

TABLE DES MATIÈRES

REMERCIEMENTS	2
LISTE DES ABREVIATIONS	7
TABLE DES ILLUSTRATIONS.....	9
INTRODUCTION.....	19
OBJECTIFS DE LA THESE	22
Fondamental	22
Clinique.....	22
Opérationnel	22
PARTIE 1. ÉTAT DE L'ART	24
CHAPITRE I. SITUATIONS MULTILOCUTEURS.....	24
I. Le parcours du son de l'oreille au cerveau.....	24
II. Ségrégation des flux sonores	26
III. Masquage informationnel et masquage énergétique	27
IV. Target-to-masker ratio.....	28
V. Séparation spatiale des locuteurs.....	32
VI. Attention auditive sélective : focalisation et inhibition	34
CHAPITRE II. NEURO-ANATOMIE FONCTIONNELLE DU TRAITEMENT DU LANGAGE	37
I. La voie dorsale et la voie ventrale	37
II. Le traitement du langage en situation d'écoute complexe : core speech network.....	38
II.1. Gyrus frontal inférieur (IFG)	39
II.1.1. L'IFG et l'effort d'écoute	40
II.1.2. L'IFG dans l'entraînement	41
II.2. Cortex préfrontal dorsolatéral.....	43
III. Du langage aux mécanismes <i>general domain</i>	44
III.1. Le réseau cingulo-operculaire.....	46
III.2. Le réseau fronto-pariétal	48
III.3. Collaboration entre les réseaux fronto-pariétal et cingulo-operculaire	49
CHAPITRE III. L'EFFORT D'ECOUTE.....	50
I. Charge cognitive	50
II. La nature multidimensionnelle de l'effort d'écoute	50
III. Facteurs influençant l'effort d'écoute	54
III.1. Ressources cognitives : attention, mémoire.....	54
III.2. Motivation	56
III.3. L'état auditif et la difficulté de son évaluation	57
III.3.1. Les lésions auditives chez le personnel militaire	57
III.3.3. L'audiométrie tonale liminaire	58
III.3.4. Otoémission	59
IV. Mesures pour évaluer l'effort d'écoute.....	64
IV.1. Évaluations subjectives.....	64
IV.2. Mesures comportementales.....	65
IV.3. Mesures psychophysologiques	66
IV.3.1. Neurophysiologique	67
IV.3.2. La spectroscopie en proche infrarouge (f/NIRS).....	67
V. Pourquoi étudier l'effort d'écoute ?	74
PARTIE 2. CONTRIBUTIONS EXPERIMENTALES.....	77
CHAPITRE I. PREMIERE ETUDE	78
I. Introduction	78
II. Matériels et méthode	80
III. Acquisition des données	81

III.1. fNIRS	81
III.2. Prétraitement données fNIRS	82
IV. Analyses	83
IV.1. Comportement – Phase initiale	83
IV.2. Comportement – Phase d’entraînement	83
IV.3. fNIRS – Phase initiale	84
IV.3.1. Activation de la tâche	84
IV.3.2. Relation entre l’activité corticale et les scores d’intelligibilité de la parole	84
IV.4. fNIRS – Phase d’entraînement	84
V. Résultats – Phase initiale	85
V.1. Comportement	85
V.2. f/NIRS	86
V.2.1. Effet de la condition	86
V.2.2. Corrélation comportement et activité corticale	86
VI. Résultats – Phase d’entraînement	88
VI.1. Comportement	88
VI.2. f/NIRS	89
VII. Discussion générale	89
CHAPITRE II. DEUXIEME ETUDE	93
I. Introduction	93
II. Matériels et méthode	94
II.1. Participants	94
II.2. Tâche d’intelligibilité	95
III Acquisition des données	97
III.1 f /NIRS	97
III.2. Prétraitement données f/NIRS	99
IV. Analyses	100
IV.1. Comportement	100
IV.1.1 f/NIRS	100
V. Résultats	100
V.1. Comportement	100
V.2. f/NIRS	102
VI. Discussion	104
CHAPITRE III. TROISIEME ÉTUDE	107
I. Introduction	107
II. Matériels et méthode	108
II.1. Participants	108
II.2. Plan expérimental	109
II.3. Mesures subjectives	110
II.3.1. Questionnaire démographique	110
II.3.2. Effort Scaling Categorical Unit - ESCU	110
II.3.3. Motivation	111
II.3.4. Speech, Spatial and Qualities Hearing	112
II.4. Mesures audiologiques	113
II.4.1. Audiométrie	113
II.4.2. Otoémissions	113
II.5. Caractéristiques des participants	113
III. Analyses	115
III.1. Tâche d’intelligibilité – CRM	115
IV. Résultats	117
IV.1. Effet de l’écartement des voix masques de 20° à 60° sur l’intelligibilité	117
IV.2. Effet de l’écartement des voix masques de 20° à 60° sur l’effort d’écoute ressenti	120
IV.3. Relation entre la variation de l’intelligibilité et la variation de l’effort	122
IV.4. Relation entre la différence d’intelligibilité liée à un écartement des voix masques de 20° à 60° et l’état auditif	122
IV.4.1. Audiométries tonales liminaires classique et haute fréquence	122
IV.4.2. Produits de distorsion acoustique (PDA)	126
IV.5. Relation entre la variation de l’effort d’écoute ressenti entre 20° à 60° et l’état auditif	129
IV.5.1. Audiométries tonales liminaires classique et haute fréquence	129
IV.5.2. Produits de distorsion acoustique (PDA)	132
IV.6. Compréhension de la voix cible dans le bruit	133

IV.7. Intelligibilité et motivation	136
IV.8. Relation entre la mesure de la motivation et les variables mesurées dans les différentes conditions expérimentales.	137
IV.9. Relation entre l'âge et les variables mesurées dans les différentes conditions expérimentales.	139
IV.10. Relation entre les réponses au SSQ et les variables mesurées pertinentes.	139
IV.11. Relation entre la variabilité interindividuelle observée à 60° de séparation et les autres mesures.	139
V. Discussion	143
SYNTHESE.....	148
PREMIERE ETUDE : L'ENTRAINEMENT A L'UTILISATION DU TMR CONDUIT A UNE AMELIORATION DE L'INTELLIGIBILITE ET A UN MOINDRE ENGAGEMENT DE RESSOURCES COGNITIVES	149
DEUXIEME ETUDE : LA SEPARATION SPATIALE DES LOCUTEURS AMELIORE L'INTELLIGIBILITE AU PRIX D'UN ENGAGEMENT PLUS GRAND DE RESSOURCES COGNITIVES.....	152
TROISIEME ETUDE : LA SEPARATION SPATIALE AMELIORE L'INTELLIGIBILITE ET REDUIT L'EFFORT D'ECOUTE RESSENTI	155
DISCUSSION GENERALE	159
POURQUOI L'EFFORT D'ECOUTE ?.....	159
L'EFFORT D'ECOUTE, UN CONCEPT MULTIFACTORIEL.....	161
MOTIVATION 162	
CONCLUSION.....	166
BIBLIOGRAPHIE	168
ANNEXE 1. ARTICLE	186
ANNEXE 2. FICHE DE RENSEIGNEMENT.....	218
RESUME :	222
ABSTRACT :	223

Liste des Abréviations

AC : Cortex Auditif

CAE : Conduit Auditif Externe

CCE : Cellule Ciliée Externe

CCI : Cellule Ciliée Interne

CRM : Coordinate Response Measure

DLPFC : DorsoLateral Prefrontal Cortex

ESCU : Effort Scaling Categorical Unit

FNIRS : Functional Near Infra-Red Spectroscopy

FUEL : Framework for Understanding Effortful Listening

HBO : Oxyhémoglobine

HBB : Desoxyhémoglobine

HRTF : Head Related Transfer Function

IFG : Inferior Frontal Gyrus

IPC : Inferior Parietal Cortex

IPL : Inférieur Parietal Lobe

ISI : Interval InterStimulus

MNI : Institut neurologique de Montréal

MOLE : Model Of Listening Engagement

MTG : Middle Temporal Gyrus

NIRS : Near Infra-Red Spectroscopy

OEA : Otoémissions Acoustiques

OEAS : Otoémissions Acoustiques Spontanées

OEAP : Otoémissions Acoustiques Provoquées

PDA : Produit de Distorsion Acoustique

PFC : Prefrontal Cortex

ROI : Regions of Interest

SSQ : Speech Spatial and Qualities Hearing

SPC : Superior Parietal Cortex

SPL : Superior Parietal Lobe

STG : Superior Temporal Gyrus

TMR : Target-to-Masker Ratio

TSA : Trauma Sonore Aigu

V1 : Visual Cortex

Table des illustrations

Figure 1 : Dessin de l'appareil auditif. L'appareil auditif humain se compose de l'oreille externe, formée par le pavillon et le conduit auditif externe (A), de l'oreille moyenne, qui contient la chaîne des osselets (B), et de la cochlée formant avec le vestibule, l'oreille interne (C) ; (Goldstein, 2008). 25

Figure 2 : La cochlée et l'organe de Corti. Les cellules sensorielles auditives se trouvent dans l'organe de Corti entre les membranes basilaire et tectoriale. Les cellules ciliées externes (CCE) ont leurs cils encastrés dans la membrane tectoriale. Ils sont cisailés lors du mouvement de la membrane basilaire, ce qui provoque l'activation des CCE qui facilite l'activation des cellules ciliées internes (CCI). Les CCI activées émettent au niveau du nerf auditif des messages nerveux codant certaines caractéristiques du signal sonore (Goldstein, 2008). 26

Figure 3 : Scores de réponses correctes (A) et activité corticale (B) en fonction de la condition du TMR. Les boîtes indiquent la médiane (barre horizontale) et l'écart interquartile (cases), les moustaches indiquent l'intervalle de 5 à 95 percentiles. Les *U-shape performers* (en rouge) montrent un score d'intelligibilité similaire pour les TMR défavorables et favorables et une [HbO2] dans le cortex préfrontal qui augmente avec le TMR. Les *non-U-shape performers* (en vert) montrent un score d'intelligibilité qui diminue avec le TMR et une [HbO2] similaire pour tous les TMR (d'après Andéol et al., 2017). 30

Figure 4 : Les scores de réponses correctes sont représentés en fonction des variations de l'activité préfrontale [HbO2] pour chaque combinaison de conditions de TMR et de séparation spatiale. La lettre "U" indique les *U-shape performers* et la lettre "N" indique les *non-U-shape performers*. Le coefficient de corrélation de Pearson et son degré de signification sont indiqués dans chaque panneau. Pour les locuteurs colocalisés, à TMR défavorable, plus l'activité préfrontale était élevée, plus la performance était élevée (d'après Andéol et al., 2017). 31

Figure 5 : Aires cérébrales de l'hémisphère gauche impliquées dans le traitement du langage (d'après Friederici & Gierhan, 2013). 37

Figure 6 : Le modèle à double voie de l'anatomie fonctionnelle du langage : la voie dorsale (en bleu) et la voie ventrale (en rose) dans les deux hémisphères. Les premières étapes du traitement du langage sont dans les régions auditives bilatérales (STG en vert et STS en jaune). Depuis ces deux régions partent les deux voies : la voie ventrale qui supporte la compréhension du langage (IFG), la voie dorsale (plus latéralisée à gauche) supporte les intégrations sensori-motrices pour la production du langage incluant les structures de la jonction temporo-pariétale (TPJ) et le lobe frontal (d'après Hickok & Poeppel, 2007). 38

Figure 7 : Les différents réseaux cérébraux impliqués dans le traitement du langage selon la qualité du stimulus, acoustiquement intelligible ou dégradé. Un signal de parole acoustiquement intelligible sollicitera l'activation du *core speech network* afin de permettre son traitement acoustique, phonologique, lexico-sémantique et de la syntaxe. En revanche un signal de parole

acoustiquement dégradé impliquera l'activation d'autres régions cérébrales notamment celles frontales et pariétales afin de permettre le monitoring de la performance et l'emploi de ressources supplémentaires en mémoire et attention (d'après Peelle, 2018). 45

Figure 8 : Modèle de l'effort d'écoute d'après Peelle, 2018. A) Le défi acoustique (composé par l'auditeur, le signal, l'environnement) engendre une demande cognitive qui est modulée par la motivation. B) Les différentes stratégies pour évaluer l'effort d'écoute (mesures comportementales, mesures de neuro-imagerie et mesures physiologiques)..... 52

Figure 9 : Model of Listening Engagement, d'après Herrmann & Johnsrude, 2020. Ce schéma illustre la relation entre l'engagement, la motivation et le vécu d'un auditeur. La motivation d'une personne et la relation entre les ressources requises et les limites des ressources (c'est-à-dire les quantités de ressources disponibles) déterminent l'engagement. Ces caractéristiques déterminent le vécu de l'auditeur, ce qui influence également la motivation..... 53

Figure 10 : Exemple de tracé audiométrique avec oreille gauche (en bleu) et oreille droite (en rouge). Capture d'écran du logiciel Audio Console Oscilla®. 59

Figure 11 : Schéma illustratif des structures et des mécanismes impliqués dans le couplage neuro-vasculaire. En détail, le traitement d'un stimulus provoque la sécrétion de glutamate. Le glutamate stimule les neurones et les astrocytes, ce qui entraîne la sécrétion d'oxyde nitrique (NO), de potassium (K⁺), d'adénosine (Ado), d'acides époxyeicosatriénoïques (EET) et de prostaglandines (PGE₂), ce qui provoque une vasodilatation artériolaire. Les astrocytes sécrètent également de l'acide arachidonique (AA), qui provoque une vasoconstriction. Un deuxième mécanisme est la stimulation des péricytes, qui entraîne une vasodilatation capillaire (depuis Hendrikx et al., 2019). 69

Figure 12 : Schéma simplifié de la réponse hémodynamique en fNIRS. Le couplage neuro-vasculaire entraîne une augmentation du flux sanguin cérébrale avec une concomitante augmentation de l'oxy-hémoglobine et une dilution de la déoxy-hémoglobine..... 70

Figure 13 : Effet de la distance d'écartement entre les optodes sur la résolution spatiale et la profondeur de la mesure en fNIRS..... 71

Figure 14 : Spectres d'absorption (coefficient d'extinction molaire) de l'hémoglobine oxygénée (HbO en rouge) et désoxygénée (HbR en bleue) entre 760 nm et 850 nm (depuis W.B. Gratzer, Med Res Council Labs, Holly Hill London et N. Kollias, Wellman Laboratories, Harvard Medical School, Boston) 72

Figure 15 : Réponse hémodynamique avec une augmentation de l'oxyhémoglobine HbO₂ avec une baisse concomitante de la désoxyhémoglobine [HHb] après une période de stimulation. 72

Figure 16 : A) Photographie du support d'optode-réseau placé sur la tête d'un participant à des fins de démonstration. (B) Représentation 2D de l'emplacement du ROI ; chaque ligne représente un canal de mesure. (C) Vues gauche (haut) et droite (bas) des cartes de sensibilité

corticale (en mm^{-1}) obtenues par simulation de Monte-Carlo dans AtlasViewer (V2.8 ; paramètres par défaut sauf pour le nombre de photons qui a été fixé à 107)..... 82

Figure 17 : Réponses correctes (Rationalized Arcsin Units) en fonction du TMR dans la phase initiale. Les boxplots montrent la médiane (barre horizontale) et l'écart interquartile (cases). Les moustaches vont de la valeur la plus basse à la valeur la plus élevée observées dans une limite de 1,5 fois l'écart interquartile. Les points indiquent les points de données individuels. Les points de données individuels correspondant au même auditeur dans différentes conditions sont reliés par des lignes grises. 85

Figure 18 : Activité corticale (coefficient beta pour HbO) pour l'IFG gauche et l'IPC gauche. * : $p < 0.05$ (FDR-corrigé) pour le nombre de ROI. Les barres verticales indiquent l'erreur standard de la moyenne..... 86

Figure 19 : Représentation 3D des canaux significativement corrélés aux scores RAU (lignes pleines ; $p < 0,05$ FDR-corrigé) : IFG gauche et droit ($r=0,17$ et $r=0,19$ respectivement), cortex auditif gauche ($r=0,18$), cortex préfrontal ($r=0,16$) et le SPC gauche ($r=-0,17$). 87

Figure 20 : Illustration de la relation entre l'activité corticale (coefficients bêta pour HbO) et l'intelligibilité de la parole (scores RAU) pour la condition défavorable. 87

Figure 21 : Scores d'intelligibilité de la parole (RAU) pour chaque TMR (Adverse, Intermédiaire, Favorable), session d'entraînement (Pre, Post), groupe d'entraînement (*driven, not-driven*). Les boxplots montrent les médianes (barres horizontales) et les intervalles interquartiles (cases). Les moustaches vont de la valeur la plus faible à la valeur la plus élevée observées comprise dans 1,5 fois l'intervalle interquartile..... 88

Figure 22 : Effets de l'entraînement des activations IFG gauche et droite pour chaque condition TMR (Adv : défavorable, Int : intermédiaire et Fav : favorable). * : $p < 0,05$ (FDR-corrigé). . 89

Figure 23 : Design expérimental des trois conditions avec la voix cible (en noir) toujours placée à 0° droit devant le participant et les voix masques (en rouge) placées sur les côtés. En détail, la condition 20-60 avec une voix cible et quatre voix masques placées à $\pm 20^\circ$ et $\pm 60^\circ$; la condition 20 avec une voix cible et deux voix masque placées à $\pm 20^\circ$; la condition 60 avec une voix cible et deux voix masque placées à $\pm 60^\circ$ 96

Figure 24 : Boîtier de réponse avec trois rangées de boutons. La première est composée par 4 couleurs (bleu, vert, rouge, blanc) ; la deuxième par les nombres de 1 à 4 ; la troisième par les nombres de 5 à 8..... 97

Figure 25 : Carte de sensibilité (en mm^{-1}) obtenue avec simulation de Monte-Carlo (haut hémisphère gauche, bas hémisphère droit). La sensibilité est la plus grande lorsqu'elle se rapproche de zéro mm^{-1} 98

Figure 26 : Représentation en 2D des ROIs. Chaque ligne représente un canal source-émetteur, chaque couleur, une région d'intérêt pour laquelle les données mesurées à travers les canaux

qui la composent ont été moyennées. Les lignes noir plus courtes représentent les short channels utilisés dans chaque hémisphère cérébral..... 99

Figure 27 : Réponses correctes (Rationalized Arcsin Units) pour les trois conditions de séparation spatiale des locuteurs. Les boxplots montrent la médiane (barre horizontale) et l'écart interquartile (cases). Les moustaches vont de la valeur la plus basse à la valeur la plus élevée observées dans une limite de 1,5 fois l'écart interquartile. Les points indiquent les points de données individuels. Les points de données individuels correspondant au même auditeur dans différentes conditions sont reliés par des lignes grises. 101

Figure 28 : Activations (beta) pour le gyrus frontal inférieur gauche pars traingularis (IIFG *: $p < 0,05$ (FDR-corrige)). Les barres verticales indiquent l'écart-type de la moyenne..... 102

Figure 29 : Activations pour le cortex pariétal inférieur droit (rIPC)*: $p < 0,05$ (FDR-corrige). Les barres verticales indiquent l'écart-type de la moyenne. 102

Figure 30 : Activations pour le cortex préfrontal dorsolatéral droit (rDLPFC) et pour le cortex préfrontal dorsolatéral gauche (IDLDFC) *: $p < 0,05$ (FDR-corrige). Les barres verticales indiquent l'écart-type de la moyenne. 103

Figure 31 : Activations pour le gyrus temporal supérieur droit (rSTG) *: $p < 0,05$ (FDR-corrige). Les barres verticales indiquent l'écart-type de la moyenne. 103

Figure 32 : Echelle ESCU à 13 catégories (traduite d'après Rennie et al., 2019). En haut à gauche, *Code sujet* représente l'identifiant d'un participant ; en haut à droite *Run* indique la condition pour laquelle l'échelle est administrée (dans l'exemple Col : co-localisée)..... 111

Figure 33 : Echelle de motivation adaptée depuis Kramer et al., 2016. Echelle à dix points, allant de '*Je n'ai jamais renoncé à écouter*' (1) à '*Pour toutes les questions, j'ai abandonné*' (10). 112

Figure 34 : Exemple d'une question posée au Speech Spatial and Qualities Hearing (SSQ) dans la version française d'après Moulin et al., 2015. Tout en haut à gauche, *1^{ère} partie : Audition de la parole* indique une de trois parties dont il est composé (Audition de la parole, Audition spatiale et Qualité de l'audition). Ensuite, dans la case à gauche : la première question de cette première partie ; à droite : l'échelle d'évaluation subjective à 10 points (de 0 : *Non, pas du tout* à 10 : *Oui, parfaitement*) plus la case *Non applicable*. 113

Figure 35 : Questionnaire démographique sur les antécédents auditifs des participants 114

Figure 36 : Résultats de l'audiométrie des participants à l'étude. Les fréquences mesurées sont représentées en abscisse ; le seuil d'intensité sonore est représenté en décibel (dB HL) en ordonnée. Chaque ligne noire indique les résultats pour chaque participant, la ligne noire foncé indique la moyenne des réponses entre les participants. La région colorée en gris, l'intervalle interquartile. 115

Figure 37 : Pourcentage de bonnes réponses en fonction des conditions (CRM.Bruit, CRM.Coloc, CRM.2060, CRM.20, CRM.60). Les boxplots montrent la médiane (barre horizontale) et l'écart interquartile (cases). Les moustaches vont de la valeur plus basse à la valeur plus élevée observées dans une limite de 1,5 fois l'écart interquartile. Les points indiquent les points de données individuels. 117

Figure 38 : Pourcentage de bonnes réponses en fonction de la condition (CRM.20 et CRM.60). Les boxplots montrent la médiane (barre horizontale) et l'écart interquartile (cases). Les moustaches vont de la valeur plus basse à la valeur plus élevée observées dans une limite de 1,5 fois l'écart interquartile. Les points indiquent les points de données individuels. Les points de données individuels correspondant au même auditeur dans différentes conditions sont reliés par des lignes grises. 118

Figure 39 : Pourcentage de bonnes réponses dans la condition CRM.60 en fonction du pourcentage de bonnes réponses dans la condition CRM.20, Chaque point représente un participant. La droite bleue représente la droite de régression. Les points correspondant aux participants dans les scores sont égaux entre les deux conditions sont sur la droite rouge. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche. 119

Figure 40 : Variation de l'intelligibilité entre 20° et 60° (Différence_intelligibilité) en fonction du pourcentage de bonnes réponses dans la condition CRM.20. Chaque point représente un participant. La droite bleue représente la droite de régression. Les points correspondant aux participants dans les scores sont égaux entre les deux conditions sont sur la droite rouge. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche. 119

Figure 41 : Score de l'échelle de l'effort d'écoute ressenti (ESCU) en fonction des conditions (CRM.Bruit, CRM.Coloc, CRM.2060, CRM.20, CRM.60). Les boxplots montrent la médiane (barre horizontale) et l'écart interquartile (cases). Les moustaches vont de la valeur plus basse à la valeur plus élevée observées dans une limite de 1,5 fois l'écart interquartile. Les points indiquent les points de données individuels. 120

Figure 42 : Score d'évaluation à l'échelle ESCU en fonction de la condition (CRM.20 et CRM.60). Les boxplots montrent la médiane (barre horizontale) et l'écart interquartile (cases). Les moustaches vont de la valeur plus basse à la valeur plus élevée observées dans une limite de 1,5 fois l'écart interquartile. Les points indiquent les points de données individuels. Les points de données individuels correspondant au même auditeur dans différentes conditions sont reliés par des lignes grises. 120

Figure 43 : Variation de la mesure de l'effort d'écoute ressenti entre 20° et 60° (Différence_Effort) en fonction de l'effort d'écoute ressenti à 20° (ESCU_20). Chaque point représente un participant. La droite bleue représente la droite de régression. Les points correspondant aux participants dans les scores sont égaux entre les deux conditions sont sur la droite rouge. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche. 121

Figure 44 : Variation de la mesure de l'effort d'écoute ressenti entre 20° et 60° (Différence_Effort) en fonction de la variation de l'intelligibilité entre 20° et 60° (Différence_intelligibilité). Chaque point représente un participant. La droite bleue représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche..... 122

Figure 45 : Variation de l'intelligibilité entre 20° et 60° (Difference_intelligibilité) en fonction des seuils audiométriques mesurés à l'oreille gauche. Chaque cadre représente une fréquence de mesure, où chaque point représente un participant. La droite bleue représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche..... 123

Figure 46 : Variabilité de l'intelligibilité entre 20° et 60° (Difference_intelligibilité) en fonction des seuils audiométriques mesurés à l'oreille droite. Chaque cadre représente une fréquence de mesure, où chaque point représente un participant. La droite bleue représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche..... 124

Figure 47 : Variation de l'intelligibilité entre 20° et 60° (Difference_intelligibilité) en fonction des mesures dérivées audiométriques. Chaque cadre représente une fréquence de mesure, où chaque point représente un participant. La droite bleue représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche. 125

Figure 49 : Variation de l'intelligibilité entre 20° et 60° (Différence_intelligibilité) en fonction de l'amplitude des PDA pour les fréquences 1000 à 5000 Hz mesurés à l'oreille gauche. Chaque cadre représente une fréquence de mesure, où chaque point représente un participant. La droite bleue représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche. 127

Figure 50 : Graphe des variables de l'analyse en composante principale. Les variables incluses sont la variation d'intelligibilité entre 20° et 60° (Différence_intelligibilité) et l'amplitude des PDA mesurés à l'oreille gauche pour différentes fréquences. Les 2 premiers axes expriment 72.82% de l'inertie totale du jeu de données ; ce qui peut être considéré comme satisfaisant. La projection des variables est également satisfaisante (extrémité du segment proche du cercle). Les PDA mesurés à 2000 et 3000 Hz semblent les plus proches de la différence d'intelligibilité tandis que les PDA mesurés à 1000, 4000 et 5000 Hz ne semblent pas être liés à cette variable. 128

Figure 51 : Variation de l'effort d'écoute ressenti entre 20° à 60° (Différence_Effort) en fonction des seuils audiométriques mesurés à l'oreille gauche de 250 à 12500 Hz. Chaque cadre représente une fréquence de mesure, où chaque point représente un participant. La droite bleu représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche dans chaque cadre..... 129

Figure 52 : Variation de l'effort d'écoute ressenti entre 20° et 60° (Différence_Effort) en fonction des seuils audiométriques mesurés à l'oreille droite. Chaque cadre représente une fréquence de mesure, où chaque point représente un participant. La droite bleu représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.	130
Figure 53 : Variation de l'effort d'écoute ressenti entre 20° et 60° (Différence_Effort) en fonction des mesures dérivées audiométriques. Chaque cadre représente une fréquence de mesure, où chaque point représente un participant. La droite bleu représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.	131
Figure 54 : Variation de l'effort d'écoute ressenti entre 20° et 60° (Différence_Effort) en fonction des amplitudes des PDA à 65 dB de l'oreille droite. Chaque cadre représente une fréquence de mesure, où chaque point représente un participant. La droite bleu représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.	132
Figure 55 : Variation de l'effort d'écoute ressenti entre 20° et 60° (Différence_Effort) en fonction des amplitudes des PDA à 65 dB de l'oreille gauche. Chaque cadre représente une fréquence de mesure, où chaque point représente un participant. La droite bleu représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.	133
Figure 57 : Corrélation positive entre l'effort d'écoute ressenti dans le bruit (CRM.Bruit) et l'effort d'écoute ressenti dans la condition ESCU.60. La droite bleu représente la droite de régression ; chaque point représente un participant. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.	134
Figure 58 : Variation de l'intelligibilité entre 20° à 60° (Différence_intelligibilité) en fonction du pourcentage de bonnes réponses dans le bruit (CRM_Bruit). Chaque point représente un participant. La droite bleu représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.	135
Figure 59 : La variation de l'effort d'écoute entre 20° et 60° (Différence_effort) en fonction du pourcentage de bonnes réponses dans le bruit dans le bruit. Chaque point représente un participant. La droite bleu représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.	135
Figure 60 : Score de l'échelle de motivation en fonction des conditions (CRM.Bruit, CRM.Coloc, CRM.2060, CRM.20, CRM.60). Les boxplots montrent la médiane (barre horizontale) et l'écart interquartile (cases). Les moustaches vont de la valeur plus basse à la valeur plus élevée observées dans une limite de 1,5 fois l'écart interquartile. Les points indiquent les points de données individuels.	136

Figure 61 : Pourcentage de bonnes réponses dans la condition avec plus grand écartement des locuteurs (c.à.d. CRM_60) en fonction de la motivation (c.-à-d., MOT_60). Chaque point représente un participant. La droite bleue représente la droite de régression. Les points correspondant aux participants dans les scores sont égaux entre les deux conditions sont sur la droite rouge. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.	137
Figure 62 : Mesure de l'effort par ESCU en fonction de la mesure de la motivation dans la condition MOT_20. Chaque point représente un participant. La droite bleue représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.	138
Figure 63 : Mesure de l'effort par ESCU en fonction de la mesure de la motivation par une adaptation de l'échelle de Kramer et al., 2016 dans la condition MOT_60. Chaque point représente un participant. La droite bleue représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.	138
Figure 64 : Arbre de classification de la variable définissant les groupes dans la condition 60°. Les participants ayant une mesure de l'effort d'écoute ressenti inférieure à 10,5 étaient pour la plupart dans le groupe avec une haute intelligibilité.	140
Figure 65 : Effort ressenti lors de la tâche d'intelligibilité dans le bruit en fonction des groupes d'intelligibilité à 60° de séparation entre les voix masques. Chaque point représente un participant. La médiane de mesure ESCU est représentée par le segment noir dans les boîtes à moustache. L'intervalle interquartile est représenté par l'extrémité inférieure et supérieure de la boîte à moustache dont les moustaches s'étendent jusqu'au 1,5 fois l'intervalle interquartile.	140
Figure 66 : Effort ressenti lors de la tâche d'intelligibilité dans la condition à 60° en fonction des sous-groupes d'intelligibilité à 60° de séparation entre les voix masques. Chaque point représente un participant. La médiane de mesure ESCU est représentée par le segment noir dans les boîtes à moustache. L'intervalle interquartile est représenté par l'extrémité inférieure et supérieure de la boîte à moustache dont les moustaches s'étendent jusqu'au 1.5 fois l'intervalle interquartile.	141
Figure 67 : Les seuils audiométriques mesurés à l'oreille gauche en fonction de deux groupe à haute intelligibilité et à basse intelligibilité. Chaque point indique un participant et chaque couleur l'appartenance au groupe (rouge : haute intelligibilité ; vert : basse intelligibilité).	142
Figure 68 : Les seuils audiométriques mesurés à l'oreille droite en fonction de deux groupes à haute intelligibilité et à basse intelligibilité. Chaque point indique un participant et chaque couleur l'appartenance au groupe (rouge : haute intelligibilité ; vert : basse intelligibilité).	142
Figure 69 : Schéma qui représente respectivement : les <i>Modèles</i> proposés dans la littérature sur l'effort d'écoute (Peelle, 2018 : en haut à gauche et Herrmann et Johsrude, 2020 : en haut à droite) les <i>Etudes de la thèse</i> et les <i>Résultats principaux</i>	165

Introduction

L'être humain est constamment immergé dans une réalité qui peut être extrêmement dynamique et riche en informations de différents types qui nous parviennent par nos sens. Cependant, même si nous sommes capables d'avoir une conscience globale du contexte, nous ne pouvons pas traiter tous les stimuli qui arrivent au même moment avec la même priorité ; ce qui est particulièrement vrai pour la modalité auditive. Parmi tous les stimuli présents dans une scène auditive, nous sélectionnons celui d'intérêt et choisissons de le traiter, parfois au prix d'un investissement coûteux en termes de ressources cognitives et parfois au détriment des autres stimuli.

Ce travail de thèse s'est focalisé sur la capacité de l'être humain à se concentrer sur une voix cible présentée parmi d'autres voix masques. Ces situations d'écoute complexe sont appelées *situations multilocuteurs*. Nous sommes souvent confrontés à ce type de situation dont un des exemples correspond aux discussions entre convives dans un restaurant. L'auditeur confronté à une situation multilocuteurs se doit souvent d'exercer un effort mental pour comprendre la voix cible sur laquelle il porte son attention en fonction des contraintes apportées par le reste de la scène auditive. L'effort mental lorsqu'il est appliqué à l'écoute a été nommé *effort d'écoute* (Peelle, 2018 ; Pichora-Fuller et al., 2016). Ce concept est complexe car multidimensionnel et dynamique. Il est encore mal appréhendé dans la littérature contemporaine, conduisant à une absence de définition claire et universelle. Différentes variables entrent en jeu dans la description de l'effort d'écoute : certaines sont liées à l'auditeur et à ses caractéristiques (état de l'audition, âge, disponibilité des ressources cognitives, motivation) d'autres sont liées à des facteurs externes tels que le contexte dans lequel se déroule l'échange (calme, bruyant), ou le message lui-même qui peut être dégradé ou clair, et complexe du fait par exemple de la présence de paroles homophones. De plus, la motivation de l'auditeur à vouloir faire face à toutes ces contraintes agit comme une variable modératrice de l'effort. On suppose généralement que plus un individu est motivé pour comprendre, plus il sera prêt à faire des efforts. Le fait de réussir ou non la tâche va générer chez l'auditeur une expérience subjective qui peut avoir une valeur positive ou négative et le pousser à persévérer dans son intention à comprendre ou, au contraire, à abandonner. L'un des aspects les plus insidieux concerne les conséquences de ces situations qui peuvent être plus ou moins importantes selon le contexte. Certes, dans un environnement protégé, les risques liés à la fatigue et à la compromission de la compréhension sont souvent négligeables, également parce que nous

avons la possibilité de demander à l'interlocuteur de répéter l'information. La situation peut au contraire devenir plus critique si, par exemple, la sécurité d'une ou de plusieurs personnes dépend de la compréhension claire et rapide de cette information. C'est le cas par exemple des travailleurs évoluant dans des systèmes à risques tels que les pilotes d'avion, les contrôleurs aériens ou le personnel militaire. Ces professionnels reçoivent une quantité d'informations critiques souvent au travers de plusieurs modalités sensorielles et sous une pression temporelle importante.

Cette situation peut devenir stressante pour les opérateurs et génératrice d'erreurs (Caldwell, 2005 ; Weeks et al., 2010). De plus, la pluralité des tâches à réaliser implique une charge cognitive élevée qui peut consommer la plupart des ressources cognitives (p. ex., capacités attentionnelles) au détriment de toute information inattendue. Par exemple, les pilotes peuvent ne pas répondre aux alarmes auditives (phénomène désigné sous le terme de *surdit  attentionnelle*) ; (Dehais et al., 2019, 2018 ; Giraudet et al., 2015 ; Scannella et al., 2013) ou encore pers v rer avec des strat gies erron es (Dehais et al., 2010). Dans ces environnements   risque, l' change d'information doit  tre fluide et compr hensible pour  viter la survenue d'accident. De plus ces environnements sont souvent bruyants, ce qui peut perturber la diffusion des informations radio. Par exemple, il est connu que certains des accidents a riens les plus tragiques (Los Rodeos Airport en 1977, John Wayne Orange Country Airport en 1981 ou Miami International Airport en 1972) ont  t  li s, au moins en partie,   une mauvaise compr hension des messages  chang s (Cushing, 1994).

Du fait de leur exposition au bruit dans leur environnement professionnel (moteur d'engins, armes   feu), certains travailleurs, en particulier militaires, peuvent d velopper des d ficits auditifs. Certaines personnes rapportent par exemple ne pas comprendre les sons de parole tout en d tectant qu'il s'agit de sons de parole : *« j'entends mais je ne comprends pas »*. De mani re int ressante, ce type de t moignage ne correspond pas n cessairement   une perte auditive d tectable   l'audiom trie tonale, qui est pourtant le test standard de contr le de l'audition. Plusieurs hypoth ses existent pour expliquer cette discordance entre audiom trie tonale et compr hension de la parole en environnement complexe, dont la d g n rescence des synapses entre les cellules cili es internes et certaines fibres du nerf auditif—appel es *synaptopathies cochl aires* (Lieberman et al., 2016). Mais il n'est actuellement pas possible de d tecter de tels dommages in vivo chez un individu ; les synaptopathies cochl aires induites par le bruit n'ayant  t  objectiv es que sur des mod les animaux (Kujawa & Liberman, 2009).

Les questions scientifiques qui ont fait l'objet de ce travail de th se ont  t  multiples. Parmi ces questions, la premi re, compte tenu de la nature multidimensionnelle de l'effort d' coute,

concernait le choix des dimensions à étudier (p. ex., engagement de ressources cognitives, effort ressenti, motivation, état auditif de l'auditeur). Une deuxième question concernait la manière dont l'implication de ces dimensions pouvait être définie et investiguée (p. ex., tâches d'intelligibilité, échelles subjectives, observations en neuroimagerie). Une troisième question concernait comment ces dimensions interagissaient ou auraient pu interagir (p. ex., via un entraînement) afin de faciliter ou non la compréhension d'une voix cible en situation d'écoute multilocuteurs. Une quatrième question concernait la possibilité de réduire l'effort d'écoute en situations multilocuteurs.

Les objectifs de cette thèse découlent de ces problématiques et sont à la croisée des intérêts fondamentaux, cliniques et opérationnels. Ces objectifs ont donné lieu à trois études complémentaires autour de la perception du langage en situation multilocuteurs pour tenter de fournir des réponses aux questions posées ci-dessus.

Objectifs de la thèse

Fondamental

Le premier objectif visait à étudier les processus cognitifs sous-jacents à la compréhension d'un message auditif en situations multilocuteurs. L'activité corticale de jeunes normo-entendants a ainsi été observée via l'outil de mesure NIRS (de l'anglais Near Infra-Red Spectroscopy ou Spectroscopie en proche infrarouge).

Nous avons ainsi pu définir les régions cérébrales impliquées dans la réalisation d'une tâche d'intelligibilité de la parole en situation multilocuteurs. Deux paramètres ont été manipulés : le rapport d'intensité entre la voix cible et la voix masque (nommé TMR pour *Target-to-Masker Ratio*) et la séparation spatiale des sources sonores (c.-à-d., entre la voix cible et les voix masque).

Clinique

L'intérêt d'ordre clinique visait à étudier les effets de la séparation spatiale des locuteurs sur l'intelligibilité et l'effort ressenti dans une tâche d'écoute multilocuteurs en fonction de l'état auditif chez une population de militaires. En effet, ces personnes exposées à des bruits intenses du fait de leur travail, présentent une altération plus ou moins modérée de leur état auditif. Les questions portaient sur la triple relation entre l'état du système auditif, l'intelligibilité dans une situation multilocuteurs et les évaluations subjectives de l'effort d'écoute ressenti pour réaliser la tâche.

Opérationnel

Le troisième objectif visait à tester des solutions pour réduire l'effort d'écoute tout en améliorant l'intelligibilité en situation multilocuteurs chez le normo-entendant et chez le personnel militaire présentant une altération modérée de son audition. Une partie de ce travail de thèse souhaitait en effet apporter des éléments contributifs à la conception et l'affinement de systèmes de communications parmi les opérateurs (personnel militaire, pilotes...) et pour les personnes présentant des troubles de l'audition.

La prise en compte du concept d'effort est primordiale dans la conception de tels systèmes afin de les rendre plus conformes aux capacités de l'opérateur et pour favoriser un échange plus sûr et efficace.

Deux solutions ont été évaluées : l'entraînement et la séparation spatiale des locuteurs.

Partie 1. État de l'art

Chapitre I. Situations multilocuteurs

Cette thèse s'intéresse à une forme particulière de situation auditive complexe dite *situation multilocuteurs*. Une situation multilocuteurs peut être définie comme une situation d'écoute dans laquelle un auditeur se focalise sur un locuteur (c.-à-d., une voix cible) qui est présent parmi plusieurs autres (c.-à-d., des voix masques).

Afin de pouvoir comprendre et suivre la voix cible, la mise en jeu de certains processus cognitifs doivent être faite. Ces derniers incluent : la **reconnaissance** des objets sonores présents dans la scène auditive, la **ségrégation** des différents flux sonores composés par ces objets, la **focalisation** sur la voix cible et l'**inhibition** des voix masques.

Avant d'examiner la manière dont ces processus se déroulent, nous présenterons dans la partie suivante les mécanismes qui permettent au son de parvenir au cerveau à partir des oreilles.

I. Le parcours du son de l'oreille au cerveau

Le parcours du son de l'oreille au cerveau peut être résumé en cinq étapes principales.

La première étape correspond à l'oreille externe composée du pavillon et du conduit auditif externe (CAE) ; (Figure 1.A). La fonction de l'oreille externe est de diriger les sons vers l'oreille moyenne. Elle joue également un rôle essentiel dans la localisation auditive en modifiant le spectre des ondes sonores selon leur angle d'incidence. Ces modifications sont reconnues par le système auditif pour situer la source sonore dans l'espace en combinaison avec d'autres indices liés aux différences de temps d'arrivée et d'intensité des ondes sonores entre les deux oreilles. Les HRTFs (de l'anglais *Head Related Transfer Function*) peuvent être capturées via la mesure de la fonction de transfert liée à la tête (Wightman & Kistler, 1989). Filtrer un signal auditif présenté sous casque avec des HRTFs produit la sensation de percevoir ce signal comme provenant d'une position précise de l'espace.

Le CAE protège la délicate membrane tympanique du froid, de la saleté et de la déshydratation, en expulsant la saleté et la poussière par le cérumen.

La deuxième étape coïncide avec l'arrivée du son à la membrane tympanique dans l'oreille moyenne (Figure 1.B). Le tympan, en vibrant, transmet les ondes sonores aux trois osselets : le

marteau, l'enclume et l'étrier. Le mouvement des trois osselets amplifie les vibrations du tympan qui atteignent l'oreille interne.

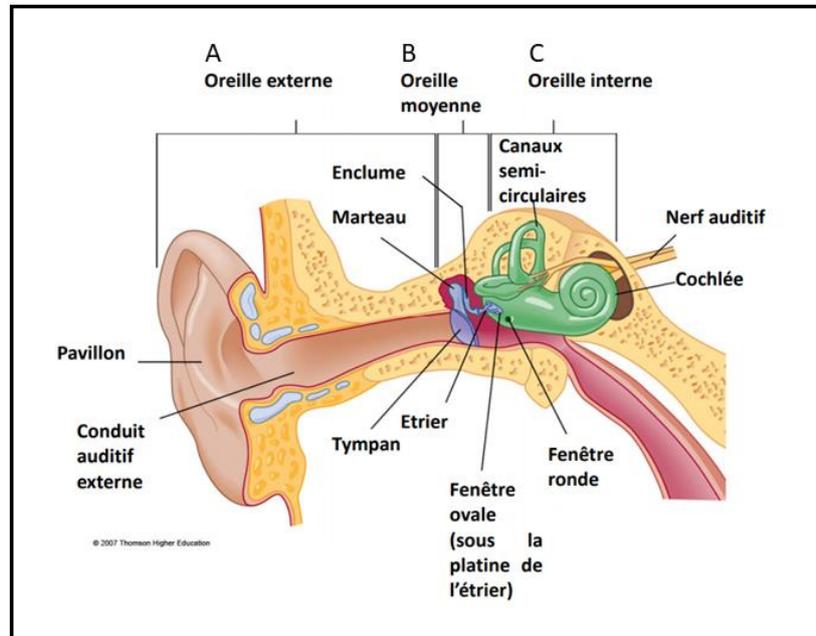


Figure 1 : Dessin de l'appareil auditif. L'appareil auditif humain se compose de l'oreille externe, formée par le pavillon et le conduit auditif externe (A), de l'oreille moyenne, qui contient la chaîne des osselets (B), et de la cochlée formant avec le vestibule, l'oreille interne (C) ; (Goldstein, 2008).

La troisième étape correspond donc à l'arrivée du son à l'oreille interne (Figure 1.C). L'oreille interne, remplie de liquides, comprend la cochlée, organe de l'audition, et le vestibule, organe de l'équilibre. Elle est composée de trois canaux contigus et parallèles, enroulés en spirale depuis la base jusqu'à l'apex : la rampe vestibulaire, le canal cochléaire, et la rampe tympanique. Les deux rampes, remplies de liquide périlymphatique, sont en continuité à l'apex de la cochlée au niveau de l'hélicotréma. À l'intérieur du canal cochléaire, plongé dans le liquide endolymphatique, se trouve l'organe de Corti qui contient les cellules sensorielles de l'audition (Figure 2) —c'est l'origine du message nerveux. Les cellules sensorielles sont de deux types : les cellules ciliées internes (CCI) et les cellules ciliées externes (CCE). On dénombre environ 3500 CCI et entre 10000 et 12000 CCE réparties respectivement sur une et trois rangées (Ashmore, 2008).

