiology, 70, 399–434.
- Carmichel, E. L., Harris, F. P., & Story, B. H. (2007). Effects of binaural electronic hearing protectors on localization and response time to sounds in the horizontal plane. *Noise & Health*, 9(37), 83–95.
- Carterette, E. C. & Friedman, M. P., Eds. (1978). Handbook of Perception. Volume IV : Hearing. Academic Press. 719 pages.
- Casali, J. G. & Robinette, M. B. (2015). Effects of user training with electronicallymodulated sound transmission hearing protectors and the open ear on horizontal localization ability. *International Journal of Audiology*, 54, S37–S45.

- Casali, J. G., Robinson, G. S., Dabney, E. C., & Gauger, D. (2004). Effect of electronic ANR and conventional hearing protectors on vehicle backup alarm detection in noise. *Human Factors*, 46, 1–11.
- Childers, D. G., Skinner, D. P., & Kemerait, R. C. (1977). The cepstrum : a guide to processing. In *Proc. IEEE*, volume **65** (pp. 1428–1443).
- Chung, K., Neuman, A. C., & Higgins, M. (2008). Effects of in-the-ear microphone directionality on sound direction identification. *Journal of the Acoustical Society of America*, 123(4), 2264–2275.
- Chung, W., Carlile, S., & Leong, P. (2000). A performance adequate computational model for auditory localization. *Journal of the Acoustical Society of America*, **107**(1), 432–445.
- CNB (2016). Le coût social des pollutions sonores. Technical report. http://www.bruit. fr/cout-social-du-bruit-en-france-57-milliards-deuros.html.
- Dancer, A., Buck, K., Hamery, P., & Parmentier, G. (1999). Hearing protection in the military environment. Noise & Health, 2(5), 1–15.
- Dobreva, M. S., O'Neill, W. E., & Paige, G. D. (2011). Influence of aging on human sound localization. *Journal of Neurophysiology*, **105**, 2471–2486.
- Dreyfus, G., Martinez, J.-M., Samuelides, M., Gordon, M. B., Badran, F., & Thiria, S. (2008). *Apprentissage statistique*. 471 pages.
- Duda, R. O., Avendano, C., & Algazi, V. (1999). An adaptable ellipsoidal head model for the interaural time difference. In *IEEE ICASSP*, *Phoenix*, AZ, USA.
- Duda, R. O. & Martens, W. L. (1998). Range dependence of the response of a spherical head model. Journal of the Acoustical Society of America, 104(5), 3048–3058.
- Durin, V., Carlile, S., Guillon, P., Best, V., & Kalluri, S. (2014). Acoustic analysis of the directional information captured by five different hearing aid styles. *Journal of the Acoustical Society of America*, 136(2), 818–828.
- Emerit, M., Rugeles, F., & Nicol, R. (2016). Transformer un jeu de hrtf en un autre à partir de données morphologiques. In *Congrès Français d'Acoustique, Le Mans, France* (pp. 185–191).
- Farina, A. (2007). Advancements in impulse response measurements by sine sweeps. In AES 122nd Convention, Vienna, Austria.
- Fastl, H. & Zwicker, E. (2007). Psychoacoustics : Facts and Models. Springer, 3rd edition. 471 pages.

- Francart, T. & Wouters, J. (2007). Perception of across-frequency interaural level differences. Journal of the Acoustical Society of America, 122(5), 2826–2831.
- Gaik, W. (1993). Combined evaluation of interaural time and intensity differences : Psychoacoustic results and computer modeling. *Journal of the Acoustical Society of America*, 94(1), 98–110.
- Gallagher, H. L., McKinley, R. L., Theis, M. A., Swayne, B. J., & Thompson, E. R. (2014). Performance assessment of passive hearing protection devices. Technical Report AFRL-RH-WP-TR-2014-0148. 39 pages.
- Gardner, B. & Martin, K. (1994). HRTF measurements of a KEMAR dummy-head microphone. Technical Report #280. http://sound.media.mit.edu/resources/KEMAR. html.
- Glasberg, B. R. & Moore, B. C. (1990). Derivation of auditory filter shapes from notchednoise data. *Hearing Research*, 47, 103–138.
- Guillon, P. (2009). Individualisation des indices spectraux pour la synthèse binaurale : recherche et exploitation des similarités inter-individuelles pour l'adaptation ou la reconstruction de HRTF. PhD thesis, Université du Maine. 378 pages.
- Hamery, P., Dancer, A., & Evrard, G. (2000). Patent US 6068079 A : Acoustic valve capable of selective and non-linear filtering of sound.
- Harris, F. J. (1978). In the use of windows for harmonic analysis with the discrete fourier transform. In *Proc. IEEE*, volume **66** (pp. 51–83).
- Hartley, R. V. L. & Fry, T. C. (1921). The binaural location of pure tones. *Physical Review*, 18(6), 431–442.
- Hartmann, W. M. (1999). How we localize sound. Physics Today, 52(11), 24-29.
- Hawley, M. L., Litovsky, R. Y., & Colburn, H. S. (1999). Speech intelligibility and localization in a multi-source environment. *Journal of the Acoustical Society of America*, 105(6), 3436–3448.
- Haykin, S. (1998). Neural Networks : A Comprehensive Foundation. Prentice Hall PTR, 2nd edition. 822 pages.
- Hebrank, J. (1976). Pinna disparity processing : A case of mistaken identity? Journal of the Acoustical Society of America, 59(1), 220–221.
- Hebrank, J. & Wright, D. (1974a). Are two ears necessary for localization of sound sources on the median plane? *Journal of the Acoustical Society of America*, **56**(3), 935–938.

- Hebrank, J. & Wright, D. (1974b). Spectral cues used in the localization of sound sources on the median plane. *Journal of the Acoustical Society of America*, 56(6), 1829–1834.
- Hofman, P. M., Van Riswick, J. G., & Van Opstal, A. J. (1998). Relearning sound localization with new ears. *Nature Neuroscience*, 1(5), 417–421.
- Hohmann, V. (2002). Frequency analysis and synthesis using a gammatone filterbank. Acta Acustica united with Acustica, 88, 433–442.
- Holdsworth, J., Nimmo-Smith, I., Patterson, R., & Rice, P. (1988). Implementing a gammatone filter bank. In Annex C of SVOS Final Report.
- Howell, K. & Martin, A. (1975). An investigation of the effects of hearing protectors on vocal communication in noise. *Journal of Sound and Vibration*, **41**(2), 181–196.
- Humes, L. E. (1983). A psychophysical evaluation of the dependence of hearing protector attenuation on noise level. *Journal of the Acoustical Society of America*, **73**(1), 297–311.
- Humes, L. E., Joellenbeck, L. M., & Durch, J. S., Eds. (2006). Noise and Military Service : Implications for Hearing Loss and Tinnitus. The National Academies Press. 339 pages.
- Ingard, U. & Ising, H. (1967). Acoustic nonlinearity of an orifice. Journal of the Acoustical Society of America, 42(1), 6–17.
- ISO 4869-1 :1990 (1990). International Standard Organization : Subjective method for the measurement of sound attenuation. Comité Technique ISO/TC 43, Acoustics.
- ISO 4869-2 :1994 (1994). International Standard Organization : Estimation of effective A-weighted sound pressure levels when hearing protectors are worn. Comité Technique ISO/TC 43, Acoustics.
- ISO 4869-3 :2007 (2007). International Standard Organization : Measurement of insertion loss of ear-muff type protectors using an acoustic test fixture. Comité Technique ISO/TC 43, Acoustics.
- ISO 7029 :2000 (2000). International Standard Organization : Distribution statistique des seuils d'audition en fonction de l'âge. Comité Technique ISO/TC 43, Acoustique.
- Itoh, M., Iida, K., & Morimoto, M. (2007). Individual differences in directional bands in median plane localization. *Applied Acoustics*, 68, 909–915.
- Jin, C., Corderoy, A., Carlile, S., & van Schalk, A. (2004). Contrasting monaural and interaural spectral cues for human sound localization. *Journal of the Acoustical Society* of America, **115**(6), 3124–3141.
- Jin, C., Schenkel, M., & Carlile, S. (2000). Neural system identification model of human sound localization. Journal of the Acoustical Society of America, 108(3), 1215–1235.

- Joubaud, T. & Zimpfer, V. (2014). First subjective experiment on sound localization with talk-through earmuffs. Technical Report ISL-R-104/2014. 33 pages.
- Joubaud, T., Zimpfer, V., Garcia, A., & Langrenne, C. (2015). Degradation of front-back spectral cues induced by tactical communication and protective systems. In *Euronoise Conference, Maastricht, The Netherlands* (pp. 1943–1947).
- Joubaud, T., Zimpfer, V., Garcia, A., & Langrenne, C. (2017). Sound localization models as evaluation tools for tactical communication and protective systems. *Journal of the Acoustical Society of America*, 141(4), 2637–2649.
- Killion, M. C. (1993). The parvum bonum, plus melius fallacy in earplug selection. In Recent Developments in Hearing Instrument Technology : 15th Danavox Symposium, Copenhagen, Denmark (pp. 415–433).
- Killion, M. C., DeVilbiss, E., & Stewart, J. (1988). An earplug with uniform 15-db attenuation. The Hearing Journal, 41(5), 14–17.
- Killion, M. C., Monroe, T., & Drambarean, V. (2011). Better protection from blasts without sacrificing situational awareness. *International Journal of Audiology*, 50(S1), S38–S45.
- Kuhl, W. & Plantz, R. (1972). The localization of a front and a rear sound source by test persons with unrestricted head movement. *Acustica*, **27**(2), 108–110.
- Kuhn, G. F. (1977). Model for the interaural time differences in the azimuthal plane. Journal of the Acoustical Society of America, 62(1), 157–167.
- Lambourg, C., Peyret, P., Dubois, F., & Gettliffe, J. (2016). Caractérisation de l'influence des picb sur l'intelligibilité de la parole. In *Congrès Français d'Acoustique, Le Mans, France* (pp. 2669–2674).
- Langendijk, E. H. & Bronkhorst, A. W. (2002). Contribution of spectral cues to human sound localization. *Journal of the Acoustical Society of America*, **112**(4), 1583–1596.
- Larcher, V. (2001). Techniques de spatialisation des sons pour la réalité virtuelle. PhD thesis, Université de Paris VI. 206 pages.
- Larcher, V. & Jot, J.-M. (1997). Techniques d'interpolation de filtres audio-numériques : Application à la reproduction spatiale des sons sur écouteurs. In Congrès Français d'Acoustique, Marseille, France.
- Lecomte, P. (2016). Ambisonie d'ordre élevé en trois dimensions : Captation, transformation et décodage adaptatif de champs sonores. PhD thesis, Conservatoire National des Arts et Métiers, Paris, France.

- Macaulay, E. J. & Hartmann, W. M. (2010). The acoustical bright spot and mislocalization of tones by human listeners. *Journal of the Acoustical Society of America*, **127**(3), 1440– 1449.
- Macpherson, E. A. & Middlebrooks, J. C. (2002). Listener weighting of cues for lateral angle : The duplex theory of sound localization revisited. *Journal of the Acoustical Society of America*, **111**(5), 2219–2236.
- Macpherson, E. A. & Middlebrooks, J. C. (2003). Vertical-plane sound localization probed with ripple-spectrum noise. Journal of the Acoustical Society of America, 114(1), 430– 445.
- Macpherson, E. A. & Sabin, A. T. (2007). Binaural weighting of monaural spectral cues for sound localization. *Journal of the Acoustical Society of America*, **121**(6), 3677–3688.
- Majdak, P., Baumgartner, R., & Laback, B. (2014). Acoustic and non-acoustic factors in modeling listener-specific performance of sagittal-plane sound localization. *Frontiers in Psychology*, 5(319).
- Mallock-Armstrong (1915). The mallock-armstrong ear defender. The British Medical Journal.
- Martin, A. (1979). Dependence of acoustic attenuation of hearing protectors on incident sound level. British Journal of Industrial Medicine, 36, 1–14.
- Martin, R. L., Paterson, M., & McAnally, K. I. (2004). Utility of monaural spectral cues is enhanced in the presence of cues to sound-source lateral angle. *Journal of the Association* for Research in Otolaryngology, 5, 80–89.
- McAnally, K. I. & Martin, R. L. (2014). Sound localization with head movement : implications for 3-d audio displays. *Frontiers in Neuroscience*, 8(210).
- McEwan, A. & Van Schaik, A. (2000). A silicon representation of the meddis inner hair cell model. In *ICSC ISA Symposia* (pp. 1544).
- McKinley, R. L. (2000). Communication and localization with hearing protectors. In *RTO HFM Lecture Series on "Damage Risk from Impulse Noise", Maryland, USA* (pp. 1–14).
- Meddis, R. (1986). Simulation of mechanical to neural transduction in the auditory receptor. *Journal of the Acoustical Society of America*.
- Méot, A. (2014). Analyse de variance : intuition, pratique, théorie. Ellipses Editions. 666 pages.
- Middlebrooks, J. C. (1992). Narrow-band sound localization related to external ear acoustics. Journal of the Acoustical Society of America, 92(5), 2607–2624.

- Middlebrooks, J. C. (1997). Spectral shape cues for sound localization. In R. H. Gilkey & T. R. Anderson (Eds.), *Binaural and Spatial Hearing in Real and Virtual Environments* chapter 4, (pp. 77–97). Psychology Press.
- Middlebrooks, J. C. (1999). Individual differences in external-ear transfer functions reduced by scaling in frequency. *Journal of the Acoustical Society of America*, **106**(3), 1480–1492.
- Middlebrooks, J. C. & Green, D. M. (1990). Directional dependence of interaural envelope delays. *Journal of the Acoustical Society of America*, 87(5), 2149–2162.
- Middlebrooks, J. C. & Green, D. M. (1991). Sound localization by human listeners. Annual Review of Psychology, 42, 135–159.
- Middlebrooks, J. C. & Knudsen, E. I. (1984). A neural code for auditory space in the cat's superior colliculus. *Journal of Neuroscience*, 4(10), 2621–2634.
- Middlebrooks, J. C., Makous, J. C., & Green, D. M. (1989). Directional sensitivity of sound-pressure levels in the human ear canal. *Journal of the Acoustical Society of America*, 86(1), 89–108.
- Morimoto, M. (2001). The contribution of two ears to the perception of vertical angle in sagittal planes. *Journal of the Acoustical Society of America*, **109**(4), 1596–1603.
- Morimoto, M. & Ando, Y. (1980). On the simulation of sound localization. Journal of the Acoustical Society of Japan, 1(3), 167–174.
- Morimoto, M., Iida, K., & Itoh, M. (2003). Upper hemisphere sound localization using head-related transfer functions in the median plane and interaural differences. Acoustical Science and Technology, 24(5), 267–275.
- Musicant, A. D. & 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.
- Neti, C., Schneider, M. H., & Young, E. D. (1992a). Maximally fault tolerant neural networks. *IEEE Transactions on Neural Networks*, 3(1), 14–23.
- Neti, C., Young, E. D., & Schneider, M. H. (1992b). Neural network models of sound localization based on directional filtering by the pinna. *Journal of the Acoustical Society* of America, 92(6), 3140–3156.
- NIOSH (2016). Method for calculating and using the NRR, NRR(SF), SNR, HML rating and APV. Technical report. https://www.cdc.gov/niosh/topics/noise/hpdcomp/ pdfs/calculation.pdf.