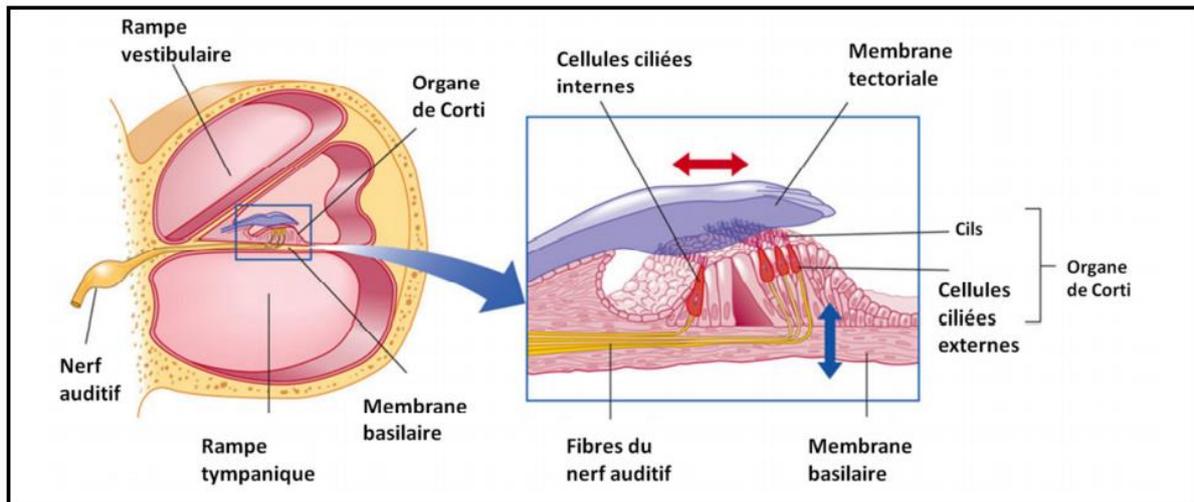


Figure 2 : La cochlée et l'organe de Corti. Les cellules sensorielles auditives se trouvent dans l'organe de Corti entre les membranes basilaire et tectoriaale. Les cellules ciliées externes (CCE) ont leurs cils encastrés dans la membrane tectoriaale. Ils sont cisailés lors du mouvement de la membrane basilaire, ce qui provoque l'activation des CCE qui facilite l'activation des cellules ciliées internes (CCI). Les CCI activées émettent au niveau du nerf auditif des messages nerveux codant certaines caractéristiques du signal sonore (Goldstein, 2008).

L'arrivée du son au tronc cérébral constitue la quatrième étape du parcours du son et le premier stade du son à l'intérieur du cerveau. À ce stade, les sons sont déchiffrés et leur durée, leur intensité et leur fréquence sont déterminées. L'étape suivante se situe au niveau du thalamus qui collecte toutes les informations perçues par les sens, à l'exception de l'odorat. Le thalamus interprète les informations et les transmet au cortex cérébral, correspondant à la dernière étape. Le message sonore est alors en grande partie déchiffré, mais le cortex cérébral ajoute une signification aux sons. Ici, les sons sont reconnus et mémorisés. Cette région du cortex cérébral, appelée cortex auditif, intègre aussi les souvenirs ou les émotions liés aux sons qui vont donner une certaine coloration idiosyncrasique en fonction de chaque expérience.

II. Ségrégation des flux sonores

Au cours du traitement de l'information sonore permettant la compréhension d'une voix cible dans des situations d'écoute complexe, la première étape consiste à séparer les différents flux sonores. Les différents flux sonores présents dans la scène auditive sont constitués de plusieurs objets sonores. Ces derniers sont reconnus et identifiés par notre cerveau grâce à une véritable analyse de la scène auditive (Bregman, 2009). En effet, ce qui n'est au départ qu'un ensemble

de fréquences sonores indistinctes et chaotiques, est transformé par le cerveau en objets sonores reconnaissables et dotés d'une signification précise. Un objet sonore peut être défini comme une estimation perceptuelle des entrées sensorielles qui proviennent d'un élément physique distinct dans le monde externe (Shinn-Cunningham, 2008).

Par exemple, la structure spectro-temporelle des objets sonores détermine la segmentation de ceux-ci sur de courtes échelles de temps (c.-à-d., la formation de syllabes). Les caractéristiques d'ordre supérieur (c.-à-d., le timbre, la hauteur, la localisation) déterminent la façon dont ces objets sonores sont organisés sur des échelles de temps plus longues qui vont former les flux sonores (Shinn-Cunningham, 2008).

Cette ségrégation des flux sonores est indispensable pour que l'auditeur puisse se concentrer sur une voix cible.

III. Masquage informationnel et masquage énergétique

L'auditeur peut se trouver dans un contexte où le bruit interfère avec la compréhension de la voix cible (appelé *speech-in-noise* en anglais pour parole dans le bruit) ou dans un contexte où d'autres messages interfèrent avec la compréhension de la voix cible (appelé *speech-in-speech* en anglais pour parole dans la parole). On parle de *masquage énergétique* dans le premier cas et de *masquage informationnel* dans le second (Brungart, 2001). Ces interférences peuvent compromettre la compréhension de la voix cible lorsqu'elles sont suffisamment fortes.

Lorsque les propriétés spectro-temporelles des sons en entrée (voix cible et masque) se superposent et deviennent inaudibles, cela crée du masquage informationnel. Dans le masquage informationnel, les informations sont audibles mais non distinguables dans la mesure où elles ne peuvent pas être séparées et attribuées à un locuteur précis. Ihlefeld & Shinn-Cunningham (2008) indiquent que dans une situation de masquage informationnel les participants reportent rarement des mots absents du mélange sonore. Les participants se trompent généralement en rapportant un mélange de mots de différents locuteurs (erreurs de mélange) ou en reportant tous les mots du mauvais locuteur (erreurs de masquage). Par ailleurs, Nakai et al. (2005) ont utilisé une tâche de *speech-in-speech* dans laquelle ils demandaient aux participants de suivre une histoire qui pouvait être masquée par le même locuteur ou par un locuteur de sexe différent. Le masquage par le même locuteur était la condition la plus difficile du fait de la similarité (c.-à-d., même genre) des locuteurs.

La ségrégation des voix repose sur la disponibilité de plusieurs indices sonores qui peuvent être à disposition de l'auditeur (Bronkhorst, 2015). Parmi ces indices, il existe :

- Les indices liés aux caractéristiques fréquentielles de la voix (c.-à-d., le genre du locuteur).
- Les indices liés aux différences d'intensité entre la voix cible (*target*, en anglais) et les autres voix masques (*masker*, en anglais), d'où le terme anglais : *target-to-masker ratio* (ou TMR ; c.-à-d., rapport cible sur masque).
- Les indices liés à la séparation spatiale des locuteurs.

La familiarité avec la voix du locuteur peut réduire le masquage informationnel (Johnsrude et al., 2013). Dans leur étude, Johnsrude et al. (2013) ont montré qu'écouter la voix de son partenaire de vie peut faciliter la ségrégation des voix dans un protocole d'écoute multilocuteurs.

Le TMR et la séparation spatiale des sources seront examinés plus en détail dans le paragraphe suivant.

Ce qu'il faut retenir

- La scène auditive est composée d'objets sonores, reconnaissables et dotés d'une signification précise, qui se regroupent en différents flux sonores.
- Plusieurs indices sonores de la scène auditive peuvent aider à ségréger les flux dont : le TMR et la séparation spatiale des locuteurs en sont deux principaux.

IV. Target-to-masker ratio

Plusieurs études ont montré que le TMR peut être un indice utile afin de ségréger les différents flux sonores (Andéol et al., 2017 ; Bronkhorst, 2015 ; Brungart & Simpson, 2005 ; Eddins & Liu, 2012). L'efficacité du TMR peut être affectée par 1) la différence de genre entre les locuteurs, car ces différences de caractéristiques vocales dominent sur les indices de ségrégation offerts par la présence du TMR ; 2) la présence de bruits supplémentaires qui peuvent couvrir le locuteur le plus silencieux et c) l'augmentation du nombre de locuteurs qui interfèrent sur la détection de la voix cible. Pour ces raisons, la première étude réalisée au cours de la thèse s'est focalisée sur l'efficacité du TMR en présence d'une seule voix masque du même sexe que la voix cible.

Il est en général plus facile de prêter attention au locuteur dont la voix est plus forte que les autres. Ainsi, intuitivement, en contexte opérationnel, une amélioration de l'intelligibilité de la voix cible devrait être obtenue en augmentant à la radio le volume du locuteur cible. Mais, comme indiqué dans l'article de Brungart & Simpson (2007), il peut être difficile en milieu opérationnel de prévoir quelle sera la voix cible. En pratique, il est fréquent que les opérateurs novices manipulent en permanence les niveaux des radios. Les opérateurs experts parviennent à utiliser le TMR comme outil de ségrégation, que la voix cible ait un niveau supérieur ou inférieur aux autres voix, du moment que la valeur absolue de la différence de niveau entre les voix est suffisante.

L'un des protocoles comportementaux les plus largement utilisés pour étudier les situations multilocuteurs est le *Coordinate Response Measure* (CRM corpus, Bolia et al., 2000). Ce corpus est composé de phrases qui respectent toutes la même structure : « Ready call sign go to color number now ». Dans la version que nous utilisons, il existe 7 identifiants différents (7 call sign : Baron, Charlie, Ringo, Eagle, Arrow, Hopper, Tiger, Laker), 4 couleurs (bleu, vert, rouge, blanc) et 8 chiffres (de 1 à 8). Généralement, l'auditeur doit indiquer la couleur et le chiffre donnés par la voix cible qui est définie par un identifiant spécifique (par exemple Baron). La voix cible peut être présentée simultanément avec une ou plusieurs voix masques et toutes indiquent à l'auditeur un couple couleur-chiffre différent. La littérature nous montre que les performances sont plus élevées lorsque la voix cible est présentée à une intensité plus élevée que les autres voix masques (Andéol et al., 2017 ; Brungart, 2001 ; Thompson et al., 2015). Toutefois, quand la voix cible est présentée à une intensité plus faible par rapport à la voix masque, les performances au sein de la population sont plus variables (Andéol et al., 2017). Il apparaît que la population se divise grossièrement en deux catégories de personnes : les participants dont la performance s'améliore au fur et à mesure que le TMR augmente (voix cible de plus en forte) et les participants dont la performance est élevée en présence d'un TMR positif mais aussi d'un TMR fortement négatif. Enfin, en présence d'un TMR intermédiaire (c.-à-d., quand les deux voix ont plus ou moins le même niveau d'intensité), la performance chute (Andéol et al., 2017 ; Brungart, 2001 ; Brungart & Simpson, 2007 ; Cooke et al., 2008 ; Thompson et al., 2015). Ces résultats suggèrent que dans les deux cas opposés—TMR positif et TMR négatif—c'est précisément la différence nette de volume entre les deux voix qui permet une meilleure ségrégation de celles-ci avec une meilleure compréhension de la voix cible, bien qu'elle soit *a priori* défavorisée dans le dernier cas.

Le fait de ne pas être en mesure de séparer les deux voix en présence d'un TMR négatif pourrait être dû au fait que certains auditeurs ne se rendent pas compte de la différence de volume entre les deux voix et se désengagent en raison de la difficulté de la tâche (Andéol et al., 2017). Andéol et al. (2017) ont enregistré l'activité du cortex préfrontal de participants ayant une performance élevée avec TMR positif et négatif (*U-shape performers*-participants avec une performance suivant une courbe en « U ») et de participants ayant une performance croissante au fur et à mesure de l'augmentation du TMR (*non-U-shape performers*) ; (Figure 3). Les premiers uniquement, montrent une plus grande implication du cortex préfrontal en présence d'un TMR négatif (comparée aux TMR positif et intermédiaire) ; (Figure 4). Considérant que le cortex préfrontal est une région clef dans la mise en œuvre des fonctions exécutives (Baddeley, 1996), son implication pourrait refléter le plus grand engagement des ressources cognitives de la part des auditeurs capables d'utiliser un TMR défavorable. Une performance plus élevée est donc liée à une plus grande activité corticale (engagement de ressources cognitives).

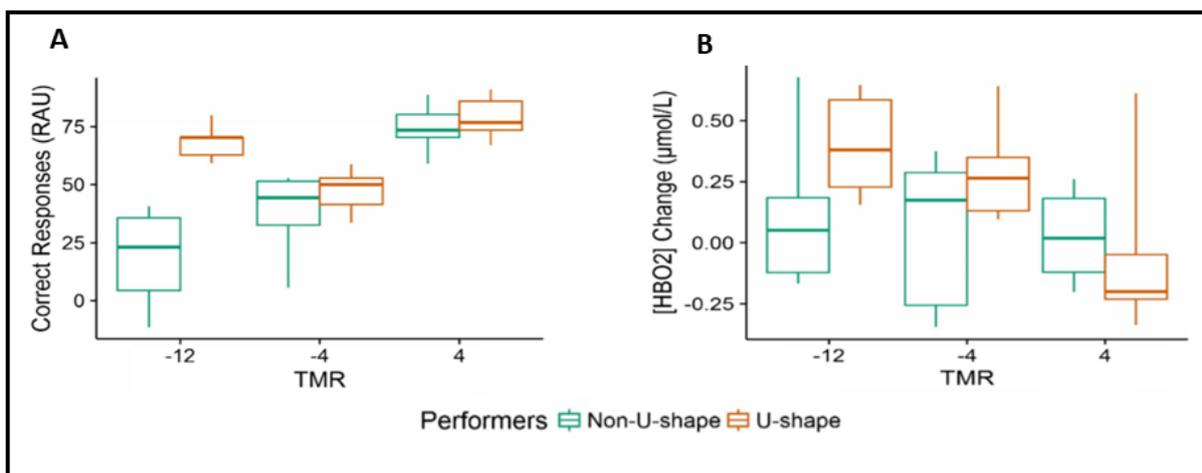


Figure 3 : Scores de réponses correctes (A) et activité corticale (B) en fonction de la condition du TMR. Les boîtes indiquent la médiane (barre horizontale) et l'écart interquartile (cases), les moustaches indiquent l'intervalle de 5 à 95 percentiles. Les *U-shape performers* (en rouge) montrent un score d'intelligibilité similaire pour les TMR défavorables et favorables et une [HbO₂] dans le cortex préfrontal qui augmente avec le TMR. Les *non-U-shape performers* (en

vert) montrent un score d'intelligibilité qui diminue avec le TMR et une [HbO2] similaire pour tous les TMR (d'après Andéol et al., 2017).

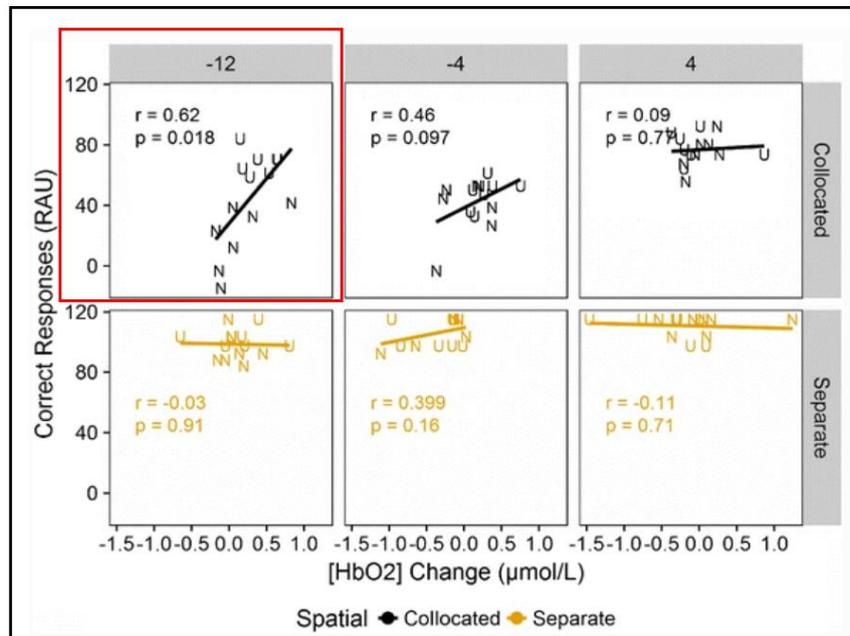


Figure 4 : Les scores de réponses correctes sont représentés en fonction des variations de l'activité préfrontale [HbO2] pour chaque combinaison de conditions de TMR et de séparation spatiale. La lettre "U" indique les *U-shape performers* et la lettre "N" indique les *non-U-shape performers*. Le coefficient de corrélation de Pearson et son degré de signification sont indiqués dans chaque panneau. Pour les locuteurs colocalisés, à TMR défavorable, plus l'activité préfrontale était élevée, plus la performance était élevée (d'après Andéol et al., 2017).

Plusieurs études montrent qu'il est possible d'améliorer les performances des participants qui rencontrent des difficultés en présence d'un TMR négatif (Thompson et al., 2015). Dans ces études, les chercheurs entraînaient des participants *via* des feedbacks sur les réponses données (Brungart & Simpson, 2004) ou des *matrices pay off* avec des pénalités selon les erreurs commises et des récompenses selon les réponses données (Thompson et al., 2015).

Ce qu'il faut retenir

- Un TMR défavorable peut être utilisé pour ségréger les flux sonores.
- Il existe une variabilité interindividuelle dans l'utilisation du TMR.
- L'utilisation du TMR en condition défavorable peut être apprise.

V. Séparation spatiale des locuteurs

La ségrégation des flux sonores est aussi possible grâce à la séparation spatiale des locuteurs dans la scène auditive (Bronkhorst, 2015). Une séparation spatiale relativement faible peut être suffisante pour conduire à une ségrégation efficace dans certaines situations. Brungart et Simpson (2007) par exemple, ont constaté qu'une séparation de deux voix de seulement 10° peut être suffisante pour maximiser les performances dans une tâche d'intelligibilité multilocuteurs. La séparation spatiale des locuteurs, en favorisant la ségrégation des sources sonores, permet donc d'améliorer l'intelligibilité de la voix cible (Andéol et al., 2017 ; Arbogast et al., 2002). Il semble que la séparation spatiale entre les voix masques et la voix cible soit un moyen efficace de faire face aux difficultés soulevées par le masquage informationnel mentionné précédemment (Freyman et al., 1999, 2001 ; Kidd et al., 1998). Les avantages procurés par cette séparation sont encore plus importants lorsque les caractéristiques acoustiques de la voix masque se rapprochent de celles de la voix cible (Arbogast et al., 2005 ; Best et al., 2012 ; Zekveld et al., 2014). Afin d'identifier la présentation optimale des informations sonores dans le domaine militaire, Brungart et Simpson (2007) comparent trois situations de présentations d'une voix cible parmi 7 locuteurs : monaurale, dichotique et avec séparation spatiale des locuteurs. Leurs résultats montrent que la séparation spatiale des locuteurs a amélioré les performances d'environ 30% pour toutes les situations d'écoute avec plus de deux locuteurs concurrents. De plus, la séparation spatiale des locuteurs a amélioré l'intelligibilité d'environ 5% par rapport à la situation dichotique. La séparation spatiale des locuteurs peut donc faciliter la réussite des opérations de communication en rendant les communications plus intelligibles. En outre, la séparation spatiale des locuteurs joue également un rôle important dans la réduction de l'engagement de certaines ressources cognitives pour réussir la tâche (Andéol et al., 2017 ; Xia et al., 2015). Xia et al. (2015), ont étudié à travers deux expériences, les effets de la séparation spatiale des locuteurs sur la charge cognitive (les auteurs définissent de cette manière ce que nous avons appelé *l'effort d'écoute* en début de document) avec un protocole de double tâche chez des personnes normo-entendantes et malentendantes. Le protocole était composé de la tâche principale d'écoute avec le CRM corpus (3 locuteurs) et la tâche secondaire consistait en une tâche de recherche visuelle. Le principe à la base des protocoles de double tâche est que plus la demande de ressources de la tâche principale est élevée, moins de ressources seront disponibles pour effectuer la tâche secondaire (Kahneman, 1973). Par conséquent, Xia et al. (2015) déduisaient à travers la réduction de la précision de recherche visuelle entre la double tâche (c.-à-d., tâche d'écoute et de recherche

visuelle ensemble) et la tâche de recherche visuelle seule, la demande cognitive de la tâche primaire. Dans les deux expériences, différents indices sonores étaient testés : condition sans indice, condition avec indice de genre, condition avec séparation spatiale des locuteurs et condition avec genre et séparation ensemble. Dans la première expérience, une séparation spatiale de 15° (la voix cible étant placée en face à 0°) a été testée tandis que dans la deuxième expérience, une séparation spatiale de 60° a été testée (la voix étant là aussi placée à 0°). Avec une séparation spatiale de 15° et des indices spatiaux et de genre, la charge cognitive déduite a été la plus faible. Toujours chez les personnes normo-entendantes, les résultats avec une séparation de 60° illustraient que la compréhension de la voix cible à travers la séparation spatiale des locuteurs permettait de soulever une demande cognitive encore moins importante par rapport à l'utilisation de la différence de genre.

Andéol et al. (2017), dans leur étude, proposaient aux participants une tâche semblable d'écoute multilocuteurs (CRM corpus, Bolia et al., 2000) avec deux voix masculines—une voix cible et une voix masque—présentées en position co-localisée (c.-à-d., les deux voix étaient placées en face à 0° par rapport à l'auditeur) ou séparée spatialement (c.-à-d., les deux voix étaient placées sur les côtés à + et -60°). Trois niveaux de TMR ont été utilisés (TMR-défavorable : -12 dB, TMR-intermédiaire : -4 dB et TMR-favorable : +4 dB). Pendant que les participants réalisaient la tâche, leur activité corticale préfrontale était enregistrée à l'aide d'une spectroscopie fonctionnelle en proche infrarouge (fNIRS). Les résultats ont montré une intelligibilité plus élevée de la voix cible dans toutes les conditions TMR avec locuteurs séparés par rapport aux conditions TMR avec locuteurs co-localisés. Concernant l'activité corticale, la condition TMR-intermédiaire avec locuteurs séparés était celle avec le moindre engagement des ressources cognitives. Cette condition semblait être celle qui a le plus bénéficié de la présence de la séparation spatiale des locuteurs. Dans cette condition, le TMR était faible (c.-à-d., TMR = -4 dB) et donc moins exploitable pour les participants. De plus, le fait que le protocole était composé de seulement deux voix masculines éliminait aussi les avantages que la présence de genres différents aurait pu apporter dans la ségrégation des voix (Xia et al., 2015). Probablement, ces deux aspects ont contribué à faire de cette condition, la plus difficile, en rendant plus évident l'effet bénéfique de la séparation spatiale des locuteurs.

Une nuance particulière concerne la continuité spatiale de la cible ; c'est-à-dire la présence d'une cible fixe au même endroit alors que les distracteurs sont placés à différents endroits à chaque essai. Cette caractéristique de la scène auditive semble importante pour permettre la

facilitation de la détection de la cible et de la compréhension de son contenu (Best et al., 2008 ; Kidd et al., 2005). Best et al. (2008) ont présenté une séquence de chiffres en présence d'autres chiffres distracteurs. Ils ont ainsi confirmé que les performances étaient meilleures quand la position du chiffre cible était la même comparée aux conditions où sa position variait, et ce malgré un écartement constant entre cible et distracteur. La continuité spatiale ne contribuerait pas seulement à éviter les coûts comportementaux (c.-à-d., identification de chiffres énoncés) associés au changement du focus de l'attention mais elle rendrait aussi la sélection spatiale plus fine essai après essai.

Ce qu'il faut retenir

- La séparation spatiale des sources sonores permet d'améliorer l'intelligibilité du langage et de réduire l'engagement des ressources cognitives (en termes d'activité du cortex préfrontal) nécessaires pour faire la tâche.

VI. Attention auditive sélective : focalisation et inhibition

Une fois que les différents flux sonores ont été ségrégués grâce aux indices sonores tels que le TMR et/ou la séparation spatiale des locuteurs, l'auditeur peut s'orienter vers la voix cible. Dans le même temps, l'auditeur doit laisser de côté les autres voix masques qui ne sont pas pertinentes pour son objectif de compréhension. L'attention sélective entre alors en jeu. A ce stade, il est important de distinguer l'attention sélective, de l'éveil ou de la vigilance. L'éveil ou la vigilance ne sont pas dirigés vers des objets précis à la différence de l'attention sélective qui est portée vers un objectif précis : dans notre cas, la compréhension de la voix cible. Une fois que l'on a décidé de diriger notre attention sur une voix cible précise, la sélection permet de se focaliser sur cette voix et d'inhiber les autres voix masques. Ces concepts peuvent être retrouvés dans la définition de l'attention donnée par le psychologue William James il y a plus d'un siècle : « *L'attention est la prise de possession par l'esprit, sous une forme claire et vive, d'un objet ou d'une suite de pensées parmi plusieurs qui sont présents simultanément (...). Elle implique le retrait de certains objets afin de traiter plus efficacement les autres* » ; (James, 2007).

Il semble donc nécessaire à la fois de se focaliser sur la voix cible mais aussi d'être en mesure d'inhiber les autres. Dans la communauté scientifique, on nomme le paradigme dans lequel ces processus sont étudiés par le terme : *cocktail party problem* (Cherry, 1953), où, tout comme dans une soirée, il est nécessaire d'inhiber les informations non pertinentes qui proviennent des

voix masques pour suivre la conversation avec une voix cible. Strauß et al. (2014) expliquent que dans ces situations notre attention se dirige vers les informations pertinentes (Fritz et al., 2007) facilitant leur traitement et améliorant la réponse neuronale associée (Lee et al., 2014). Au même moment, l'inhibition sélective des masques aide à limiter les ressources cognitives déployées pour un traitement complet et approfondi de ceux-ci (Fuxe & Snyder, 2011 ; Melara et al., 2002). Dans cette optique, les auteurs énoncent que l'interférence des voix masques pourrait être le résultat à la fois d'une attention insuffisante portée à la voix cible et d'une mauvaise inhibition du bruit et/ou des voix masques. Par exemple, durant une soirée avec beaucoup de personnes, suivre la conversation avec une voix cible requiert l'acte volontaire et délibéré de garder une attention constante afin de comprendre son discours. Mais malgré l'attention portée sur une voix cible, le bruit d'un verre qui tombe sur le sol et se brise, capte automatiquement notre attention. Si le bruit de ce verre est capable d'attirer notre attention, tous les autres sons peuvent également être traités dans une certaine mesure, même s'ils ne correspondent pas aux objectifs de l'auditeur. Qu'il s'agisse d'un stimulus délibérément choisi ou d'un stimulus qui, par sa saillance, nous « frappe », les deux seront traités plus en détail par rapport aux autres, ce qui nécessite le recrutement des ressources centrales qui restent malgré tout limitées (Desimone & Duncan, 1995).

Différents modèles ont été proposés pour expliquer la manière dont l'attention sélective se met en place. Les *modèles structurels*, d'une part, considèrent l'attention comme un *filtre* qui ne laisserait passer que certaines informations en laissant les autres de côté (Deutsch & Deutsch, 1963 ; Johnston & Heinz, 1978 ; Treisman, 1960) ; les *modèles des ressources*, quant à eux, considèrent l'attention comme une sorte de *réservoir d'énergie* à distribuer en fonction du nombre de tâches à accomplir et des priorités de chacune (Kahneman, 1973). Il existe donc une capacité limitée de ressources mentales qui peuvent être allouées à l'exécution de tâches ; des différences individuelles dans la capacité maximale d'exploitation des ressources. De plus, la quantité de ressource allouée aux tâches augmente au fur et à mesure que les tâches deviennent plus difficiles ou exigeantes (Pichora-Fuller et al., 2016). Les auteurs Pichora-Fuller et al. (2016) partent du modèle de Kahneman (1973) et l'adaptent aux situations d'écoute complexe et au concept d'effort d'écoute. Lorsque la complexité de la situation d'écoute augmente (par exemple par la baisse d'acuité auditive et/ou le bruit de fond), l'effort perceptif nécessaire à la reconnaissance de la parole occupe toutes les ressources cognitives disponibles. Les ressources qui restent peuvent donc être insuffisantes pour d'autres traitements (par exemple l'encodage des informations en mémoire ou la compréhension de phrases syntaxiquement complexes) ; (Wingfield, 2016).

La sélection des informations importantes est donc nécessaire étant donné les limites de la capacité à traiter toutes les informations sensorielles en entrée.

Les situations multilocuteurs nous offrent un exemple clair de la manière dont nous devons sélectionner les objets d'intérêt parmi tous les objets présents, en mettant de côté les objets qui ne sont pas urgents ou primordiaux à traiter, nous conduisant à une adaptation constante à notre environnement. Cependant, cette capacité doit aussi être flexible, de manière à changer notre focus attentionnel par rapport à de nouvelles priorités émergeant dans la scène auditive.

Ce qu'il faut retenir

- La ségrégation de flux sonores est propédeutique aux deux processus de *focalisation* de l'attention sur la cible d'intérêt et *d'inhibition* des messages non pertinents.

Chapitre II. Neuro-anatomie fonctionnelle du traitement du langage

I. La voie dorsale et la voie ventrale

Avec l'avènement des méthodologies de neuro-imagerie cérébrale (IRM, TEP, fNIRS), il est désormais possible d'observer l'activité corticale sous-jacente au traitement du langage. Différentes études montrent que les zones les plus impliquées dans le traitement du langage sont les suivantes : l'aire de Broca, dans le gyrus frontal inférieur (inferior frontal gyrus ; IFG), l'aire de Wernicke, dans le gyrus temporal supérieur (STG), ainsi que des parties du gyrus temporal moyen (MTG) et des gyrus pariétal et angulaire inférieurs dans le lobe pariétal (voir une illustration Figure 5) ; (Friederici, 2011 ; Peelle, 2018).

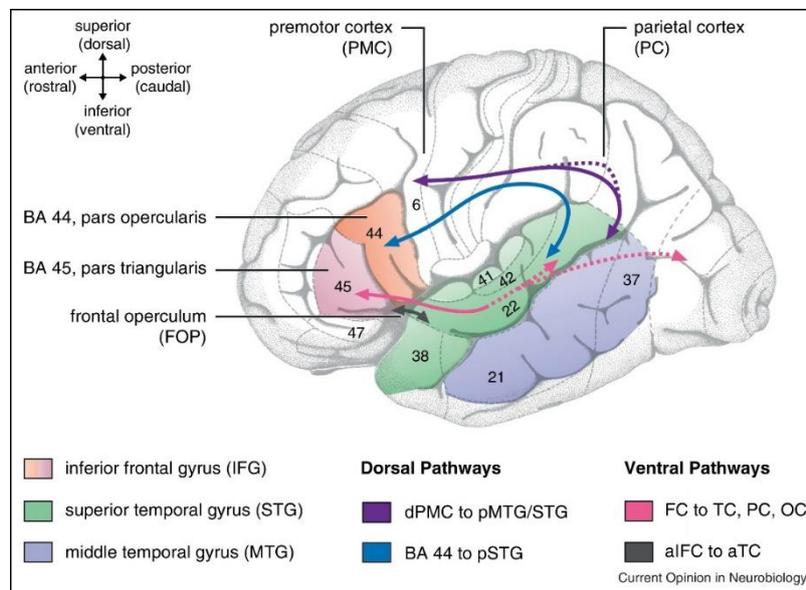


Figure 5 : Aires cérébrales de l'hémisphère gauche impliquées dans le traitement du langage (d'après Friederici & Gierhan, 2013).

Sur le modèle des études réalisées sur la vision, Gregory Hickok et David Poeppel ont proposé un modèle d'anatomie fonctionnelle du traitement du langage qui compte une voie ventrale et une voie dorsale (Hickok & Poeppel, 2004, 2007). Ce modèle comprend le gyrus temporal supérieur (STG) et le sulcus temporal supérieur (STS) comme régions de départ du traitement du langage. À partir de ces régions de départ, deux voies sont mises en évidence : la voie ventrale et la voie dorsale. La voie ventrale est impliquée dans le traitement sémantique du son (c.-à-d. la compréhension du son) tandis que la voie dorsale est impliquée dans le traitement du son selon des représentations articulatoires essentielles pour la production du langage (Figure

6). En détail, la voie ventrale part du STG et MTG et arrive jusqu'à l'IFG (Chang et al., 2015 ; Duffau et al., 2014). Elle est responsable de la reconnaissance du langage et de la représentation conceptuelle (Chang et al., 2015). La voie dorsale, quant à elle, parvient au cortex pré-moteur et à la région pariéto-temporale. La voie dorsale est impliquée dans l'intégration sensorimotrice du langage (Hickok & Poeppel, 2004 ; Rauschecker & Scott, 2009).

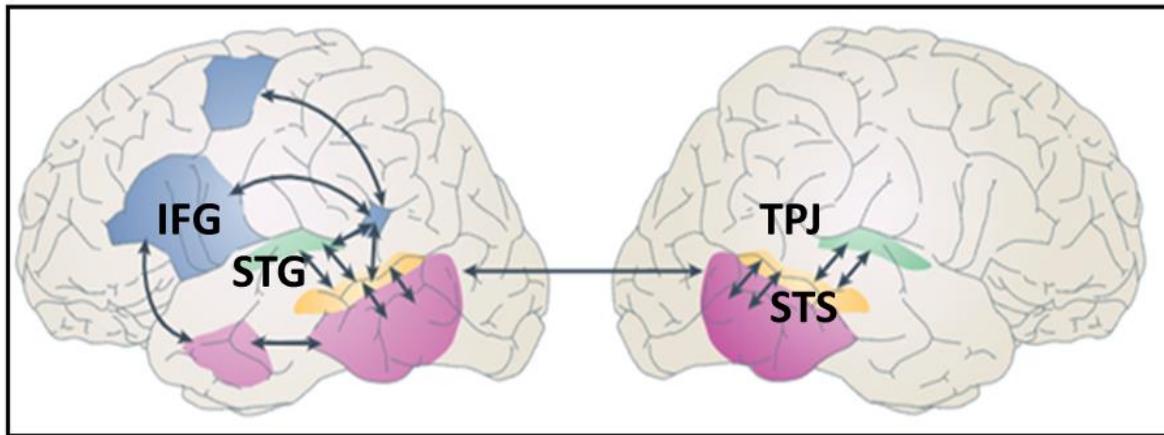


Figure 6 : Le modèle à double voie de l'anatomie fonctionnelle du langage : la voie dorsale (en bleu) et la voie ventrale (en rose) dans les deux hémisphères. Les premières étapes du traitement du langage sont dans les régions auditives bilatérales (STG en vert et STS en jaune). Depuis ces deux régions partent les deux voies : la voie ventrale qui supporte la compréhension du langage (IFG), la voie dorsale (plus latéralisée à gauche) supporte les intégrations sensori-motrices pour la production du langage incluant les structures de la jonction temporo-pariétale (TPJ) et le lobe frontal (d'après Hickok & Poeppel, 2007).

II. Le traitement du langage en situation d'écoute complexe : core speech network

Le traitement d'un message intelligible dans les situations d'écoute optimale sollicite principalement une activation des parties antérieures et postérieures du STG et du STS (Abrams et al., 2013 ; Davis & Johnsrude, 2003 ; Peelle, 2018 ; Scott et al., 2000). Dans une véritable hiérarchie de traitement, les régions les plus proches du cortex auditif présentent une sensibilité aux caractéristiques acoustiques du son à tel point que leur activation est présente indépendamment de la complexité de la tâche (Wild et al., 2012) ainsi, elles constituent le **Core Speech Network** du traitement d'un message souvent complété par l'IFG (Peelle, 2018). Le

traitement de plus haut niveau du langage est davantage représenté dans les cortex temporaux postérieurs (par exemple, le cortex d'association périsylvien) et le lobe frontal (en particulier l'IFG et le STG bilatéral ; Peelle, 2018). En revanche, la présence d'un message dégradé par exemple par le bruit (speech-in-noise) ou par d'autres messages (speech-in-speech) va entraîner le recrutement de régions corticales supplémentaires : frontales et pariétales (Alain et al., 2018 ; Defenderfer et al., 2017 ; Lawrence et al., 2018 ; Peelle, 2018 ; Scott et al., 2004) ; pour une revue sur les régions cérébrales impliquées dans le masquage informationnel et énergétique voir (Evans et al., 2016 ; Scott & McGettigan, 2013).

Dans les prochains paragraphes, l'attention sera portée sur les régions frontales et pariétales principalement impliquées : l'IFG, le DLPFC et le cortex pariétal.

II.1. Gyrus frontal inférieur (IFG)

Le gyrus frontal inférieur est un gyrus du lobe frontal du cortex cérébral, situé au-dessous du sillon frontal inférieur. Il se décompose en trois sous-parties anatomiques : *la pars triangularis* (BA 45), *la pars orbitalis* (BA 47) et *la pars opercularis* (BA 44). L'IFG, notamment l'IFG gauche (appelé aussi *Aire de Broca*) est considérée comme l'aire principale liée au traitement du langage, impliquée à la fois dans la production et dans la perception de la parole (Hickok & Poeppel, 2004 ; Liakakis et al., 2011) mais aussi dans la mémoire de travail verbal (Buchsbaum & D'Esposito, 2008 ; Davis et al., 2011 ; Eisner et al., 2010 ; Nixon et al., 2004). Différentes études montrent des fonctions différentes, bien que partiellement superposées, pour chacune de ces trois parties (Abrams et al., 2013 ; Skipper et al., 2007 ; Wang et al., 2021). Généralement, *la pars opercularis* de l'IFG serait plus liée à la production de la parole (Eickhoff et al., 2009), tandis que *la pars triangularis* serait plus liée au traitement sémantique de la parole (Skipper et al., 2007) de manière à être considérée comme le principal lieu de l'intelligibilité de la parole dans l'IFG (Abrams et al., 2013). Enfin, *la pars orbitalis* gérerait la dynamique entre langage et expression émotionnelle au cours de la communication (Belyk et al., 2017).

L'IFG est impliqué dans plusieurs aspects du traitement du langage : phonologique (Nixon et al., 2004), syntaxique (Ben-Shachar et al., 2004) et sémantique (Obleser & Kotz, 2010 ; Rodd et al., 2010).

II.1.1. L'IFG et l'effort d'écoute

La littérature scientifique a montré une implication importante de l'IFG gauche en situation d'écoute complexe. L'activation de cette région était plus importante lorsque le signal est dégradé mais encore intelligible comparé aux situations dans lesquelles le signal est parfaitement intelligible, ou à des conditions non intelligibles de bruit (Davis & Johnsrude, 2003 ; Lawrence et al., 2018 ; Wijayasiri et al., 2017 ; Wild et al., 2012, 2012). L'activité de l'IFG présenterait une relation non linéaire avec l'intelligibilité du stimulus à traiter. Cette activation suivrait une activation dite en *U-inversée* par rapport à la difficulté de la tâche. Cette relation non monotone a également été observée dans d'autres marqueurs objectifs de l'effort d'écoute, notamment la dilatation pupillaire (Zekveld et al., 2014). Les conditions d'intelligibilité intermédiaire, comparées aux conditions d'intelligibilité parfaite ou impossible, nécessitent un effort pour accomplir la tâche. Dans les conditions d'intelligibilité intermédiaire, l'auditeur reste encore motivé et se sent capable de réussir la tâche (Eckert et al., 2016 ; Pichora-Fuller et al., 2016; Vaden et al., 2013). En revanche, si la tâche devient trop difficile (quasi impossible) ou si le coût pour compléter la tâche est jugé trop élevé (Pichora-Fuller et al., 2016), l'activation se réduit, probablement à cause d'un désengagement de la part de l'auditeur. En outre, l'activité de l'IFG a été montrée comme dépendante de l'attention portée à la parole (Wijayasiri et al., 2017 ; Wild et al., 2012). À cet égard, dans l'étude de Wild et al. (2012), les auteurs ont utilisé une tâche de reconnaissance de phrases dont l'intelligibilité était manipulée et qui étaient présentées en présence de distracteurs (auditifs ou visuels). Ils ont étudié comment la distraction (par rapport à l'attention) pouvait affecter le traitement des stimuli. L'hypothèse était que si le traitement de la parole dégradée nécessitait plus de ressources attentionnelles que la parole intelligible, alors la distraction devrait interférer davantage avec le traitement de la parole dégradée qu'avec celui de la parole intelligible. Les auteurs ont montré que la parole intelligible était traitée même sans y diriger volontairement l'attention mais que diriger l'attention améliorerait considérablement le traitement de la parole dégradée. Les régions frontales, y compris l'IFG gauche, n'étaient engagées que lorsque les auditeurs étaient attentifs à la parole. De plus elles présentaient une activité plus élevée à la parole dégradée qu'à la parole intelligible. Les auteurs ont suggéré que cette activité corticale pouvait être considérée comme un marqueur neuronal de l'effort d'écoute.

II.1.2. L'IFG dans l'entraînement

Le rôle de l'IFG a également été montré dans des études qui ont utilisé l'entraînement avec des tâches d'écoute (Bless et al., 2014 ; Eisner et al., 2010 ; Qi et al., 2019). L'activité de l'IFG était modulée par l'entraînement, certaines études montraient une réduction de l'activité de l'IFG après l'entraînement (Qi et al., 2019) tandis que d'autres montraient une augmentation de l'activité (Eisner et al., 2010). Les différences en termes de protocoles expérimentaux utilisés (c.-à-d., stimuli différents, durée d'entraînement différente) pourraient expliquer ces différences de résultats. Eisner et al. (2010), dans leur étude en IRMf, ont étudié la plasticité neuronale associée à l'apprentissage de phrases dégradées. Les résultats comportementaux ont montré une grande variabilité individuelle à la fois dans la capacité d'apprendre à comprendre la parole dégradée et dans la capacité de mémoire de travail phonologique. Les résultats de l'IRMf ont montré que des régions situées sur le côté gauche du sillon temporal supérieur (STS) et du gyrus frontal inférieur (IFG) étaient sensibles à la capacité d'apprentissage, mais seule l'activité de l'IFG était corrélée à la variabilité interindividuelle des scores d'intelligibilité et de la capacité de mémoire phonologique. De plus, avec l'entraînement, une augmentation de la connectivité fonctionnelle entre les régions frontales (IFG gauche) et pariétales (gyrus angulaire gauche) a aussi été observée (Eisner et al., 2010).

Dans une autre l'étude, Qi et al. (2019) ont mesuré l'activité corticale pendant des tâches de discrimination de sons vocaux (mandarin) et de sons sinusoïdaux, avant et après un cours intensif de mandarin (c'est-à-dire l'écoute, la lecture et l'écriture de mandarin) chez des personnes de langue maternelle anglaise. Ils ont montré que les participants présentant une diminution plus importante de l'activité de l'IFG droit avaient une meilleure rétention à long terme du mandarin.

La réduction de l'activité de l'IFG suite à l'entraînement a également été démontrée dans des études utilisant des tâches autres que des tâches d'écoute (Chavan et al., 2015).

Dans une étude réalisée en IRMf, Chavan et al. (2015) proposaient une tâche de type Go/NoGo pour étudier les effets d'un entraînement de deux semaines sur les mécanismes de contrôle inhibiteur frontal top-down. Ils ont constaté que les améliorations comportementales étaient associées à une diminution de l'activité dans les IFGs bilatéraux. Ils ont supposé que cette diminution pouvait refléter un engagement plus efficient des populations neuronales impliquées dans le processus d'inhibition pour réaliser la tâche (Bless et al., 2014 ; Curtin et al., 2019 ; Fairclough et al., 2018 ; Putkinen & Saarikivi, 2018). En outre, les auteurs ont révélé que

l'entraînement conduisait aussi à des changements de l'anatomie cérébrale avec une augmentation du volume de la matière grise dans la *pars orbitalis* de l'IFG droit et des modulations de la microstructure de la matière blanche dans la *pars triangularis* de l'IFG droit. Les interprétations possibles de ces résultats, c'est-à-dire la réduction de l'activité cérébrale après un programme d'entraînement, proviennent souvent des prémisses de la théorie de l'efficacité neuronale (c.-à-d., *neural efficiency theory*, Curtin et al., 2019 ; Fairclough et al., 2018). La théorie de l'efficacité neuronale a été énoncée pour la première fois en 1988 quand Haier et ses collègues, dans une étude TEP, ont trouvé une relation inverse entre le métabolisme du glucose dans le cerveau pendant l'exécution d'un test cognitif d'intelligence (c.-à-d., le test des Matrices Progressives de Raven) et les scores obtenus. Les participants ayant obtenu les scores les plus élevés étaient les mêmes que ceux dont la consommation d'énergie cérébrale était la plus faible. A partir de là, les études se sont multipliées et différentes variables modérant la relation entre l'intelligence et l'activation cérébrale ont été prises en compte telles que le sexe, le type et la complexité de la tâche, la quantité ainsi que la qualité des programmes d'apprentissage (Neubauer & Fink, 2009) et la complexité de la tâche (Fairclough et al., 2018).

Fairclough et ses collègues, dans leur article de 2018, ont souligné comment les questions scientifiques étudiées par cette théorie se croisent avec celles soulevées par la neuro-ergonomie qui s'interroge sur les interactions entre l'activation neurophysiologique d'un opérateur en situation opérationnelle, la demande de la tâche et l'efficacité de la performance (Parasuraman, 2003).

Les auteurs Neubauer et Fink dans leur review (2009) ont porté aussi l'attention sur l'importance de prendre en compte avec les données de performance des tâches comportementales et d'imagerie cérébrale ainsi que les mesures subjectives de l'effort pour bien comprendre les interactions entre toutes les variables en jeu.

Ce qu'il faut retenir

- L'activité de l'IFG gauche est fréquemment associée au traitement du langage en situation d'écoute complexe.
- L'IFG est impliqué dans plusieurs aspects du traitement du langage : phonologique, syntaxique et sémantique. Il est aussi impliqué dans la mémoire de travail verbale.
- L'activité de l'IFG suit une activation à loi en *U-inversée* par rapport à la difficulté de la tâche.
- La réduction de l'activité cérébrale de l'IFG suite à l'entraînement s'insère dans le cadre de la *neural efficiency theory*.

II.2. Cortex préfrontal dorsolatéral

D'un point de vue anatomique, le DLPFC est une grande région qui comprend les portions latérales des aires BA 9, 44, 45 et 46. D'un point de vue fonctionnel, il correspond aux portions dorsolatérales de BA 9 et BA 46 majoritairement dans le gyrus frontal moyen (MFG) et une partie de l'IFG (Scannella, 2011). Outre les différentes fonctions qui lui sont attribuées (p. ex., contrôle de l'inhibition, attention, mémoire, planification), le DLPFC est associé au traitement du langage et notamment lorsque la parole est dégradée par rapport à une parole claire (Hertrich et al., 2021).

Le DLPFC est fréquemment associé à des tâches nécessitant l'implication de fonctions du système exécutif (Ferbinteanu, 2019 ; Panikratova et al., 2020 ; Xiong & Newman, 2021). Parmi ces fonctions, figurent l'inhibition et la sélection avec une distinction entre le DLPFC gauche qui représenterait le siège de processus d'inhibition/sélection des informations non pertinentes/pertinentes en entrée et le DLPFC droit qui traiterait des mêmes types de processus, mais au niveau de la réponse motrice (voir Scannella, 2011, pour une synthèse). Son implication a été ainsi retrouvée dans des tâches d'inhibition classiques comme la stop-signal task (Rubia et al., 2003).

Hertrich et al. (2021), dans une revue récente, mettent aussi en évidence les différentes fonctions du DLPFC liées au traitement du langage. L'activité du DLPFC serait associée aux différents aspects de la pragmatique linguistique (c.-à-d., la discipline qui étudie le langage et

son usage chez les communicants) comprenant le traitement des phrases simples et complexes, la cohérence du discours, l'interaction entre auditeur et locuteur pendant une conversation, la prosodie, le bilinguisme, la compréhension du sens non littéral des messages et les mécanismes prédictifs de type top-down (pour une revue détaillée, lire Hertrich et al., 2021). Ces derniers mécanismes sont particulièrement visibles dans des situations d'écoute complexe (speech-in-noise ou situations multilocuteurs) où l'auditeur, afin de comprendre, essaye de prédire le contenu du message. Le DLPFC est aussi impliqué dans la mémoire de travail verbale (D'Esposito & Postle, 2015 ; White & Langdon, 2021). Dans les situations d'écoute complexe, la mémoire est aussi fortement sollicitée afin de permettre la compréhension de stimuli dégradés (Peelle, 2018).

Ce qu'il faut retenir

- L'activité du DLPFC est associée au traitement du langage notamment aux différents aspects de la pragmatique linguistique.
- Le DLPFC est associé à des tâches qui, pour être accomplies, nécessitent l'implication de fonctions du système exécutif (inhibition/sélection).

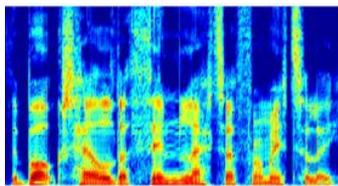
III. Du langage aux mécanismes *general domain*

Le recrutement de toutes les régions cérébrales précédemment évoquées, sous-tendu par une connectivité structurelle et fonctionnelle, semble donc être nécessaire afin de pouvoir comprendre en situation d'écoute complexe. De fait, ces situations nécessitent la mise en place de plusieurs mécanismes qui comprennent, entre autres, des mécanismes attentionnels, de mémoire, d'inhibition, de détection des erreurs (Andéol et al., 2017 ; Peelle et al., 2009 ; Rodd et al., 2010). Par exemple, l'activité de l'IFG a été aussi associée à la prise de décision dans un paradigme de speech-in-noise (Binder et al., 2004). Binder et al. (2004) dans leur étude en IRMf montraient une distinction fonctionnelle entre les mécanismes sensoriels et décisionnels qui sous-tendent l'identification auditive des stimuli dans le bruit. En particulier, l'activation de régions proches du cortex auditif primaire prédisait la précision de l'identification des sons tandis que l'activation de l'IFG gauche prédisait le temps de réponse à la tâche. L'activation de l'IFG pouvait donc refléter la mise en place de mécanismes liés à la prise de décision plutôt que

l'intelligibilité elle-même. En outre, l'activité de l'IFG a été associée au contrôle inhibiteur (Chavan et al., 2015 ; Hartmann et al., 2016 ; Manuel et al., 2013 ; Qi et al., 2019).

Ces mécanismes ne seraient pas spécifiquement liés à des tâches langagières, mais aussi à d'autres tâches comme les tâches de Stroop ou de Flanker qui requièrent la mise en place de processus d'inhibition et de sélection, de mémoire et de prise de décision. Pour cette raison, ils font partie des processus dits *general domain* qui se mettent en place indépendamment des caractéristiques du stimulus, mais qui peuvent requérir la mise en place des mêmes types de mécanismes (Peelle, 2018). Peelle dans sa revue (Peelle, 2018) souligne comment deux réseaux de type *general domain* interviendraient afin de réaliser la tâche d'écoute en situation complexe : le réseau cingulo-operculaire et le réseau fronto-pariétal (Figure 7) dont l'IFG, le DLPFC et le cortex pariétal font partie.

Acoustically clear speech



Degraded speech

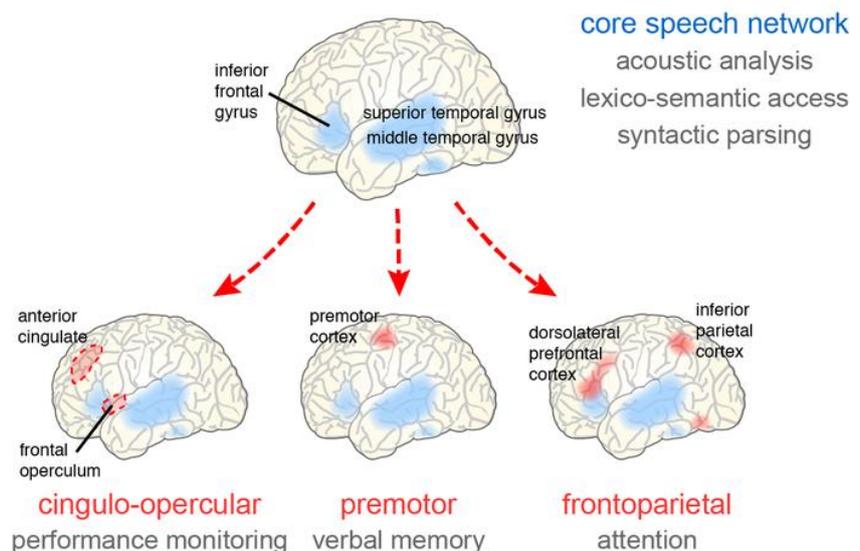
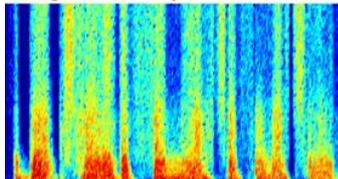


Figure 7 : Les différents réseaux cérébraux impliqués dans le traitement du langage selon la qualité du stimulus : acoustiquement intelligible ou dégradé. Un signal de parole acoustiquement intelligible sollicitera l'activation du *core speech network* afin de permettre son traitement acoustique, phonologique, lexico-sémantique et syntaxique. En revanche, un signal de parole acoustiquement dégradé impliquera l'activation d'autres régions cérébrales, notamment frontales et pariétales, afin de permettre le monitoring de la performance et l'emploi de ressources supplémentaires en mémoire et attention (d'après Peelle, 2018).

Le système cingulaire-operculaire assure l'accomplissement de la tâche grâce à l'intention du participant de rester engagé tandis que le système fronto-pariétal assure la réalisation de la tâche

à travers la focalisation de l'attention sur les stimuli pertinents, l'inhibition des stimuli non pertinents et donc la sélection de la réponse appropriée à la tâche (Eckert et al., 2016). Dans ces situations d'écoute complexe, un auditeur, afin de comprendre, doit en même temps vouloir rester engagé dans la tâche et être suffisamment attentif afin de se focaliser sur son message d'intérêt (p. ex., une voix cible) et inhiber les messages non pertinents (p. ex., voix masques, bruit).

Ce qu'il faut retenir

- Le *core speech network* est composé principalement par le cortex auditif bilatéral et l'IFG.
- Les situations d'écoute complexe requièrent un recrutement de régions cérébrales supplémentaires (frontales et pariétales).
- Le traitement du langage en situation d'écoute complexe implique la mise en place de mécanismes de type *general domain* (attention, mémoire, contrôle inhibiteur).
- Le réseau cingulo-operculaire et le réseau fronto-pariétal interviennent dans le traitement du langage en situation d'écoute complexe.

Les paragraphes suivants traitent plus particulièrement de la description et du rôle des deux réseaux cérébraux fonctionnels impliqués dans le traitement du langage en situation d'écoute complexe : le réseau cingulo-operculaire et le réseau fronto-pariétal, suivi de la description du rôle de l'ensemble des régions clefs du traitement langagier, qu'il soit spécifique au langage ou plus général.

III.1. Le réseau cingulo-operculaire

Le réseau cingulo-operculaire comprend bilatéralement, le cortex cingulaire dorsal, l'IFG et l'insula antérieure qui sont tous fortement interconnectés (Catani et al., 2012). Cette connectivité permet un contrôle continu (avec un ajustement possible) des performances en fonction de la difficulté de la tâche et de la conscience de sa réalisation effective (Kerns, 2006 ; Shenhav et al., 2013).

Une plus grande activation de ce réseau est donc en lien avec l'intention du participant de rester engagé dans une tâche complexe en surveillant sa propre performance et en corrigeant les erreurs commises. À cette fin, l'activité de ce réseau représente une véritable activité de régulation top-down de la performance. Par exemple, l'activation de ce réseau émerge aussi dans la tâche de Stroop (Leung, 2000 ; Peterson et al., 1999) dans laquelle il est nécessaire de surveiller sa propre performance en essayant d'inhiber les réponses automatiques suscitées par lecture d'un mot pour se focaliser sur la dimension de sa « couleur », moins automatique que la lecture. À ce propos, dans la revue de Peelle (2018), l'auteur propose un lien possible entre les processus attentionnels d'inhibition et les processus attentionnels de contrôle de la performance. Par ailleurs, certaines études montrent que les participants qui ont de bons scores à la tâche de Stroop sont les mêmes à avoir de bonnes performances à des tâches de compréhension de la parole en situation d'écoute complexe (Sommers & Danielson, 1999) mettant en évidence, une fois encore, l'utilité d'un contrôle inhibiteur efficace de type domaine général.

L'activité du réseau cingulo-operculaire serait donc primordiale chez l'auditeur qui se rend compte, en contrôlant sa performance, des erreurs commises, et qui cherche intentionnellement à les corriger afin de réaliser la tâche de façon optimale. Par exemple, il a été montré que ce réseau est plus activé lorsque l'auditeur est activement engagé dans une tâche de reconnaissance de mots plutôt que lors d'une tâche d'écoute passive (Fiez et al., 1995), ou lorsqu'on lui demande de porter son attention sur un stimulus spécifique (Wild et al., 2012). En revanche, si la complexité de la tâche est trop élevée, l'auditeur se désengage et l'activation de ce réseau s'affaiblit de manière concomitante (Zekveld et al., 2006, 2012). De fait, l'activité cingulo-operculaire est en même temps liée à l'amélioration de la performance et à un effort accru (Botvinick et al., 2009). Afin de rester engagé dans la tâche, le participant doit considérer que ses efforts en valent la peine. L'activité cingulo-operculaire est donc présente lorsque les conditions d'écoute sont difficiles, mais suffisantes pour la reconnaissance de la parole. La composante intentionnelle sous-tend donc l'activation de ce réseau qui reste sollicité malgré la complexité de la tâche, indiquant une volonté forte de l'individu et donc la motivation à surveiller ses propres performances et à s'améliorer, en d'autres termes à rester engagé dans la tâche (Shenhav et al., 2013 ; Vaden et al., 2013).

Ce qu'il faut retenir

- Le réseau cingulo-operculaire reflète l'intention de contrôler ses propres performances en corrigeant les erreurs commises afin de s'améliorer.
- L'activité cingulo-operculaire est présente lorsque les conditions d'écoute sont complexes, mais suffisantes pour la reconnaissance de la parole.

III.2. Le réseau fronto-pariétal

Le système fronto-pariétal comprend le sillon frontal inférieur/sillon central, le cortex préfrontal dorsolatéral/gyrus frontal moyen, le sillon interpariétal et les régions du lobule pariétal inférieur (Dosenbach et al., 2008). L'activité de ce système favorise la focalisation sur le message d'intérêt et l'inhibition des messages non pertinents en générant une réponse appropriée à la tâche (Kerns, 2006 ; Obleser et al., 2007).

Au sein du système fronto-pariétal, il y aurait différentes régions qui seraient activées en fonction des caractéristiques du stimulus (localisation spatiale ou caractéristiques spectrales) vers lequel l'attention est dirigée (Corbetta et al., 2000 ; Hill & Miller, 2010 ; Shomstein & Yantis, 2006). Hill & Miller (2010), en utilisant une tâche d'écoute multilocuteurs, proposait aux participants de distinguer la voix cible selon sa localisation spatiale ou selon ses caractéristiques spectrales. Les participants étaient prévenus grâce à la présence d'un indice qui annonçait vers quelle caractéristique porter l'attention afin de se focaliser sur la voix cible. Ce faisant, ils ont pu dissocier le contrôle attentionnel exercé sur la tâche, de la sélection de la voix cible. Concernant le contrôle de l'attention, un réseau fronto-pariétal latéralisé à gauche constitué du sillon précentral dorsal et du lobule pariétal supérieur était activé lors du traitement spatial du stimulus tandis que l'IFG était impliqué dans le traitement des caractéristiques spectrales du stimulus. Pendant la sélection de la voix cible, le sillon intrapariétal gauche était impliqué lors du traitement spatial du stimulus tandis que le STS bilatéral, avec dominance à droite, était impliqué dans le traitement des caractéristiques spectrales de la voix cible.

III.3. Collaboration entre les réseaux fronto-pariétal et cingulo-operculaire

Les deux réseaux collaborent afin d'assurer la réussite de la tâche. L'activité de l'un sans l'autre n'est pas suffisante pour atteindre l'objectif fixé, qui dans ce cas est celui de comprendre le message cible.

Le succès de la tâche sera remis en question si l'auditeur est capable de rester focalisé sur le message cible, mais est incapable de surveiller sa propre performance, ou si l'auditeur est capable de surveiller sa propre performance, mais est incapable de focaliser sur la cible et d'inhiber les distracteurs (Eckert et al., 2016). Des études montrent que la collaboration entre les deux réseaux serait sous-tendue par une connectivité fonctionnelle et structurelle entre les deux (Eckert et al., 2016; Taren et al., 2011). De fait, le cortex cingulaire antérieur et le DLPFC présentent une connectivité fonctionnelle au repos qui semble être soutenue par des trajets de fibres directs entre les deux régions (Taren et al., 2011).

En outre, des études qui évaluent la dynamique de la performance (essai après essai) démontrent que l'activité du cingulum antérieur prédit positivement la mesure de l'activité du réseau fronto-pariétal lors d'un essai ultérieur (Kerns, 2006 ; Walsh et al., 2011). De plus, des activités plus importantes dans le cingulum antérieur et dans les régions du réseau fronto-pariétal sont associées à une réponse plus rapide et plus précise lors de l'essai en cours (Kerns, 2006).

Ce qu'il faut retenir

- Le système fronto-pariétal favorise la focalisation sur le message d'intérêt et l'inhibition des messages non pertinents en générant une réponse appropriée à la tâche.
- Les deux réseaux, cingulo-operculaire et fronto-pariétal, collaborent afin d'assurer la réussite de la tâche. L'activité de l'un sans l'autre n'est pas suffisante pour atteindre l'objectif fixé.

Chapitre III. L'effort d'écoute

I. Charge cognitive

La notion de charge cognitive est souvent présente dans la littérature scientifique où ce concept est fréquemment associé à celui de charge mentale (Chanquoy et al., 2007). Ce concept est employé dans les disciplines de la psychologie, des neurosciences et de la neuro-ergonomie. La neuro-ergonomie prend en compte l'individu (l'opérateur) dans son système de travail (p. ex., un pilote dans son cockpit) et vise à déterminer si l'opérateur est capable de travailler en sécurité et de manière efficiente dans son environnement. Le scientifique Raja Parasuraman ouvre les portes à cette nouvelle thématique scientifique dans le domaine de l'industrie aérospatiale (Parasuraman, 2003). Selon lui, étudier l'individu immergé dans son contexte de travail peut permettre de concevoir des systèmes de travail qui sont mieux adaptés au fonctionnement humain. La neuro-ergonomie est née de cette approche en ce sens qu'elle étudie la manière dont les processus cognitifs, émotionnels et physiques d'un individu peuvent jouer un rôle décisif dans les situations de travail. Prendre en compte la notion de charge cognitive devient alors important. La notion de « charge » laisse présager une interaction stricte entre les ressources à disposition de l'individu, à un moment donné, et le type et le niveau de complexité intrinsèque à une tâche. Le type de tâche peut exiger de l'individu la mise en place de ressources spécifiques, par exemple des ressources plutôt physiques, cognitives ou émotionnelles.

Aborder la définition de la charge mentale n'est pas facile en raison de la variété des différentes approches théoriques existantes. Dans leur ouvrage, les chercheurs Chanquoy et al. (2007) résument les différentes contributions théoriques à la définition de la charge mentale, en la centrant sur trois points principaux : l'individu, la tâche qu'il doit réaliser et son environnement.

Dans le paragraphe suivant, l'intérêt sera porté sur la charge cognitive liée à l'audition que l'on peut traduire par **l'effort d'écoute**.

II. La nature multidimensionnelle de l'effort d'écoute

Bien que la communauté scientifique travaille depuis longtemps sur la notion d'effort d'écoute, ce concept échappe encore à une définition claire et universelle. Une des raisons réside dans sa nature multifactorielle. L'effort d'écoute fait référence à un concept dynamique : on parle

d'effort d'écoute spécifique à une situation donnée dans laquelle l'individu est dans un certain état physique et psychologique. La situation change, l'individu change, l'effort d'écoute change alors aussi. Certains auteurs considèrent l'effort d'écoute comme le vécu subjectif d'un auditeur qui fait face à une situation d'écoute complexe (Johnsrude & Rodd, 2015 ; Krueger et al., 2017). D'autres considèrent l'effort d'écoute comme un acte mental, un recrutement de ressources cognitives pour réaliser la tâche (Peelle, 2018 ; Pichora-Fuller et al., 2016). Dans la présente section, nous présentons une synthèse des connaissances et concepts associés à l'effort d'écoute dans le but d'en extraire la définition la plus consensuelle possible.

Un bon point de départ peut se trouver chez les auteurs Pichora-Fuller et al. (2016) qui dans leur modèle intitulé *Framework for Understanding Effortful Listening* (FUEL), définissent l'effort d'écoute ainsi : “*the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a task, with listening effort applying more specifically when tasks involve listening*” (Pichora-Fuller et al., 2016, page 5S). Cette définition emploie le terme *délibéré*. Ainsi, nous pouvons entrevoir un autre concept clé lié à la définition de l'effort d'écoute : celui de la motivation. Plus l'individu est motivé, plus il sera possible de tolérer le poids que la demande cognitive engendre. Les auteurs continuent en énonçant que la réussite des objectifs d'écoute peut dépendre du déploiement d'une plus grande énergie cognitive lorsque la qualité du signal disponible pour l'auditeur est sous-optimale. Cependant, il n'y a aucune garantie que l'augmentation de l'énergie cognitive soit suffisante pour résoudre tous les problèmes d'écoute. Dans certaines situations, lorsque les auditeurs ne peuvent ou ne veulent pas maintenir un niveau d'effort suffisamment élevé, ils peuvent ressentir de la fatigue et/ou décider d'abandonner la tâche en cours pour éviter de se fatiguer. Dans d'autres situations, la récompense de l'engagement immersif dans la communication peut avoir l'effet inverse, dans la mesure où certains auditeurs trouvent que les avantages intellectuels et sociaux de l'écoute et de la conversation augmentent la motivation et ajoutent de la valeur à l'effort déployé. À long terme, si l'écoute dans les activités quotidiennes exige fréquemment plus d'efforts que ce que les auditeurs sont capables ou désireux de faire, ils peuvent ressentir du stress chronique et se retirer de l'interaction sociale, avec des conséquences négatives sur la cognition, la santé générale, le bien-être et la qualité de vie (Pichora-Fuller et al., 2016).

Peelle (2018), dans sa revue de littérature, se concentre sur la demande accrue en ressources cognitives engendrée par les situations d'écoute complexe. À partir du modèle FUEL présenté ci-dessus, l'auteur décrit comment les défis acoustiques soulevés par des situations d'écoute

complexe (p. ex., *acoustic challenge*) peuvent entraîner une demande cognitive et il fournit un cadre de réflexion sur les processus cognitifs spécifiques qui pourraient être impliqués dans la compréhension des stimuli acoustiquement dégradés. En détail, le défi acoustique est défini comme la demande cognitive composée par des difficultés propres à l'auditeur (c.-à-d. troubles de l'audition, âge...), des difficultés propres au stimulus à déchiffrer (c.-à-d. accent étranger, articulation) et des difficultés propres à l'environnement (c.-à-d. bruit, situations multilocuteurs). À la suite du modèle FUEL, il décrit comment un défi acoustique accru entraîne une plus grande demande cognitive, qui, modulée par la motivation de l'auditeur, conduit à un effort d'écoute accru (Peelle, 2018) ; (Figure 8).

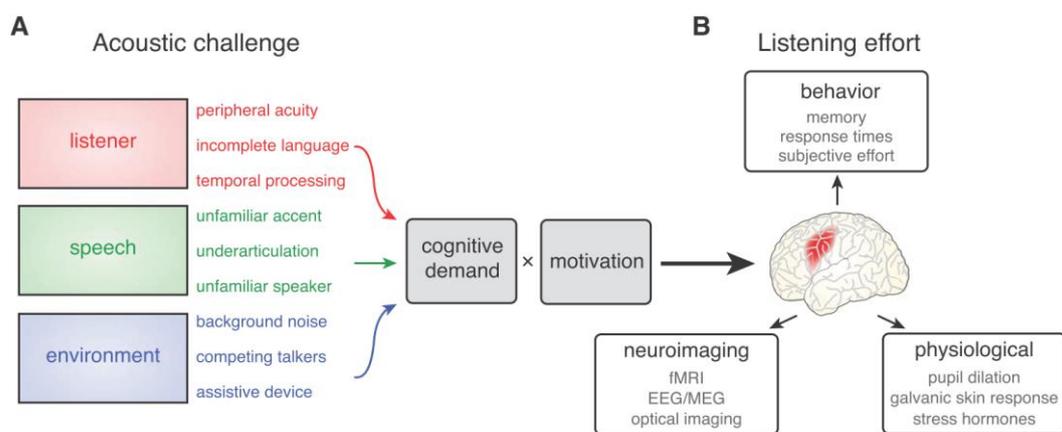


Figure 8 : Modèle de l'effort d'écoute d'après Peelle, 2018. A) Le défi acoustique (composé par l'auditeur, le signal, l'environnement) engendre une demande cognitive qui est modulée par la motivation. B) Les différentes stratégies pour évaluer l'effort d'écoute (mesures comportementales, mesures de neuro-imagerie et mesures physiologiques).

Pour aller plus loin, Herrmann & Johnsrude (2020) ont conceptualisé un nouveau modèle appelé MOLE (*Model of Listening Engament*). Les auteurs partent du constat qu'il existe une confusion quant à la définition de l'effort d'écoute. La confusion naît du fait que pour certains auteurs l'effort d'écoute est considéré comme l'expérience subjective de la difficulté ressentie (Johnsrude & Rodd, 2015 ; Krueger et al., 2017) tandis que pour d'autres auteurs la notion d'effort d'écoute fait référence à un processus ou à un acte mental dans lequel l'effort est investi (McGarrigle et al., 2014 ; Peelle, 2018 ; Pichora-Fuller et al., 2016). Les auteurs soulignent également le fait que les matériaux expérimentaux utilisés ne correspondent pas toujours à ceux de la vie quotidienne (utilisation de phrases répétitives, mots isolés, syllabes). La qualité des stimuli utilisés ne reflète pas, par exemple, des vécus de curiosité, de découverte, surprise ou

même aussi d'ennui, qu'une vraie conversation peut susciter. Les auteurs se demandent ce qui pousse l'auditeur à rester engagé dans la tâche malgré la complexité (notamment pour les personnes malentendantes), et ce qui le pousse à se désengager. Les concepts d'engagement, de motivation et de vécu subjectif deviennent centraux dans ce modèle (Figure 9).

Le « *listening engagement* » est défini comme : *"un recrutement automatique ou volontaire de ressources exécutives et d'autres ressources cognitives, lorsque la compréhension de la parole sert un objectif de communication important : l'engagement permet le contrôle cognitif et le maintien d'une implication consciente ou explicite dans une activité d'écoute telle qu'une conversation "*. Les auteurs supposent que chaque individu disposerait d'une quantité limitée de ressources cognitives. Dans une situation d'écoute donnée, des ressources spécifiques peuvent être recrutées. Au même moment une quantité définie peut être nécessaire pour garantir la compréhension de ce qui est écouté. D'autres ressources peuvent rester inutilisées et prêtes pour répondre aux autres demandes. Ce déploiement correspond à l'engagement qui est en relation directe avec la motivation vue comme une force qui pousse l'auditeur à décider de recruter ou non ses ressources pour réaliser la tâche et atteindre ses objectifs. Cet engagement génère un vécu subjectif chez l'auditeur. Ce vécu peut être un vécu d'effort, de frustration, d'ennui, mais aussi de la joie, de plaisir à pouvoir comprendre. La qualité de ce vécu subjectif, à son tour, pourra alimenter la motivation de l'auditeur pour continuer d'engager des ressources cognitives ou pourra, au contraire, pousser l'auditeur à y renoncer (Herrmann & Johnsrude, 2020). L'apport des neurosciences et des sciences cognitives devient alors nécessaire afin d'évaluer et quantifier l'engagement des ressources cognitives dans la compréhension des mécanismes sous-jacents l'écoute en situations complexes.

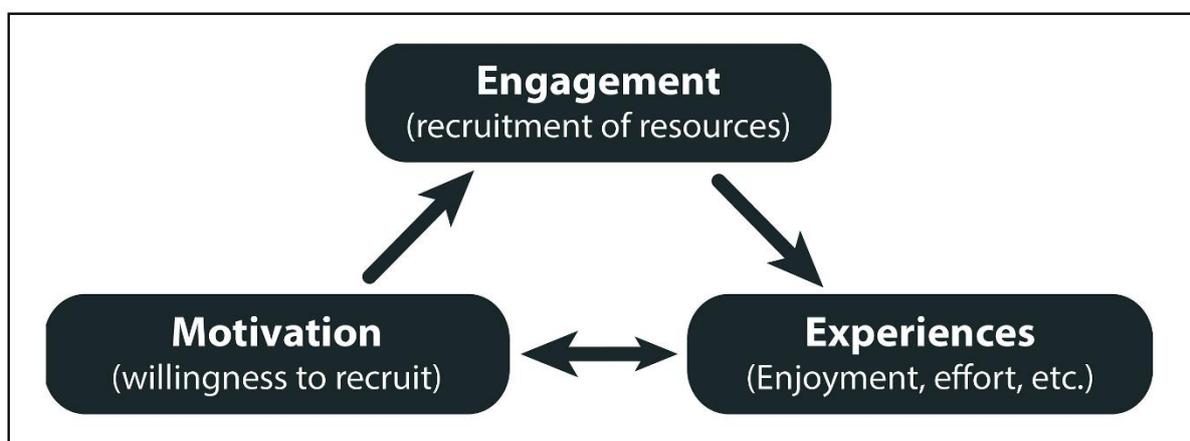


Figure 9 : Model of Listening Engagement, d'après Herrmann & Johnsrude, 2020. Ce schéma illustre la relation entre l'engagement, la motivation et le vécu d'un auditeur. La motivation

d'une personne et la relation entre les ressources requises et les limites des ressources (c'est-à-dire les quantités de ressources disponibles) déterminent l'engagement. Ces caractéristiques déterminent le vécu de l'auditeur, ce qui influence également la motivation.

Il apparaît donc que la notion d'effort d'écoute peut être modélisée de différentes manières. Néanmoins, tous les modèles soulignent l'importance de prendre en compte plusieurs dimensions (subjectives et objectives) pour extraire une évaluation globale de l'effort d'écoute.

Ce qu'il faut retenir

- Le concept d'effort d'écoute est multifactoriel et dynamique (*la situation change, l'individu change, l'effort d'écoute change aussi*).
- L'effort d'écoute est le résultat de l'interaction de différentes variables, dont celles concernant l'auditeur, en termes de ressources et d'expériences subjectives, et celles concernant la complexité du contexte dans lequel l'auditeur se trouve.

III. Facteurs influençant l'effort d'écoute

III.1. Ressources cognitives : attention, mémoire

Parmi les composantes essentielles qui interviennent dans les tâches d'écoute complexes et qui peuvent influencer l'effort d'écoute figurent les ressources cognitives dont dispose l'auditeur. Les ressources cognitives de l'auditeur sont limitées (Kahneman, 1973). Une tâche d'écoute complexe peut nécessiter l'engagement de plusieurs fonctions exécutives parmi lesquelles la mémoire de travail, le contrôle de l'inhibition et de l'interférence, ainsi que l'attention sélective (Lemke & Besser, 2016 ; Pichora-Fuller et al., 2016 ; Wingfield, 2016). Lorsque nous nous trouvons dans une situation d'écoute complexe, nous devons être capables de nous concentrer sur le message qui nous intéresse, d'inhiber les messages qui créent des interférences et, en même temps, de rester flexibles et ouverts à d'éventuels changements de situation (p. ex., le locuteur cible change). Toutes ces ressources sont impliquées afin d'assurer la réussite de la tâche d'écoute. Il existe différents modèles pour expliquer le processus d'allocation de ressources dans la perception de la parole (Francis & Love, 2020).

Pour donner un exemple, il existe des modèles dits *active models* (Heald & Nusbaum, 2014) qui prévoient que lorsque le signal est dégradé, l'attention sélective doit s'ajuster de manière

adaptative sur les propriétés du signal qui sont les plus robustes et/ou les plus accessibles. Au même moment, des interprétations potentielles du stimulus sont transférées de la mémoire à long terme vers la mémoire de travail pour être comparées. Ces processus entraîneraient un effort sous la forme d'une demande croissante de ressources cognitives en termes de mémoire et d'attention et la complexité du stimulus à traiter déterminera un engagement plus ou moins important de ces ressources.

Le recrutement des ressources cognitives permettant de comprendre un message dans une situation complexe peut être déficitaire à cause de la présence de troubles de l'audition.

Par exemple, dans les travaux de Piquado et al. (2012), des auditeurs malentendants effectuant une tâche de rappel d'histoires écoutées présentaient de moins bonnes performances de mémoire par rapport aux auditeurs ayant une audition normale. De plus, ils ont montré que les performances s'amélioraient lorsque les participants ralentissaient la vitesse à laquelle les histoires étaient racontées. Ces résultats soutiennent l'hypothèse selon laquelle l'effort d'écoute associé à une acuité auditive réduite peut ralentir les opérations de traitement et augmenter les demandes en mémoire de travail, avec des effets négatifs conséquents sur la précision du rappel narratif.

Une autre possibilité d'évaluer l'importance de disposer de ressources cognitives provient des études sur les lésions cérébrales. Krause & Richards (2014), ont exploré les différences entre les adultes avec et sans lésion cérébrale traumatique (Traumatic Brain Injury, TBI) par rapport à la précision du traitement de la parole, la vitesse de traitement et l'effort dans diverses conditions d'écoute (speech-in-noise et speech-in-speech). Les résultats de cette étude montrent des similitudes entre les groupes en ce qui concerne la précision de la répétition dans les conditions d'écoute, mais des différences entre les groupes en ce qui concerne la vitesse et l'effort : le groupe TBI a montré une vitesse de traitement plus lente (c.-à-d., des temps de réponse plus lents) que le groupe contrôle et a également rapporté un effort significativement plus important dans la condition multilocuteursmultilocuteurs.

Ce qu'il faut retenir

- Les situations d'écoute complexe requièrent un engagement supplémentaire de ressources cognitives (mémoire, attention), qui sont limitées.

III.2. Motivation

Il est de plus en plus clair que pour faire face à des situations d'écoute complexe il faut considérer la variable motivationnelle. Selon le *Framework for Understanding of Effortful Listening* (FUEL) de Pichora-Fuller et al. (2016), une écoute performante dans des conditions complexes nécessite de la motivation et des efforts.

Eckert et al. (2016), partant des hypothèses des études d'économie comportementale, tentent de comprendre quand et pourquoi l'auditeur décide de ne pas écouter. La motivation dépend du coût ou de l'effort pour écouter. Les auteurs fournissent un exemple de la façon dont l'effort peut être modélisé dans deux scénarios d'écoute différents : avoir une conversation avec un proche et écouter une publicité. En présence d'une difficulté importante de la tâche, si la motivation à faire l'effort n'est pas assez forte, l'auditeur cessera de faire l'effort et donc d'écouter. Vis-à-vis de la difficulté de la tâche, l'auditeur est plus enclin à faire l'effort d'écouter un de ses amis qu'une publicité qui ne l'intéresse pas. De plus, il peut aussi arriver qu'en présence d'une situation simple, la motivation se perde. Par exemple, l'effet du caractère agréable ou non de l'écoute a rarement été évaluée. Le plaisir dans l'échange motive et aide l'auditeur ; il réduit le coût d'écoute sans nécessairement réduire l'effort d'écoute (Matthen, 2016). Comprendre pourquoi un auditeur décide d'arrêter d'écouter devient une question importante. Par exemple, les personnes malentendantes sont parmi les personnes les plus affectées par les situations d'écoute complexe. Ces situations génèrent un vécu de frustration, d'effort, de perte de la motivation qui peut conduire au stress et même à la dépression (Arlinger, 2003 ; Pichora-Fuller et al., 2015). De plus, des recherches ont été réalisées pour étudier l'effort d'écoute dans l'environnement scolaire et la perte de motivation à écouter. Ces études sont importantes afin de concevoir des activités éducatives et d'apprentissage optimales, tant sur le plan du contenu que sur les moyens de le diffuser. Par exemple, Rudner et al. (2018) ont étudié des enfants d'école primaire et ont observé les effets combinés du bruit de fond (plusieurs enfants qui parlent), de la qualité de la voix (une voix dysphonique : rauque) et des indices visuels, sur la compréhension et l'effort d'écoute des enfants. Parmi les variables, la dysphonie ne réduisait pas de manière significative les scores de compréhension, mais elle était considérée comme désagréable et rendait l'écoute difficile, probablement en réduisant la motivation à écouter.

Ce qu'il faut retenir

- La motivation est une variable modératrice de l'effort d'écoute.
- Pour rester engagé dans la tâche, l'auditeur doit penser ou ressentir que *l'effort en vaut la peine*.

III.3. L'état auditif et la difficulté de son évaluation

III.3.1. Les lésions auditives chez le personnel militaire

Le personnel militaire est particulièrement à risque de lésions auditives du fait de son exposition aux bruits intenses tels que ceux générés par les armes à feu ou les véhicules terrestres ou aériens. Comme rapporté dans les travaux de Tepe et al. (2017), les acouphènes et les pertes auditives sont les deux handicaps les plus communément pris en charge par le ministère des Anciens Combattants du gouvernement des États-Unis (VA, Veteran Affairs). Pour citer un autre exemple, un quart de toutes les blessures chez les marines au cours de l'Opération *Iraqi Freedom* est représenté par le traumatisme sonore aigu (TSA) dû à une explosion d'une bombe artisanale ou d'une mine (Gondusky & Reiter, 2005). Un TSA est défini comme l'« apparition ou l'aggravation d'au moins un symptôme audio-vestibulaire (acouphène, hypoacousie, vertige...) dans les suites immédiates d'une exposition sonore jugée excessive » ; (Breyton et al, 2018).

Les lésions auditives provoquées par un TSA peuvent concerner plusieurs structures de la cochlée :

- Les cellules ciliées externes (CCE), entre 10000 et 12000, qui jouent le rôle d'amplificateur cochléaire ;
- Les cellules ciliées internes (CCI), environ 3500, qui transforment le signal mécanique en un signal électrique, réalisant ainsi la mécano-transduction ;
- Les synapses entre les cellules ciliées internes et le nerf auditif, premières étapes de la transmission du signal électrique codant la stimulation sonore vers les structures auditives centrales.

De même, différents mécanismes physiopathologiques peuvent être en jeu dans le TSA, tels que l'apparition d'un hydrops endolymphatique¹, une réaction de stress oxydatif ou une réaction inflammatoire. Schématiquement, les lésions peuvent être induites soit par le traumatisme mécanique du fait de la stimulation acoustique, soit par les modifications métaboliques (excitotoxicité, stress oxydatif, réaction inflammatoire). Ces situations peuvent provoquer des déficiences de l'audition périphérique et contribuer à réduire les informations auditives ou induire des déficits centraux qui peuvent concourir à compromettre le traitement de l'information auditive. Généralement, ces déficiences peuvent réduire la conscience de la situation en dégradant le seuil de détection des sons, la précision de la localisation et l'intelligibilité de la parole, menaçant la bonne réussite de la mission et parfois la survie des personnes (Tepe et al., 2017).

Généralement, la cause la plus commune de la perte auditive liée au bruit est l'altération ou la destruction des CCE (Medina-Garin et al., 2016). Ces dommages entraînent classiquement une modification des seuils audiométriques.

Ce qu'il faut retenir

- Les militaires sont à risque de lésions auditives concernant différentes structures du système auditif, en particulier la cochlée.

III.3.2. Les atteintes des CCE et leur évaluation

III.3.3. L'audiométrie tonale liminaire

L'audiométrie demeure actuellement le « gold standard » dans l'évaluation du système auditif. La personne testée via l'audiométrie est invitée à détecter des bips d'intensité variée et à les signaler chaque fois qu'elle les perçoit. L'audiométrie tonale utilise une région du spectre de fréquences sonores entendues par un être humain (théoriquement de 20 à 20000 Hz, en pratique les fréquences testées vont généralement de 250 à 8000 Hz) et enregistre les pertes éventuelles. Cliniquement, on considère une déficience comme légère lorsqu'une perte est comprise entre 20 et 40 dB et une déficience sévère lorsque la perte est ≥ 70 dB (Figure 10).

¹ Phénomène physiopathologique également présent dans la maladie de Menière et caractérisée par un élargissement de l'espace endolymphatique

	125	250	500	750	1K	1.5K	2K	3K	4K	6K	8K
CA GAUCHE	5	10	15		5		20	10	0	15	0
CA DROITE	10	15	10		10		0	5	15	5	5

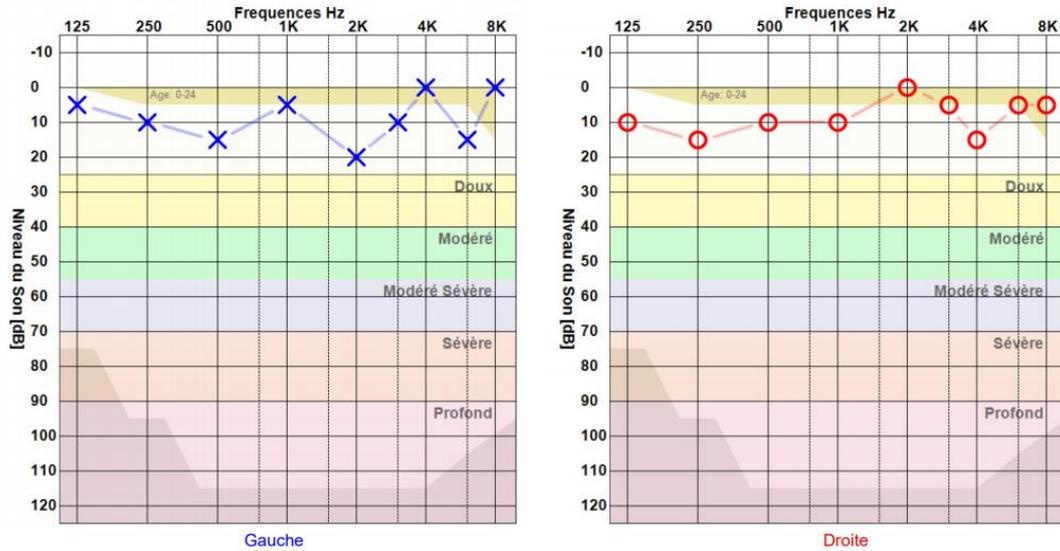


Figure 10 : Exemple de tracé audiométrique avec oreille gauche (en bleu) et oreille droite (en rouge). Capture d’écran du logiciel Audio Console Oscilla®.

Malgré son statut de « gold standard », l’audiométrie tonale liminaire souffre de plusieurs limites dont celle d’être un test comportemental sensible tant à l’expérience de l’opérateur qu’à d’éventuels biais cognitifs de l’auditeur (p. ex., état de vigilance, peur de la sanction en termes d’aptitude au travail, mise en place de stratégies destinées à améliorer artificiellement les résultats). De plus, ce type de test comportemental ne permet pas de connaître quelles sont les structures de l’oreille interne lésées lors du TSA, ni quels sont les mécanismes physiopathologiques en cours lors de la prise en charge (p. ex., hydrops endolymphatique).

Les otoémissions permettent de s’affranchir de certaines de ces limites tout en étant sensibles également à l’état des CCE.

III.3.4. Otoémission

Les otoémissions acoustiques (OEA) évaluent la fonction de l’amplification cochléaire sous-tendue par les CCE. Les OEA sont des ondes sonores produites par la cochlée et constituent un

indicateur de l'intégrité et de l'électromotilité des CCE. Les OEA peuvent être enregistrées à l'aide d'un microphone qui est placé dans le CAE. Une altération des CCE provoquera une absence ou une réduction de l'amplitude des OEA. Différents types d'OEA existent selon la présence ou non d'une stimulation externe (c.-à-d., otoémissions acoustiques spontanées : OEAS ; otoémissions acoustiques provoquées : OEAP). Une sous-catégorie est représentée par le produit de distorsions acoustiques (PDA). Une sonde composée par un microphone et de deux haut-parleurs est placée dans le conduit externe de l'oreille. Les haut-parleurs envoient deux sons de fréquences proches l'une de l'autre et de niveaux égaux ou légèrement asymétriques. Ce procédé conduit à stimuler les CCE au niveau d'une zone très étroite de la cochlée, et donc d'y détecter d'éventuelles altérations au point que certains auteurs ont parlé d'audiométries *objectives* (Mosnier et al., 2014).