- Niquette, P. A. (2009). Uniform hearing protection for musicians. In *Hearing loss in musicians : prevention and management* chapter 6, (pp. 63–74). Plural Publishing Inc., San Diego, USA.
- Noble, W. G. & Russell, G. (1972). Theoretical and practical implications of the effects of hearing protection devices on localization ability. Acta Oto-Laryngologica, 74, 29–36.
- Norin, J. A., Emanuel, D. C., & Letowski, T. R. (2011). Speech intelligibility and passive, level-dependent earplugs. *Ear & Hearing*, **32**(5), 642–649.
- Ocklenburg, S., Hirnstein, M., Hausmann, M., & Lewald, J. (2010). Auditory space perception in left- and right-handers. *Brain and Cognition*, 72, 210–217.
- Oppenheim, A. V., Schafer, R. W., & Buck, J. R. (1999). Discrete-Time Signal Processing. Prentice Hall, Inc., 2nd edition. 896 pages.
- Parmentier, G., Dancer, A., Buck, K., Kronenberger, G., & Beck, C. (2000). Artificial head (ATF) for evaluation of hearing protectors. Acta Acustica united with Acustica, 86(5), 847–852.
- Patel, J. (2008). Musicians' hearing protection A review. Technical Report RR664. 62 pages.
- Patterson, R. D. & Holdsworth, J. (1996). A functional model of neural activity patterns and auditory images. Advances in Speech, Hearing and Language Processing, 3 Part B, 547–563.
- Rabinowitz, W. M., Maxwell, J., Shao, Y., & Wei, M. (1993). Sound localization cues for a magnified head : Implications from sound diffraction about a rigid sphere. *Presence*, 2(2), 125–129.
- Lord Rayleigh (1907). XII. On our perception of sound direction. *Philosophical Magazine*, 13(74), 214–232.
- Lord Rayleigh (1909). On the perception of the direction of sound. Proceedings of the Royal Society of London, 81, 61–64.
- Rubak, P. & Johansen, L. G. (1999). Active hearing protector with improved localization performance. In *Inter-Noise 99, Lauderdale, FL, USA*.
- Sällberg, B., Sattar, F., & Claesson, I. (2007). On a method for improving impulsive sounds localization in hearing defenders. *EURASIP Journal on Audio, Speech, and Music Processing*, (pp. 1–7).
- Savioja, L., Huopaniemi, J., Lokki, T., & Väänänen, R. (1999). Creating interactive virtual acoustic environments. Journal of the Audio Engineering Society, 47(9), 675–705.

- Searle, C. L., Braida, L. D., Cuddy, D. R., & Davis, M. F. (1975). Binaural pinna disparity : another auditory localization cue. *Journal of the Acoustical Society of America*, 57(2), 448–455.
- Shaw, E. (1974). Transformation of sound pressure level from the free field to eardrum in the horizontal plane. *Journal of the Acoustical Society of America*, **56**(6), 1848–1861.
- Shaw, E. A. G. & Teranishi, R. (1968). Sound pressure generated in an external-ear replica and real human ears by a nenear point source. *Journal of the Acoustical Society* of America, 44(1), 240–249.
- Shaw, Shaw, E. G. (1997). Acoustical features of the human external ear. In R. H. Gilkey & T. R. Anderson (Eds.), *Binaural and Spatial Hearing in Real and Virtual Environments* chapter 2, (pp. 25–47). Psychology Press.
- Shinn-Cunningham, B. G., Santarelli, S., & Kopco, N. (2000). Tori of confusion : binaural localization cues for sources within reach of a listener. *Journal of the Acoustical Society* of America, **107**(3), 1627–1636.
- Simpson, B. D., Bolia, R. S., McKinley, R. L., & Brungart, D. S. (2005). The impact of hearing protection on sound localization and orienting behavior. *Human Factors*, 47(1), 188–198.
- Slaney, M. (1993). An efficient implementation of the Patterson-Holdsworth auditory filter bank. Technical report. 42 pages.
- Slattery, W. H. I. & Middlebrooks, J. C. (1994). Monaural sound localization : acute versus chronic unilateral impairment. *Hearing Research*, 75, 38–46.
- Sra, S. (2012). A short note on parameter approximation for von Mises-Fisher distributions and a fast implementation of  $I_s(x)$ . Computational Statistics, **27**(1), 177–190.
- Stern, R., Wang, D., & Brown, G. (2005). Binaural sound localization. In Computational Auditory Scene Analysis chapter 5. John Wiley & Sons, Inc., Hoboken NJ.
- Sumner, C. J., Lopez-Poveda, E. A., O'Mard, L. P., & Meddis, R. (2002). A revised model of the inner-hair cell and auditory-nerve complex. *Journal of the Acoustical Society of America*, **111**(5), 2178–2188.
- Suter, A. H. (1989). The effects of hearing protectors on speech communication and the perception of warning signals. Technical Report 2-89. 36 pages.
- Talcott, K. A. (2011). An in-field experiment on the effects of hearing protection/enhancement devices and military vehicle noise on auditory localization of a gunshot in azimuth. PhD thesis, Faculty of the Virginia Polytechnic Institute and State University. 202 pages.

- Talcott, K. A., Casali, J. G., Keady, J. P., & Killion, M. C. (2012). Azimuthal auditory localization of gunshots in a realistic fiel environment : Effects of open-ear versus hearing protection-enhancement devices (HPEDs), military vehicle noise, and hearing impairment. *International Journal of Audiology*, **51**, S20–S30.
- Terroir, J. (2016). Dangerosité des bruit impulsionnels en milieu professionnel : réglementation, connaissances et mesures. In Congrès Français d'Acoustique, Le Mans, France (pp. 2653–2659).
- Thurlow, W. R., Mangels, J. W., & Runge, P. S. (1967). Head movement during sound localization. Journal of the Acoustical Society of America, 42(2), 489–493.
- Trapeau, R. & Schönwiesner, M. (2011). Relearning sound localization with digital earplugs. Canadian Acoustics, 39(3), 116 – 117.
- Trapeau, R. & Schönwiesner, M. (2015). Adaptation to shifted interaural time differences changes encoding of sound location in human auditory cortex. *NeuroImage*, **118**, 26–38.
- Van den Bogaert, T., Carette, E., & Wouters, J. (2011). Sound source localization using hearing aids with microphones placed behind-the-ear, in-the-canal and in-the-pinna. *International Journal of Audiology*, 50(3), 164–176.
- Van den Bogaert, T., Klasen, T. J., Moonen, M., Van Deun, L., & Wouters, J. (2006). Horizontal localization with bilateral hearing aids : without is better than with. *Journal* of the Acoustical Society of America, **119**(1), 515–526.
- Van Wanrooij, M. M. & Van Opstal, A. J. (2005). Relearning sound localization with a new ear. *Journal of Neuroscience*, 25(22), 5413–5424.
- Vause, N. L. & Grantham, D. W. (1999). Effects of earplugs and protective headgear on auditory localization ability in the horizontal plane. *Human Factors*, 41(2), 282–294.
- Wallach, H. (1939). On sound localization. Journal of the Acoustical Society of America, 10(4), 270–274.
- Wallach, H., Newman, E. B., & Rosenzweig, M. R. (1949). The precedence effect in sound localization. The American Journal of Psychology, 62(3), 315–336.
- Wallis, R. & Lee, H. (2015). Directional bands revisited. In AES 138th Convention, Warsaw, Poland.
- Wiener, F. M. (1947). Sound diffraction by rigid sphere and circular cylinders. Journal of the Acoustical Society of America, 19(3), 444–451.
- Wightman, F. L. & Kistler, D. J. (1989a). Headphone simulation of free-field listening. I : Stimulus synthesis. Journal of the Acoustical Society of America, 85(2), 858–867.