Dans une population de normo-entendants, la variabilité interindividuelle dans une tâche de compréhension de la parole dans le bruit pourrait être expliquée en partie par l'état de leurs CCE. Ainsi une corrélation est retrouvée entre la mesure des otoémissions et la compréhension de la parole dans le bruit (Parker, 2020) ; probablement du fait d'un élargissement des filtres auditifs (Badri et al., 2011) à cause des CCE qui ne parviennent plus à amplifier sélectivement les fréquences composant le signal de parole.

Cette variabilité chez le normo-entendant illustre le fait que la normalité des seuils auditifs peut être retrouvée chez certains auditeurs présentant des déficits de perception auditive supraliminaire tels que la localisation auditive ou la compréhension de la parole dans le bruit. Ainsi, il est retrouvé dans la littérature le vocable de « pertes auditives cachées » ou « hidden hearing loss ».

Les pertes auditives cachées ou synaptopathies cochléaires et leur évaluation

Historiquement, les pertes auditives cachées désignaient des atteintes supposées centrales ou cérébrales de l'audition chez des auditeurs se plaignant de difficulté auditive (voir par exemple le syndrome de King-Kopetzky : Pryce et al., 2010) nommée également dysfonction auditive obscure (Stephens & Zhao, 2000). Aujourd'hui ce vocable tend à désigner les synaptopathies cochléaires.

En effet, des études animales montrent que l'exposition à des bruits continus peut altérer les synapses présentes entre la cochlée et le nerf auditif. Ces *synaptopathies cochléaires* conduisent

à une perte d'efficacité du traitement auditif, et ce même en l'absence de lésion des cellules ciliées. Des études indiquent que jusqu'à la moitié des synapses pourraient être perdues après une exposition à certains niveaux de bruit, sans aucun dommage permanent des cellules ciliées et/ou sans élévation permanente des seuils auditifs (Jenson et al., 2015 ; Kujawa & Liberman, 2009). Les synapses les plus atteintes seraient celles codant pour les sons d'intensité modérée à élevée (Furman et al., 2013). Ces résultats d'études animales pourraient faire la lumière sur certains problèmes d'audition chez l'homme qui ne sont toujours pas éclaircis à ce jour. Certaines personnes déclarent avoir des problèmes pour *comprendre* ce qui se dit dans des situations bruyantes ou lorsque de nombreuses personnes parlent, sans avoir de problèmes à les *entendre*. Les déclarations de ces personnes sont du type : « *J'entends, mais je ne comprends pas* » (Moser & Starr, 2016). Ces difficultés ne sont cependant pas corroborées par des déficits ou des altérations détectables en audiométrie tonale du fait de la préservation des CCE, CCI et des synapses codant pour les sons faibles. Chez l'homme, la difficulté vient évidemment du fait qu'il n'existe actuellement aucune technique permettant de détecter la perte de synapses de manière aussi détaillée et *in vivo*.

Les études corroborant l'hypothèse de l'existence des synaptopathies chez l'homme ont principalement été produites par le même laboratoire ayant publié l'étude princeps chez les souris (Kujawa & Liberman, 2009). Une des études les plus citées portait sur un effectif de 34 auditeurs répartis selon un questionnaire en un groupe d'individus dits « à haut risque » : 22 personnes qui étaient principalement de jeunes amateurs de musique, et un groupe « à bas risque » : 12 personnes étudiant en sciences (audiologie). Les résultats ont montré des différences significatives tant au niveau de tâches comportementales de compréhension de mots dans le bruit, qu'en électrocochléographie (voir encadré ci-dessous) et qu'au niveau des seuils audiométriques pour les très hautes fréquences (> 8 kHz) ; (Liberman et al., 2016).

Electrocochléographie

L'électrocochléographie consiste à recueillir l'activité électrophysiologique produite lors de stimulations acoustiques au niveau de la cochlée, du nerf auditif, et des premiers relais du tronc cérébral.

Elle permet de mesurer l'amplitude de différentes ondes correspondant à différents niveaux du tractus auditif ascendant depuis la synapse entre les cellules ciliées internes et le nerf auditif (onde I), jusqu'au colliculus inférieur (onde V).

La stimulation acoustique est délivrée via des écouteurs insérés au sein du conduit auditif externe à l'aide d'un embout en mousse recouvert d'une feuille d'or pour assurer une conductivité optimale, ces écouteurs servant d'électrode de recueil (électrocochléographie extratympanique non invasive). L'activité électrique est recueillie par deux électrodes placées au niveau du front.

Une autre étude menée à plus grande échelle sur 138 jeunes auditeurs (<36 ans) n'a toutefois pas permis de mettre en évidence de différence significative entre trois groupes d'individus déterminés selon la dose de bruit reçue au cours de leur vie. Les mesures employées étaient variées avec une dizaine de tâches comportementales incluant des mesures de seuils psychophysiques, des scores de compréhension de la parole dans le bruit, des questionnaires d'audition subjective. Ces mesures comportementales étaient associées à des mesures électrophysiologiques comme l'électrocochléographie (Prendergast et al., 2016). La dose de bruit était évaluée par un questionnaire attaché à rendre compte de l'exposition professionnelle et non professionnelle. Les auteurs indiquent un facteur 17 entre le groupe le plus exposé et celui le moins exposé. Il existe aussi un seuil auditif à très haute fréquence (16 kHz) plus élevé dans le groupe le plus exposé. Seuls les résultats en termes d'audiométrie hautes fréquences sont concordants entre ces deux études.

Ces divergences illustrent la difficulté à transposer à l'humain les résultats obtenus chez l'animal. Ainsi, Dobie et Humes (2017) soulignent que les résultats obtenus chez l'animal doivent être considérés avec précaution étant donné les différences entre espèces en termes de susceptibilité au bruit. Ils alertent également sur le fait que contrairement aux rongeurs, l'agression sonore chez l'humain pourrait provoquer en premier lieu une atteinte des CCE et donc une élévation du seuil auditif avant de provoquer une synaptopathie (Dobie & Humes, 2017).

Une autre divergence avec les études animales concerne l'évaluation de la dose reçue. Cette donnée est cruciale, or chez l'homme, dans les études précédemment citées, elle est évaluée

uniquement par un questionnaire. Des questionnaires sont aussi utilisés pour évaluer l'histoire de l'auditeur avec le bruit. Trois mesures de l'exposition au bruit par questionnaire ont été utilisées dans les études concernant la recherche des synaptopathies cochléaires chez l'être humain : le *Noise Exposure Questionnaire* (NEQ – Johnson et al., 2017 ; Bernard et al., 2019), le *Lifetime Exposure of Noise and Solvents Questionnaires* (LENS-Q ; Bramhall et al., 2017 ; Griest et al., 2015 ; Gordon et al., 2017) and the *Noise Exposure Structured Interview* (NESI ; Guest et al., 2018). Bien que cela soit évalué de façon différente, le principe à la base de ces questionnaires vise à quantifier l'exposition annuelle au bruit de l'auditeur dans le cadre de son travail et de ses activités de loisir, néanmoins leur variabilité illustre la difficulté à atteindre cet objectif.

Au final, le concept de synaptopathie reste particulièrement débattu chez l'humain (Bramhall et al., 2019). Ainsi, bien qu'il aurait été souhaitable de quantifier le niveau de synaptopathie cochléaire chez nos auditeurs pour prendre en compte son effet sur l'effort d'écoute, la mesure semble difficile en pratique. La mesure la plus directe aurait été l'électrocochléographie cependant chez l'homme la grande variabilité interindividuelle dans l'amplitude de l'onde II rend cette mesure difficilement exploitable (Plack et al., 2016). Néanmoins, quantifier l'état auditif par l'audiométrie tonale uniquement n'est pas acceptable. Notre stratégie a alors consisté à évaluer l'état auditif par un ensemble de mesures :

- L'audiométrie tonale liminaire du fait de statut de gold standard
- L'audiométrie vocale dans le bruit du fait de sa capacité à rendre compte de déficit non directement explicable par l'audiométrie tonale (Phatak et al., 2019).
- L'audiométrie très haute fréquence du fait des résultats cohérents retrouvés à travers les études indiquant qu'elle puisse être un biomarqueur global de l'état cochléaire (Smith et al., 2019).
- Les otoémissions du fait de leur caractère objectif et de leur relation avec les capacités supraliminaires (Parker, 2020).

Ce qu'il faut retenir

- Présence potentielle, mais difficilement objectivable des synaptopathies cochléaires chez le personnel militaire exposé aux agressions sonores provoquées par le bruit.
- Nécessité d'étendre l'évaluation de l'état auditif à d'autres mesures que l'audiométrie tonale liminaire : audiométrie vocale dans le bruit, audiométrie très haute fréquence, otoémissions.

IV. Mesures pour évaluer l'effort d'écoute

Différentes méthodes existent d'ores et déjà pour tenter de mesurer cet effort comme les évaluations subjectives, les mesures comportementales (p. ex., tâche d'écoute dichotique ou tâches d'intelligibilité du langage), les mesures psychophysiologiques (activité cardiaque, dilatation pupillaire) et neurophysiologiques (activité cérébrale) ; (voir Peelle, 2018, pour une synthèse).

IV.1. Évaluations subjectives

Qu'elle soit en lien avec le langage ou non, la complexité d'une tâche peut générer des sentiments subjectifs d'effort et de fatigue qui doivent être pris en compte dans les différentes évaluations. À cet égard, c'est donc l'auditeur lui-même qui estime le niveau d'effort ressenti pour effectuer la tâche à l'aide d'échelles ou de questionnaires (Johnson et al., 2015 ; van Esch et al., 2013).

De manière plus spécifique, l'Effort Scaling Categorical Unit (ESCU) est une échelle dédiée à l'évaluation subjective dans les tâches auditives (Rennies et al., 2019 ; Rennies & Kidd, 2018). Cette échelle a été traduite en français et elle est composée de 13 niveaux d'évaluations de l'effort d'écoute à travers sept catégories nommées : *pas d'effort* (1 ESCU), *très peu d'effort* (3 ESCU), *peu d'effort* (5 ESCU), *effort modéré* (7 ESCU), *effort considérable* (9 ESCU), *beaucoup d'effort* (11 ESCU) et *effort extrême* (13 ESCU) et *que du bruit* (14 ESCU). À ces catégories labellisées, six catégories intermédiaires entre chacune s'ajoutaient.

Ils existent aussi des questionnaires visant à observer les difficultés au quotidien chez l'auditeur. Un exemple est le Speech, Spatial and Qualities Hearing (SSQ) ; (Gatehouse & Noble, 2004) qui comporte des questions pour comprendre les difficultés au quotidien du participant dans des situations bruyantes et multilocuteurs. La troisième étude décrite dans ce manuscrit de thèse contient une version française adaptée de l'échelle d'effort ESCU (Rennies et al., 2019 ; Rennies & Kidd, 2018) et l'échelle de motivation (Kramer et al., 2016). Une version française courte du questionnaire SSQ a aussi été utilisée (Moulin et al., 2015).

Il convient toutefois de souligner qu'outre l'utilité des mesures subjectives de l'effort d'écoute, celles-ci présentent des inconvénients. Bien que ces mesures offrent la possibilité de connaître quel est le vécu de l'individu, elles demeurent des mesures subjectives et donc liées à l'interprétation que chaque auditeur fait de son ressenti. Des différences individuelles peuvent exister dans l'évaluation personnelle des seuils d'effort. Pour surmonter ces difficultés d'interprétation, des mesures comportementales sont utilisées et/ou associées. Dans le paragraphe suivant, nous allons illustrer les mesures comportementales les plus couramment utilisées.

IV.2. Mesures comportementales

Les mesures comportementales associées à l'étude de l'effort d'écoute permettent de compléter l'approche subjective en ce sens qu'elles visent à évaluer la performance factuelle (et non ressentie) d'un auditeur dans des situations d'écoute complexe (Gagné et al., 2017).

Généralement, les protocoles comportementaux choisis proposent des tâches d'intelligibilité dans lesquelles l'auditeur est invité à essayer de comprendre des phrases présentées à différents niveaux d'intelligibilités (plus ou moins dégradées) ou encore à prêter attention à la cible d'intérêt présentée parmi des distracteurs. Les réponses correctes ou les temps de réaction/réponse sont les variables dépendantes permettant d'évaluer l'effort à travers la performance.

Une des limites de ces protocoles comportementaux est de ne représenter que la sortie de la chaîne perceptivo-motrice, ce qui rend impossible l'évaluation des étapes successives du traitement auditif, ni de savoir quelle est la composante de l'effort d'écoute qui a été mise en jeu. Par exemple, des performances élevées peuvent refléter à la fois une facilité à réaliser la tâche, ou un engagement important, mais efficace de ressources cognitives.

Les protocoles en double tâche sont utilisés afin d'apporter une granulosité plus fine dans la compréhension des facteurs limitants l'intelligibilité (Gagné et al., 2017). Ces protocoles présupposent qu'un individu dispose d'une capacité limitée en ressources cognitives. Les protocoles de double tâche prévoient la réalisation de deux tâches : une tâche principale et une tâche secondaire (Xia et al., 2015). Le participant fait tout d'abord la première tâche, après la deuxième et ensuite les deux ensembles. L'effort d'écoute est alors pensé comme la différence entre la performance à la deuxième tâche (qui devient la condition de base) et la performance à la première tâche en présence de la deuxième (p. ex., réduction de bonnes réponses et/ou temps de réponse plus longs). Si la tâche principale a une demande importante en ressources cognitives alors elle laissera peu de ressources disponibles pour la réalisation de la tâche secondaire. La demande cognitive sera calculée comme la différence de performance de la tâche secondaire en concomitance ou non avec la tâche principale. Le caractère limité des ressources cognitives conduirait à les consacrer prioritairement à la réussite de la tâche principale.

Pour une description détaillée sur les protocoles de double tâche, nous conseillons la lecture de la revue de Gagné et al., 2017. Dans cette revue, les auteurs décrivent, entre autres, la grande variabilité d'approches méthodologiques appliquées avec les tâches secondaires et fournissent un résumé général des résultats obtenus en relation avec certaines variables telles que l'âge de l'auditeur, l'état de l'audition et la modalité de présentation du signal sonore (audio ou audiovisuel).

IV.3. Mesures psychophysiologiques

Comme nous l'avons abordé dans les paragraphes précédents, les situations d'écoute complexes nécessitent un déploiement de ressources supplémentaires par rapport à celles qui seraient suffisantes en conditions d'écoute optimale. Différentes techniques psychophysiologiques sont aujourd'hui à disposition des chercheurs. Elles se concentrent sur l'investigation des réponses du système nerveux autonome lors des situations d'écoute complexe. Parmi celles-ci, on retrouve : la dilatation pupillaire (Kramer et al., 2016 ; Zekveld et al., 2014), la réponse électrodermale et l'activité cardiaque (Mackersie & Calderon-Moultrie, 2016).

Généralement, les études ont montré qu'avec l'augmentation de l'effort mental, le diamètre de la pupille s'élargit (Verney et al., 2004), la conductance cutanée augmente et la variabilité de la fréquence cardiaque se réduit (Mackersie & Calderon-Moultrie, 2016). Ces patterns de

réponse du système nerveux autonome sont fréquemment liés à des conditions de stress ou de mal-être (Mackersie & Calderon-Moultrie, 2016). À cet égard, il existe également des méthodes pour quantifier la présence des hormones du stress (p. ex., biomarqueurs physiologiques sensibles au stress : niveaux de cortisol et d'alpha-amylase depuis la salive ; Zekveld et al., 2019). Généralement ces mesures physiologiques sont associées à des mesures comportementales.

IV.3.1. Neurophysiologique

Les méthodes neurophysiologiques se concentrent sur l'investigation des effets de l'effort d'écoute sur le système nerveux central. Parmi ces méthodes on retrouve : l'imagerie par résonance magnétique fonctionnelle (IRMf) ; (Rodd et al., 2010 ; Wild et al., 2012), l'électroencéphalographie/EEG (Blessner et al., 2017), la TEP (Salvi et al., 2002) et plus récemment, la spectroscopie en proche infrarouge (f/NIRS) ; (Andéol et al., 2017 ; Lawrence et al., 2018). Les études réalisées au moyen de ces techniques se sont concentrées sur l'observation du réseau fronto-temporo-pariétal que comme évoqué dans le paragraphe *Chapitre II. Neuro-anatomie fonctionnelle du traitement du* est impliqué dans le traitement du langage en situation d'écoute complexe. Dans le paragraphe suivant, un focus particulier sera fait sur une de ces méthodologies d'observation neurophysiologique : la spectroscopie en proche infrarouge, qui a été utilisée dans le cadre de cette thèse.

IV.3.2. La spectroscopie en proche infrarouge (f/NIRS)

La spectroscopie proche infrarouge (« Near-InfraRed Spectroscopy » ou NIRS) est une technique non invasive d'imagerie optique fonctionnelle. Elle mesure le métabolisme cérébral concomitant au déroulement d'une tâche à travers la quantification des changements relatifs en concentration de l'oxy-hémoglobine (oxyHb) et déoxy-hémoglobine (déoxyHb) qui sont deux chromophores de l'hémoglobine. À la différence d'autres tissus cérébraux, l'oxyHb et la déoxyHb ont la propriété d'absorber la lumière dans le spectre proche infrarouge. La fNIRS utilise cette propriété pour inférer l'activité corticale locale. L'idée que les tissus cérébraux peuvent être transparents à la lumière remonte en 1831, quand le chercheur Richard Bright en examinant, à l'aide d'une chandelle, le crâne d'un patient souffrant d'hydrocéphalie s'aperçoit

qu'il était semi-transparent. La première fois que l'activité corticale est mesurée à travers ce principe d'absorption de la lumière traversant le crâne est en 1977 (Jöbsis, 1977) et c'est à partir des années 1980 que les premières études cliniques utilisant la spectroscopie en proche infrarouge sont publiées (Ferrari & Quaresima, 2012 ; Meek et al., 1998).

Le principe de base est le même que celui de l'IRM fonctionnelle : le couplage neuro-vasculaire. Ce mécanisme physiologique correspond au fait que l'activation neuronale, produite, par exemple, lors de la réalisation d'une tâche cognitive, engendre une demande métabolique dont la réponse est une augmentation du flux sanguin cérébral local (ou *Cerebral Blood Flow*, CBF en anglais). Ce mécanisme de couplage neuro-vasculaire est assuré par la présence d'une *unité neurovasculaire* composée du muscle lisse vasculaire, du neurone et de la cellule gliale astrocyte. Lors de l'activation neuronale, le glutamate est libéré, ce qui amène les neurones et les astrocytes à transmettre des signaux afin de réguler le débit sanguin cérébral. La combinaison de différents effets mécaniques et chimiques règle les mécanismes de vasodilatation et vasoconstriction afin de permettre l'oxygénation corticale rapide. La mise en place du mécanisme du couplage neuro-vasculaire est bien expliquée dans l'article d'Hendrikx et al., 2019 que nous citons ici. Les signaux chimiques connus comprennent les prostaglandines, l'oxyde nitrique et l'adénosine, sécrétés à la fois par les neurones et les astrocytes, qui provoquent une vasodilatation de la cellule musculaire lisse. En outre, les astrocytes sécrètent également d'autres vasodilatateurs comme le potassium et les acides époxyeicosatriénoïques. Outre ces vasodilatateurs, les astrocytes sont également connus pour sécréter de l'acide arachidonique, qui a un effet vasoconstricteur. Tous ces médiateurs ont un impact direct sur le muscle lisse des artérioles et contrôlent donc le flux sanguin cérébral de manière directe. En plus de ces mécanismes présents dans l'unité neurovasculaire, un deuxième mécanisme a été récemment décrit par lequel les péricytes autour des capillaires provoquent également une vasodilatation, jouant probablement aussi un rôle dans la distribution locale du flux (Hendrikx et al., 2019 ; Huneau et al., 2015 ; Phillips et al., 2016 ; Figure 11).

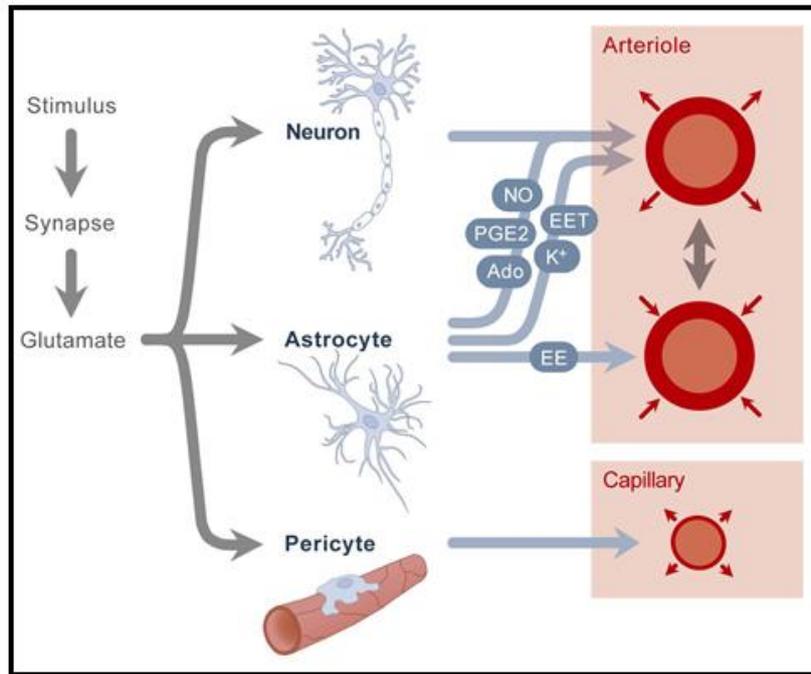


Figure 11 : Schéma illustratif des structures et des mécanismes impliqués dans le couplage neuro-vasculaire. En détail, le traitement d'un stimulus provoque la sécrétion de glutamate. Le glutamate stimule les neurones et les astrocytes, ce qui entraîne la sécrétion d'oxyde nitrique (NO), de potassium (K⁺), d'adénosine (Ado), d'acides époxyeicosatriénoïques (EET) et de prostaglandines (PGE₂), ce qui provoque une vasodilatation artériolaire. Les astrocytes sécrètent également de l'acide arachidonique (AA), qui provoque une vasoconstriction. Un deuxième mécanisme est la stimulation des péricytes, qui entraîne une vasodilatation capillaire (depuis Hendrikx et al., 2019).

Lors de l'activité neuronale, il y a une augmentation concomitante de l'oxyHb et une dilution relative de la déoxyHb qui donne lieu à la réponse hémodynamique (Figure 12).

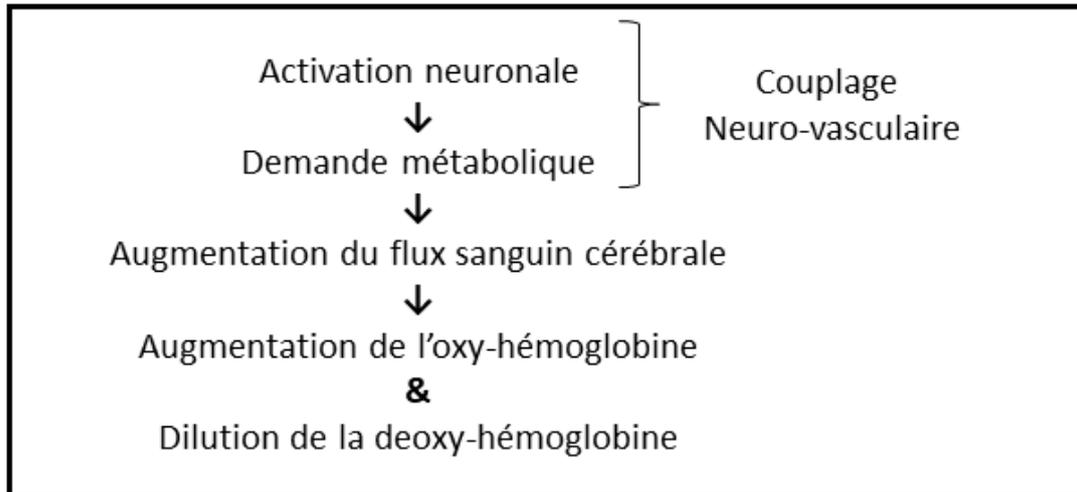


Figure 12 : Schéma simplifié de la réponse hémodynamique en fNIRS. Le couplage neuro-vasculaire entraîne une augmentation du flux sanguin cérébrale avec une concomitante augmentation de l'oxy-hémoglobine et une dilution de la déoxy-hémoglobine.

L'enregistrement se fait à l'aide d'un certain nombre de capteurs appelés optodes (nommées en analogie aux électrodes de l'EEG, mais pour la lumière). Ces optodes sont placées sur le cuir chevelu et elles se distinguent entre sources et détecteurs. Source et détecteur sont toujours couplés et forment un canal de mesure. Les sources émettent de la lumière de faible puissance dans le proche infrarouge, au niveau du crâne, et en correspondance de l'aire cérébrale choisie pour l'observation. La lumière suit une trajectoire typique dite en « forme de banane » (en 2D) ou de chapeau de champignon (en 3D) et peut rejoindre le cortex cérébral à différents niveaux de pénétration selon l'écartement entre la source et le détecteur. Une distance source - détecteurs petite (p. ex., < 1cm) permet d'accéder à la mesure de l'activité systémique de la couche la plus superficielle du cortex cérébral (Ekkekakis, 2009; Hoshi, 2003) tandis qu'une distance plus grande (p. ex., > 3cm) permet d'aller plus en profondeur, mais au détriment d'une plus faible résolution spatiale (Figure 13). En pratique, un compromis, profondeur/résolution spatial a été empiriquement établi avec une distance inter-optodes de l'ordre de 3 cm. Cette distance permet d'atteindre la majeure partie des différentes couches corticales avec une résolution grossière de l'ordre du cm³. Différentes méthodes existent pour envoyer la lumière proche infrarouge à travers le scalp. La méthode NIRS, la plus courante, correspond à l'utilisation d'une onde continue (de l'anglais *continuous wave*, CW). Cette méthode exige que la lumière soit envoyée en continu et à une intensité constante. Cela permet de mesurer l'atténuation globale d'intensité de la lumière qui traverse la tête. Une deuxième méthode est la

NIRS résolue en temps, dans laquelle la lumière n'est pas continue, mais pulsée de manière brève. Cela permet de mesurer le profil temporel de la lumière transmise. Dans une troisième méthode, la NIRS résolue en fréquence, la lumière reste constante et continue, mais l'intensité est modulée. Cela permet de quantifier l'atténuation globale, l'atténuation de l'indice de modulation et le décalage de phase (Ferrari & Quaresima, 2012 ; Mandrick, 2013).

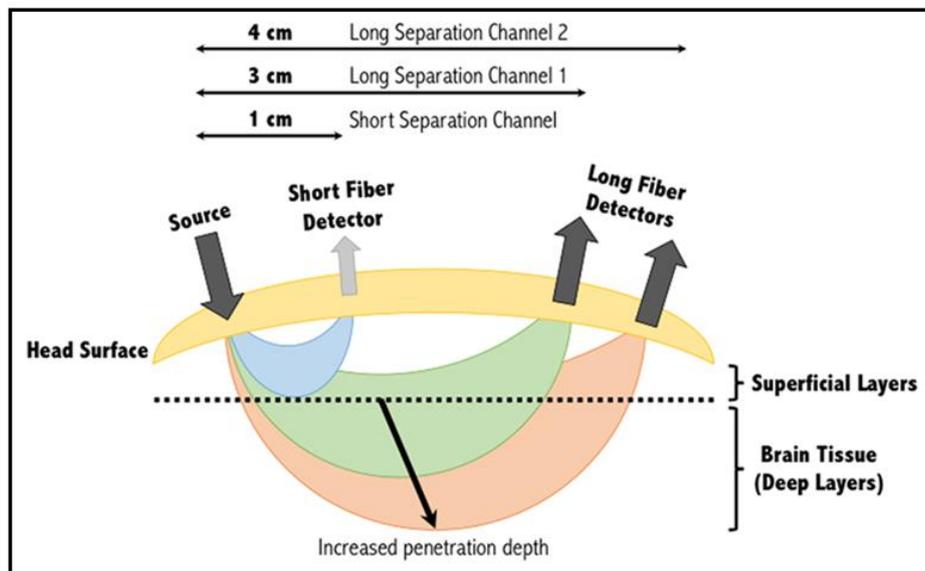


Figure 13 : Effet de la distance d'écartement entre les optodes sur la résolution spatiale et la profondeur de la mesure en fNIRS.

Lors d'une activation neuronale, le débit et le volume sanguin augmentent pour apporter aux neurones un supplément d'oxygène et de glucose. Cette activation engendre une modification de la concentration d'hémoglobine liée à une augmentation de la circulation sanguine locale (Ogawa et al., 1992). La réponse hémodynamique est représentée par une augmentation de [oxyHb] avec une diminution relative de [déoxyHb] ; (Huppert et al., 2006 ; Obrig et al., 2000 ; Perrey, 2008 ; Schroeter et al., 2006). La composition des tissus biologiques traversés conduit à différentes perturbations de la lumière au fil de son parcours. Ainsi, les photons subissent des phénomènes d'absorption, de dispersion, de réflexion et de transmission provoqués par les différences structurelles et morphologiques des tissus biologiques (Maikala, 2010 ; Villringer & Chance, 1997). Toutefois, une fois considéré que ces effets sont relativement stables pour un même individu au cours d'une même session de mesure, c'est la variation de concentration de l'hémoglobine (ses deux états) qui va induire la majorité des modifications d'absorption de la lumière émise. Les photorécepteurs détectent alors la quantité de lumière restante qui

correspond au total émis moins la quantité absorbée par l'hémoglobine. Une dernière étape consiste alors à utiliser les propriétés différentielles d'absorption de l'oxyHb et de la déoxyHb pour en déduire leur concentration respective. En effet les deux états de l'hémoglobine n'absorbent pas l'infrarouge de la même façon (Figure 14 et Figure 15).

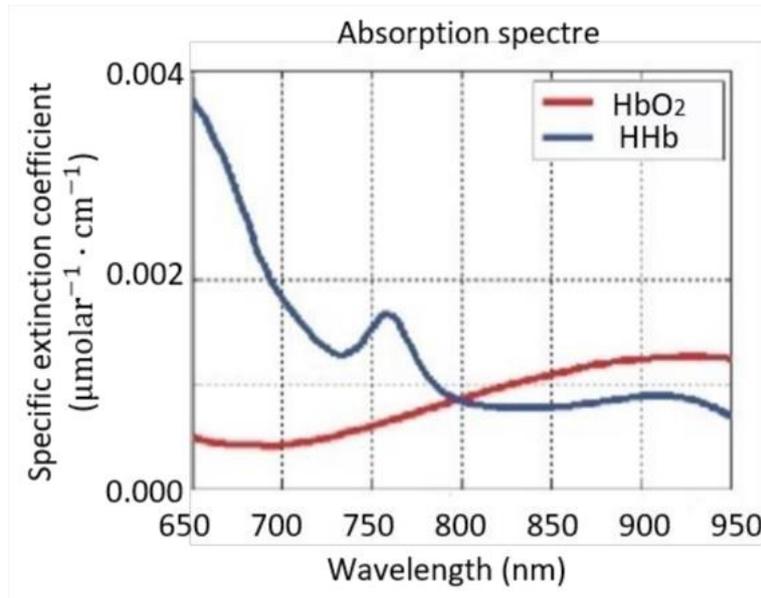


Figure 14 : Spectres d'absorption (coefficient d'extinction molaire) de l'hémoglobine oxygénée (HbO en rouge) et déoxygénée (HbR en bleue) entre 760 nm et 850 nm (depuis Roldán et al., 2021).

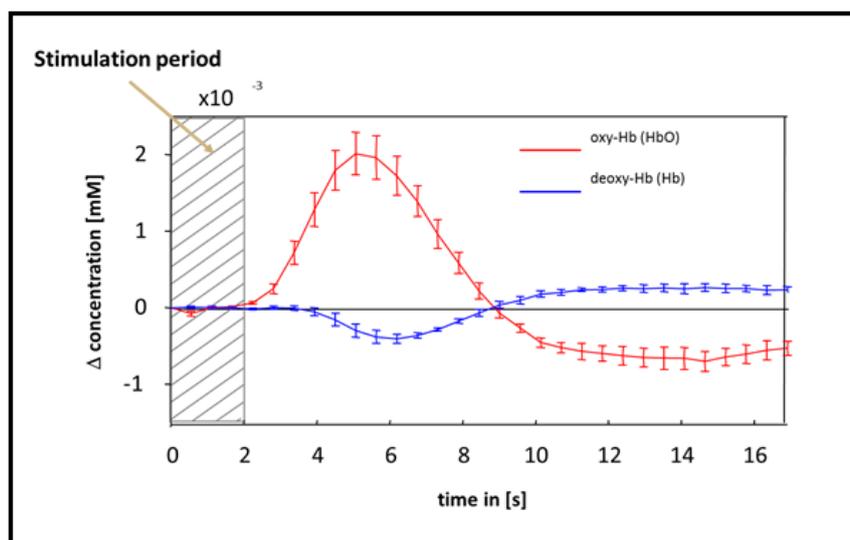


Figure 15 : Réponse hémodynamique avec une augmentation de l'oxyhémoglobine HbO₂ avec une baisse concomitante de la déoxyhémoglobine [HHb] après une période de stimulation.

Ainsi il est possible de résoudre un système d'équations à 2 inconnues ([HBO] et [HBR]). L'équation qui permet de calculer ces concentrations provient de la loi de Beer-Lambert représentée par l'équation ci-dessous (Swinehart, 1962) :

$$A = \log(I/I_0)$$

$$A = c \times \varepsilon_\lambda \times l$$

Où A représente l'absorption, I l'intensité de lumière initiale, I₀ l'intensité mesurée, c la concentration du chromophore mesurée, ε_λ le coefficient d'absorption de ce chromophore et l la distance parcourue.

Cette équation a été modifiée pour tenir compte de l'inhomogénéité des tissus biologiques (Kocsis et al., 2006) :

$$A = (c \times \varepsilon_\lambda \times l \times DPF) + G$$

Dans cette équation, le facteur de trajet différentiel (DPF) a été introduit pour tenir compte du trajet non linéaire dans les tissus biologiques et un facteur de dispersion G a été ajouté.

À travers la NIRS, il est possible de déduire les aires cérébrales les plus sollicitées pendant une tâche. L'intérêt de la communauté scientifique envers l'utilisation de la NIRS est croissant (Pinti et al., 2020). Elle est utilisée dans plusieurs domaines, depuis les neurosciences et la psychologie cognitive jusqu'aux sciences du sport (Albinet et al., 2014 ; Ferreri et al., 2014). L'utilisation de la NIRS est fréquente dans l'exploration du concept d'effort et de charge de travail. Elle peut être utilisée avec différentes populations y compris les populations saines (dès les nouveau-nés, Peng & Hou, 2021) et les populations cliniques souffrant de maladies neurologiques et psychiatriques (Irani et al., 2007) ou souffrant de troubles auditifs, y compris chez les porteurs d'implants cochléaires (Pinti et al., 2019).

Elle constitue une mesure non invasive de l'activité corticale, aisément transportable et facile à installer et utiliser. La simple émission de lumière qui permet de révéler les changements métaboliques lors d'une tâche fait de la NIRS un outil de mesure silencieux, sûr et sans risque pour le participant. Premièrement, en comparaison à l'IRMf, la fNIRS n'émet aucun son

susceptible d'interférer avec les tâches auditives. Deuxièmement, alors que l'IRMf utilise les propriétés magnétiques de l'hémoglobine pour évaluer sa présence, la fNIRS utilise les propriétés optiques de celle-ci et notamment son absorption de lumière dans le proche infrarouge (longueurs d'onde comprises entre 780 et 1400 nm). Toujours en comparaison avec l'IRMf, la NIRS fournit des mesures complètes du métabolisme cérébral avec les changements de concentrations de l'oxyHb et de la déoxyHb et la somme de ceux-ci qui correspond au volume sanguin régional cérébral (rCBV) tandis que l'IRM donne exclusivement une information sur la concentration de déoxyHb.

En outre, les coûts d'acquisition et d'opération sont beaucoup moins élevés que ceux de l'IRMf ou de la tomographie par émission de positrons (TEP) et proche de ceux d'un EEG.

Enfin, le rapport signal sur bruit est satisfaisant ce qui fait de cet outil une technique d'imagerie cérébrale optimale pour des enregistrements pendant le déroulement d'une tâche qui requiert du mouvement (c.-à-d. le sport, jouer de la musique). De plus, la NIRS offre aussi la possibilité de co-enregistrer l'activité cérébrale avec l'EEG ou l'IRMf, ce qui permet d'associer plusieurs techniques d'imagerie et d'obtenir des données plus complètes. La limitation majeure à l'utilisation de la NIRS concerne l'impossibilité d'étudier des structures plus profondes du cerveau étant donné que seules les couches superficielles du cerveau peuvent être étudiées.

V. Pourquoi étudier l'effort d'écoute ?

L'étude de l'effort d'écoute dans des situations d'écoute complexes est d'actualité et commune à plusieurs disciplines. Par exemple, une étude récente a étudié l'impact du port de masque facial dans la compréhension du langage en situation d'écoute complexe (Rahne et al., 2021). Les résultats montrent comment l'atténuation acoustique de la pression sonore par le tissu du masque réduit la capacité de perception de la parole, en particulier dans les situations bruyantes, augmentant l'effort d'écoute. Les masques peuvent également entraver le processus de compréhension de la parole en dissimulant les mouvements de la bouche, ce qui interfère avec la lecture labiale. L'effet acoustique des masques faciaux combiné avec l'impossibilité de lecture labiale pourrait avoir un impact important sur la communication dans la vie quotidienne, y compris chez les personnes ayant une audition normale.

L'étude de l'effort d'écoute et des répercussions associées est donc d'actualité. Un autre champ d'investigation est celui des personnes souffrant de problèmes d'audition. Ces dernières figurent parmi les personnes les plus affectées par les situations d'écoute complexe. De fait, les

problèmes d'audition rendent la communication plus difficile (Pichora-Fuller et al., 2016 ; Pichora-Fuller & Souza, 2003), contribuant à a) augmenter l'effort perçu, b) démotiver et pousser l'auditeur à renoncer à comprendre et c) à augmenter chez l'auditeur le risque d'exclusion sociale. Les problèmes qui semblent se limiter à la difficulté de compréhension dans une situation d'écoute complexe peuvent donc affecter d'autres aspects de la vie de l'auditeur, altérant le bien-être de la personne. À cet égard, l'OMS rapporte que les liens sociaux et l'engagement dans la communication représentent des facteurs clés pour une vieillesse en bonne santé (OMS, 2002). Plusieurs études ont récemment eu pour objet la recherche d'une association entre la perte auditive liée à l'âge et le déclin cognitif. Ainsi, la perte auditive liée à l'âge serait une cible potentielle pour les stratégies de prévention de la démence (Chern & Golub, 2019 ; Deal et al., 2017 ; Quaranta et al., 2014). La nature de cette association n'est pas encore totalement comprise, des mécanismes communs (par exemple, la maladie micro-vasculaire) et/ou des mécanismes causaux (p. ex., l'épuisement de la réserve cognitive et l'isolement social) pourraient expliquer cette relation (pour une description exhaustive, voir la revue de Chern & Golub, 2019).

Une autre catégorie de personnes qui souffrent des effets de l'effort d'écoute est celle des travailleurs en environnements bruyants (*open space*, écoles). Des études montrent que travailler dans des bureaux ouverts en présence d'autres collègues qui parlent peut entraîner des distractions, des difficultés de concentration, de la fatigue et des altérations de la mémoire de travail (Renz, 2020 ; Roelofsen, 2008). L'étude approfondie de ces situations peut permettre de concevoir un environnement de travail acoustique confortable tout en permettant à l'entreprise/organisme d'assurer le bien-être des travailleurs et d'améliorer la productivité (Rudner et al., 2018 ; Visentin et al., 2018).

Les militaires sont aussi des professionnels travaillant dans des environnements bruyants. Ces personnes, en raison de leur exposition aux bruits intenses, présentent des risques de développer des troubles de l'audition. C'est notamment le cas des situations de combats, des entraînements, où l'exposition au bruit est grande. L'étude des mécanismes qui sous-tendent dans cette population la compréhension d'un message dans des contextes difficiles pourrait permettre d'identifier des pistes pour les aider au quotidien.

D'un point de vue appliqué, l'étude de l'effort d'écoute permet de concevoir des outils de travail (p. ex., systèmes de radiocommunication, lieux de travail insonorisés...) mieux adaptés aux capacités d'un opérateur en prenant en compte l'individu lui-même.

Reconnaître que l'effort d'écoute peut être une composante importante et, dans certains cas, décisive de la vie des personnes, a stimulé les travaux scientifiques dans le domaine, en recherchant des stratégies et en étudiant des interventions visant à le minimiser.

Pour se faire, il est toutefois nécessaire de reconnaître la nature multidimensionnelle de l'effort d'écoute et d'examiner chacune de ses composantes.

Partie 2. Contributions expérimentales

Chapitre I. Première étude

La première étude présentée a donné lieu à un article scientifique, en langue anglaise, qui a été inclus en Annexe 1.

Dans ce chapitre, un résumé de l'article incluant l'introduction, un rappel de la méthode, les résultats et la discussion générale sont fournis en français.

I. Introduction

Ce premier travail visait à étudier les mécanismes neuronaux impliqués dans les situations multilocuteurs, où l'auditeur cherche à se focaliser sur une voix cible en présence d'une autre voix (voix masque) qui en perturbe la compréhension.

Dans une tâche d'intelligibilité en situation multilocuteurs, la plupart des erreurs consistent dans le report de la voix masque plutôt que de la voix cible (Brungart, 2001). Ce résultat montre à quel point l'inhibition de la voix masque est déterminante pour réussir cette tâche (Johnsrude & Rodd, 2015). La ségrégation des différentes voix, l'inhibition de la voix masque et la focalisation sur la voix cible, constituent donc les prérequis pour assurer la bonne compréhension de cette dernière.

Pour faciliter la ségrégation des différentes voix, un auditeur peut utiliser différents indices. Parmi ces indices figurent : les caractéristiques spectrales de la voix liées au genre du locuteur, la localisation des sources sonores dans l'espace ou la différence d'intensité sonore entre les locuteurs (Andéol et al., 2017 ; Brungart & Simpson, 2005 ; Cherry, 1953). Cette différence est connue sous le terme anglais de *Target-to-Masker ratio* ou TMR. Concernant ce dernier indice, une grande variabilité interindividuelle est retrouvée au niveau des réponses comportementales lorsque deux voix sont présentées simultanément. Ainsi, si pour certains auditeurs, la performance décroît de façon monotone avec le TMR (c.-à-d., performances moins bonnes en TMR défavorable), pour d'autres, la performance avec un TMR défavorable est paradoxalement plus élevée que celle obtenue avec un TMR intermédiaire (c.-à-d., quand l'intensité sonore des deux voix est proche) ; (Andéol et al., 2017 ; Brungart & Simpson, 2007 ; Cooke et al., 2008 ; Thompson et al., 2015). Utiliser la différence d'intensité entre les deux

voix pour se focaliser sur la voix cible serait donc une stratégie perceptive bénéfique, même dans la condition où la voix cible est moins forte que la voix masque. Toutefois, la variabilité interindividuelle observée montre que seulement certains auditeurs arrivent à la mettre en place. À cet égard, plusieurs études montrent que des changements de stratégies perceptives conduisant à des améliorations de performance dans les tâches d'intelligibilité de la parole en situation multilocuteurs peuvent être obtenus : par exemple, à l'aide d'un entraînement par instructions et/ou d'un feedback spécifique (Brungart & Simpson, 2004 ; Thompson et al., 2015). Au niveau neuronal, il apparaît que la variabilité interindividuelle en condition de TMR défavorable corrèle avec l'activité corticale du cortex préfrontal : les auditeurs avec la performance la plus élevée sont les auditeurs dont l'activité du cortex préfrontal est la plus importante (Andéol et al., 2017). L'activité corticale plus élevée au niveau du cortex préfrontal pourrait refléter la mise en place de processus attentionnels qui permettent de sélectionner la cible d'intérêt et d'inhiber la voix masque. De fait, l'intelligibilité de la voix cible peut être influencée par de multiples facteurs comme les différences acoustiques entre les voix, la variabilité interindividuelle dans la perception de ces différences acoustiques et dans les capacités d'attention et de mémoire. Ces facteurs peuvent influencer le degré d'implication dans l'exécution de la tâche des différentes régions du cerveau au-delà du cortex préfrontal. En effet, la compréhension d'un message en situation d'écoute complexe requiert la mise en place d'un large réseau cérébral qui comprend les régions fronto-temporo-pariétales (Evans et al., 2016 ; Fiedler et al., 2019 ; Hill & Miller, 2010 ; Szalárdy et al., 2019 ; Tóth et al., 2019 ; Zhang et al., 2018). Au sein de ce réseau, le gyrus frontal inférieur gauche (en anglais IFG) est une des régions le plus souvent impliquées.

Du fait des résultats de la littérature, nous avons utilisé une tâche d'écoute multilocuteurs avec différents TMR afin d'étudier :

- Les différences interindividuelles parmi des auditeurs normo-entendants
- Les changements intra-individuels d'intelligibilité (des auditeurs avec des performances plus basses) après des sessions d'entraînement.

À l'aide d'un appareil de spectroscopie fonctionnelle en proche infrarouge (fNIRS) ; (Quaresima & Ferrari, 2019), nous avons pu étudier les changements de l'activité corticale liés spécifiquement aux changements de stratégies perceptives induits par les sessions d'entraînement.

La présente étude a consisté en deux expériences. Dans une première expérience (*phase initiale*), des participants initialement naïfs et non entraînés ont effectué une tâche d'intelligibilité en situation multilocuteurs pendant que leur activité corticale était enregistrée.

Nous avons fait l'hypothèse que les scores d'intelligibilité de la parole des participants dans la condition TMR défavorable seraient corrélés à l'activité du cortex préfrontal (Andéol et al., 2017), mais aussi à celle d'un plus large réseau : le réseau fronto temporo-pariétal (Alain et al., 2018 ; Peelle, 2018 ; Wild et al., 2012).

Dans la deuxième expérience (*phase d'entraînement*), les participants qui avaient obtenu les scores les plus bas d'intelligibilité dans la première expérience ont été sélectionnés et entraînés à exploiter la différence d'intensité sonore entre les voix (en d'autres termes, le TMR). L'entraînement consistait soit en des instructions explicites et des feedbacks, soit simplement en une répétition de la tâche. Nous avons fait les hypothèses suivantes : a) l'intelligibilité de la parole devrait être améliorée par l'entraînement ; b) l'activité de l'IFG diminuerait de manière bilatérale suite à l'entraînement, reflétant une meilleure efficacité dans l'inhibition de la voix masque (Golestani & Zatorre, 2004 ; Qi et al., 2019 ; Sohoglu & Davis, 2016) ; c) pour les conditions dans lesquelles l'engagement de ressources cognitives était initialement faible (c.-à-d., TMR défavorable), l'activité corticale dans l'IFG gauche pourrait aussi augmenter après l'entraînement, reflétant un plus grand engagement des ressources cognitives pour l'exécution de la tâche (Eisner et al., 2010).

II. Matériels et méthode

La première expérience était propédeutique à la deuxième expérience et seulement certains participants étaient sélectionnés pour en faire partie.

Dans la première expérience, 74 participants, tous normo-entendants, ont été invités à effectuer une tâche d'écoute en situation multilocuteurs pendant que leur activité corticale était enregistrée à l'aide d'une spectroscopie fonctionnelle en proche infrarouge (fNIRS). La tâche d'écoute consistait à détecter des mots-clés (une couleur et un nombre) dictés par une voix cible (CRM corpus, Bolia et al., 2000). La voix cible était présentée en même temps qu'une autre voix, appelée voix masque, qui contribuait à rendre la situation d'écoute complexe et à perturber la détection des mots-clés de la voix cible. La voix cible et la voix masque pouvaient être placées à trois niveaux différents d'intensité sonore relative (c.-à-d., trois niveaux de target-to-masker ratio, TMR). Les trois niveaux correspondaient respectivement à trois conditions expérimentales appelées : TMR-favorable (TMR = +4dB), TMR-intermédiaire (TMR = -4dB), TMR-défavorable (TMR = -12dB ; dans les figures cette condition est appelée *Adverse*).

Après cette première expérience, seulement certains participants ont été invités à poursuivre avec la deuxième expérience. Afin de sélectionner les participants, nous avons considéré la condition TMR-défavorable, où l'écart entre les deux voix est important, pour évaluer si l'utilisation de cet indice pouvait être apprise par ceux qui n'avaient pas réussi à l'utiliser lors de la première expérience. La sélection s'est donc faite sur les participants qui avaient obtenu un score inférieur à 50 % de bonnes réponses dans la condition TMR-défavorable.

Vingt-sept participants ont ainsi été sélectionnés, et ont été répartis dans deux groupes d'entraînement différents. Le premier groupe (n=14), appelé *driven group*, recevait des instructions explicites concernant le type de TMR présent. Le deuxième groupe (n=13), appelé *not driven group*, a simplement répété la tâche telle qu'elle avait été effectuée lors de la première expérience, sans instructions ni feedback sur la performance.

La deuxième expérience durait quatre jours au total. Les trois premiers jours étaient consacrés au programme d'entraînement et ne comprenaient pas d'enregistrements de l'activité corticale. Le quatrième jour, la division des groupes disparaissait et les 27 participants effectuaient la tâche de manière identique à la première expérience avec enregistrement de l'activité corticale.

III. Acquisition des données

III.1. fNIRS

L'activité corticale a été évaluée à l'aide d'une spectroscopie en proche infrarouge-fNIRS (Nirx <https://nirx.net/nirscout/>). Quinze émetteurs ($\lambda_1=760$ nm ; $\lambda_2=850$ nm) et 23 détecteurs ont été placés sur le scalp du participant donnant lieu à 44 canaux de mesure (Figure 16.A). L'espacement source-détecteur a été maintenu à environ 30 mm à l'aide d'espaceurs en plastique. L'échantillonnage de fréquence était de 7,81 Hz. Les données ont été acquises à l'aide du logiciel NIRSTAR v14.2. Le positionnement des canaux a été effectué en utilisant le système international 10-20 EEG. Les 15 sources étaient positionnées à Fpz, AF7, AF8, F3, F4, FC5, FC6, T7, T8, TPP7h, TPP8h, P3, Pz, P4, Oz ; les 23 détecteurs étaient positionnés à Fp1, Fp2, AF3, AF4, F5, F6, FT7, FC3, FC4, FT8, C5, C6, TP7, TP8, P5, P1, P2, P6, PO1, POZ, PO2, O1, O2.

Les cartes de sensibilité corticale ont été calculées à l'aide d'un algorithme de simulation de Monte-Carlo dans Atlas Viewer (V2.8) ; (Figure 16.C). Les paramètres ont été utilisés par défaut, à l'exception du nombre de photons, qui a été fixé à $1E7$ (Aasted et al., 2015). Treize régions cérébrales d'intérêt (ROI) ont été définies (Figure 16.B), correspondant au : Cortex

Préfrontal (PFC), Cortex Préfrontal Dorsolatéral (DLPFC), Gyrus Frontal Inférieur (IFG), Cortex Auditif (AC), Lobe Pariétal Supérieur (SPL), Lobe pariétal Inférieur (IPL) et Cortex Visuel (V1). La résolution spatiale de la fNIRS permettait une zone d'observation du cerveau d'environ 1 cm² autour du centre du système de coordonnées de l'Institut neurologique de Montréal (MNI) ; (Evans et al., 1993).

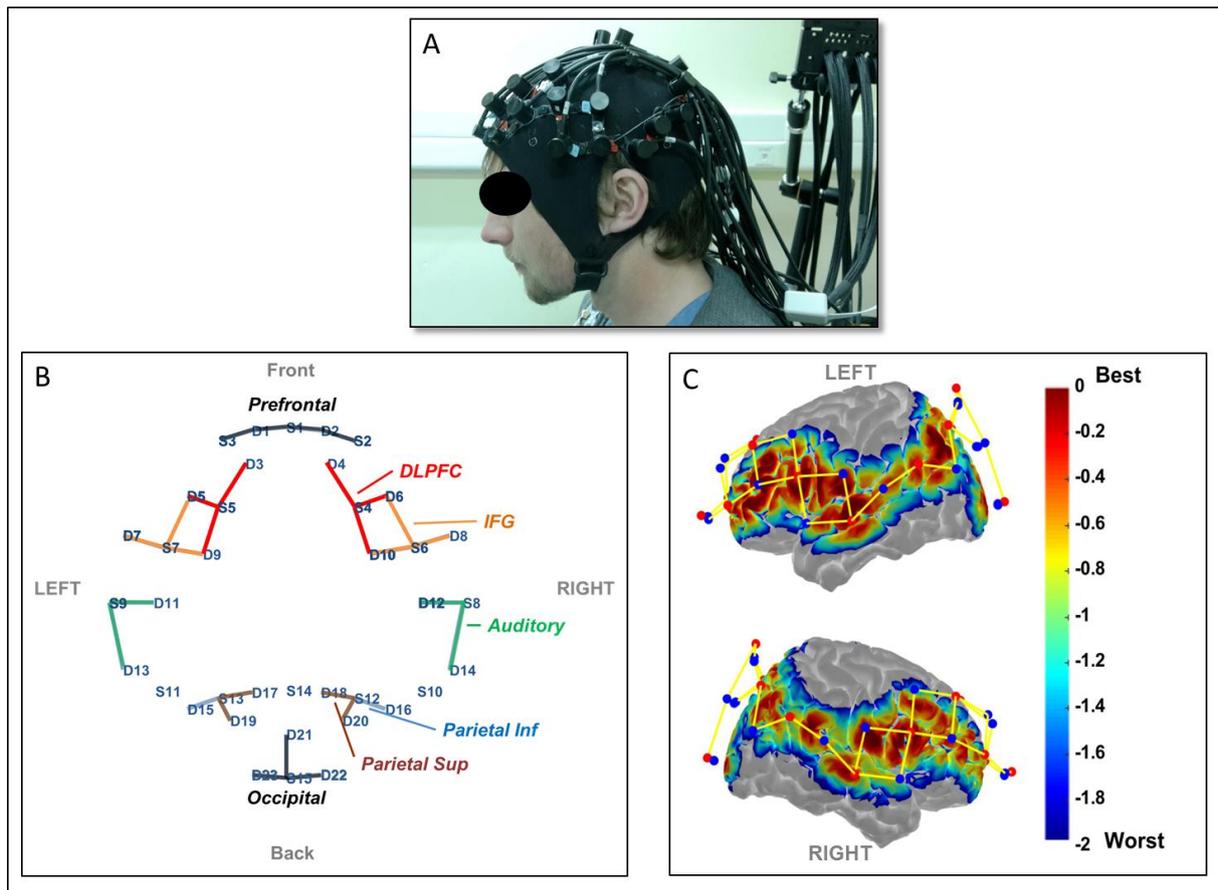


Figure 16 : A) Photographie du support d'optode-réseau placé sur la tête d'un participant à des fins de démonstration. (B) Représentation 2D de l'emplacement du ROI ; chaque ligne représente un canal de mesure. (C) Vues gauche (haut) et droite (bas) des cartes de sensibilité corticale (en mm⁻¹) obtenues par simulation de Monte-Carlo dans AtlasViewer (V2.8 ; paramètres par défaut sauf pour le nombre de photons qui a été fixé à 107).

III.2. Prétraitement données fNIRS

Les données ont été traitées à l'aide du logiciel NIRS AnalyzIR basée sur Matlab (Santosa et al., 2018) V.615, avec Matlab R2018b. Tout d'abord, les données fNIRS brutes ont été importées dans Matlab, et un tableau démographique a été créé sur la base des entrées

démographiques de NIRstar et des résultats de la performance des tâches. La durée du stimulus a été fixée à 2s, ce qui correspond à la durée d'une phrase du CRM. Les données brutes ont ensuite été converties en densité optique et soumises à une réparation de distribution dérivée temporelle (TDDR) pour la correction des mouvements (Fishburn et al., 2019). Ces auteurs ont montré que cette correction est efficace pour corriger les artefacts de mouvement dans les données fNIRS. Les changements de concentration de l'hémoglobine oxygénée et désoxygénée ont été calculés en utilisant la loi de Beer-Lambert modifiée avec un facteur de trajet partiel (ppf) de 0,1 pour les deux longueurs d'onde. Conformément aux recommandations de Pinti et al. (2019) afin de minimiser les effets de la composante systémique, les données ont été corrigées par un filtre passe-bande de Butterworth du quatrième ordre avec une fréquence passe-haut de 0,01 Hz et une fréquence passe-bas de 0,09 Hz. Une dernière étape de prétraitement a consisté en un filtrage par analyse des composantes principales (ACP) qui a été utilisé pour filtrer les deux principales composantes systémiques (c.-à-d., l'activité cardiaque et les ondes de Mayer).

IV. Analyses

IV.1. Comportement – Phase initiale

Une réponse était considérée comme correcte lorsque la couleur et le nombre étaient correctement indiqués. Le taux de réponses correctes pour chaque condition a été transformé en utilisant la transformation RAU (Rationalized Arcsin Unit) pour normaliser les données (Studebaker, 1985).

Un test préliminaire de sphéricité de Mauchly s'est avéré significatif pour le facteur TMR ($\chi^2=11,24$; $p<0,01$), c'est pourquoi une analyse de variance multivariée (MANOVA) a été utilisée pour analyser les scores RAU, avec les TMR (défavorable vs intermédiaire vs favorable) comme facteur *within subject*. Toutes les analyses statistiques comportementales ont été réalisées à l'aide du logiciel STATISTICA© (V10) de StatSoft en un seuil significatif a été fixé à 0,05.

IV.2. Comportement – Phase d'entraînement

Les tests de sphéricité de Mauchly sur les scores RAU ont donné des résultats significatifs ($\chi^2=17,43$; $p<0,001$ pour le TMR, et $\chi^2=8,56$; $p<0,05$ pour l'interaction TMR-session). Par

conséquent, des MANOVA deux par trois ont été utilisées pour analyser les scores RAU, avec la session (avant vs après l'entraînement) et les TMR (défavorable vs intermédiaire vs favorable) comme facteur *within subject* et le groupe (groupe non entraîné vs groupe entraîné) comme facteur *between subject*. Toutes les comparaisons post-hoc ont été effectuées à l'aide des tests HSD (honestly significant difference) de Tukey.

IV.3. fNIRS – Phase initiale

IV.3.1. Activation de la tâche

L'activation de la tâche a été quantifiée en convoluant le signal fNIRS avec une fonction de réponse hémodynamique canonique (HRF) pour chaque condition (TMR défavorable, intermédiaire et favorable). En suivant Barker et al. (2013) les coefficients bêta ont ensuite été estimés à l'aide d'un algorithme des moindres carrés autorégressifs itérativement repondérés (AR-IRLS) ; cet algorithme a été utilisé pour tenir compte de la présence de corrélations sérielles dans les données fNIRS et pour contrôler les erreurs de type 1. Les bêtas estimés pour les trois conditions ont été soumis à un modèle robuste à effets mixtes, avec la condition comme effet fixe et le sujet comme effet aléatoire. Enfin, les différences statistiques des bêtas pondérés entre les conditions ont été évaluées à l'aide de tests t multiples à deux échantillons. Les comparaisons multiples entre ROI ont été contrôlées à l'aide de la procédure de Benjamini-Hochberg (Benjamini & Hochberg, 1995), et les valeurs p ont été corrigées à l'aide du taux de fausse découverte (FDR).

IV.3.2. Relation entre l'activité corticale et les scores d'intelligibilité de la parole

La relation entre les scores d'intelligibilité de la parole et les activations corticales a été évaluée en utilisant des modèles à effets mixtes robustes de second niveau pour chaque condition séparément, avec la condition comme effet fixe et les scores d'intelligibilité de la parole comme régresseurs.

IV.4. fNIRS – Phase d'entraînement

Les analyses fNIRS ont été réalisées de manière identique à celles de la phase initiale. Les analyses comportementales n'ayant montré aucune différence entre les groupes, tous les

résultats fNIRS présentés ci-après se réfèrent à l'ensemble de la population (*driven group* vs *not driven group*).

V. Résultats – Phase initiale

V.1. Comportement

Les résultats comportementaux montrent un effet du TMR [$F(2, 146)=90,63, p<0,001$; voir Figure 17].

Les tests post hoc montrent que le taux de bonnes réponses (Rationalized Arcsine Unit – RAU) est plus élevé dans la condition favorable ($M=76,03$; $SD=14,93$) comparée aux conditions intermédiaire ($M=31,56$; $SD=26,85$) et défavorable ($M=35,37$; $SD=28,95$) ; ($p<0,001$ dans les deux cas). Aucune différence significative entre les conditions défavorable et intermédiaire n'a été trouvée ($p=0,5$).

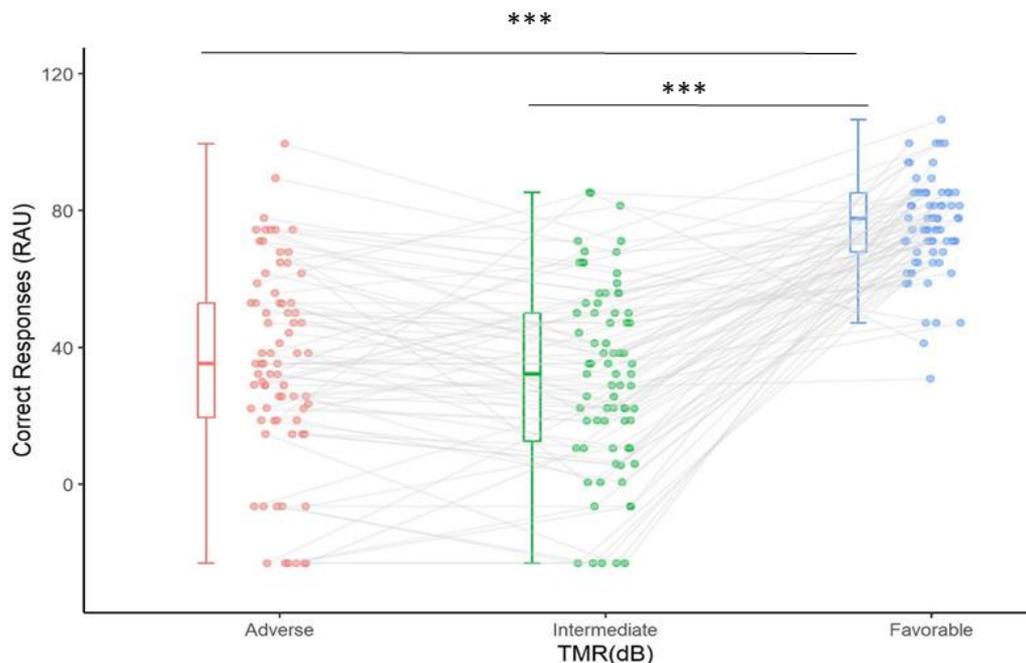


Figure 17 : Réponses correctes (Rationalized Arcsin Units) en fonction du TMR dans la phase initiale. Les boxplots montrent la médiane (barre horizontale) et l'écart interquartile (cases). Les moustaches vont de la valeur la plus basse à la valeur la plus élevée observées dans une limite de 1,5 fois l'écart interquartile. Les points indiquent les points de données individuels. Les points de données individuels correspondant au même auditeur dans différentes conditions sont reliés par des lignes grises.

V.2. f/NIRS

V.2.1. Effet de la condition

L'évaluation de l'activité corticale en fonction de la condition (c.-à-d., du TMR) a été faite sur l'hémoglobine oxygénée (HbO). Les résultats ont montré une plus grande activation de l'IFG gauche dans la condition favorable comparée à la condition défavorable ($p < 0,05$; FDR-corrigé). Le cortex pariétal inférieur (IPC) gauche montrait une plus grande activation dans la condition défavorable comparée à la condition favorable ($p < 0,05$; FDR-corrigé). Les contrastes significatifs entre les conditions favorable et défavorable sont illustrés dans la Figure 18.

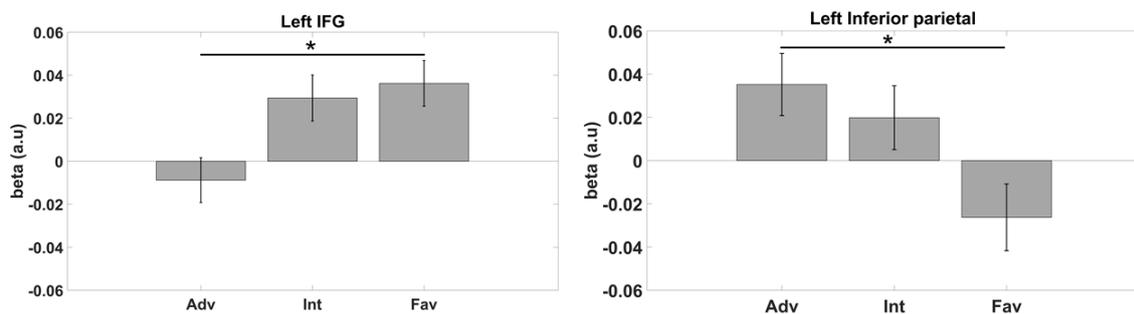


Figure 18 : Activité corticale (coefficient beta pour HbO) pour l'IFG gauche et l'IPC gauche. * : $p < 0,05$ (FDR-corrigé) pour le nombre de ROI. Les barres verticales indiquent l'erreur standard de la moyenne.

V.2.2. Corrélation comportement et activité corticale

Une régression robuste a été effectuée afin de tester la relation entre l'activité corticale et la performance. La régression robuste a été faite en utilisant les coefficients betas pondérés individuels (p. ex., les mesures de l'activation corticale) et le score d'intelligibilité (RAU). Dans la condition défavorable, une corrélation positive a été retrouvée entre ces deux variables. Cette corrélation est présente au sein d'un large réseau neuronal qui inclut le cortex préfrontal (gauche : $r = 0,18$; $p < 0,05$; droit : $r = 0,17$; $p < 0,05$), le cortex auditif gauche ($r = 0,18$; $p < 0,05$), l'IFG gauche ($r = 0,17$; $p < 0,05$) et l'IFG droit ($r = 0,17$; $p < 0,05$). Une corrélation négative a été trouvée entre l'activation du cortex pariétal supérieur (SPC) gauche et le RAU en condition défavorable ($r = -0,17$; $p < 0,05$) ; (voir Figure 19 et Figure 20). Aucune corrélation significative n'est présente entre les mêmes variables dans les conditions intermédiaires et favorables.

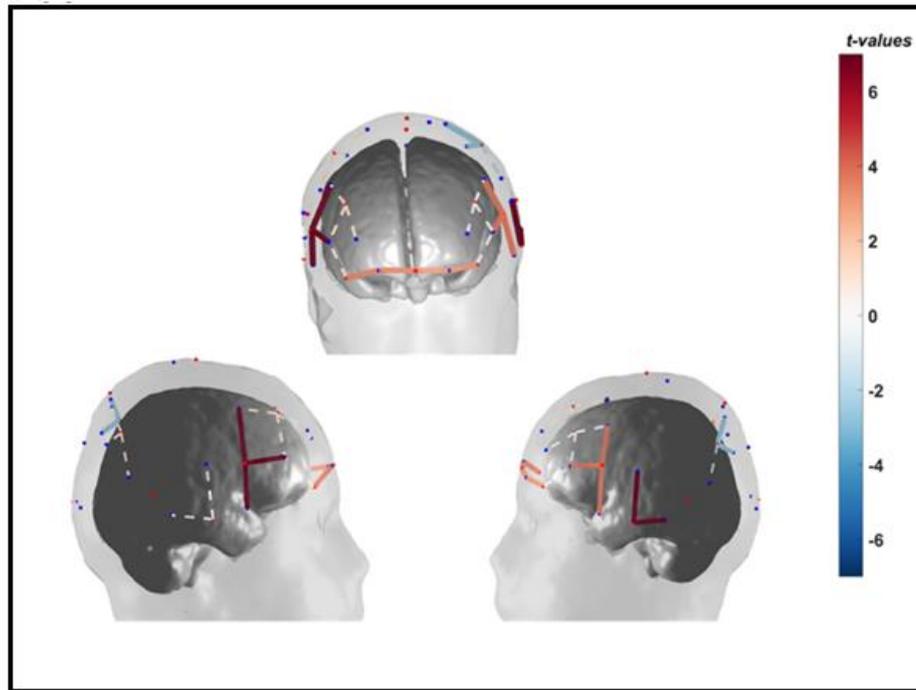


Figure 19 : Représentation 3D des canaux significativement corrélés aux scores RAU (lignes pleines ; $p < 0,05$ FDR-corrigé) : IFG gauche et droit ($r = 0,17$ et $r = 0,19$ respectivement), cortex auditif gauche ($r = 0,18$), cortex préfrontal ($r = 0,16$) et le SPC gauche ($r = -0,17$).

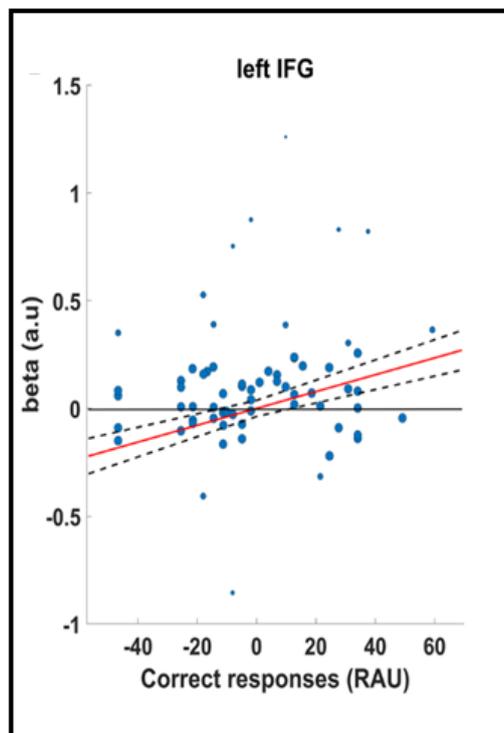


Figure 20 : Illustration de la relation entre l'activité corticale (coefficients bêta pour HbO) et l'intelligibilité de la parole (scores RAU) pour la condition défavorable.

VI. Résultats – Phase d’entraînement

VI.1. Comportement

Aucun effet significatif du groupe d’entraînement ou d’interaction entre les groupes et les autres facteurs n’a été retrouvé ($p=0,09$ dans le meilleur des cas ; Figure 21). Ces résultats indiquent que les deux types d’entraînement ont conduit à des performances post-entraînement comparables. Une interaction significative était toutefois présente entre la condition et la séance d’entraînement ($F(2,52)=12.14$; $p<0,001$, voir Figure 21). Ce résultat indique que l’entraînement n’a pas eu le même effet dans toutes les conditions. En effet, bien que les comparaisons post-hoc montrent un effet significatif de l’entraînement pour les trois conditions ($p<0,001$ pour tous les TMR) cette même interaction montre aussi un effet de l’entraînement plus petit dans la condition favorable comparé aux deux autres conditions ($p<0,001$ pour chaque comparaison).

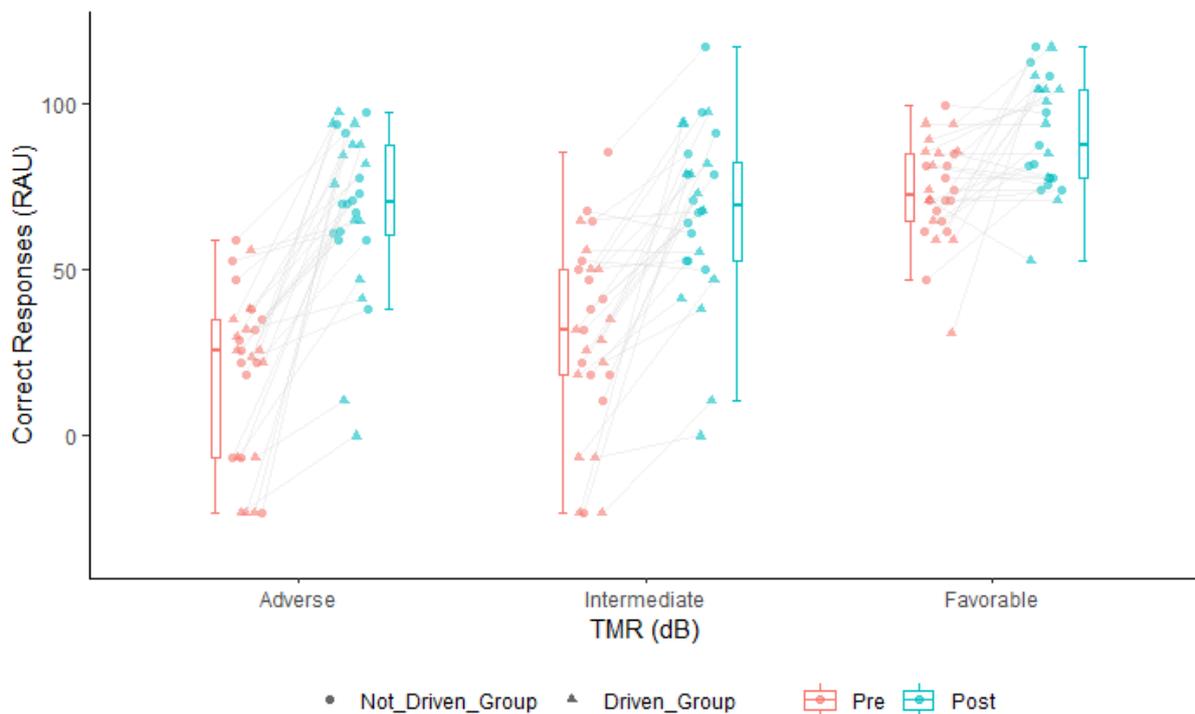


Figure 21 : Scores d’intelligibilité de la parole (RAU) pour chaque TMR (Adverse, Intermédiaire, Favorable), session d’entraînement (Pre, Post), groupe d’entraînement (*driven*, *not-driven*). Les boxplots montrent les médianes (barres horizontales) et les intervalles interquartiles (cases). Les moustaches vont de la valeur la plus faible à la valeur la plus élevée observées comprise dans 1,5 fois l’intervalle interquartile.

VI.2. f/NIRS

Les t-tests réalisés sur les betas HbO montrent une diminution de l'activité des IFGs bilatérales entre le pré et le post-entraînement ($p < 0,05$; FDR-corrigé) dans la condition défavorable. Ces contrastes révèlent aussi une diminution de l'activation de l'IFG droit dans la condition intermédiaire. Dans la condition favorable, aucune région ne montre un changement d'activation (Figure 22).

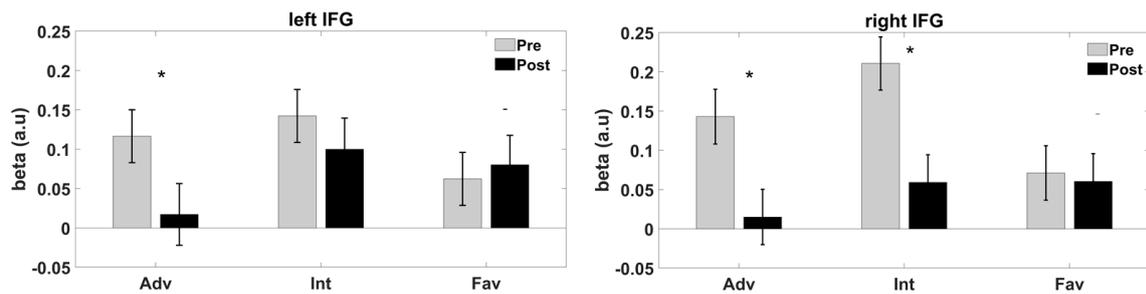


Figure 22 : Effets de l'entraînement des activations IFG gauche et droite pour chaque condition TMR (Adv : défavorable, Int : intermédiaire et Fav : favorable). * : $p < 0,05$ (FDR-corrigé).

VII. Discussion générale

La présente étude avait pour but d'étudier les mécanismes neuronaux qui sous-tendent la capacité des auditeurs à prêter attention de manière sélective à la voix d'un locuteur et à la comprendre en présence d'une voix concurrente.

Pour répondre à cette question, deux approches complémentaires (*phase initiale et phase d'entraînement*) ont été utilisées, donnant lieu à deux expériences différentes.

Dans une première expérience, les corrélations entre la performance comportementale et l'activité corticale ont été évaluées à travers trois conditions, correspondant à trois niveaux de TMR (défavorable, intermédiaire, favorable). Des corrélations significatives ont été observées uniquement pour la condition défavorable, indiquant une activité corticale plus importante lorsque la performance était plus élevée. Il est important de noter que pour certains participants, la performance était plus faible dans cette condition que dans les deux autres conditions alors que pour d'autres la performance était plus élevée dans la condition défavorable que dans la condition intermédiaire. Une interprétation plausible est que ces derniers ont probablement utilisé la différence d'intensité entre la voix cible et la voix masque pour ségréger les deux voix.

Pour tenter d'appréhender les mécanismes cérébraux en lien avec cette faculté à utiliser ce type d'indice sonore, nous avons utilisé une approche complémentaire au sein d'une seconde expérience. Dans cette approche, des sessions d'entraînement (*driven group vs not driven group*) ont été utilisées pour améliorer les performances des participants qui avaient initialement obtenu de faibles résultats dans la condition de TMR défavorable lors de la première expérience (<50 % de bonnes réponses). L'entraînement a tout d'abord conduit à une nette amélioration de la performance de manière comparable entre les deux groupes (répétition seule vs consignes + répétition + feedback). Les participants peuvent avoir développé leur propre stratégie pour réaliser la tâche (Thompson et al., 2015) probablement basée sur le contrôle inhibiteur mis en place pour contraster l'interférence produite par la présence de la voix masquée. D'ailleurs, les améliorations observées peuvent être dues, au moins en partie, à un apprentissage de la différenciation des caractéristiques des voix des locuteurs par les participants (Johnsrude et al., 2013). Les participants peuvent donc avoir appris à utiliser le TMR. Les résultats obtenus dans la phase initiale et dans la phase d'entraînement confirment l'hypothèse proposée dans des études antérieures (Alain et al., 2018), selon laquelle l'IFG gauche est impliqué dans les situations d'écoute complexes.

Toutefois, et selon nos hypothèses initiales, l'activation croissante de l'IFG gauche avec l'augmentation du TMR, observée dans la première expérience pouvait être liée :

- 1) Aux modifications acoustiques liées à l'augmentation du TMR (un paramètre sonore physique) ;
- 2) À l'augmentation de l'intelligibilité de la parole (une mesure de la performance perceptive) ;
- 3) À l'augmentation de l'efficacité des processus neuronaux liés à l'inhibition de la voix masquée ;
- 4) À l'augmentation de l'engagement de ressources cognitives.

Les résultats de la deuxième expérience concernant l'augmentation de l'intelligibilité de la parole avec l'entraînement tout en conservant les mêmes TMR, excluent la première explication ci-dessus (1). En outre, la diminution de l'activité de l'IFG gauche dans la condition de TMR défavorable après l'entraînement est contre une relation positive directe entre l'intelligibilité de la parole et l'activation de l'IFG gauche (2). De plus, si l'effet d'engagement de ressources cognitives était la principale explication, on pourrait s'attendre à ce que la corrélation entre l'activité de l'IFG gauche et l'intelligibilité de la parole trouvée dans la première expérience soit interprétée comme le résultat d'un faible engagement des ressources cognitives dans les conditions où la performance était la plus faible. Dans ces conditions, les participants auraient simplement "arrêté d'essayer". En conséquence, l'entraînement aurait conduit à augmenter

l'activité corticale en augmentant l'engagement chez les auditeurs avec de faibles performances (4), mais au lieu de cela, une diminution de l'activité corticale a été observée.

Ainsi, l'hypothèse d'une modification de l'activité neuronale liée à l'inhibition de la voix masque (3) devient l'explication la plus plausible à l'ensemble des résultats de la première et deuxième expérience.

Par ailleurs, l'entraînement conduisait à une diminution de l'activité de l'IFG droit dans la condition intermédiaire. L'IFG droit fait partie du réseau fronto-basal latéralisé à droite qui comprend les fonctions exécutives liées au contrôle inhibiteur (Chavan et al., 2015). La condition intermédiaire était la condition dans laquelle le TMR était le plus proche et donc peut-être pour les participants la condition plus difficile pour exploiter le TMR. Il est possible qu'avec l'entraînement, les participants aient amélioré leurs stratégies afin de réaliser correctement la tâche. L'acquisition de stratégies similaires à celles employées dans la condition défavorable a pu conduire dans la condition intermédiaire à une réduction de l'activité de l'IFG, notamment l'IFG droit impliqué dans le contrôle de l'inhibition.

En associant l'apprentissage à une activité neuronale plus faible, car plus efficace (Bless et al., 2014 ; Curtin et al., 2019 ; Fairclough et al., 2018 ; Putkinen & Saarikivi, 2018), la théorie de l'efficacité neuronale (en anglais *neural efficiency theory*) fournit un cadre pertinent pour la double observation dans la présente étude d'une meilleure performance et d'une activité corticale plus faible en termes d'engagement de ressources cognitives. Conformément aux hypothèses de cette théorie, Bless et al., 2014 ont constaté que l'entraînement à une tâche dichotique entraînait une meilleure performance et une activité corticale plus faible dans les régions cérébrales associées au traitement auditif sélectif. Une diminution de l'activité neuronale résultant de l'entraînement comporte probablement une économie de ressources cognitives qui peuvent être utilisées ultérieurement pour effectuer des tâches concurrentes. Cela pourrait être évalué, par exemple, dans de futures études utilisant un paradigme de double tâche. Enfin, la tâche multilocuteurs utilisée dans cette étude implique l'inhibition de la voix masque pour se concentrer sur la voix cible, ce qui suggère que les bénéfices de l'entraînement peuvent provenir d'un meilleur contrôle inhibiteur. En effet, le contrôle inhibiteur est particulièrement pertinent ici puisque la voix masque est plus forte dans la condition de TMR défavorable et beaucoup plus facile à rapporter que la voix cible. Par conséquent, l'auditeur doit lutter contre cette tendance naturelle pour obtenir une performance correcte. Néanmoins, un rôle plus large du contrôle inhibiteur en situation multilocuteurs doit être confirmé.

Le rôle du contrôle inhibiteur en situation multilocuteurs pourrait n'être crucial que lorsque la voix masque est présentée à une intensité beaucoup plus élevée que la voix cible. De futures

études pourraient cibler le rôle du contrôle inhibiteur grâce à des protocoles spécifiques. Par exemple, une expérience pourrait renforcer le contrôle inhibiteur en entraînant des participants à une tâche non auditive, et en les testant par la suite dans une tâche d'intelligibilité en situation multilocuteurs. Réciproquement, une autre expérience pourrait tester l'effet de l'entraînement via un protocole en situation multilocuteurs sur une mesure non auditive du contrôle inhibiteur (p. ex., tâche de Flanker ou tâche Go/NoGo).

Les résultats actuels indiquent que l'amélioration de l'intelligibilité dans les situations multilocuteurs peut être médiée par une capacité cognitive, à savoir le contrôle inhibiteur. Cette capacité n'étant pas spécifique à l'audition, elle pourrait être entraînée par des tâches non auditives (par exemple, des jeux sérieux) et profiter ensuite aux patients présentant des lésions auditives cliniques ou sub-cliniques. De plus, comme le suggère la diminution de l'activité de l'IFG gauche suite à l'entraînement, un tel entraînement devrait conduire à une moindre consommation de ressources cognitives et donc être plus bénéfique pour les personnes âgées ou les travailleurs soumis à des contraintes cognitives.

Chapitre II. Deuxième étude

I. Introduction

Dans une situation multilocuteurs, un auditeur peut avoir des difficultés à comprendre la voix cible en présence des voix masques. La première étude a abordé la question de la capacité d'apprentissage d'un auditeur à utiliser le TMR comme indice pour sélectionner et comprendre une voix cible. L'apprentissage de l'utilisation du TMR a en effet amélioré l'intelligibilité de la voix cible et réduit l'effort d'écoute en termes d'engagement de ressources cognitives nécessaires pour effectuer la tâche.

Comme le montre la littérature, le TMR n'est pas le seul indice dont dispose un auditeur pour distinguer les différents flux sonores dans une scène auditive complexe. La séparation spatiale des locuteurs a aussi montré son efficacité sur l'amélioration de l'intelligibilité (Bronkhorst, 2015 ; Brungart & Simpson, 2005) et sur la réduction des ressources cognitives engagées pour y parvenir (Andéol et al., 2017 ; Xia et al., 2015). En particulier, Xia et al. (2015) ont montré que le fait d'ajouter des indices spatiaux (c.-à-d., séparation spatiale de 15°) aux indices de genre pouvait réduire l'effort d'écoute. Les mêmes auteurs dans une deuxième expérience ont montré qu'une séparation spatiale plus large des locuteurs (c.-à-d., séparation spatiale de 60°) permettait d'obtenir un effort d'écoute plus faible que celui obtenu avec les indices de genre. Dans une autre étude, Andéol et al. (2017), montraient une amélioration de l'intelligibilité de la voix cible en concomitance d'une réduction de l'effort d'écoute avec une séparation spatiale de 60°. La réduction de l'effort d'écoute se traduisait par un moindre engagement de ressources cognitives préfrontales enregistrées à l'aide d'une fNIRS.

Dans la présente étude, les effets de la séparation spatiale dans une tâche multilocuteurs étaient étudiés chez de jeunes normo-entendants. L'objectif était de dépasser les observations d'Andéol et al. (2017) limitées au cortex préfrontal, en explorant les régions cérébrales fronto-temporo-pariétales dont la littérature montre l'implication dans les situations d'écoute complexe (Alain et al., 2018 ; Peelle, 2018 ; Chapitre I. Première étude). Dans le but d'approfondir le rôle de l'IFG, nous avons défini de façon plus détaillée les trois parties dont il est composé et nous avons adapté le montage expérimental de la fNIRS pour les enregistrer : la *pars orbitalis* (BA 47), la *pars triangularis* (BA 45) et la *pars opercularis* (BA 44). L'observation en fNIRS a donc compris l'IFG et ses trois parties, d'autres régions frontales dont le DLPFC, les régions temporales comprenant le STG et le MTG, et les régions pariétales comprenant l'IPC (Alain et al., 2018 ; Peelle, 2018).

En outre, du fait de l'objectif opérationnel de la thèse, l'efficacité de la séparation spatiale des locuteurs dans la compréhension de la voix cible et l'effort d'écoute ont été évalués avec un nombre plus élevé de locuteurs. Des solutions techniques sont en effet recherchées pour permettre une meilleure intelligibilité sans induire d'effort significatif dans des systèmes de communication avec un grand nombre d'interlocuteurs. La littérature montre une diminution de l'intelligibilité de la voix cible au fur et à mesure de l'augmentation du nombre de voix masque (Brungart, 2001 ; Humes et al., 2017 : sans séparation spatiale des locuteurs ; Pinto et al., 2020 : avec séparation spatiale des locuteurs). Par exemple, Pinto et al. (2020) dans leur étude montraient que les effets de la séparation spatiale étaient plus bénéfiques dans la condition avec 2 locuteurs qu'avec 4 locuteurs.

Dans la présente étude, les participants étaient invités à réaliser une tâche d'écoute multilocuteurs avec différents degrés de séparation des locuteurs. En détail, la voix cible restait tout au long de l'expérience à 0° par rapport à l'auditeur et elle pouvait être présentée avec deux voix masque placées à 60° à droite et à gauche (condition 60) ; avec deux voix masque placées à 20° à droite et à gauche (condition 20) et avec quatre voix masque placées à 60° et à 20° à droite et à gauche (condition 20-60).

Ainsi, nos hypothèses étaient de vérifier si a) un degré d'écartement plus grand (condition 60) facilitait la ségrégation des locuteurs en augmentant l'intelligibilité de la voix cible (Andéol et al., 2017 ; Xia et al., 2015), si b) cette meilleure intelligibilité était sous-tendue par un engagement plus important de ressources cognitives (première étude) et enfin c) si la séparation spatiale des locuteurs pouvait garantir la compréhension de la voix cible en présence d'un nombre élevé de locuteurs.