- Wightman, F. L. & Kistler, D. J. (1989b). Headphone simulation of free-field listening. II : Psychophysical validation. *Journal of the Acoustical Society of America*, 85(2), 868–878.
- Wightman, F. L. & Kistler, D. J. (1997a). Factors affecting the relative salience of sound localization cues. In R. H. Gilkey & T. R. Anderson (Eds.), *Binaural and spatial hearing* in real and virtual environment chapter 1, (pp. 1–23). Psychology Press.
- Wightman, F. L. & Kistler, D. J. (1997b). Monaural sound localization revisited. Journal of the Acoustical Society of America, 101(2), 1050–1063.
- Wightman, F. L. & Kistler, D. J. (1999). Resolution of front-back ambiguity in spatial hearing by listener and source movement. *Journal of the Acoustical Society of America*, 105(5), 2841–2853.
- Wilde, G. L. & Humes, Larry, E. (1987). Measurement of the attenuation characteristics of nonlinear hearing protective devices using the auditory brain stem response. *Journal* of the Acoustical Society of America, 81(3), 730–733.
- Woodworth, R. S. & Schlosberg, G. (1962). Experimental Psychology.
- Zakarauskas, P. & Cynader, M. S. (1993). A computational theory of spectral cue localization. Journal of the Acoustical Society of America, 94(3), 1323–1331.
- Zannin, P. H. T. & Gerges, S. N. Y. (2006). Effects of cup, cushion, headhead force, and foam lining on the attenuation of an earmuff. *International Journal of Industrial Ergonomics*, 36, 165–170.
- Zhang, M., Zhang, W., Kennedy, R. A., & Abhayapala, T. D. (2009). HRTF measurement on KEMAR manikin. In Acoustics Conference, Adelaide, Australia.
- Zhang, P. X. & Hartmann, W. M. (2010). On the ability of human listeners to distinguish between front and back. *Hearing Research*, 260, 30–46.
- Zimpfer, V. & Sarafian, D. (2014). Impact of hearing protection devices on sound localization performance. Frontiers in Neuroscience, 8(135).
- Zimpfer, V., Sarafian, D., Buck, K., & Hamery, P. (2012). Spatial localization of sounds with hearing protection devices allowing speech communication. In Acoustics Conference, Nantes, France (pp. 527–532).
- Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands. *Journal* of the Acoustical Society of America, **33**(2), 1961.
- Zündorf, I. C., Karnath, H.-O., & Lewald, J. (2011). Male advantage in sound localization at cocktail parties. *Cortex*, 47, 741–749.

# Annexes

# ANNEXE A

# Systèmes de coordonnées polaires verticales et interaurales

# 1 Rappel sur les coordonnées polaires verticales

Tout d'abord, nous ne nous intéressons ici qu'au repérage de points sur la sphère unité, nous nous affranchissons donc du paramètre r.

Considérons le repère cartésien  $(\vec{x}, \vec{y}, \vec{z})$  présenté en Figure A.1, avec  $\vec{x}$  représentant l'axe horizontal orienté vers le devant,  $\vec{y}$  représentant l'axe interaural orienté vers la droite et  $\vec{z}$  représentant l'axe vertical orienté vers le haut. Les coordonnées polaires verticales  $(\theta, \phi)$ , respectivement l'azimut et l'élévation, sont définies par :

$$\begin{cases} x = \cos\theta\sin\phi \\ y = \sin\theta\sin\phi \\ z = \cos\phi \end{cases}$$
(A.1)

avec  $\theta \in [0, 2\pi[$  et  $\phi \in [0, \pi]$ .

# 2 Coordonnées polaires interaurales

Ce système de coordonnées polaires permet de définir avec un seul angle les cônes de confusion dus aux indices de localisation binauraux. Il est défini par les angles  $\alpha$ , dit angle latéral, et  $\beta$ , dit angle polaire (voir Figure A.1). Un cône de confusion, ou un plan sagittal dans le cas où l'on considère la distance r, est défini à  $\alpha$  constant.

En reprenant le même système de coordonnées cartésiennes que dans la section précédent, nous avons cette fois :

$$\begin{cases} x = \cos\beta\cos\alpha \\ y = \sin\alpha \\ z = \sin\beta\cos\alpha \end{cases}, \tag{A.2}$$

avec  $\alpha \in [-\frac{\pi}{2}, \frac{\pi}{2}]$  et  $\beta \in [0, 2\pi[.$ 



FIG. A.1 – Coordonnées polaires verticales (a) et interaurales (b)

# 3 Lien entre coordonnées polaires verticales et interaurales

Il peut être utile de connaître le lien direct entre ces deux systèmes de coordonnées. Les Équations A.2 et A.1 nous permettent d'écrire le système d'équations suivant :

$$\begin{cases} \cos\beta\cos\alpha &= \cos\theta\sin\phi\\ \sin\alpha &= \sin\theta\sin\phi\\ \sin\beta\cos\alpha &= \cos\phi \end{cases}$$
(A.3)

# 3.1 Connaissant les coordonnées polaires verticales

Si l'on connaît les coordonnées polaires verticales  $\theta$  et  $\phi$ , il vient aisément de l'Équation A.3 que  $\alpha = \arcsin(\sin\theta\sin\phi)$ . L'utilisation de l'arc sinus ne pose ici aucun problème, car  $\alpha$  est compris entre  $-\frac{\pi}{2}$  et  $\frac{\pi}{2}$ .

Avant de calculer  $\beta$ , considérons le cas où cos  $\alpha = 0$  (la justification apparaît dans la suite de cette partie). Dans ce cas,  $\alpha$  est soit égal à  $-\frac{\pi}{2}$  soit à  $\frac{\pi}{2}$ . Le cône de confusion qu'il définit est alors réduit à un unique point,  $\beta$  peut donc prendre n'importe quelle valeur. Par convention, on notera  $\beta = 0$  dans ce cas particulier.

Si  $\cos \alpha \neq 0$ , on déduit de l'équation A.3 :

$$\cos \beta = \frac{\cos \theta \sin \phi}{\cos \alpha} \,. \tag{A.4}$$

En faisant un dessin, on remarque que, si l'élévation est inférieure ou égale à  $\frac{\pi}{2}$ , alors l'angle polaire est compris entre 0 et  $\pi$ . Il est donc possible d'utiliser l'arc cosinus.