II. Matériels et méthode

II.1. Participants

Vingt-huit volontaires (8 femmes, âge moyen=24,75±3,83) ont participé à cette étude. Tous les participants étaient des étudiants de l'ISAE-SUPAERO et de l'Université de Toulouse, ils étaient tous normo-entendants avec un seuil auditif au-dessous de 20 dB HL pour les fréquences de 125 Hz to à 8 kHz. L'audiométrie a été enregistrée avec le logiciel Audio Console Oscilla®.

L'étude a été approuvée par le Comité de Protection des Personnes sud-ouest et outre-mer II, IDRCB 2017-A00859-44). Avant de commencer l'étude, les participants remplissaient un formulaire de consentement éclairé. Le montant de leur indemnisation pour les contraintes de participation était de 20 euros.

II.2. Tâche d'intelligibilité

Les participants étaient invités à effectuer une tâche adaptée à partir du « Coordinate Response Measure » (CRM) corpus ; (Bolia et al., 2000). Comme dans la première étude, le protocole du CRM corpus était constitué d'une liste de phrases, chacune composée d'un des huit noms de code « call sign » (parmi : Baron, Charlie, Ringo, Eagle, Arrow, Hopper, Tiger, Laker), une couleur (parmi : bleu, vert, rouge et blanc) et un chiffre (de 1 à 8). La structure d'une phrase était : "Ready *Call Sign* go to *Color Number* Now". La phrase désignée comme étant la voix cible était celle avec le call sign *Baron*. Toutes les autres phrases ayant un call sign autre que *Baron* devaient être considérées comme des voix masques. Les participants devaient suivre les indications de la voix cible, qui leur demandait d'associer une couleur à un chiffre. Ils pouvaient ensuite indiquer la couleur et le chiffre sur un boîtier de réponse. Tous les locuteurs (c.-à-d., voix cible et voix masques) étaient masculins.

Trois conditions expérimentales ont été testées. Chaque condition expérimentale correspondait à une configuration spatiale différente de locuteurs. La voix cible était toujours placée à 0°, droit devant l'auditeur. Dans la première condition, appelée *Condition 20*, deux voix masques étaient présentées à +20° et -20° par rapport à la voix cible. Dans la deuxième condition, appelée *Condition 60*, deux voix masques étaient présentées à +60° et -60° par rapport à la voix cible. Dans la troisième condition, appelée *Condition 20-60*, quatre voix masque étaient présentées à +20°, -20°, +60° et -60° par rapport à la voix cible (Figure 23).

L'expérience commençait avec 8 présentations isolées de la voix cible. Les participants étaient invités à reporter sur le boîtier de réponse le couple couleur-chiffre énoncé par la voix cible. Ces présentations permettaient de familiariser les participants avec la voix cible sans présentation des voix masques. Ensuite, 8 présentations de chaque condition (c.-à-d., Condition 20, Condition 60, Condition 20-60) étaient proposées pour faciliter la bonne compréhension de la tâche. Après cette première phase d'entraînement et de familiarisation, pour chaque condition, 32 essais étaient proposés pour un total de 96 essais. Les 32 essais par condition

étaient divisés en deux mini-blocs de 16 essais et un total de 6 mini-blocs de 16 essais chacun était donc présenté. Chaque stimulus (une phrase complète) durait 2 secondes. L'intervalle inter-stimulus (ISI) était fixé à 18 secondes pour tenir compte de la durée de la réponse hémodynamique.

L'objectif pour les participants était de rester focalisés sur la voix cible tout en négligeant les voix masques. Les participants ne recevaient pas de consignes de vitesse de réponse, mais ils étaient invités à rester concentrés et focalisés le plus possible.

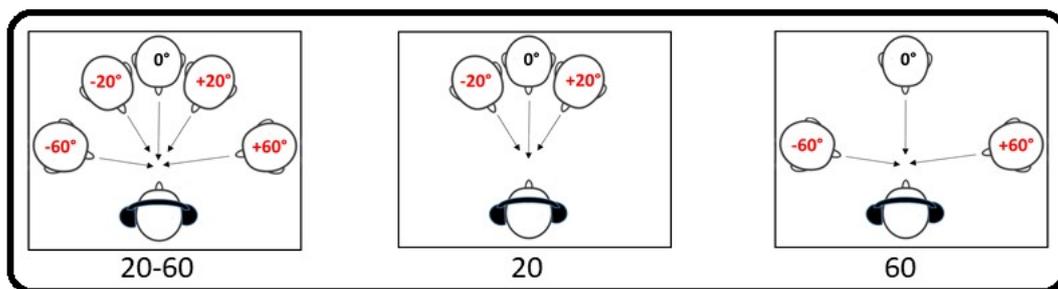


Figure 23 : Design expérimental des trois conditions avec la voix cible (en noir) toujours placée à 0° droit devant le participant et les voix masques (en rouge) placées sur les côtés. En détail, la condition 20-60 avec une voix cible et quatre voix masques placées à $\pm 20^\circ$ et $\pm 60^\circ$; la condition 20 avec une voix cible et deux voix masque placées à $\pm 20^\circ$; la condition 60 avec une voix cible et deux voix masque placées à $\pm 60^\circ$.

Les stimuli ont été générés à l'aide d'un script écrit en Matlab® (2018b). Les fichiers audio (.wav) ont été transmis de l'ordinateur (Dell Optiflex 990) à un processeur de signal numérique (RX8 ; Tucker Davis Technologies) via fibre optique pour être filtrés en temps réel avec des HRTF mesurés avec une tête artificielle (Neuman KU-100). Le signal a été amplifié par un tampon de casque (HB7, TDT) et présenté à l'aide d'un casque (Beyer Dynamics DT770) à un niveau global d'environ 70 dB SPL (puissance efficace) aux deux oreilles. Le signal combiné (voix cible et voix masque) a été déplacé sur une plage de 6 dB (de 67 dB SPL à 73 dB SPL, par pas de 1 dB). Les réponses comportementales ont été enregistrées à l'aide d'un boîtier de réponse avec trois rangées de boutons (la première rangée était dédiée aux couleurs, la deuxième rangée aux nombres d'un à quatre et la troisième rangée aux nombres de cinq à huit). Le boîtier de réponse était relié au processeur en temps réel et les réponses étaient enregistrées avec Matlab (Figure 24).



Figure 24 : Boîtier de réponse avec trois rangées de boutons. La première est composée par 4 couleurs (bleu, vert, rouge, blanc) ; la deuxième par des chiffres de 1 à 4 ; la troisième par des chiffres de 5 à 8.

III Acquisition des données

III.1 f /NIRS

L'activité corticale était enregistrée avec un appareil de spectroscopie en proche infrarouge (fNIRS, NirScout Nirx <https://nirx.net/nirscout/>). Afin de permettre l'enregistrement, 24 optodes ont été utilisées : 12 sources à deux longueurs d'onde ($\lambda_1=760$ nm ; $\lambda_2=850$ nm) et 12 récepteurs, placées sur le scalp pour réaliser un total de 34 canaux de mesure. L'espace entre les optodes était d'environ 30 mm. Pour assurer le maintien correct de cette distance, des espaceurs en plastique étaient utilisés entre les sources et les détecteurs. La fréquence d'échantillonnage était de 5,2 Hz. Les données étaient acquises avec le logiciel NIRSTAR (v14.2). Les optodes étaient placés sur le scalp des participants selon le système 10-20 EEG. Les 12 sources étaient placées sur F8, T8, FC6, CP6, F4, C4, F7, T7, FC5, CP5, F3, C3 ; les 12 détecteurs étaient placés sur FT8, TP8, F6, C6, FC4, CP4, FT7, TP7, F5, C5, FC3, CP3. La carte de sensibilité a ensuite été élaborée par la méthode de simulation de Monte-Carlo avec l'Atlas Viewer (V2.8) ; (Figure 25).

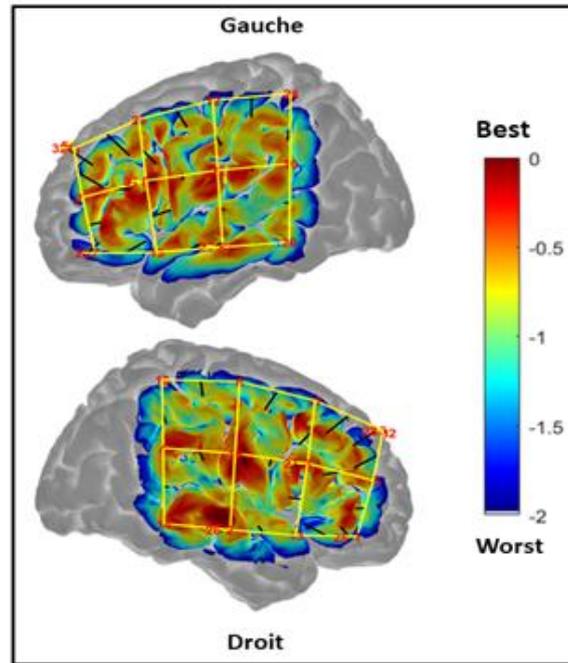


Figure 25 : Carte de sensibilité (en mm^{-1}) obtenue avec simulation de Monte-Carlo (haut hémisphère gauche, bas hémisphère droit). La sensibilité est la plus grande lorsqu'elle se rapproche de zéro mm^{-1} .

Les paramètres ont été sélectionnés par défaut exception faite du nombre de photons qui était fixé à 10^7 (Aasted et al., 2015). À l'aide de ce montage, huit régions cérébrales d'intérêt (ROIs) ont été définies (Figure 26). Elles étaient respectivement : le gyrus frontal inférieur (IFG) comprenant ses trois sous-parties : pars orbitalis, pars opercularis et pars triangularis ; le cortex préfrontal dorsolatéral (DLPFC) ; le gyrus temporal supérieur (STG) ; le gyrus temporal moyen (MTG), le MTG-STG (région à cheval entre le STG et le MTG) et le cortex pariétal inférieur (IPC). La dénomination des régions a été effectuée en tenant compte des coordonnées MNI (Montreal National Institut) des centroïdes des canaux de mesures et d'un rayon de 5mm autour de ces centroïdes, reportés dans l'atlas MNI-Talairach en ligne de Bioimage (<https://bioimagesuiteweb.github.io/webapp/mni2tal.html>).

Pour rappel, la résolution de la fNIRS suppose un volume d'observation de l'ordre du cm^3 autour du centroïde de chaque canal de mesure.

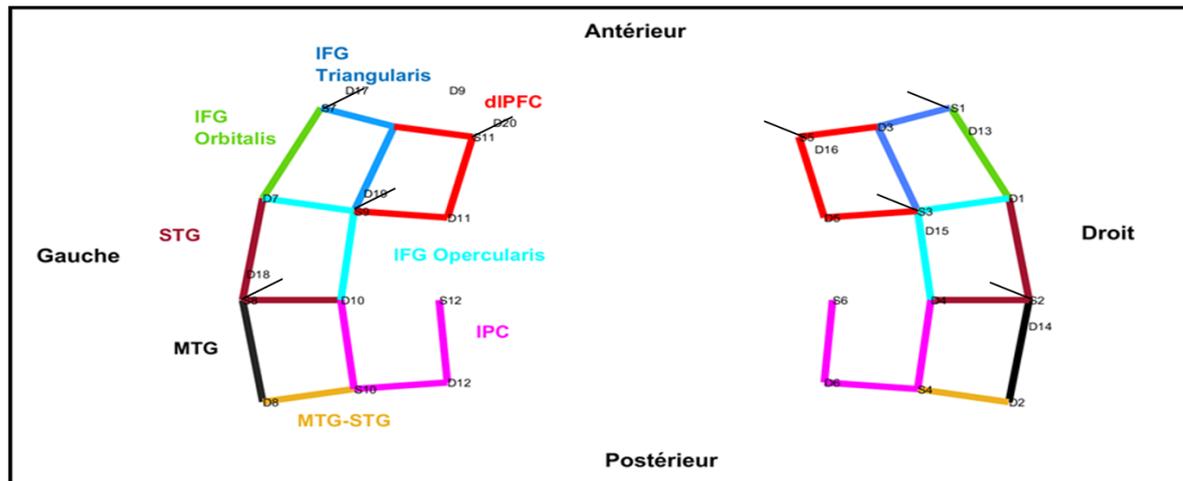


Figure 26 : Représentation en 2D des ROIs. Chaque ligne représente un canal source-émetteur, chaque couleur, une région d'intérêt pour laquelle les données mesurées à travers les canaux qui la composent ont été moyennées. Les lignes noir plus courtes représentent les short channels utilisés dans chaque hémisphère cérébral.

III.2. Prétraitement données f/NIRS

Les données ont été traitées avec le logiciel NIRS AnalyzIR V.615 (Santosa et al., 2018) avec Matlab R2018b. Initialement, les données fNIRS brutes ont été importées dans Matlab. Les données brutes ont ensuite été converties en densité optique et soumises à une correction par la mesure de 8 canaux de courte distance (<8mm) en utilisant la fonction NtbxSSR fournie par NIRX. Les 8 canaux étaient respectivement les canaux : S1-D13, S2-D14, S3-D15, S5-D16 (dans l'hémisphère droit) et les canaux : S7-D17, S8-D18, S9-D19, S11-D20 (dans l'hémisphère gauche) ; (Figure 26).

Cette méthode s'est montrée particulièrement efficace pour réduire la pollution du signal neurophysiologique par les composantes systémiques (rythme cardiaque, ondes de Mayer). Les données optiques corrigées ont ensuite subi une correction des mouvements par la méthode de TDDR pour *Temporal Derivative Distribution Repair* (Fishburn et al., 2019). Ces auteurs ont montré que cette correction est particulièrement efficace pour corriger les artefacts de mouvement dans les données fNIRS. Les changements de concentration de l'hémoglobine oxygénée et désoxygénée ont été calculés en utilisant la loi de Beer-Lambert modifiée avec un facteur de trajet partiel (ppf) de 0,1 pour les deux longueurs d'onde. Conformément aux recommandations de Pinti et al. (2019) afin de minimiser les effets résiduels de la composante systémique, les données ont été corrigées par un filtre passe-bande de type Butterworth d'ordre quatre avec une fréquence passe-haut de 0,01 Hz et une fréquence passe-bas de 0,09 Hz.

IV. Analyses

IV.1. Comportement

Une réponse était considérée comme correcte quand la couleur et le chiffre étaient correctement reportés. Le nombre de réponses correctes a été transformé en utilisant le Rationalized Arcsin Unit (RAU) ; (Studebaker, 1985).

Pour analyser les différentes performances à travers les conditions expérimentales, une Analyse de Variance Multivariée (MANOVA) a été effectuée avec les trois conditions (Condition 20 vs Condition 60 vs Condition 20-60) comme facteur *within subject*.

IV.1.1 f/NIRS

L'effet de la condition a été quantifié en convoluant le signal fNIRS avec une fonction de réponse hémodynamique (HRF) canonique pour chaque participant et chaque condition (Condition 20 vs Condition 60 vs Condition 20-60). La durée du stimulus a été fixée à 2 s, ce qui correspond à la durée de chaque phrase du CRM. Suivant Barker et al. (2013) les coefficients bêta ont ensuite été estimés à l'aide d'un algorithme de moindres carrés autorégressifs itérativement pondérés (AR-IRLS) ; cet algorithme a été utilisé pour tenir compte de la présence de corrélations sérielles dans les données fNIRS et pour contrôler les erreurs de type 1. Les bêtas estimés pour les trois conditions ont ensuite été soumis à un modèle robuste à effets mixtes pour les analyses de groupe, avec la condition comme effet fixe et le sujet comme effet aléatoire. Enfin, les différences statistiques des bêtas pondérés entre les conditions ont été évaluées à l'aide de tests de t multiples à deux échantillons. Les comparaisons multiples entre ROIs ont été contrôlées à l'aide de la procédure de Benjamini-Hochberg (Benjamini & Hochberg, 1995), et les valeurs p ont été corrigées à l'aide du taux de fausse découverte (FDR).

V. Résultats

V.1. Comportement

Les résultats comportementaux montrent un effet significatif de la condition [$F(2, 54)=105,93$; $p<0,001$; Figure 27]. Les comparaisons post hoc montrent des différences significatives entre toutes les conditions ($p < 0,001$). Dans la condition 60, la performance est la plus élevée ($M =$

65,77 ; SD = 30,06) comparée à la condition 20 (M = 43.60 ; SD = 23.62 ; $p < 0,001$) et à la condition 2060 (M = 15,87 ; SD = 16,53 $p < 0,001$).

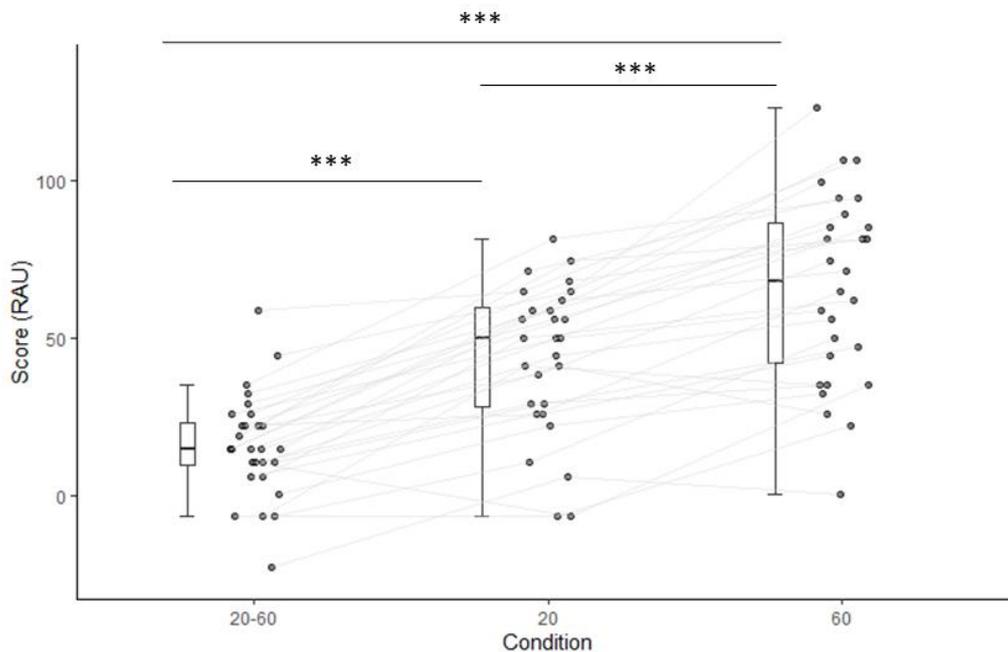


Figure 27 : Réponses correctes (Rationalized Arcsin Units) pour les trois conditions de séparation spatiale des locuteurs. Les boxplots montrent la médiane (barre horizontale) et l'écart interquartile (cases). Les moustaches vont de la valeur la plus basse à la valeur la plus élevée observées dans une limite de 1,5 fois l'écart interquartile. Les points indiquent les points de données individuels. Les points de données individuels correspondant au même auditeur dans différentes conditions sont reliés par des lignes grises.

V.2. f/NIRS

Une activation plus importante a été observée pour la condition 60 comparée à la condition 20 dans l'IFG *pars triangularis* gauche uniquement ($p < 0,05$; FDR-corrigé) ; (Figure 28).

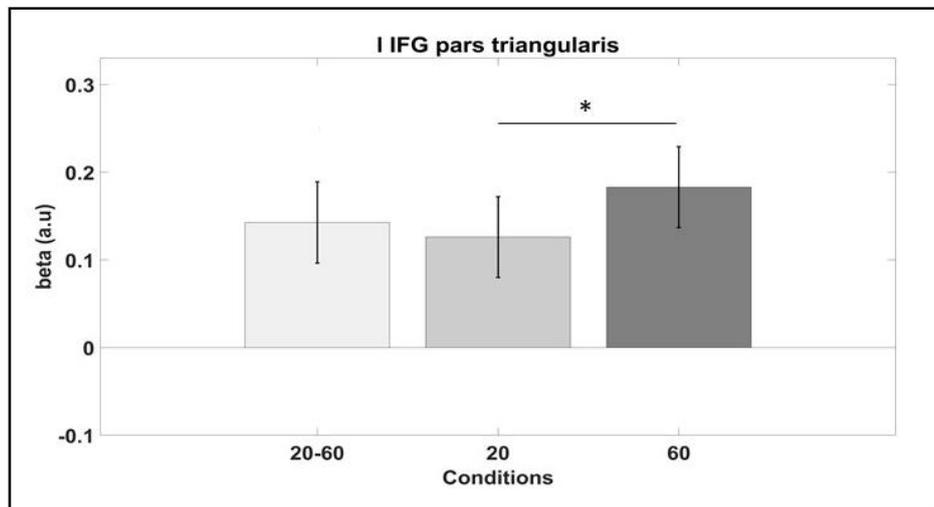


Figure 28 : Activations (beta) pour le gyrus frontal inférieur gauche pars traingularis (lIFG * : $p < 0,05$ (FDR-corrigé)). Les barres verticales indiquent l'écart-type de la moyenne.

Une activation plus importante est observée pour la condition 20-60 comparée à la condition 20 dans l'IPC droit ($p < 0,05$; FDR-corrigé) et à la condition 60 ($p < 0,01$; FDR-corrigé) ; (Figure 29).

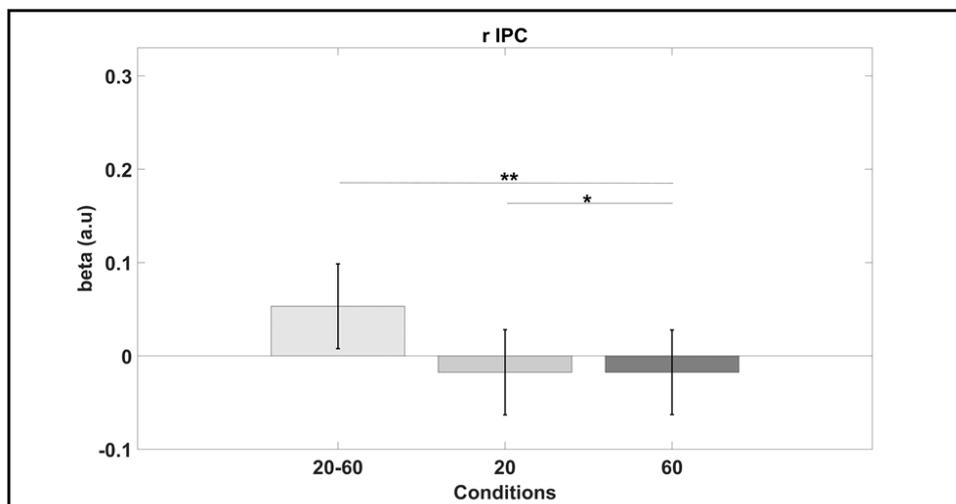


Figure 29 : Activations pour le cortex pariétal inférieur droit (rIPC)*: $p < 0,05$ (FDR-corrigé). Les barres verticales indiquent l'écart-type de la moyenne.

Une activation plus importante est observée pour la condition 60 comparée à la condition 20-60 dans le DLPFC droit ($p < 0,01$; FDR-corrigé). Dans le DLPFC gauche, toutes les conditions sont différentes entre elles avec une plus grande activation dans la condition 60 (Condition 20 : $p < 0,001$ FDR-corrigé; Condition 20-60 : $p < 0,05$ FDR-corrigé) ; (Figure 30).

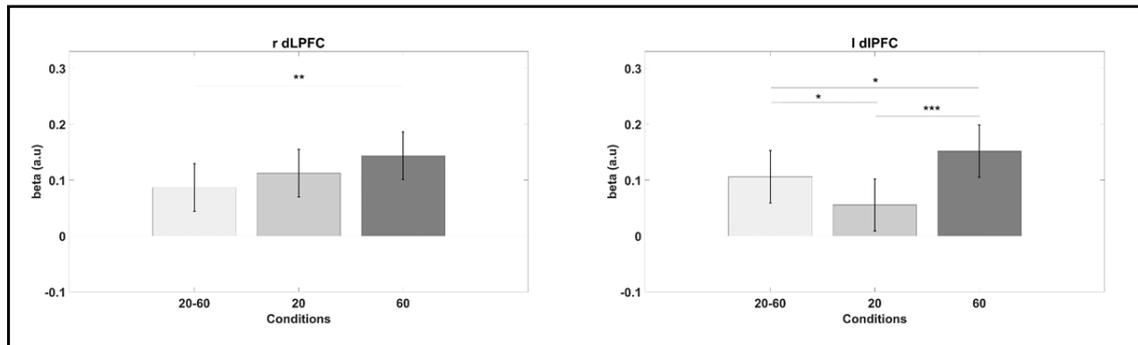


Figure 30 : Activations pour le cortex préfrontal dorsolatéral droit (rDLPFC) et pour le cortex préfrontal dorsolatéral gauche (lDLPFC) *: $p < 0,05$ (FDR-corrigé). Les barres verticales indiquent l'écart-type de la moyenne.

Une activation plus importante est observée dans le STG pour la condition 60 et pour la condition 20 comparée à la condition 20-60 (respectivement Condition 20 : $p < 0,001$; FDR-corrigé et Condition 60 : $p < 0,01$ FDR-corrigé) ; (Figure 31).

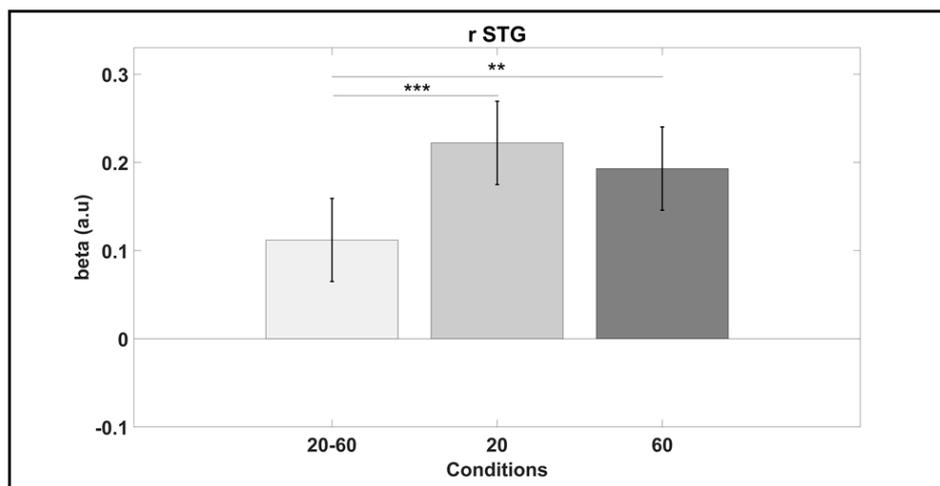


Figure 31 : Activations pour le gyrus temporal supérieur droit (rSTG) *: $p < 0,05$ (FDR-corrigé). Les barres verticales indiquent l'écart-type de la moyenne.

VI. Discussion

La présente étude visait à étudier les effets de la séparation spatiale des locuteurs sur l'intelligibilité de la voix cible et sur l'effort d'écoute dans une situation d'écoute multilocuteurs, en termes de ressources cognitives engagées dans la tâche d'écoute.

Dans cette étude, différentes configurations spatiales ont été testées tout en faisant varier le nombre de locuteurs présents dans la scène auditive. La voix cible restait toujours au même endroit tandis que différentes voix masques, de 2 à 4, entouraient l'auditeur dans un espace qui couvrait un arc de -60° à $+60^\circ$.

Premièrement, les résultats ont montré que ce n'est pas seulement la séparation spatiale qui garantit une intelligibilité plus élevée de la voix cible, mais aussi le degré d'écartement entre les locuteurs et le nombre de ceux-ci. De fait, la condition 60 avec trois locuteurs placés à un plus grand écartement était la condition avec la plus haute intelligibilité. De plus, la performance se réduisait à partir de la condition avec trois locuteurs (dont deux masques) placés à un degré d'écartement moins grand (condition 20) jusqu'à se réduire drastiquement dans la condition avec 5 locuteurs qui recouvraient différents degrés d'écartements (condition 20-60). L'interprétation la plus probable est qu'un plus grand écart parmi les locuteurs permet à la fois de bien ségréger les flux sonores et aussi de se focaliser sur la voix cible en inhibant les autres voix masque (Andéol et al., 2017 ; Xia et al., 2015), mais que ce bénéfice est limité à un nombre de locuteurs réduit.

Deuxièmement, au niveau cérébral, un large réseau cérébral était impliqué dans la réalisation de la tâche et incluait les régions fronto-temporo-pariétales (Alain et al., 2018 ; Peelle, 2018). En détail, la condition 60, qui était la condition avec la plus haute intelligibilité, sollicitait l'activation de l'IFG pars triangularis gauche, le DLPFC bilatéral et le STG gauche, confirmant ainsi le fait de l'implication d'un large réseau de régions cérébrales. Dans cette étude, un point particulier parmi les intérêts de recherche consistait à approfondir le rôle de l'IFG gauche qui avait déjà émergé dans la première étude et dans la littérature scientifique (Abrams et al., 2013 ; Alain et al., 2018 ; Peelle, 2018 ; Chapitre I. Première étude). Ainsi, nous avons trouvé que la condition 60 montrait une activation significativement plus importante de l'IFG triangularis gauche par rapport à la condition 20. L'activité de l'IFG triangularis gauche a été déjà montrée comme étant liée à l'intelligibilité du langage (Abrams et al., 2013), mais aussi à la capacité apprise à inhiber dans une tâche d'inhibition comme le Go/NoGo (Chavan et al., 2015). En outre, une étude récente, qui utilisait un protocole d'écoute multilocuteurs, a montré

que les activations de l'IFG *pars triangularis* gauche était en corrélation avec les améliorations comportementales dues à la présence de la séparation spatiale des locuteurs où les locuteurs étaient placés à $\pm 90^\circ$ et 0° (Wang et al., 2021). Il semblerait donc qu'il existe un lien spécifique entre l'activité de l'IFG *pars triangularis* gauche et l'utilisation de l'indice de la séparation spatiale des locuteurs.

Cette étude a également mis en évidence l'implication des DLPFC bilatéraux dans la condition avec la meilleure intelligibilité atteinte. Cette condition était celle avec la plus grande séparation spatiale des locuteurs. Cette séparation spatiale facilitait probablement un meilleur contrôle inhibiteur des voix masques au sein de l'IFG. Le DLPFC est considéré comme une des régions impliquées dans le contrôle de l'inhibition (Scannella, 2011).

Ce résultat semble s'aligner avec les résultats de la première étude présentée dans ce manuscrit, dans laquelle, il existe un lien entre une performance plus élevée et un plus grand engagement cognitif chez les auditeurs non entraînés. Probablement, le plus grand écartement entre les locuteurs a permis de mieux ségréger les voix et de se focaliser sur la voix cible via l'inhibition des voix masques.

Bien que la séparation spatiale facilite la compréhension, elle n'évite pas à l'auditeur de devoir mobiliser davantage de ressources cognitives pour réussir la tâche. Toutefois, ce que nous ne pouvons pas révéler ici, c'est à quel *coût personnel* l'auditeur déploie des ressources cognitives tout en assurant une bonne compréhension. Par *coût personnel*, nous entendons l'investissement subjectif de l'auditeur, par exemple le niveau d'effort perçu et/ou la motivation investie afin de réaliser la tâche, qu'il aurait fallu mesurer à l'aide de questionnaires. La littérature scientifique sur l'effort d'écoute (p. ex., MOLE de Herrmann & Johnsrude, 2020 ou FUEL de Pichora-Fuller et al., 2016) souligne l'importance de la motivation pour faire face aux défis soulevés par la tâche d'écoute et pour générer un vécu subjectif de valence positive (c.-à-d., joie) ou négative (c.-à-d., frustration, ennui) qui poussera l'auditeur à poursuivre dans la réalisation de la tâche ou à renoncer. Pour cette raison, évaluer l'effort subjectif ressenti par l'auditeur avec des mesures subjectives semble important et devrait être intégré dans les études futures.

Troisièmement, malgré leur séparation spatiale, un accroissement du nombre de locuteurs de 3 à 5 a eu un impact négatif sur la performance (condition 20-60). Pinto et al. (2020) souligne que les meilleures performances sont celles avec un nombre limité de locuteurs. Dans notre étude ce nombre serait de trois locuteurs avec un degré important de séparation spatiale.

Au niveau cérébral la condition 20-60 qui est la condition avec la moindre intelligibilité ne semble pas engager de ressources cognitives importantes dans les régions frontales et pariétales contrairement à la condition 60. La région cérébrale la plus impliquée était l'IPC droit. Plusieurs études chez des sujets sains et chez des patients avec des lésions cérébrales ont montré que la localisation d'un son dans un espace sollicite l'activation de régions postérieures parmi lesquelles l'IPC bilatéral (Lewald et al., 2018 ; Tissieres et al., 2019). Les participants ont probablement essayé d'engager des ressources cognitives dans cette condition pour exploiter le plus possible le seul indice sonore présent pour ségréger les voix. Cet engagement ne conduisait pourtant pas à atteindre une intelligibilité élevée. Ces résultats semblent donc révéler que la condition 60, qui compte le plus grand écartement entre les locuteurs permet d'avoir une meilleure intelligibilité, mais avec un engagement plus important des ressources cognitives. Cependant, il reste à déterminer le vécu subjectif de cet engagement. Ainsi, l'étude suivante s'est attachée à mesurer ce vécu.

De plus, l'objectif opérationnel et clinique de cette thèse a conduit à inclure des militaires porteurs de lésions auditives plus ou moins prononcées. Étant donné qu'une plus grande séparation spatiale des locuteurs améliore l'intelligibilité de la voix cible chez les personnes normo-entendantes, la troisième étude s'est intéressée aux effets de la séparation spatiale chez les militaires modérément mal entendants avec la prise en compte de mesures subjectives de l'effort d'écoute et de la motivation.

Chapitre III. Troisième Étude

I. Introduction

La présente étude était complémentaire à la deuxième. Cette dernière a confirmé l'efficacité de la séparation spatiale des locuteurs afin d'améliorer l'intelligibilité d'une voix cible en situation multilocuteurs avec, toutefois, un plus grand engagement de ressources cognitives. Cependant, dans cette deuxième étude, le *coût personnel* investi afin d'attendre ces résultats était inconnu du fait du manque d'évaluations subjectives. De plus, la population d'intérêt était celle de jeunes normo-entendants, ici l'intérêt de recherche a été porté vers une population différente : une population de militaires en carrière modérément mal entendants.

Contrairement aux jeunes normo-entendants, la population militaire est exposée à des intensités sonores fortes et peut présenter une audition périphérique dégradée plus fréquemment. En effet, les militaires peuvent être souvent exposés à des situations de travail bruyantes comme le bruit des engins motorisés ou être exposés à des tirs ou à des explosions lors d'entraînements ou lors de combats. Des lésions auditives peuvent être la conséquence de telles surexpositions sonores. Dans les armées, l'état auditif est classiquement évalué par l'audiométrie tonale liminaire. Cependant, cette évaluation de l'état auditif est insuffisante, car peu corrélée aux mesures supraliminaires comme la compréhension de la parole dans le bruit (Phatak et al., 2019), ou la localisation auditive (Ruggles & Shinn-Cunningham, 2011). De plus, l'audiométrie tonale est considérée comme une mesure subjective et opérateur-dépendant (p. ex., de nombreux opérateurs ne testent pas des intensités inférieures à 0 dB HL qui n'est pourtant que la valeur moyenne attendue dans une population normo-entendante et non pas une valeur minimale absolue). Les résultats qui sont fournis peuvent entraîner une inaptitude à servir, redoutée par le militaire dont certains développeraient des stratégies plus ou moins efficaces pour améliorer artificiellement le résultat, mais qui concourent plutôt à perturber la mesure. Enfin, la mesure de l'audiométrie ne permet pas la détection de pertes auditives cachées : d'origine périphérique (synaptopathie cochléaire : Kujawa et Liberman, 2015 ; Liberman et al., 2016) ou centrale (Pryce et al., 2010 ; Zhao, 2000).

Pour ces raisons, dans la présente étude une première partie a été dédiée à une évaluation plus précise de l'état auditif des participants. L'évaluation de l'état auditif a consisté en une audiométrie tonale liminaire complétée par une évaluation de l'état des CCE (c.-à-d., produits de distorsion acoustique des otoémissions, PDA) et un test de compréhension de la parole dans

le bruit. En outre, il était explicitement indiqué aux participants que ces mesures n'avaient aucun lien avec leur dossier médical et donc ne pouvaient avoir aucune conséquence sur leur aptitude à servir dans les armées.

Sur la base des résultats de la deuxième étude, l'intérêt de recherche a été porté sur l'effet de la séparation spatiale des locuteurs sur l'intelligibilité chez les militaires, en intégrant l'étude des composantes subjectives de l'effort d'écoute. Plus précisément, la réalisation de la tâche était associée aux évaluations subjectives des participants concernant leur effort d'écoute ressenti et la motivation investie afin de réussir la tâche. De plus, compte tenu de l'importance de prendre en compte chez les militaires l'état du système auditif, les effets de la séparation spatiale des locuteurs sur l'intelligibilité et sur l'effort d'écoute ressenti ont aussi été étudiés en tenant compte de leur état auditif de manière plus précise.

Les hypothèses de la troisième étude étaient donc les suivantes : a) une séparation spatiale de 20° serait probablement insuffisante pour obtenir une intelligibilité optimale dans cette population à l'audition altérée (Xia et al., 2015) ; b) une augmentation de la séparation spatiale de 20° à 60° devrait améliorer l'intelligibilité avec une réduction l'effort d'écoute ressenti et une plus grande motivation à réaliser la tâche. La question posée était de savoir, d'une part, si la séparation spatiale des locuteurs améliorait l'intelligibilité dans cette population particulière, ce qui pouvait être une piste pour l'élaboration de futurs systèmes de communication, et d'autre part quels étaient les effets de la séparation spatiale sur l'effort d'écoute, et enfin quels étaient les effets de l'état auditif sur les variations de l'intelligibilité et de l'effort ressenti.

L'objectif étant, à terme, de proposer des solutions pour améliorer l'intelligibilité tout en réduisant l'effort d'écoute ressenti pour une population travaillant dans des situations à risque où une charge cognitive élevée peut compromettre la sécurité.

II. Matériels et méthode

II.1. Participants

Trente-cinq volontaires ont participé à cette expérience (7 femmes, $M=38.19\pm 7.81$). Les participants étaient des militaires de carrière qui pouvaient avoir été exposés à des bruits intenses.

L'étude a été approuvée par le Comité de Protection des Personnes Sud-Ouest et Outre-Mer II, IDRCB 2017-A00859-44). Avant de commencer l'étude, les participants lisaient un document d'information sur l'expérience et leur consentement écrit était récolté. Ils étaient indemnisés à hauteur de 20 euros. L'étude s'est déroulée sur le site de la Section Technique de l'Armée de Terre (STAT) à Versailles et sur le site de l'Institut de Recherche Biomédicale des Armées (IRBA) à Brétigny sur Orge. Tous les participants étaient aptes au service dans leur spécialité.

II.2. Plan expérimental

La tâche d'intelligibilité utilisée a été adaptée à partir de la deuxième expérience (c.-à-d., CRM corpus) ; (voir II.2. Tâche d'intelligibilité dans le Chapitre II. Deuxième étude pour informations détaillées). La voix cible restait celle avec l'identifiant (call sign) Baron. Toutes les voix utilisées étaient des voix masculines.

Avant de commencer l'expérience, une courte séance d'entraînement était réalisée avec 10 essais comportant uniquement la voix cible (c.-à-d., *Ready Baron go to Color Number now*). Cette séance d'entraînement permettait de vérifier la bonne compréhension de la tâche et elle constituait aussi un critère d'exclusion pour les participants (trois erreurs dans le report de la couleur ou du chiffre entraînaient l'exclusion du participant). Ensuite, un test de compréhension de la voix cible dans le bruit était réalisé. Ce test nous permettait d'avoir une mesure supplémentaire de l'état de l'audition du participant et pouvait être considéré comme une audiométrie vocale dans le bruit. En détail, la voix cible était placée à -9 dB par rapport à un bruit présenté (correspondant à la condition CRM.Bruit). Ce test était suivi par une condition de contrôle consistant dans la présentation de trois voix co-localisées parmi lesquels la voix cible était présentée. Trois locuteurs étaient présentés à 0° droit devant l'auditeur ; parmi ces trois locuteurs était présente la voix cible (condition CRM.COL). Le test de compréhension de la voix cible dans le bruit et la condition de contrôle co-localisée étaient suivis par une courte séance d'entraînement préparatoire au recueil effectif des données. Elle était constituée de trois conditions de séparation spatiale des locuteurs (égales à la deuxième étude voir II.2. Tâche d'intelligibilité). La séance d'entraînement proposait 2 essais pour chacune des trois conditions avec un feedback auditif à la fin de chaque essai indiquant l'exactitude de la réponse donnée. À travers les trois conditions, la voix cible était toujours placée à 0° droit devant l'auditeur. Dans la première condition la voix cible était présentée avec deux voix masques respectivement à +60° et -60° (c.-à-d., CRM.60). Dans la deuxième condition, la voix cible était présentée avec

deux voix masques à +20° et -20° (c.-à-d., CRM.20). Dans la troisième condition spatialisée, la voix cible était présentée avec toutes les autres voix masques à +20°, -20°, +60° et -60° (c.-à-d., CRM.2060). Toutes les voix ont été présentées au même niveau d'intensité sonore (rapport signal/bruit égal à 0 dB).

Le test de compréhension de la voix cible dans le bruit et la condition de contrôle co-localisée étaient présentés avec deux blocs homogènes de 32 essais chacun (c.-à-d., CRM.Bruit : 32 essais de la voix cible dans le bruit ; CRM.COL: 32 essais de la voix cible et de deux voix masques co-localisées). L'ordre de présentation de ceux-ci n'était pas randomisé.

Les trois conditions expérimentales CRM.60, CRM.20, CRM.2060 étaient toujours constituées par 32 essais chacune, mais elles étaient divisées en 6 mini-blocs de 16 essais.

À la fin de chaque bloc d'essai (c.-à-d., CRM.Bruit, CRM.COL, CRM.20, CRM.60, CRM.2060), le participant évaluait son effort d'écoute sur une échelle d'effort d'écoute (c.-à-d., ESCU) et sa motivation sur une échelle de motivation adaptée depuis celle de Kramer et al., 2016.

II.3. Mesures subjectives

II.3.1. Questionnaire démographique

Le questionnaire comprenait plusieurs questions sur les antécédents auditifs du participant. Les questions concernaient le fait d'avoir reçu un diagnostic de perte auditive, le fait d'avoir ressenti des troubles auditifs après une exposition à un bruit intense, le fait d'avoir été exposé au bruit lors d'activités professionnelles (type de travail effectué et pendant combien de temps) et le fait d'avoir été exposé au bruit lors d'activités de loisir (loisirs, musique, discothèques). Pour une lecture détaillée, voir le document ANNEXE 2. FICHE DE RENSEIGNEMENT.

II.3.2. Effort Scaling Categorical Unit - ESCU

L'échelle pour évaluer l'effort d'écoute a été traduite en français à partir de l'Effort Scaling Categorical Unit (ESCU) qui est une échelle catégorielle d'évaluation de l'effort d'écoute ressenti pour réaliser une tâche auditive (Rennies et al., 2019 ; Rennies & Kidd, 2018). L'échelle que nous avons utilisée était composée de 13 catégories dont sept avec les labels

suyvants : *pas d'effort* à faire pour répondre (1 ESCU), *très peu d'effort* (3 ESCU), *peu d'effort* (5 ESCU), *effort modéré* (7 ESCU), *effort considérable* (9 ESCU), *beaucoup d'effort* (11 ESCU), *effort extrême* (13 ESCU) et *que du bruit* (14 ESCU). Six catégories intermédiaires s'ajoutaient entre chacune des catégories labellisées (Figure 32).

Une échelle ESCU format papier était donnée au participant après chaque bloc par l'expérimentateur. Le participant déclarait son choix à haute voix après chaque condition (c.-à-d., CRM.Bruit, CRM.COL, CRM.20, CRM.60, CRM.2060). L'expérimentateur notait ensuite la réponse donnée par le participant.



Figure 32 : Échelle ESCU à 13 catégories (traduite d'après Rennies et al., 2019). En haut à gauche, *Code sujet* représente l'identifiant d'un participant ; en haut à droite *Run* indique la condition pour laquelle l'échelle est administrée (dans l'exemple Col : co-localisée).

II.3.3. Motivation

L'échelle de la motivation a été adaptée et traduite en français à partir de l'échelle de motivation de Kramer et al., 2016. L'échelle de motivation est une échelle à dix points, allant de '*Je n'ai jamais renoncé à écouter*' (1) à '*Pour toutes les questions, j'ai abandonné*' (10).

Entre 1 et 10, 8 points étaient présentés sans aucun label (Figure 33). Une échelle de la motivation format papier était donnée au participant par l'expérimentateur. Le participant déclarait son choix à haute voix après chaque condition (c.-à-d., CRM.Bruit, CRM.COL, CRM.20, CRM.60, CRM.2060). L'expérimentateur notait ensuite la réponse donnée par le participant.

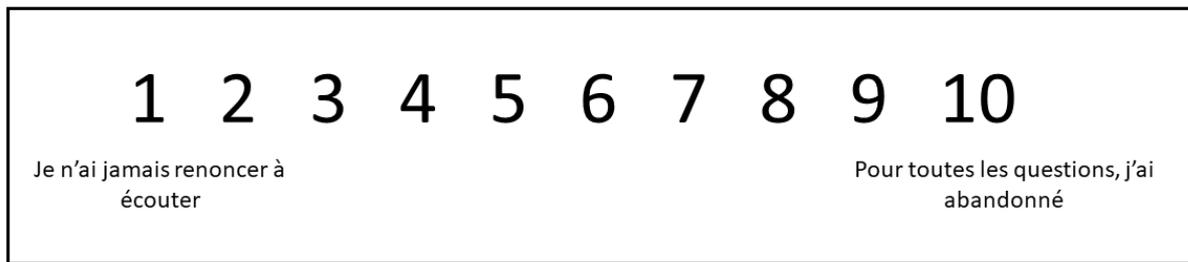


Figure 33 : Échelle de motivation adaptée depuis Kramer et al., 2016. Échelle à dix points, allant de ‘*Je n’ai jamais renoncé à écouter*’ (1) à ‘*Pour toutes les questions, j’ai abandonné*’ (10).

II.3.4. Speech, Spatial and Qualities Hearing

Afin de recueillir des informations sur la capacité auditive des participants, le Speech, Spatial and Qualities of Hearing Scale (SSQ) a également été administré aux participants. Il s’agit d’une version française adaptée par Moulin et al., 2015 depuis le Speech, Spatial and Qualities of Hearing Scale (Gatehouse & Noble, 2004). Le questionnaire est composé de trois parties : audition de la parole, audition spatiale et qualité de l’audition. Chaque partie regroupe 15 questions concernant les capacités auditives des participants dans le cadre de situations quotidiennes diverses. Ces quinze questions sont jugées subjectivement par le participant sur une échelle de 0 à 10. L’exemple montré dans la Figure 34 reporte une question de la première partie. Cette question concerne la capacité pour un auditeur de suivre la conversation avec un locuteur dans une pièce avec téléviseur allumé. La réponse correspondant au chiffre 0 indique l’impossibilité de faire cette tâche (c.-à-d., ‘*non, pas de tout*’), tandis que la réponse correspondant au chiffre 10 indique une capacité complète à réaliser la tâche (c.-à-d., ‘*oui, parfaitement*’). Si l’auditeur, dans ce cas, se sent capable de suivre environ la moitié de la conversation dans une telle situation, il placera sa réponse vers le milieu de l’échelle. Toutes les questions sont pensées comme pertinentes au regard de la vie quotidienne de l’auditeur. Si l’auditeur estime qu’elles ne le sont pas, il coche la case ‘*non applicable*’. Dans ce cas l’auditeur est invité par écrit à expliquer en quelques mots la raison pour laquelle la question n’est pas pertinente (Figure 34).

1ère partie : Audition de la parole

<p>1. Vous discutez avec une autre personne dans une pièce dans laquelle un téléviseur est allumé. Pouvez-vous suivre les propos de votre interlocuteur sans baisser le son du téléviseur ?</p>	<p style="text-align: center;"><i>Non, pas du tout</i> <i>Oui, parfaitement</i></p> <p style="text-align: center;"> ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- </p> <p style="text-align: center;">0 1 2 3 4 5 6 7 8 9 10 <input type="checkbox"/></p> <p style="text-align: right;"><i>Non applicable</i></p>
---	--

Figure 34 : Exemple d’une question posée au Speech Spatial and Qualities Hearing (SSQ) dans la version française d’après Moulin et al., 2015. Tout en haut à gauche, *1^{re} partie : Audition de la parole* indique une de trois parties dont il est composé (Audition de la parole, Audition spatiale et Qualité de l’audition). Ensuite, dans la case à gauche : la première question de cette première partie ; à droite : l’échelle d’évaluation subjective à 10 points (de 0 : *Non, pas du tout* à 10 : *Oui, parfaitement*) plus la case *Non applicable*.

II.4. Mesures audiologiques

II.4.1. Audiométrie

L’audiométrie tonale liminaire a été effectuée avec un audiomètre ELIOS (© 2020 Echodia) dans une cabine audiométrique.

L’audiométrie tonale liminaire a été réalisée pour les fréquences : 250, 500, 1000, 2000, 3000, 4000, 6000, 8000, et 12500 Hz (seuil très haute fréquence).

II.4.2. Otoémissions

L’évaluation du système auditif se poursuivait par les enregistrements des produits de distorsion acoustiques des otoémissions (PDA). Les PDA ont été mesurées sur les fréquences 1000, 1500, 2000, 3000, 4000, 5000 Hz à deux niveaux d’intensité 65/65 dB et 55/55 dB pour chaque oreille droite et gauche ($F1 = 1,2 * F2$). Le participant était invité à s’allonger sur une civière médicale et à se détendre complètement pour éviter tous mouvements ou bruits externes dérangeants en vue de l’enregistrement.

Les valeurs des produits de distorsion inférieures à -10 dB ou dont le rapport signal sur bruit était inférieur à 3 n’étaient pas incluses dans les analyses.

II.5. Caractéristiques des participants

Antécédents d’exposition au bruit

Le questionnaire démographique a permis d'avoir une évaluation des antécédents d'exposition au bruit dans notre population d'intérêt (Figure 35).

La plupart des participants (23 sur 31) a déclaré avoir travaillé dans des environnements bruyants. Seulement une petite partie (3 sur 31) a été diagnostiquée comme ayant eu un traumatisme sonore aigu.

Questionnaire démographique	Participants affectés (sur 31)
Diagnostic de perte d'audition ou trouble associé (acouphènes, otites répétées)	7
Diagnostic traumatisme sonore aigu	3
Douleur dans l'oreille suite à un bruit intense	8
Sifflement dans l'oreille suite à un bruit intense	13
Difficulté à entendre suite à un bruit intense	4
Travaillé dans un environnement bruyant	23
Loisirs bruyants (tir, course automobile)	10

Figure 35 : Questionnaire démographique sur les antécédents auditifs des participants.

Audiométrie

Les résultats de l'audiométrie tonale liminaire ont montré une large variabilité interindividuelle entre les participants avec en moyenne des pertes auditives pour les hautes fréquences (c.-à-d., 8000 et 12500 Hz) ; (Figure 36).

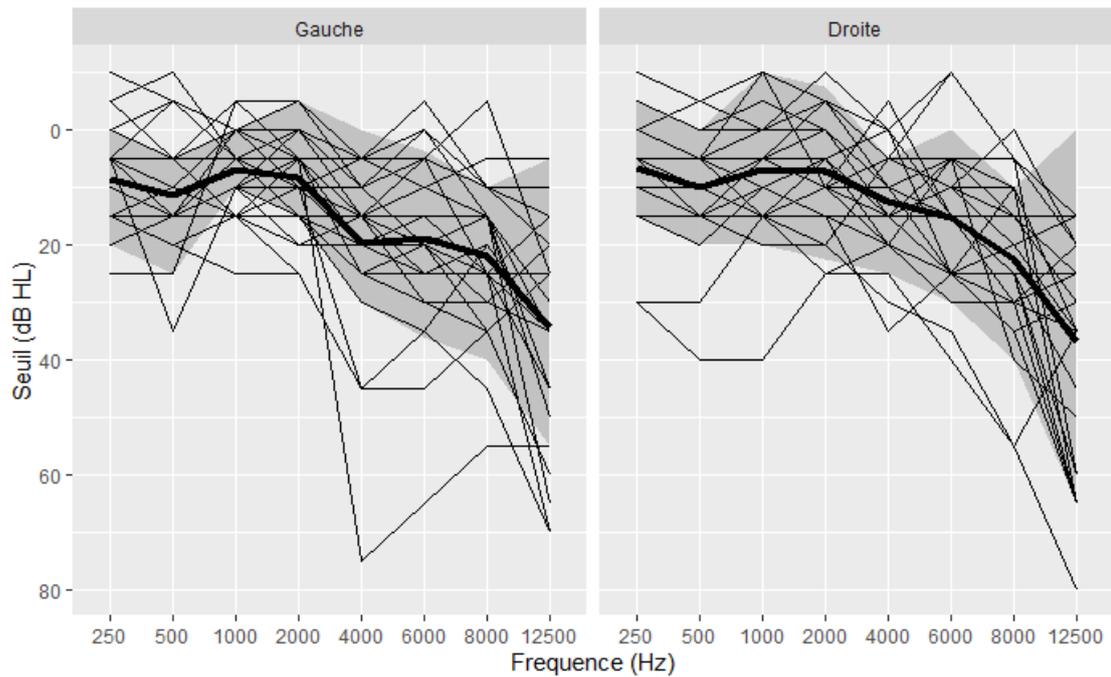


Figure 36 : Résultats de l'audiométrie des participants à l'étude. Les fréquences mesurées sont représentées en abscisse ; le seuil d'intensité sonore est représenté en décibel (dB HL) en ordonnée. Chaque ligne noire indique les résultats pour chaque participant, la ligne noire foncée indique la moyenne des réponses entre les participants. La région colorée en gris, l'intervalle interquartile.

III. Analyses

III.1. Tâche d'intelligibilité – CRM

Une réponse était considérée correcte quand la couleur et le chiffre étaient correctement reportés. Quatre participants ont été exclus des analyses à cause du haut nombre d'erreurs faites dans la condition avec la voix cible seule (> 3 erreurs sur les 10 chiffres et couleurs présentés). Le pourcentage de réponses correctes n'a pas été transformé, comme pour les deux études précédentes, en utilisant le Rationalized Arcsin Unit (RAU) ; (Studebaker, 1985), car pour certaines conditions les distributions étaient clairement éloignées d'une distribution binomiale (voir Figure 37 pour la symétrie de la distribution des données).

Pour analyser les différentes performances à travers les conditions expérimentales, des tests non paramétriques ont été réalisés : test de Friedman pour les cinq conditions (c.-à-d., conditions CRM.Bruit, CRM.COL, CRM.20, CRM.60, CRM.2060) comme facteur *within-subject*. Les

comparaisons deux à deux des conditions ont été réalisées avec des tests de Wilcoxon appariés avec correction FDR.

Pour évaluer si l'amélioration de l'intelligibilité avec la séparation spatiale des locuteurs est fonction de l'état auditif, nous avons considéré la différence d'intelligibilité entre la condition CRM.20 et la condition CRM.60 (c.-à-d., Différence_Intelligibilité) et les mesures de l'état auditif (c.-à-d., audiométries et PDA). Ensuite, des corrélations de Spearman ont été réalisées entre a) la Différence_Intelligibilité et les mesures de l'audiométrie et b) la Différence_Intelligibilité et les PDA. Les mêmes évaluations ont été faites pour l'effort d'écoute (Différence_Effort).

Toutes les fréquences de l'audiométrie ont été prises en compte ainsi que les dérivées des mesures audiométriques :

- Moyenne des seuils audiométriques des fréquences 500 à 4000 Hz pour chaque oreille (Moy_OG et Moy_OD) et à la meilleure oreille (Best_Ear) ;
- Moyenne des seuils audiométriques des fréquences 4000 à 12500 Hz pour chaque oreille (Moy_HF4_12_OD et Moy_HF4_12_OG) et à la meilleure oreille (Best_4_12) et à la moins bonne (Worse_4_12) ;
- Moyenne des seuils audiométriques des fréquences 500 à 2000 Hz pour chaque oreille (Moy_BF_5_2000_OD et Moy_BF_5_2000_OG) et à la meilleure oreille (Best_5_2K) et la moins bonne (Worse_5_2K) ;
- Moyenne des seuils audiométriques des fréquences 4000 et 8000 Hz pour chaque oreille (Moy4_8_OD et Moy4_8_OG) et à la meilleure oreille (Best_EarHF) ;
- Moyenne des seuils audiométriques des fréquences 8000 et 12500 Hz pour chaque oreille (MoyHF_OD et MoyHF_OG) ;
- Valeur absolue de la différence interaurale de la moyenne des seuils audiométriques des fréquences 500 à 4000 Hz (Asym) ;
- Valeur absolue de la différence interaurale de la moyenne des seuils audiométriques des fréquences 4000 à 12500 Hz (AsymHF).

Concernant les PDA, nous avons retiré les mesures à 55 dB du fait du nombre important de mesures non exploitables (amplitude < 10 dB) et bruitées (SNR < 3).

Les analyses statistiques ont été réalisées avec le logiciel R (version 4.0.3), et les packages *rstatix* pour les tests statistiques, *Factoshiny* et *FactomineR* pour l'ACP, *rpart* et *partykit* pour

l'arbre décisionnel, *ggpubr* pour les graphes en nuages de points avec corrélations de Spearman associées.

IV. Résultats

IV.1. Effet de l'écartement des voix masques de 20° à 60° sur l'intelligibilité

La séparation spatiale des locuteurs améliorait l'intelligibilité de la voix cible. L'intelligibilité variait significativement entre toutes les conditions expérimentales : test de Friedman ($\chi^2=45,68$; $p<0,001$) ; (Figure 37).

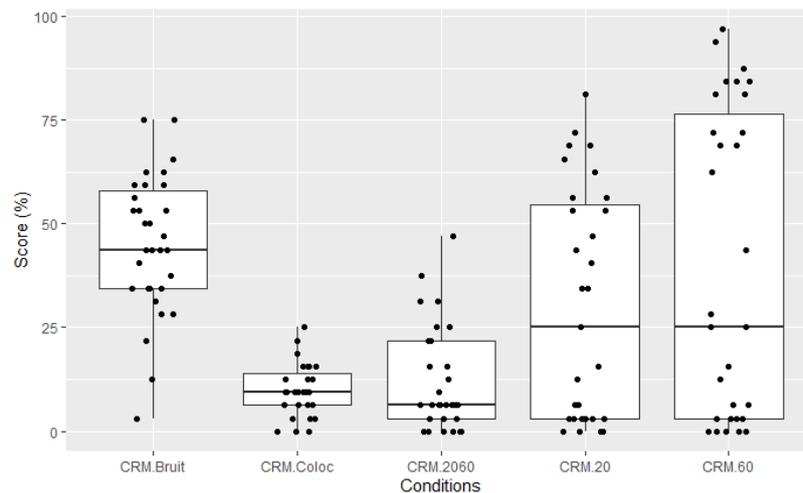


Figure 37 : Pourcentage de bonnes réponses en fonction des conditions (CRM.Bruit, CRM.Coloc, CRM.2060, CRM.20, CRM.60). Les boxplots montrent la médiane (barre horizontale) et l'écart interquartile (cases). Les moustaches vont de la valeur plus basse à la valeur plus élevée observées dans une limite de 1,5 fois l'écart interquartile. Les points indiquent les points de données individuels.

L'augmentation de la séparation spatiale des locuteurs de 20° à 60° pouvait augmenter significativement l'intelligibilité des participants ($V=44$; $p<0,001$) ; (Figure 38).

La médiane restait cependant semblable dans les deux conditions, probablement du fait d'un nombre important d'auditeurs (environ la moitié) dont l'intelligibilité demeurait faible, voir nulle, entre les deux conditions.

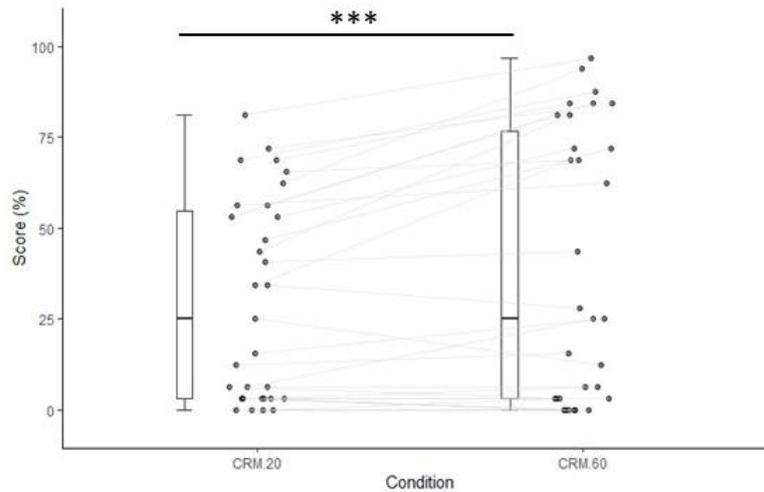


Figure 38 : Pourcentage de bonnes réponses en fonction de la condition (CRM.20 et CRM.60). Les boxplots montrent la médiane (barre horizontale) et l'écart interquartile (cases). Les moustaches vont de la valeur plus basse à la valeur plus élevée observées dans une limite de 1,5 fois l'écart interquartile. Les points indiquent les points de données individuels. Les points de données individuels correspondant au même auditeur dans différentes conditions sont reliés par des lignes grises.

Une large variabilité interindividuelle a été mise en évidence dans les deux conditions. Dans la condition CRM.60 deux groupes de participants semblaient se distinguer : un groupe à *haute intelligibilité* et un groupe à *basse intelligibilité*. Le groupe à haute intelligibilité dans la condition CRM.60 était constitué d'auditeurs dont la performance était la plus élevée aussi dans la condition CRM.20 ($R=0,96$; $p<0,001$) ; (Figure 39). Ce sont ces derniers dont la performance s'est améliorée le plus avec l'écartement des voix masques de 20° à 60° (Difference_Intelligibilité), ($R=0,61$; $p<0,001$) ; (Figure 40).

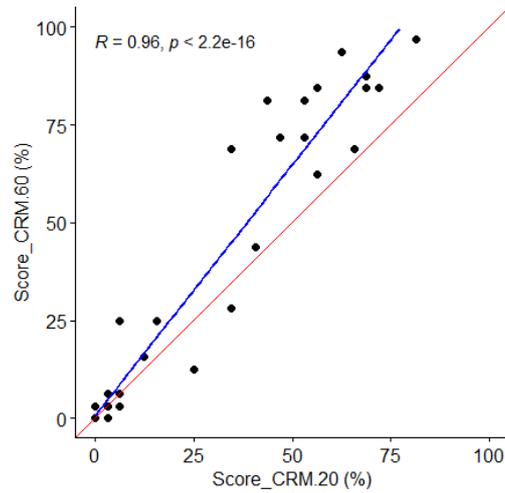


Figure 39 : Pourcentage de bonnes réponses dans la condition CRM.60 en fonction du pourcentage de bonnes réponses dans la condition CRM.20, chaque point représente un participant. La droite bleue représente la droite de régression. Les points correspondant aux participants dans les scores sont égaux entre les deux conditions sont sur la droite rouge. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.

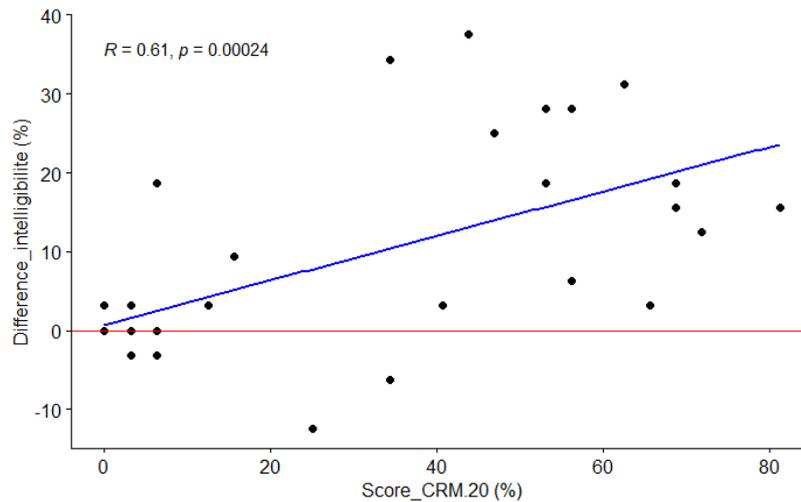


Figure 40 : Variation de l'intelligibilité entre 20° et 60° (Différence_Intelligibilité) en fonction du pourcentage de bonnes réponses dans la condition CRM.20. Chaque point représente un participant. La droite bleue représente la droite de régression. Les points correspondant aux participants dans les scores sont égaux entre les deux conditions sont sur la droite rouge. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.

IV.2. Effet de l'écartement des voix masques de 20° à 60° sur l'effort d'écoute ressenti

La séparation spatiale des locuteurs a conduit à une réduction de l'effort d'écoute ressenti. L'effort d'écoute ressenti variait significativement entre les conditions expérimentales : test de Friedman ($\chi^2=30,17$; $p<0,001$) ; (Figure 41).

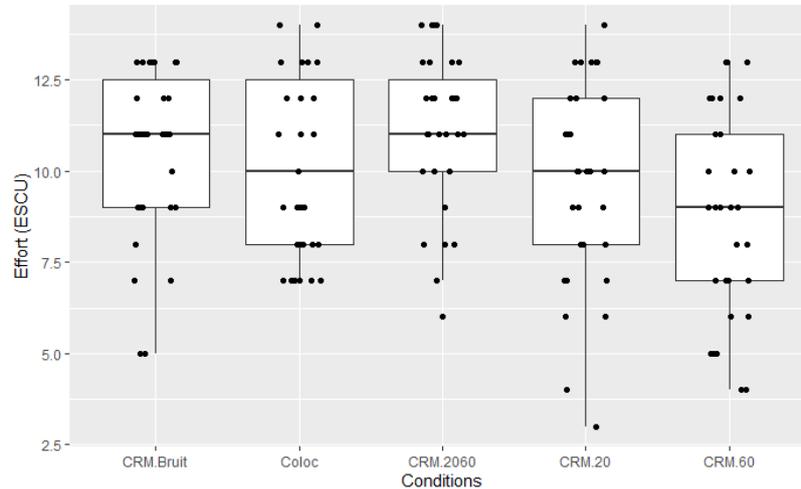


Figure 41 : Score de l'échelle de l'effort d'écoute ressenti (ESCU) en fonction des conditions (CRM.Bruit, CRM.Coloc, CRM.2060, CRM.20, CRM.60). Les boxplots montrent la médiane (barre horizontale) et l'écart interquartile (cases). Les moustaches vont de la valeur plus basse à la valeur plus élevée observées dans une limite de 1,5 fois l'écart interquartile. Les points indiquent les points de données individuels.

L'écartement des voix masques de 20° à 60° a eu également un effet sur l'effort d'écoute ressenti en conduisant à une réduction de ce dernier ($V=266,5$; $p<0,02$) ; (Figure 42).

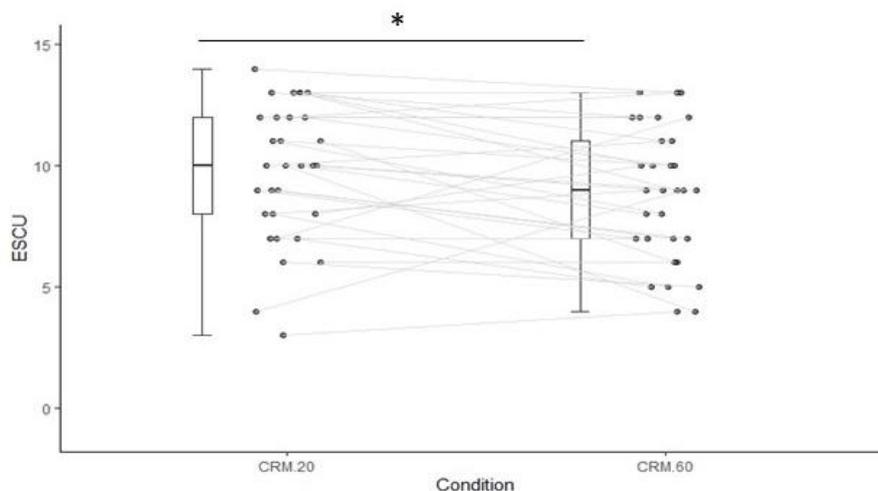


Figure 42 : Score d'évaluation à l'échelle ESCU en fonction de la condition (CRM.20 et CRM.60). Les boxplots montrent la médiane (barre horizontale) et l'écart interquartile (cases).

Les moustaches vont de la valeur plus basse à la valeur plus élevée observées dans une limite de 1,5 fois l'écart interquartile. Les points indiquent les points de données individuels. Les points de données individuels correspondant au même auditeur dans différentes conditions sont reliés par des lignes grises.

La relation entre la mesure ESCU.20 et la variation de l'effort d'écoute ressenti entre 20° et 60° (Difference_Effort) n'est pas significative ($p=0,054$). La tendance serait une réduction de l'effort d'écoute ressenti plus importante chez les participants pour lesquels la mesure était la plus élevée à ESCU.20. L'écartement de 20° à 60° serait alors favorable aux participants faisant le plus d'effort à 20° de séparation spatiale (Figure 43).

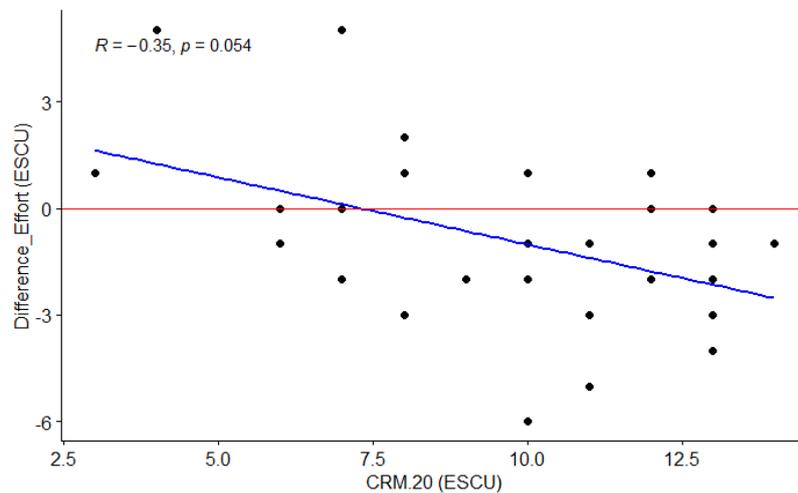


Figure 43 : Variation de la mesure de l'effort d'écoute ressenti entre 20° et 60° (Difference_Effort) en fonction de l'effort d'écoute ressenti à 20° (ESCU_20). Chaque point représente un participant. La droite bleue représente la droite de régression. Les points correspondant aux participants dans les scores sont égaux entre les deux conditions sont sur la droite rouge. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.

IV.3. Relation entre la variation de l'intelligibilité et la variation de l'effort

Les participants qui ont le plus amélioré leur intelligibilité avec l'écartement des voix masques de 20° à 60° (Différence_Intelligibilité) sont également ceux dont la mesure subjective de l'effort d'écoute ressenti a le plus diminué entre ces deux conditions (Différence_Effort), ($R=-0,57$; $p<0,001$) ; (Figure 44).

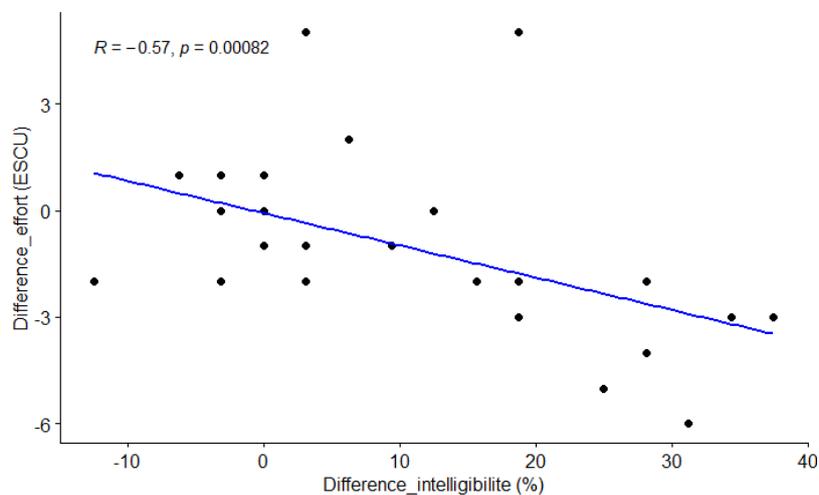


Figure 44 : Variation de la mesure de l'effort d'écoute ressenti entre 20° et 60° (Différence_Effort) en fonction de la variation de l'intelligibilité entre 20° et 60° (Différence_intelligibilité). Chaque point représente un participant. La droite bleue représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.

IV.4. Relation entre la différence d'intelligibilité liée à un écartement des voix masques de 20° à 60° et l'état auditif.

IV.4.1. Audiométries tonales liminaires classique et haute fréquence

Aucune corrélation significative n'a été observée entre la variation de l'intelligibilité entre 20° et 60° (Différence_Intelligibilité) et les seuils audiométriques mesurés à l'oreille gauche et à l'oreille droite, y compris les seuils mesurés à 12500 Hz ($p=0,09$ dans le meilleur des cas) ;

(Figure 45 et Figure 46). Aucune corrélation significative n'a été observée pour les mesures dérivées des seuils audiométriques ($p=0,11$ dans le meilleur des cas) ; (Figure 47).

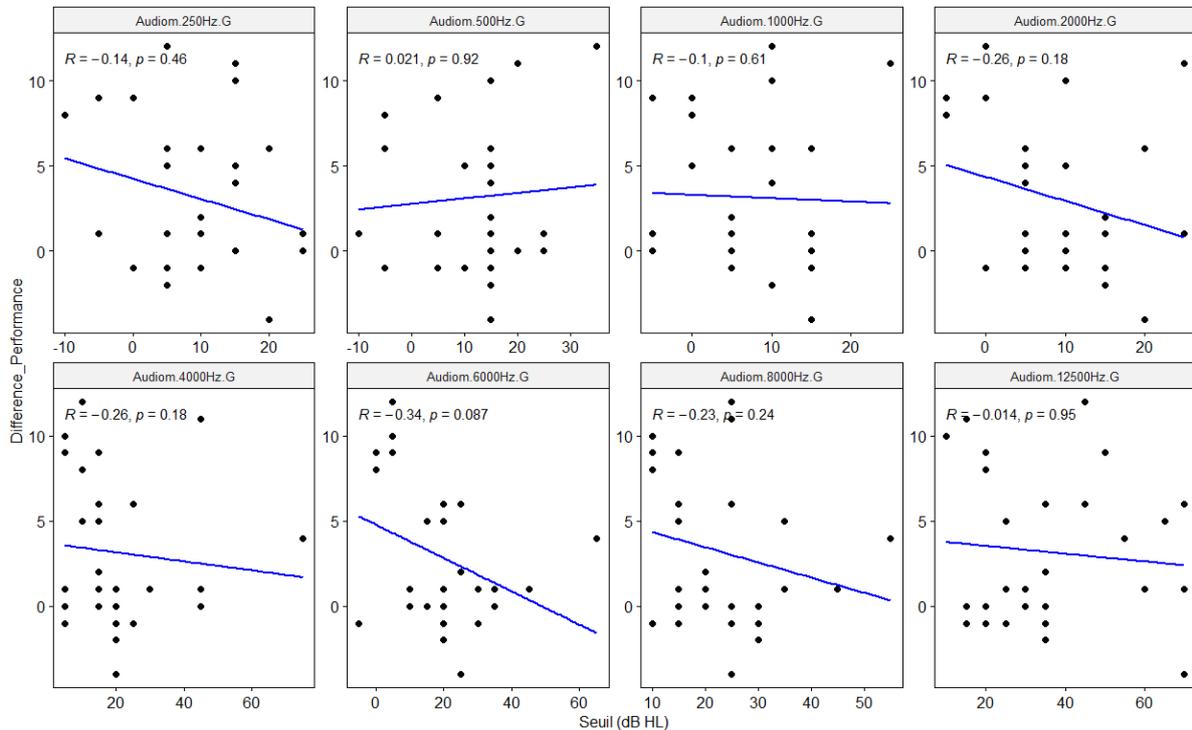


Figure 45 : Variation de l'intelligibilité entre 20° et 60° (Difference_Intelligibilité) en fonction des seuils audiométriques mesurés à l'oreille gauche. Chaque cadre représente une fréquence de mesure, où chaque point représente un participant. La droite bleue représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.

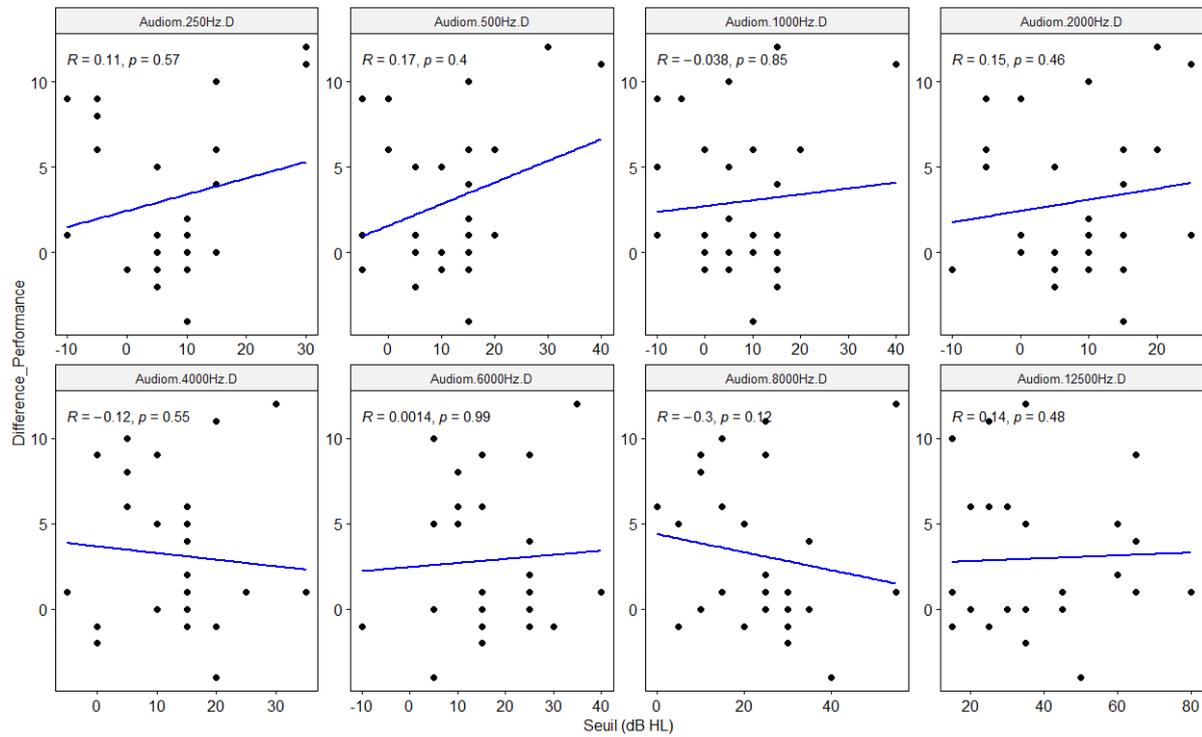


Figure 46 : Variabilité de l'intelligibilité entre 20° et 60° (Difference_Intelligibilité) en fonction des seuils audiométriques mesurés à l'oreille droite. Chaque cadre représente une fréquence de mesure, où chaque point représente un participant. La droite bleue représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.

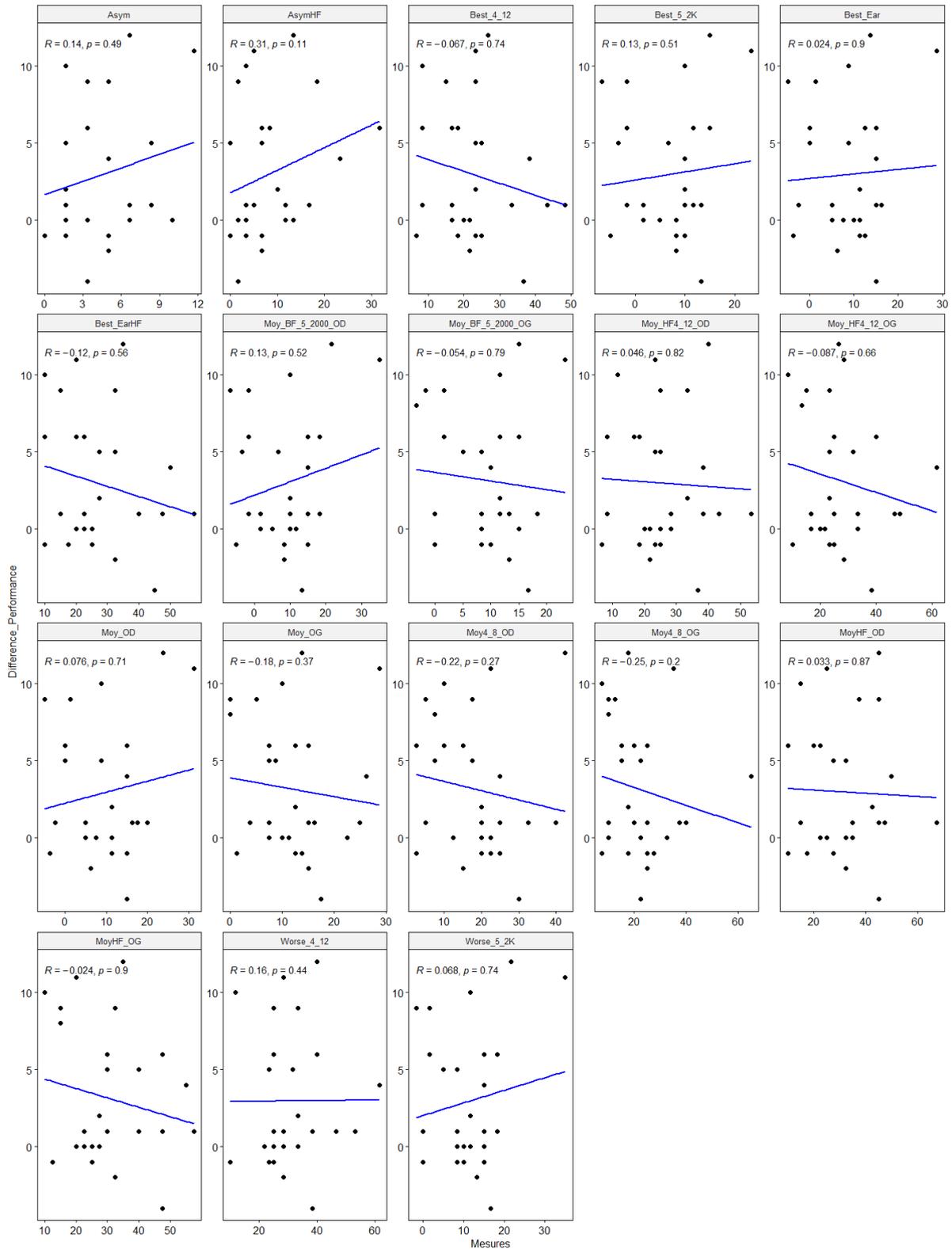


Figure 47 : Variation de l’intelligibilité entre 20° et 60° (Difference_Intelligibilité) en fonction des mesures dérivées audiométriques. Chaque cadre représente une fréquence de mesure, où chaque point représente un participant. La droite bleue représente une fréquence de mesure, où chaque point représente un participant. La droite bleue représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.

IV.4.2. Produits de distorsion acoustique (PDA)

Aucune corrélation significative n'est observée entre la variation de l'intelligibilité entre 20° et 60° (Différence_Intelligibilité) avec les PDA mesurés à l'oreille droite ($p=0,071$ dans le meilleur des cas) ; (Figure 48).

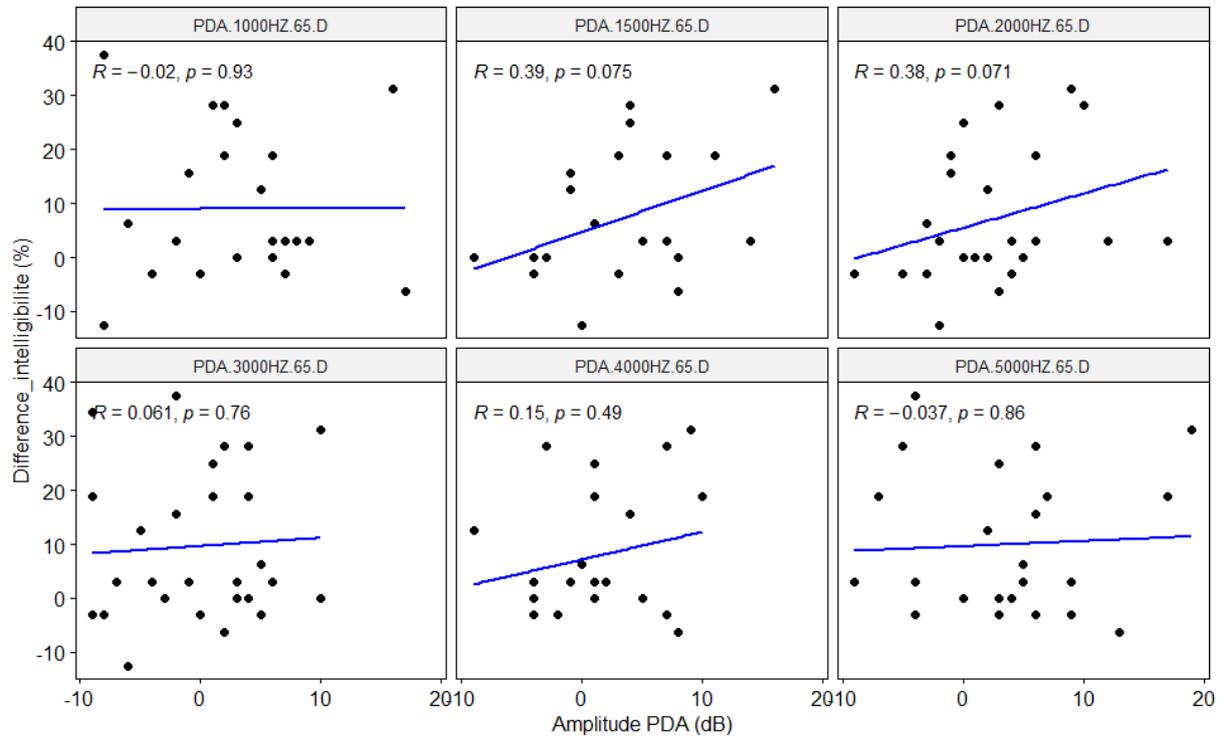


Figure 48 : Variation de l'intelligibilité entre 20° et 60° (Différence_Intelligibilité) en fonction des amplitudes des PDA à 65 dB de l'oreille droite. Chaque cadre représente une fréquence de mesure, où chaque point représente un participant. La droite bleue représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.

Pour les PDA mesurés à l'oreille gauche, une corrélation est observée entre la variation de l'intelligibilité entre 20° et 60° (Différence_Intelligibilité) et l'amplitude des PDA mesurés à 2000 Hz ($R=0,57$; $p<0,01$) ; (Figure 49).