Pour résumer, les coordonnées polaires interaurales sont données en fonction des coor-

données polaires verticales par le système d'équations :

$$\begin{cases} \alpha = \arcsin(\sin\theta\sin\phi) \\ \beta = 0 \qquad si\cos\alpha = 0 \\ \beta = \arccos(\frac{\cos\theta\sin\phi}{\cos\alpha}) \qquad si\phi \le \frac{\pi}{2} \\ \beta = 2\pi - \arccos(\frac{\cos\theta\sin\phi}{\cos\alpha}) \qquad sinon \end{cases}$$
(A.5)

### 3.2 Connaissant les coordonnées polaires interaurales

Si l'on connaît les coordonnées polaires interaurales  $\alpha$  et  $\beta$ , on remarque depuis l'Équation A.3 que  $\phi = \arccos(\sin \beta \cos \alpha)$ . L'utilisation de l'arc cosinus ne pose pas de problème, car  $\phi$  est compris entre 0 et  $\pi$ .

Considérons le cas où sin  $\phi = 0$ .  $\phi$  est alors égale à 0 ou à  $\pi$  et représente donc les pôles du cercle unité. Dans ce cas, l'azimut  $\theta$  peut prendre n'importe quelle valeur. Par convention, on notera  $\theta = 0$  si sin  $\phi = 0$ .

Si  $\sin \phi \neq 0$ , alors :

$$\cos\theta = \frac{\cos\beta\cos\alpha}{\sin\phi}.\tag{A.6}$$

On remarque que, si l'angle latéral  $\alpha \ge 0$ , alors l'azimut est compris entre 0 et  $\pi$  et l'arc cosinus peut être employé.

Pour résumer, les coordonnées polaires verticales sont données en fonction des coordonnées polaires interaurales par l'équation :

$$\begin{cases}
\phi = \arccos(\sin\beta\cos\alpha) \\
\theta = 0 & si\sin\phi = 0 \\
\theta = \arccos(\frac{\cos\beta\cos\alpha}{\sin\phi}) & si\alpha \ge 0 \\
\theta = 2\pi - \arccos(\frac{\cos\beta\cos\alpha}{\sin\phi}) & sinon
\end{cases}$$
(A.7)
# ANNEXE B

# La fonction gammatone et sa transformée de Fourier

## 1 La fonction gammatone

La fonction gammatone est le produit d'une distribution Gamma (non normalisée) et d'un son pur (sinusoïde). Il s'agit d'un modèle de filtre auditif largement utilisé pour simuler le filtrage fréquentiel réalisé par la cochlée. Sa réponse impulsionnelle est donnée par :

$$\forall t \ge 0 \quad g(t) = t^{n-1} e^{-2\pi f_b t} \cos(2\pi f_c t) \,, \tag{B.1}$$

avec n l'ordre du filtre (en général n = 4),  $f_c$  sa fréquence centrale et  $f_b$  sa fréquence de coupure.

La fréquence de coupure du filtre est en fait déduite de la fréquence centrale en utilisant les Equivalent Rectangular Bandwidths (ERB), estimées ci-dessous avec les paramètres de Glasberg et Moore :

$$ERB = \frac{f_c}{EarQ} + minBW, \qquad (B.2)$$

avec EarQ = 9,26449 et minBW = 24,7. La fréquence de coupure du filtre gammatone est ensuite définie comme  $f_b = 1,019ERB$ . La Figure B.1 donne la réponse impulsionnelle d'un filtre gammatone avec une fréquence centrale de 1 kHz.

## 2 Transformée de Fourier

La transformée de Fourier d'une fonction g(t) est définie comme suit :

$$G(f) = \mathcal{F}[g(t)](f) = \int_{-\infty}^{\infty} g(t)e^{-2\pi i f t} \mathrm{d}t.$$
 (B.3)

En utilisant la propriété stipulant que la transformée de Fourier d'un produit est le produit de convolution des transformées de Fourier, on remarque pour la fonction gammatone



FIG. B.1 – Fonction gammatone avec une fréquence centrale à 1 kHz

que :

$$G(f) = \mathcal{F}\left[\cos(2\pi f_c t)\right] * \mathcal{F}\left[t^{n-1}e^{-2\pi f_b t}\right].$$
(B.4)

Sachant que  $\mathcal{F}[\cos(2\pi f_c t)] = \frac{1}{2}(\delta(f - f_c) + \delta(f + f_c))$ , il ne nous reste qu'à déterminer la transformée de Fourier de la distribution Gamma.

#### 2.1 Transformée de la distribution Gamma

En supposant la distribution Gamma nulle pour des valeurs de t négatives, on cherche :

$$I(f) = \int_0^\infty t^{n-1} e^{-2\pi (f_b + if)t} dt.$$
 (B.5)

Cette intégrale est majorée en module par l'intégrale de son module, qui converge pour n > 0 et  $f_b > 0$ . Notre intégrale existe donc et s'écrit :

$$I(f) = \lim_{A \to \infty} \lim_{\epsilon \to 0} \int_{\epsilon}^{A} t^{n-1} e^{-2\pi (f_b + if)t} \mathrm{d}t \,. \tag{B.6}$$

Appelons  $I_l(f)$  l'intégrale de  $\epsilon$  à A. Posons  $re^{i\theta} = 2\pi(f_b + if)$ . On remarque que  $f_b > 0$ implique  $\theta \in ]-\frac{\pi}{2}; \frac{\pi}{2}[$ . Procédons ensuite au changement de variable :  $z = (2\pi f_b + i2\pi f)t$ .

$$I_l(f) = (2\pi f_b + i2\pi f)^{-n} \int_{re^{i\theta}\epsilon}^{re^{i\theta}A} z^{n-1} e^{-z} \mathrm{d}z$$
(B.7)

Considérons à présent le lacet rectifiable  $\gamma = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)$  avec :

- $-\gamma_1$  le segment allant de  $re^{i\theta}\epsilon$  à  $re^{i\theta}A$ ;
- $-\gamma_2$  l'arc allant de  $re^{i\theta}A$  à rA;
- $-\gamma_3$  le segment allant de rA à  $r\epsilon$ ;
- $\gamma_4$  l'arc allant de  $r\epsilon$  à  $re^{i\theta}\epsilon$ .

La fonction complexe  $z^{n-1}e^{-z}$  ne possède aucun point singulier à l'intérieur du domaine limité par  $\gamma$ , donc :

$$\int_{\gamma} z^{n-1} e^{-z} \mathrm{d}z = 0.$$
 (B.8)

En conclusion, nous obtenons :

$$\int_{\gamma_1} z^{n-1} e^{-z} dz = -\int_{\gamma_2} z^{n-1} e^{-z} dz - \int_{\gamma_3} z^{n-1} e^{-z} dz - \int_{\gamma_4} z^{n-1} e^{-z} dz \,. \tag{B.9}$$

Calculons ces trois intégrales.

#### 2.1.1 Intégrale sur $\gamma_2$

L'intégrale sur  $\gamma_2$  de l'Équation (B.9) se calcule en réalisant le changement de variable  $z = rAe^{i\phi}$ , ainsi :

$$\int_{\gamma_2} z^{n-1} e^{-z} \mathrm{d}z = \int_{\theta}^{0} i(rA)^n e^{-rA(\cos\phi + i\sin\phi) + in\phi} \mathrm{d}\phi \,. \tag{B.10}$$

En remarquant que quelque soit  $\phi$ , on a  $e^{-rA\cos\phi} \leq e^{-rA\cos\theta}$ , on peut majorer le module de notre intégrale de la manière suivante :

$$\left| \int_{\gamma_2} z^{n-1} e^{-z} \mathrm{d}z \right| \le (rA)^n \left| \int_{\theta}^0 e^{-rA\cos\phi} \mathrm{d}\phi \right| \le (rA)^n e^{-rA\cos\theta} |\theta| \,. \tag{B.11}$$

Rappelons que notre objectif sera de faire tendre A vers l'infini. On a finalement :

$$\lim_{A \to \infty} \int_{\gamma_2} z^{n-1} e^{-z} \mathrm{d}z = 0.$$
 (B.12)

#### **2.1.2** Intégrale sur $\gamma_3$

Pour le calcul de l'intégrale sur  $\gamma_3$  de l'Équation (B.9), on garde cette fois le signe – pour avoir :

$$\int_{\gamma_3} -z^{n-1} e^{-z} dz = \int_{r\epsilon}^{rA} z^{n-1} e^{-z} dz.$$
 (B.13)

Le passage à la limite donne directement :

$$\lim_{\epsilon \to 0} \lim_{A \to \infty} \int_{\gamma_3} -z^{n-1} e^{-z} dz = \Gamma(n) = (n-1)!.$$
 (B.14)

#### **2.1.3** Intégrale sur $\gamma_4$

À l'instar de la Section 2.1.1, on réalise le changement de variable  $z = r\epsilon e^{i\phi}$  pour obtenir :

$$\int_{\gamma_4} z^{n-1} e^{-z} dz = \int_0^\theta i(r\epsilon)^n e^{-r\epsilon(\cos\phi + i\sin\phi) + in\phi} d\phi, \qquad (B.15)$$

dont on peut majorer le module par  $(r\epsilon)^n e^{-r\epsilon\cos\theta}|\theta|$ . On évalue finalement la limite :

$$\lim_{\epsilon \to 0} \int_{\gamma_4} z^{n-1} e^{-z} dz = 0.$$
 (B.16)

#### 2.1.4 Conclusion

On déduit des Sections 2.1.1, 2.1.2 et 2.1.3 la limite :

$$\lim_{\epsilon \to 0} \lim_{A \to \infty} \int_{r\epsilon e^{i\theta}}^{rAe^{i\theta}} z^{n-1} e^{-z} dz = \Gamma(n) .$$
 (B.17)

Finalement, nous avons trouvé la transformée de Fourier de la distribution Gamma :

$$I(f) = \Gamma(n)(2\pi f_b + i2\pi f)^{-n}.$$
 (B.18)

#### 2.2 Calcul final

Grâce au résultat de la Section 2.1 et à l'Équation (B.4), nous allons à présent pouvoir déterminer la transformée de Fourier de la fonction Gammatone. Pour cela, rappelons d'abord que pour une fonction H(f) :

$$[H(f) * \delta(f - f_c)](f) = \int_{-\infty}^{\infty} H(f - x)\delta(x - f_c)dx = H(f - f_c).$$
(B.19)

La transformée de Fourier du cosinus et la linéarité du produit de convolution nous permettent d'obtenir notre résultat final :

$$G(f) = \frac{\Gamma(n)}{2} \left(2\pi f_b + i2\pi (f - f_c)\right)^{-n} + \frac{\Gamma(n)}{2} \left(2\pi f_b + i2\pi (f + f_c)\right)^{-n} .$$
(B.20)

La Figure B.2 représente le module (en décibels) du spectre de la fonction gammatone précédemment représentée dans le domaine temporel (cf. Figure B.1).



FIG. B.2 – Gain d'un filtre gammatone avec une fréquence centrale à 1 kHz

# ANNEXE C

# Estimation des paramètres d'une distribution de von Mises

## 1 Définition

La distribution de von Mises a une fonction de densité de probabilité donnée par :

$$f(\theta|\mu,\kappa) = \frac{e^{\kappa\cos(\theta-\mu)}}{2\pi I_0(\kappa)},$$
(C.1)

avec  $I_0$  la fonction de Bessel modifiée d'ordre 0. f prend comme variable  $\theta$ , un angle en radian. Elle est centrée en  $\mu$  (angle en radians) et  $\kappa$  représente la concentration de la distribution autour de  $\mu$ . Le paramètre  $\kappa$  est similaire à l'inverse de la variance d'une loi normale exprimée en « radian<sup>2</sup> » :

$$\kappa \approx \frac{1}{\sigma^2}$$
(C.2)

L'avantage de l'utilisation de cette distribution pour des données circulaires par rapport à l'utilisation d'une loi normale classique provient de sa  $2\pi$ -périodicité : quand une loi normale interprèterait une grosse différence entre des résultats à 0° et 359°, la distribution de von Mises fait le lien entre 0° et 360°.

L'estimation de ses paramètres est toutefois plus compliquée, principalement pour la concentration  $\kappa$ . Une méthode efficace est toutefois proposée dans (Sra, 2012).

## 2 Estimation de la moyenne

Soit un échantillon de N observations  $\theta_n$  issues d'une distribution de von Mises. On considère la variable  $z_n = e^{i\theta_n}$ . On définit sa moyenne  $\bar{z}$  comme :

$$\bar{z} = \frac{1}{N} \sum_{n=1}^{N} z_n \,.$$
 (C.3)

Alors l'argument de  $\bar{z}$  constitue un estimateur de la moyenne  $\mu$  compris entre  $-\pi$  et  $\pi$ .

## 3 Estimation de la concentration

On considère à présent la statistique  $\bar{R}$  égale au module de la moyenne  $\bar{z}$  :

$$\bar{R} = \sqrt{\bar{z}\bar{z^*}} \tag{C.4}$$

 $\overline{R}$  constitue un estimateur du rapport  $\frac{I_1(\kappa)}{I_0(\kappa)}$ . L'approximation la plus simple de  $\kappa$  est alors donnée par la formule de (Banerjee et al., 2005) :

$$\kappa_0 = \frac{\bar{R}(2 - \bar{R}^2)}{1 - \bar{R}^2} \tag{C.5}$$

D'après (Sra, 2012), cette approximation peut être améliorée en réalisant quelques itérations de la méthode de Newton pour résoudre  $\frac{I_1(\kappa)}{I_0(\kappa)} - \bar{R} = 0$ . Deux itérations sont suffisantes pour obtenir une bonne estimation de  $\kappa$  sans consommer trop de temps de calcul. En définissant  $A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}$ , on a :

$$\kappa_{1} = \kappa_{0} - \frac{A(\kappa_{0}) - \bar{R}}{1 - A(\kappa_{0})^{2} - \frac{1}{\kappa_{0}} A(\kappa_{0})} \\ \kappa_{2} = \kappa_{1} - \frac{A(\kappa_{1}) - \bar{R}}{1 - A(\kappa_{1})^{2} - \frac{1}{\kappa_{1}} A(\kappa_{1})}$$
(C.6)

# ANNEXE D

# Décomposition en composantes latérale et polaire

Dans le cadre de l'amélioration du modèle d'appariement, nous définissons une méthode de décomposition de probabilités dans la dimension azimutale en un produit d'une composante polaire et d'une composante latérale.

## 1 Définition

Nous considérons une distribution de probabilité P dépendant de l'azimut  $\theta$  (dépendance en  $\theta_i$  dans le cas du modèle d'appariement). Par définition, on a :

$$\sum_{\theta=0}^{2\pi} P(\theta) = 1.$$
 (D.1)

L'azimut  $\theta$  représente une position sur le plan horizontal, correspondant à une élévation  $\phi$  égale à  $\frac{\pi}{2}$ . Comme nous l'avons vu dans l'Annexe A, cette position peut être exprimée à l'aide d'un angle latéral  $\alpha$  et d'un angle polaire  $\beta$ . Un azimut  $\theta$  compris entre 0 et  $2\pi$  correspond à un angle latéral compris entre  $-\frac{\pi}{2}$  et  $\frac{\pi}{2}$  et à un angle polaire égal à 0 ou  $\pi$  (mod  $2\pi$ ). On écrit alors de manière équivalente  $P(\theta) = P(\alpha, \beta)$ .

Nous supposons que, dans le cadre de la localisation sonore, la perception de l'angle polaire est indépendante de l'angle latéral. On peut donc écrire :

$$P(\alpha, \beta) = P(\alpha) P(\beta).$$
 (D.2)

Cette équation représente la décomposition d'une probabilité dans la dimension azimutale en un produit d'une composante latérale et d'une composante polaire.

Le calcul de la composante latérale (sur le plan horizontal) se fait de façon à obtenir une valeur indépendante de l'angle polaire, soit :

$$P(\alpha) = P(\alpha, \beta = 0) + P(\alpha, \beta = \pi).$$
(D.3)

Les deux positions utilisées pour ce calcul sont illustrées sur la Figure D.1, dans le cas d'un angle latéral  $\alpha = 22,5^{\circ}$ . Sur l'axe interaural ( $\alpha = \pm \pi$ ), la probabilité est déjà indépendante de l'angle polaire, donc  $P(\alpha) = P(\theta)$ . La composante latérale de l'Équation D.2 représente une distribution de probabilité dans la dimension latérale. On a donc :

$$\sum_{\alpha = -\frac{\pi}{2}}^{\frac{\pi}{2}} P(\alpha) = 1.$$
 (D.4)

Comme il est représenté sur la Figure D.1 dans le cas où  $\beta = 0$ , la composante polaire de la décomposition se calcule selon l'équation :

$$P(\beta) = \sum_{-\frac{\pi}{2} < \alpha < \frac{\pi}{2}} P(\alpha, \beta) .$$
 (D.5)

Les positions correspondant à l'axe interaural ne sont pas prises en compte, puisqu'elles appartiennent aux deux hémisphères avant et arrière. Pour cette raison, la composante polaire doit être normalisée afin de représenter une distribution de probabilité dans la dimension polaire. On a donc finalement  $P(\beta = 0) + P(\beta = \pi) = 1$ .



FIG. D.1 – Décomposition de la probabilité au point d'azimut  $\theta = 22,5^{\circ}$  en composantes latérale ( $\alpha = 22,5^{\circ}$ ) et polaire ( $\beta = 0^{\circ}$ ).

Pour finir, si une probabilité dans la dimension azimutale est décomposée en composantes latérale et polaire, sa reconstruction dans la dimension azimutale selon l'Équation D.2 n'est pas forcément exacte. Nous avons en effet supposé l'indépendance des deux composantes. Il aurait donc fallu, par exemple, écrire :

$$P(\alpha, \beta) = P(\alpha) P(\beta|\alpha), \qquad (D.6)$$

où la composante polaire est exprimée « sachant »  $\alpha$ . Toutefois, l'observation des données de l'expérience subjective a montré que l'hypothèse d'indépendance est justifiée.

## 2 Exemple

Dans le cadre de l'optimisation du modèle d'appariement, nous utilisons les résultats moyens de l'expérience subjective de localisation en condition d'écoute normale. Ils sont transformés en probabilités dans la dimension azimutale pour chaque position de source  $\theta_i$ . Pour chaque position testée  $\theta_j$ , on obtient la probabilité de localisation  $P(\theta_i, \theta_j)$ , représentée sur la Figure D.2a. Elle peut être notée de manière équivalente  $P(\theta_i, \alpha, \beta)$ . Cette probabilité est décomposée en composantes latérale et polaire, respectivement représentées pour chaque position de source  $\theta_i$  sur les Figures D.2b et D.2c.

La première figure montre que les performances de localisation sont bonnes (probabilité importante sur la diagonale principale). Ceci se traduit tout d'abord par une bonne perception de l'angle latéral (Figure D.2b). Dans cette dimension, on peut observer les différences de performances de localisation entre les sources réelles et virtuelles. Enfin, la composante polaire montre que très peu d'erreurs avant-arrière sont commises. Pour les positions proches du plan médian, la probabilité de localiser la source sonore dans le bon hémisphère est très haute. En revanche, la probabilité de confusion avant-arrière augmente quand la source sonore se rapproche de l'axe interaural, qui correspond à la séparation entre les hémisphères avant et arrière.



FIG. D.2 – Exemple de décomposition d'une distribution dans la dimension azimutale (en haut) en une composante latérale (à gauche) et une composante polaire (à droite).

# ANNEXE E

# Matrices de probabilité de localisation sonore

Dans notre étude, nous avons jugé que les modèles permettant d'obtenir les meilleures prédictions des performances de localisation sonore sont le modèle d'appariement, basé sur la décomposition latérale/polaire avec des paramètres dépendant de la position de la source, et le réseau de neurones entraîné à partir des données de l'expérience subjective avec un seuil d'apprentissage faible (E = 0,01).

Les distributions de probabilités résultant de l'application des deux modèles de localisation sonore aux mesures de DTF avec ou sans TCAPS peuvent être représentées sous la forme de matrices ( $16 \times 16$ ), dont les valeurs sont codées sur une échelle de couleur. La présente annexe répertorie ces figures pour chaque condition d'écoute, chaque modèle et chaque tête artificielle utilisée :

- la Figure E.1 présente les prédictions avec et sans TCAPS du modèle d'appariement pour la tête artificielle ISL;
- la Figure E.2 présente les prédictions avec et sans TCAPS du modèle d'appariement pour le mannequin KEMAR;
- la Figure E.3 présente les prédictions avec et sans TCAPS du réseau de neurones pour la tête artificielle ISL;
- la Figure E.4 présente les prédictions avec et sans TCAPS du réseau de neurones pour le mannequin KEMAR.



FIG. E.1 – Sorties du modèle d'appariement, optimisé pour la tête artificielle ISL, pour chaque couple de positions cible (Azimut réel) et test (Azimut testé), dans les sept conditions d'écoute.



FIG. E.2 – Sorties du modèle d'appariement, optimisé pour le mannequin KEMAR, pour chaque couple de positions cible (Azimut réel) et test (Azimut testé), dans les sept conditions d'écoute.



FIG. E.3 – Sorties du réseau de neurones, optimisé pour la tête artificielle ISL, pour chaque couple de positions cible (Azimut réel) et test (Azimut testé), dans les sept conditions d'écoute.



FIG. E.4 – Sorties du réseau de neurones, optimisé pour le mannequin KEMAR, pour chaque couple de positions cible (Azimut réel) et test (Azimut testé), dans les sept conditions d'écoute.

# Glossaire

#### Angle latéral

Dans un système de coordonnées polaires interaurales, l'angle latéral désigne l'écart par rapport au plan médian..

#### Angle polaire

Dans un système de coordonnées polaires interaurales, l'angle polaire désigne une position sur le cône de confusion défini par un angle latéral.

#### Binaural

Qui se rapporte aux deux oreilles. Un indice de localisation binaural, comme l'ITD, est déterminé à partir des signaux reçus aux deux oreilles..

#### Cepstre

Transformée d'un signal dans un domaine analogue au domaine temporel, à la base de l'analyse cepstral. Le cepstre est généralement défini comme la transformée de Fourier inverse du logarithme naturel du module de la transformée de Fourier du signal initial. Anagramme de spectre..

#### Contralatéral

Qui se trouve du côté opposé du corps. On appelle oreille contralatérale l'oreille se trouvant du côté opposé à la source sonore..

#### Cône de confusion

Surface iso-ITD et iso-ILD, pour laquelle ces indices ne permettent pas une localisation exacte. Il s'agit d'une approximation d'une hyperboloïde en champ lointain..

#### Gammatone

Un filtre gammatone est défini par sa réponse impulsionnelle égale au produit d'une fonction Gamma et d'un cosinus. Ce type de filtre est utilisé pour simuler la sélection fréquentielle dans la cochlée..

#### Homoscédasticité

Critère d'homogénéité des variances entre plusieurs échantillons. Il s'agit d'une condition d'application de l'analyse de variance (ANOVA)..

#### Interaural

Entre les deux oreilles. Une différence interaurale est calculée comme une différence entre les deux oreilles. L'axe interaural est l'axe passant par les deux oreilles.

#### Ipsilatéral

Qui se trouve du même côté du corps. Dans ce mémoire, on désigne comme oreille ipsilatérale l'oreille se trouvant du même côté que la source sonore..

### Liftrage

Processus de filtrage d'un cepstre. Anagramme de filtrage..

#### Monaural

Qui ne dépend que d'une oreille. Un indice de localisation monaural (les indices spectraux dans ce mémoire) est obtenu à partir du signal reçu à une seule oreille.

#### Perceptron

Modèle basique de neurone pour les réseaux de neurones. Il s'agit d'un neurone de McCulllogh-Pitts dont la fonction d'activation est une sigmoïde..

#### Psychoacoustique

Branche de l'acoustique reliant les caractéristiques physiques des sons qui arrivent aux oreilles et la perception de ces sons par l'auditeur..

#### Situational awareness

Connaissance, ou conscience, de la situation. Dans notre contexte, ce terme anglais désigne l'ensemble des capacités utilisées par l'auditeur pour percevoir et interpréter son environnement sonore..

#### Talk-through

Désigne une protection auditive active dont l'atténuation dépend du niveau sonore. Un système talk-through est généralement complémentaire aux systèmes de radiocommunication..

#### Tonotopie cochléaire

Organisation spatiale de la cochlée pour la séparation fréquentielle..

le cnam

## Thomas JOUBAUD Prédiction objective de l'effet des systèmes tactiques de communication et protection sur les performances de localisation sonore



#### Résumé :

Dans de nombreuses situations civiles ou militaires, la protection de l'audition du personnel est cruciale. La perception et l'interprétation de l'environnement sonore par l'auditeur doivent cependant être préservées. Les Systèmes Tactiques de Communication et Protection (TCAPS) sont des protections auditives qui, à la fois, protègent suffisamment les oreilles de l'auditeur contre les bruits dangereux, et préservent l'intelligibilité, permettant ainsi la communication vocale à bas niveau. Des études précédentes ont cependant démontré que les TCAPS continuent de détériorer la perception de l'environnement sonore de l'auditeur, en particulier sa capacité à localiser les sources sonores. Sur le plan horizontal, la dégradation des indices acoustiques empêchant, en temps normal, l'auditeur de confondre les sources avant et arrière, en est la principale explication. Dans ce travail de thèse, une expérience subjective de localisation sonore est conduite avec six TCAPS : deux bouchons d'oreille passifs, deux bouchons d'oreilles actifs et deux casques actifs. Si aucun protecteur ne permet de retrouver les performances de la condition d'écoute normale, l'expérience permet de classifier les TCAPS du point de vue de la localisation sonore : les performances des bouchons passifs sont meilleures que celles des bouchons actifs, et les casques actifs engendrent les plus mauvaises performances. Dans le cadre de la conception et de l'évaluation des TCAPS, une méthode prédisant leur dégradation des performances de localisation sonore, basée sur des mesures électroacoustiques, serait plus adaptée que des expériences comportementales très chronophages. Dans ce contexte, deux méthodes basées sur les Fonctions de Transfert Relatives à la Tête (HRTF) mesurées sur tête artificielle sont étudiées : un processus d'appariement et un réseau de neurones à trois couches. Ils sont optimisés pour reproduire les performances de localisation humaine en condition d'écoute normale. Les méthodes sont ensuite appliquées aux HRTF mesurées avec les six TCAPS, et prédisent des probabilités de localisation en fonction de la position. Comparé aux résultats de l'expérience subjective, le réseau de neurones prédit des performances réalistes avec les bouchons d'oreille, mais surestime les erreurs avec les casques. Le modèle d'appariement prédit correctement les performances de localisation. Toutefois, la vraisemblance de ses distributions de probabilité avec les observations subjectives demeure plus faible que celle du réseau de neurones. Pour finir, les deux méthodes développées dans cette étude sont indépendantes de la tête artificielle utilisée, et peuvent être utilisées pour évaluer non seulement des prototypes de TCAPS, mais aussi des prothèses auditives.

#### Mots clés :

Protection auditive, localisation sonore, psychoacoustique, HRTF, ITD, ILD, indices spectraux, réseau de neurones

#### Abstract :

In many civilian or military situations, hearing protection is of major importance. The listener's acoustical situational awareness must however also be preserved. Tactical Communication and Protective Systems (TCAPS) are hearing protection devices that sufficiently protect the listener's ears from hazardous sounds and preserve speech intelligibility, thus allowing low-level speech communication. However, previous studies demonstrated that TCAPS still deteriorate the listener's situational awareness, in particular the ability to locate sound sources. On the horizontal plane, this is mainly explained by the degradation of the acoustical cues normally preventing the listener from making front-back confusions. In the present PhD work, a behavioral sound localization experiment is conducted with six TCAPS: two passive and two active earplugs, and two active earmuffs. The performance in open ear condition is not retrieved with any protector, but the experiment ranks the TCAPS by type : passive earplugs lead to better performance than active earplugs, and active earmuffs induce the worst performance. As part of TCAPS development and assessment, a method predicting the protector-induced degradation of the sound localization capability, and based on electroacoustic measurements, would be more suitable than time-consuming behavioral experiments. In this context, two methods based on Head-Related Transfer Functions (HRTFs) measured on an artificial head are investigated : a template-matching model and a three-layer neural network. They are optimized to fit human sound localization performance in open ear condition. The methods are applied to the HRTFs measured with the six TCAPS, providing position-dependent localization probabilities. Compared with the behavioral results, the neural network predicts realistic performances with earplugs, but overestimates errors with earmuffs. The template-matching model predicts human performance well. However, the likelihood of the resulting probability distributions with the behavioral observations is lower than that of the neural network. Finally, both methods developed in this study are independent of the artificial head used, and can be applied to assess not only TCAPS prototypes, but also hearing aids.

#### Keywords :

Hearing protection, sound localization, psychoacoustic, HRTF, ITD, ILD, spectral cues, neural network