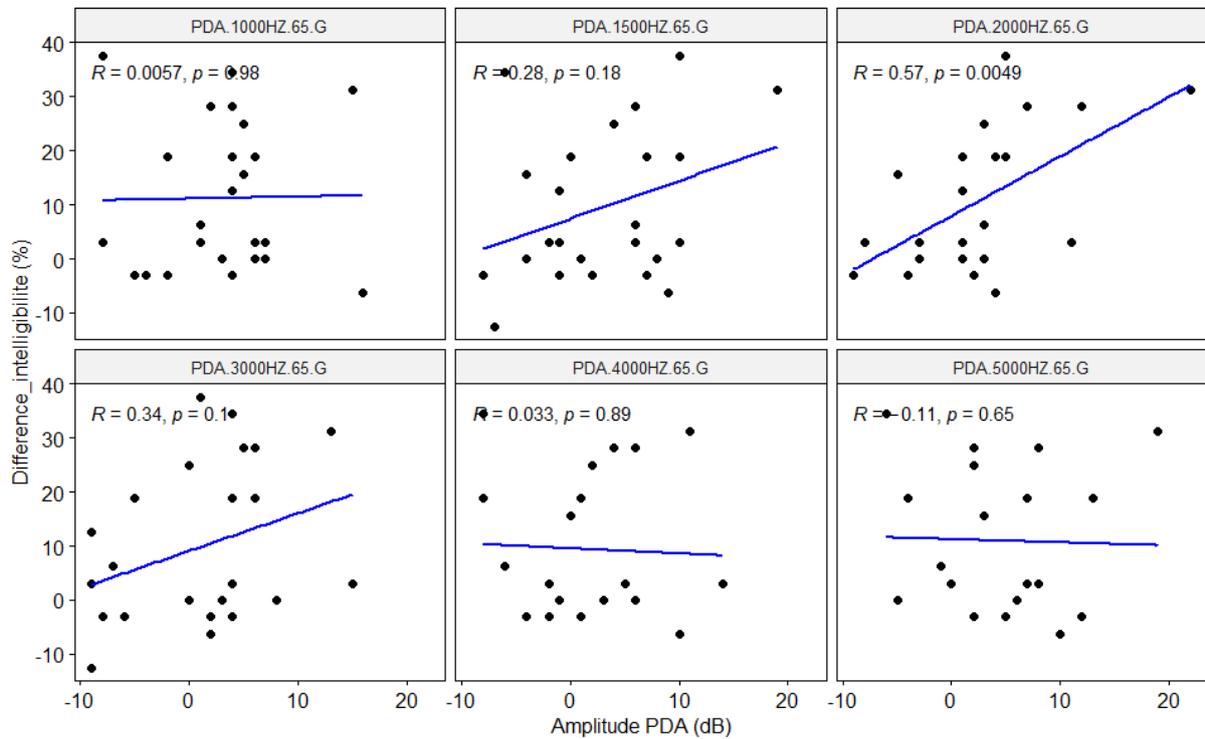


Figure 49 : Variation de l'intelligibilité entre 20° et 60° (Différence_intelligibilité) en fonction de l'amplitude des PDA pour les fréquences 1000 à 5000 Hz mesurés à l'oreille gauche. Chaque cadre représente une fréquence de mesure, où chaque point représente un participant. La droite bleue représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.

Une analyse en composante principale réalisée sur la variation de l'intelligibilité entre 20° et 60° (Différence_Intelligibilité) et les PDA mesurés à l'oreille gauche confirme la plus grande proximité de la différence d'intelligibilité avec les PDA mesurés à 1500, 2000 et 3000 Hz, comparativement aux autres fréquences (Figure 50).

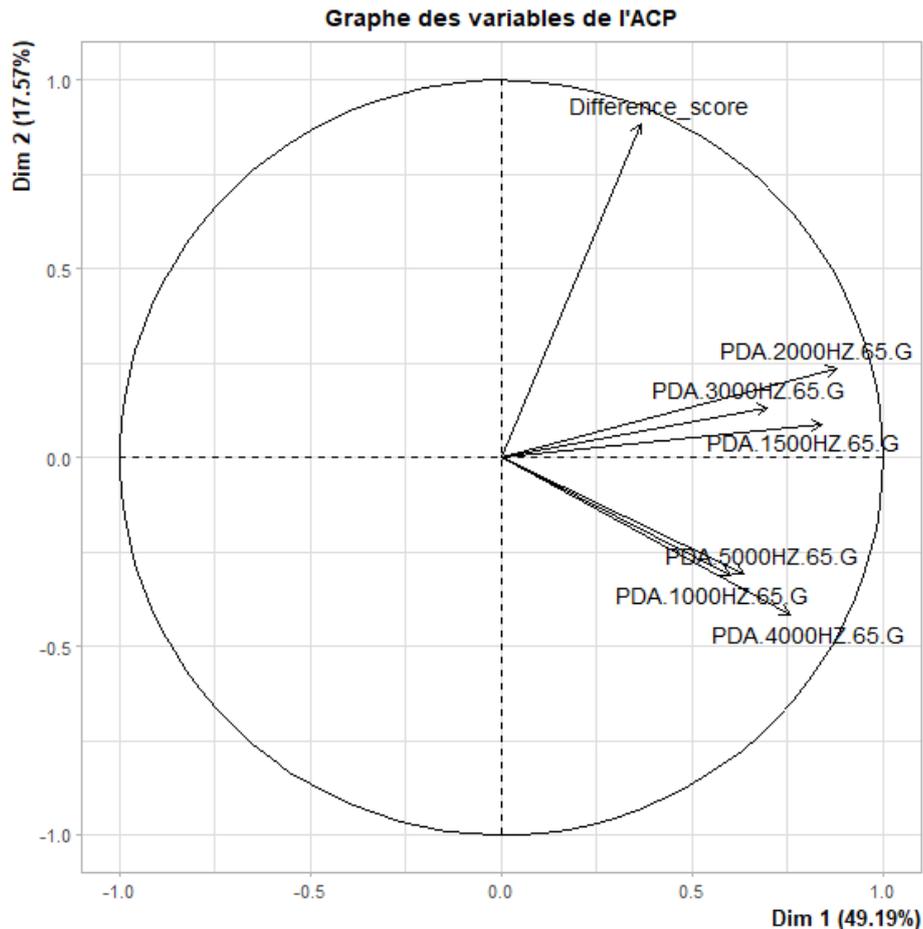


Figure 50 : Graphe des variables de l'analyse en composante principale. Les variables incluses sont la variation d'intelligibilité entre 20° et 60° (Différence_Intelligibilité) et l'amplitude des PDA mesurés à l'oreille gauche pour différentes fréquences. Les 2 premiers axes expriment 72.82 % de l'inertie totale du jeu de données ; ce qui peut être considéré comme satisfaisant. La projection des variables est également satisfaisante (extrémité du segment proche du cercle). Les PDA mesurés à 2000 et 3000 Hz semblent les plus proches de la différence d'intelligibilité tandis que les PDA mesurés à 1000, 4000 et 5000 Hz ne semblent pas être liés à cette variable.

IV.5. Relation entre la variation de l'effort d'écoute ressenti entre 20° à 60° et l'état auditif

IV.5.1. Audiométries tonales liminaires classique et haute fréquence

Une corrélation significative a été retrouvée entre la variation de l'effort d'écoute ressenti entre 20° à 60° et les seuils audiométriques mesurés à 6000 Hz ($R=0,39$; $p<0,05$) et 8000 Hz ($R=0,4$; $p<0,05$) à l'oreille gauche. Au plus le seuil audiométrique est bas au plus la réduction de l'effort est importante (Figure 51). Par conséquent, les participants avec un état auditif plus intègre étaient aussi les participants qui ressentait une réduction de l'effort entre la condition 20° et la condition 60°. En d'autres termes, ces participants avec un seuil audiométrique bas étaient les mêmes qui arrivaient à bénéficier le plus de l'écartement des voix masques. À noter, que nous n'avons pas trouvé de relation entre le seuil audiométrique mesuré à 12500 Hz et la différence d'effort ($p=0,9$).

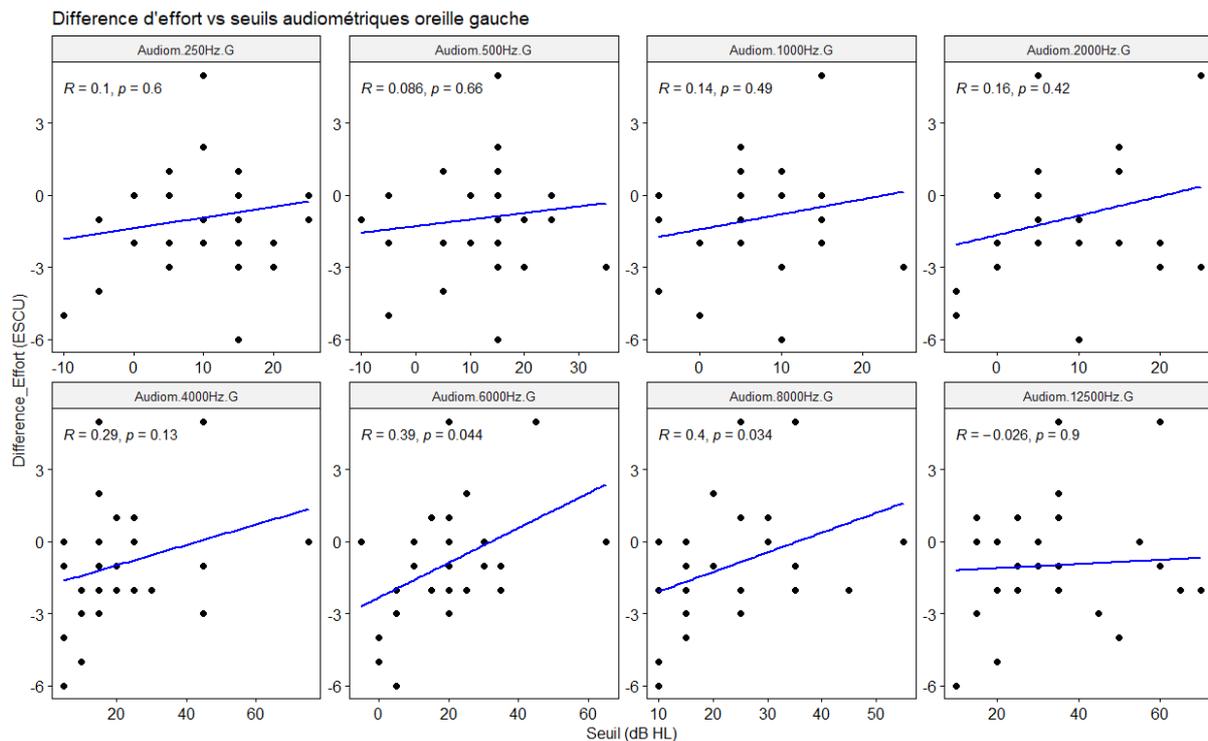


Figure 51 : Variation de l'effort d'écoute ressenti entre 20° à 60° (Différence_Effort) en fonction des seuils audiométriques mesurés à l'oreille gauche de 250 à 12500 Hz. Chaque cadre représente une fréquence de mesure, où chaque point représente un participant. La droite bleu représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche dans chaque cadre.

Aucune corrélation n'a été observée au niveau de l'oreille droite ($p=0,1$ dans le meilleur cas) ; (Figure 52) ni pour les mesures dérivées des seuils audiométriques ($p=0,063$) ; (Figure 53).

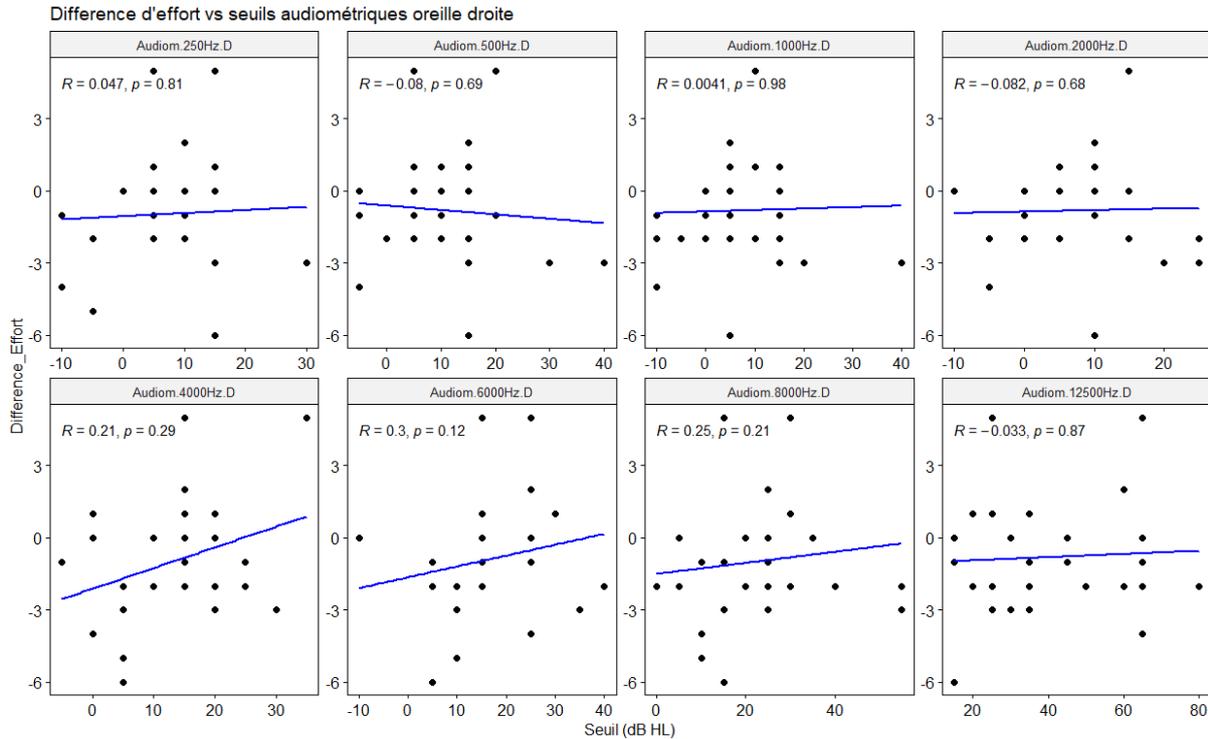


Figure 52 : Variation de l'effort d'écoute ressenti entre 20° et 60° (Différence_Effort) en fonction des seuils audiométriques mesurés à l'oreille droite. Chaque cadre représente une fréquence de mesure, où chaque point représente un participant. La droite bleu représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.

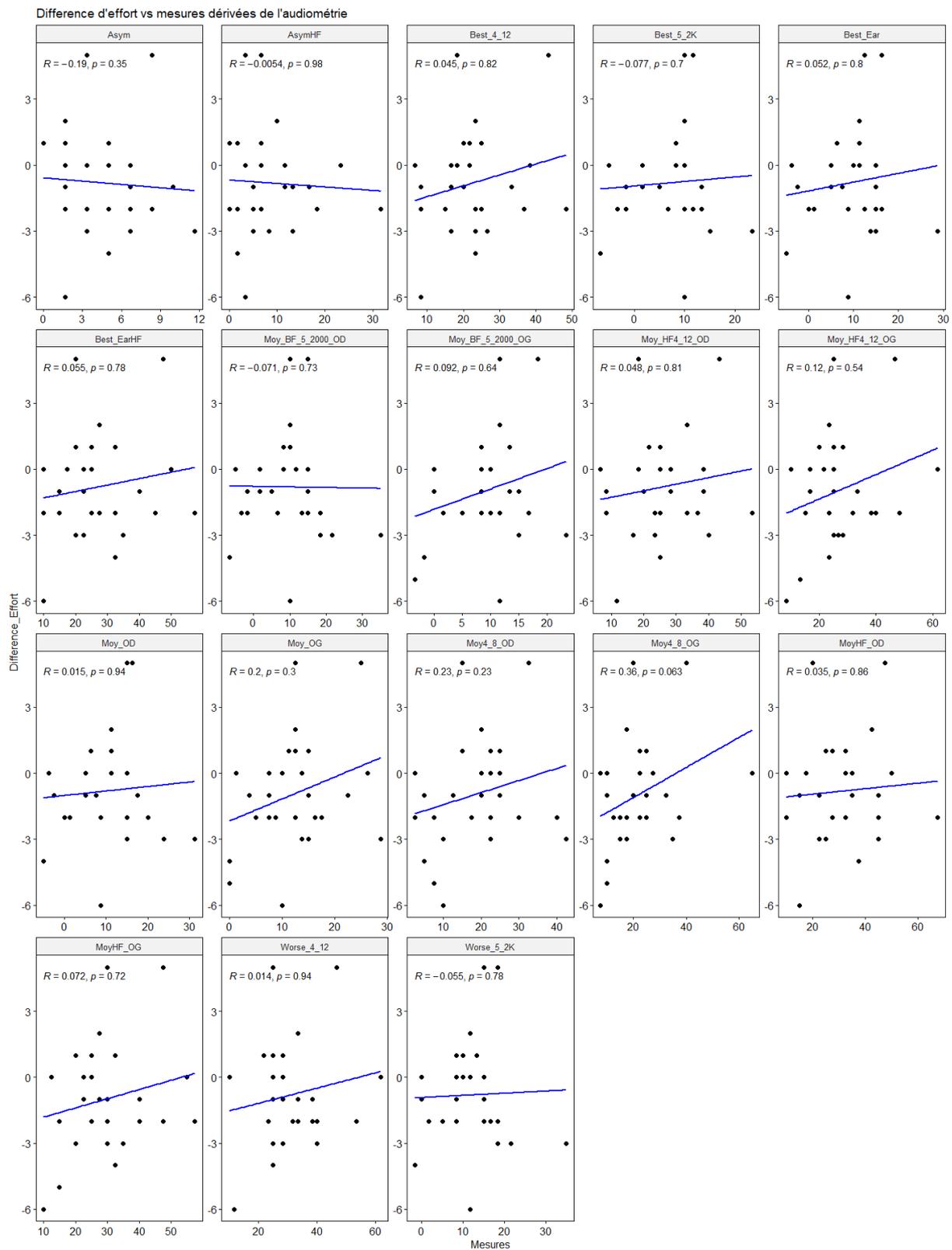


Figure 53 : Variation de l'effort d'écoute ressenti entre 20° et 60° (Différence_Effort) en fonction des mesures dérivées audiométriques. Chaque cadre représente une fréquence de mesure, où chaque point représente un participant. La droite bleue représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.

IV.5.2. Produits de distorsion acoustique (PDA)

Aucune corrélation significative n'a été observée entre la variation de l'effort d'écoute ressenti entre 20° à 60° et les mesures des PDA à 65 dB ni à droite ($p=0,2$ dans le meilleur cas) ; (Figure 54), ni à gauche ($p=0,1$) ; (Figure 55).

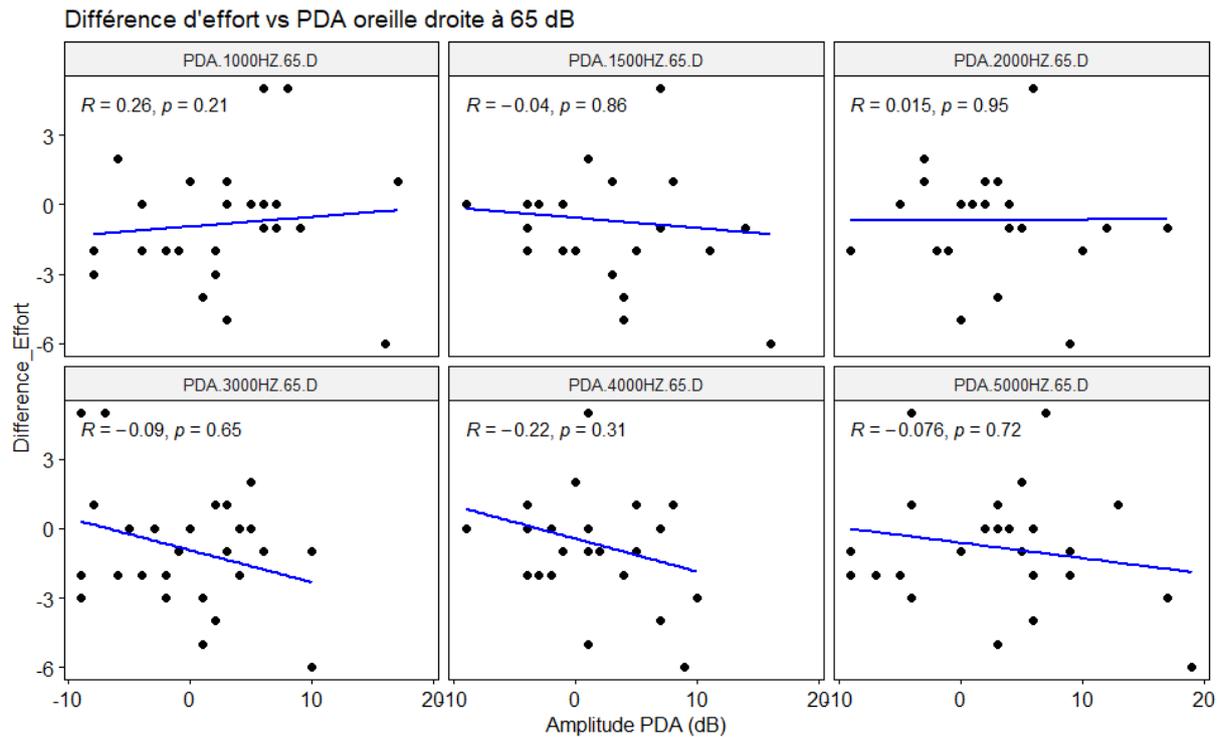


Figure 54 : Variation de l'effort d'écoute ressenti entre 20° et 60° (Différence_Effort) en fonction des amplitudes des PDA à 65 dB de l'oreille droite. Chaque cadre représente une fréquence de mesure, où chaque point représente un participant. La droite bleue représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.

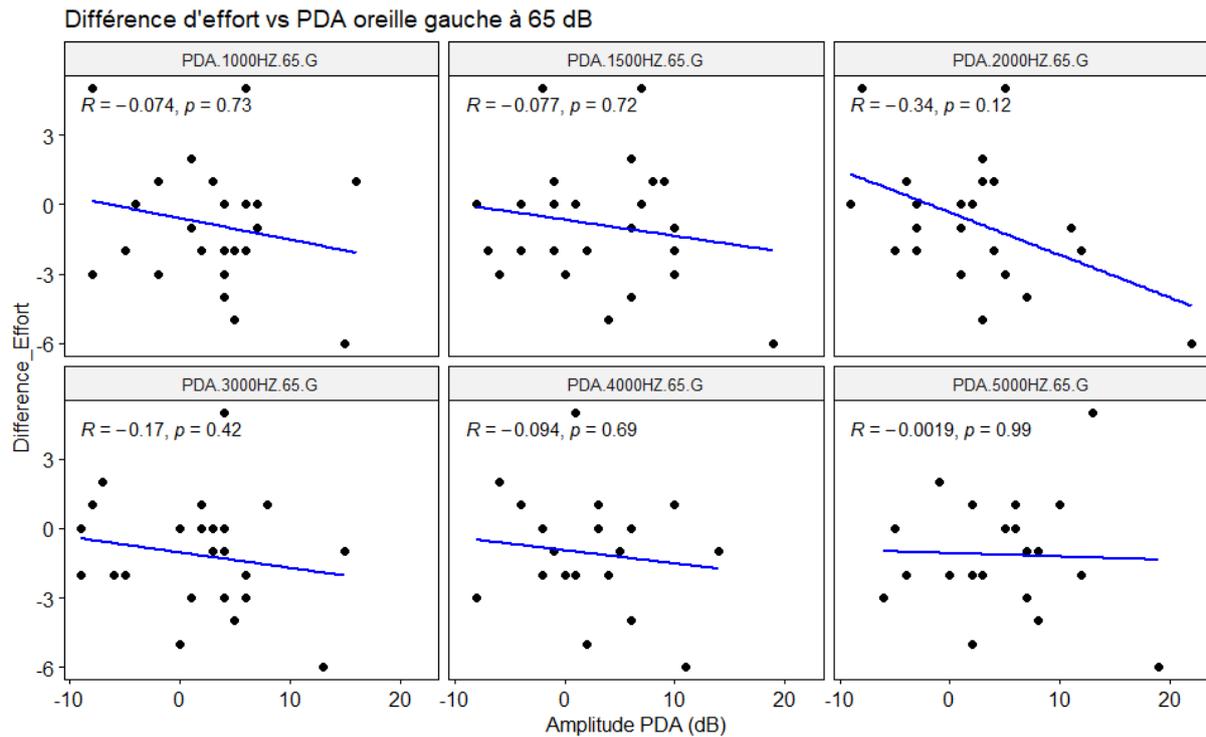


Figure 55 : Variation de l'effort d'écoute ressenti entre 20° et 60° (Différence_Effort) en fonction des amplitudes des PDA à 65 dB de l'oreille gauche. Chaque cadre représente une fréquence de mesure, où chaque point représente un participant. La droite bleue représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.

IV.6. Compréhension de la voix cible dans le bruit

Une corrélation positive a été retrouvée entre l'intelligibilité obtenue dans la condition CRM.60 et le test de compréhension de la voix cible dans le bruit (CRM.Bruit) ; ($R=0,53$; $p<0,01$) ; (Figure 56). Les participants qui réussissaient le mieux à comprendre la voix cible dans le bruit étaient ceux qui réussissaient le mieux à la comprendre la voix cible avec un écartement important entre les locuteurs.

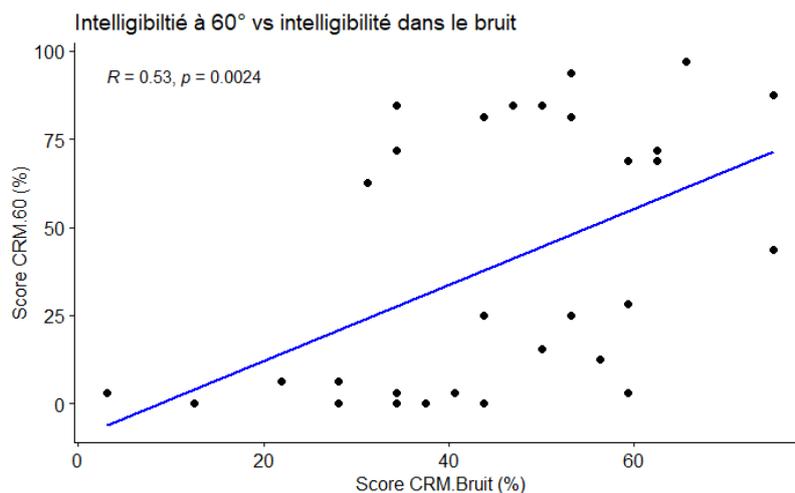


Figure 56 : Corrélation positive entre l’intelligibilité obtenue en condition CRM.Bruit (en abscisse) et l’intelligibilité obtenue dans la condition CRM.60 (en ordonnée). La droite bleue représente la droite de régression ; chaque point représente un participant. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.

De plus, une corrélation positive a été mise en évidence entre l’effort ressenti dans le test de compréhension dans le bruit (c.-à-d., BRUIT.ESCU) et dans la condition avec un plus grand écartement entre les locuteurs (c.-à-d., ESCU.60) ; ($R=0,57$; $p<0,001$). Les participants qui font le plus d’effort à comprendre la voix cible dans le bruit sont ceux qui font le plus d’effort à comprendre la voix cible dans la condition avec le plus grand écartement entre les locuteurs (Figure 57).

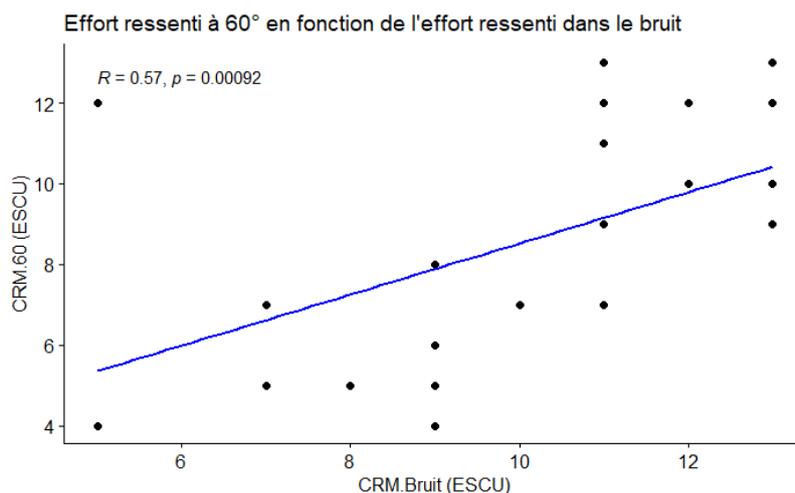


Figure 57 : Corrélation positive entre l’effort d’écoute ressenti dans le bruit (CRM.Bruit) et l’effort d’écoute ressenti dans la condition ESCU.60. La droite bleue représente la droite de régression ; chaque point représente un participant. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.

Il n'a pas été observé de corrélation significative entre le pourcentage de bonnes réponses dans le bruit et la variation de l'intelligibilité entre 20° à 60° ($R=0,3$; $p=0,09$) ; (Figure 58).

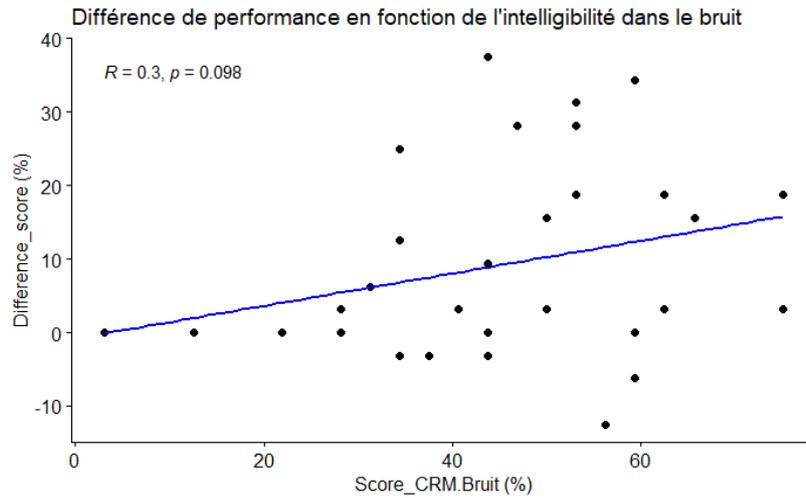


Figure 58 : Variation de l'intelligibilité entre 20° à 60° (Différence_intelligibilité) en fonction du pourcentage de bonnes réponses dans le bruit (CRM_Bruit). Chaque point représente un participant. La droite bleue représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.

Il n'a pas été observé de corrélation significative entre le pourcentage de bonnes réponses dans le bruit et la variation de l'effort d'écoute entre 20° et 60° ($p=0,16$) ; (Figure 59).

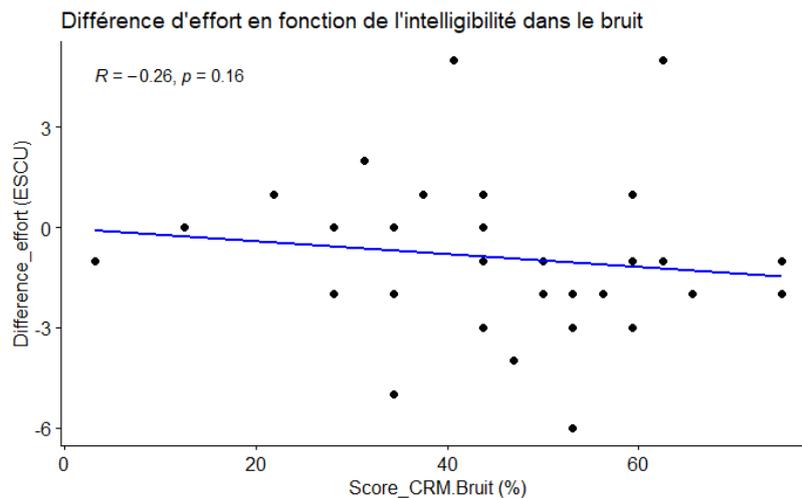


Figure 59 : La variation de l'effort d'écoute entre 20° et 60° (Différence_Effort) en fonction du pourcentage de bonnes réponses dans le bruit dans le bruit. Chaque point représente un participant. La droite bleue représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.

IV.7. Intelligibilité et motivation

Aucune variation significative de la motivation à travers l'ensemble des conditions testées n'a été trouvée (c.-à-d., conditions CRM.Bruit, CRM.COL, CRM.20, CRM.60, CRM.2060) ; (Figure 60).

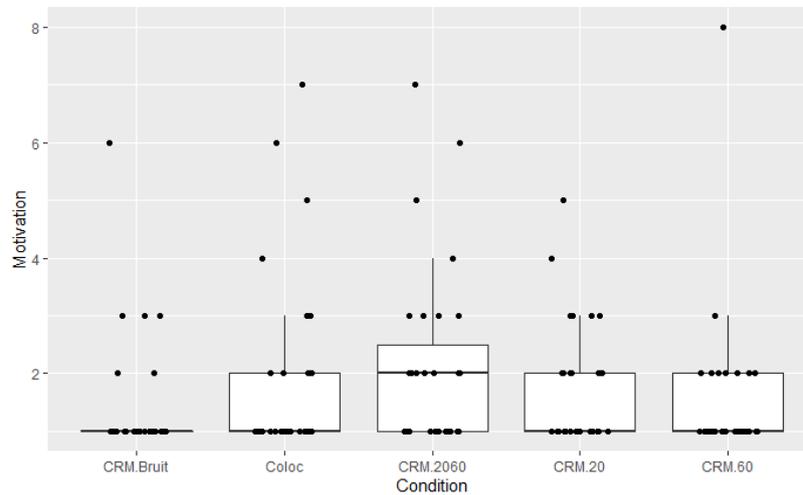


Figure 60 : Score de l'échelle de motivation en fonction des conditions (CRM.Bruit, CRM.Coloc, CRM.2060, CRM.20, CRM.60). Les boxplots montrent la médiane (barre horizontale) et l'écart interquartile (cases). Les moustaches vont de la valeur plus basse à la valeur plus élevée observées dans une limite de 1,5 fois l'écart interquartile. Les points indiquent les points de données individuels.

La meilleure intelligibilité atteinte dans la condition avec plus grand écartement des locuteurs (c.-à-d., CRM_60) n'était pas significativement corrélée à une plus grande motivation (c.-à-d., MOT_60) ; ($R=-0,13$; $p=0,47$) ; (Figure 61).

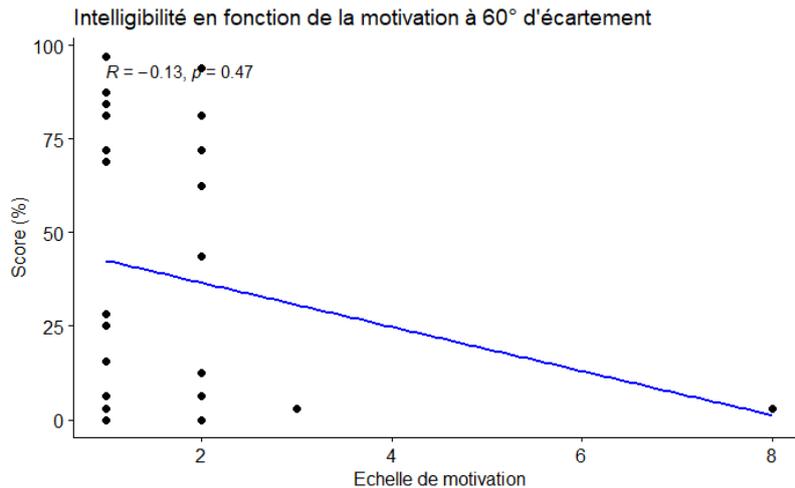


Figure 61 : Pourcentage de bonnes réponses dans la condition avec plus grand écartement des locuteurs (c.-à-d. CRM_60) en fonction de la motivation (c.-à-d., MOT_60). Chaque point représente un participant. La droite bleue représente la droite de régression. Les points correspondant aux participants dans les scores sont égaux entre les deux conditions sont sur la droite rouge. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.

IV.8. Relation entre la mesure de la motivation et les variables mesurées dans les différentes conditions expérimentales.

Aucune relation significative entre la mesure subjective de l'effort et la mesure subjective de la motivation n'a été mise en évidence quelle que soit la condition. À titre d'exemple, les figures Figure 62 et Figure 63 montrent la relation entre motivation et effort pour les conditions à 20° (c.-à-d. MOT_20 : $R=0,17$; $p=0,36$) et 60° (c.-à-d., MOT_60 : $R=0,19$; $p=0,3$) d'écartement. Ces figures illustrent également la faible variabilité interindividuelle : la plupart des participants évaluaient leur motivation entre 1 et 2 sur l'échelle proposée (pour rappel l'échelle était à 10 points dont seulement le premier et le dixième avaient un label : 1 = 'Je n'ai jamais renoncé à écouter' et 10 = 'Pour toutes les questions, j'ai abandonné').

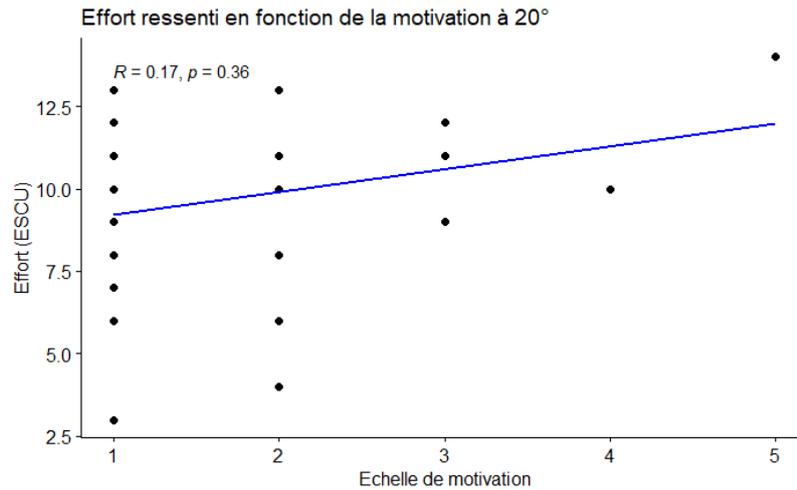


Figure 62 : Mesure de l'effort par ESCU en fonction de la mesure de la motivation dans la condition MOT_20. Chaque point représente un participant. La droite bleue représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.

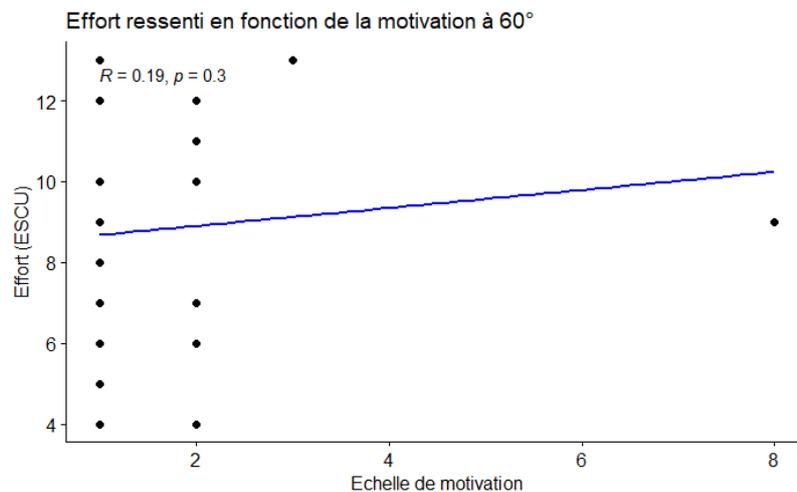


Figure 63 : Mesure de l'effort par ESCU en fonction de la mesure de la motivation par une adaptation de l'échelle de Kramer et al., 2016 dans la condition MOT_60. Chaque point représente un participant. La droite bleue représente la droite de régression. Le coefficient de corrélation de Spearman et sa significativité sont indiqués en haut à gauche.

IV.9. Relation entre l'âge et les variables mesurées dans les différentes conditions expérimentales.

Il n'a pas été retrouvé de corrélation significative entre l'âge des participants et la mesure des scores d'intelligibilité, de l'effort ressenti ou de la motivation dans les différentes conditions expérimentales.

IV.10. Relation entre les réponses au SSQ et les variables mesurées pertinentes.

Les corrélations entre les réponses au SSQ et les mesures jugées comme pertinentes à associer (par exemple items relatif à la compréhension de la parole dans la parole et scores dans les conditions expérimentales correspondantes) n'ont pas permis de mettre en évidence de corrélation significative.

IV.11. Relation entre la variabilité interindividuelle observée à 60° de séparation et les autres mesures.

La Figure 38 montre que dans la condition avec un écartement de 60° entre les voix masques et la voix cible, il existe une distribution des participants en deux groupes selon que leur performance soit supérieure à 60 % ou inférieure ou égale à 25 %, soit un écart de performance de 35 % entre les deux groupes. Seuls deux participants n'appartiennent pas à ces deux groupes.

Afin d'explorer cette variabilité interindividuelle, un arbre décisionnel a été utilisé de manière à pouvoir rechercher une relation entre l'appartenance à ces groupes et un ensemble de variables catégorielles et continues, que cette relation soit linéaire ou non. Les variables utilisées comme prédicteurs étaient les mesures de l'état auditif : seuils audiométriques de 250 à 12500 Hz, PDA, intelligibilité de la parole dans le bruit, l'effort ressenti et motivation pour toutes les conditions expérimentales, l'âge, les réponses au SSQ et les réponses au questionnaire sur les antécédents auditifs.

La variable qui a prédisait le mieux cette variabilité était l'effort ressenti dans le bruit. En d'autres termes, les personnes qui avaient ressenti un effort majeur dans le test de compréhension de la voix cible dans le bruit étaient celles qui avaient une faible intelligibilité en condition 60 (Figure 64) ; (Figure 65).

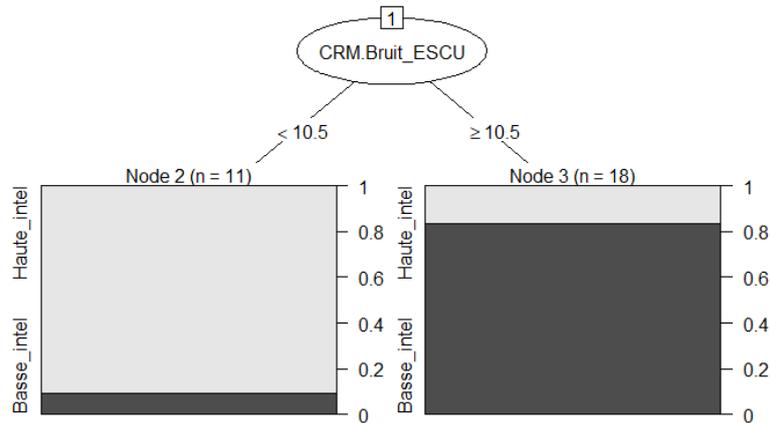


Figure 64 : Arbre de classification de la variable définissant les groupes dans la condition 60°. Les participants ayant une mesure de l’effort d’écoute ressenti inférieure à 10,5 étaient pour la plupart dans le groupe avec une haute intelligibilité.

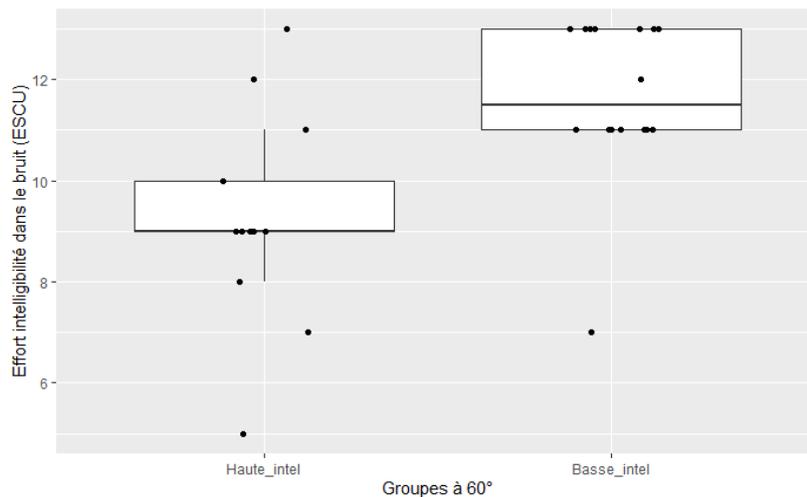


Figure 65 : Effort ressenti lors de la tâche d’intelligibilité dans le bruit en fonction des groupes d’intelligibilité à 60° de séparation entre les voix masques. Chaque point représente un participant. La médiane de mesure ESCU est représentée par le segment noir dans les boîtes à moustache. L’intervalle interquartile est représenté par l’extrémité inférieure et supérieure de la boîte à moustache dont les moustaches s’étendent jusqu’au 1,5 fois l’intervalle interquartile.

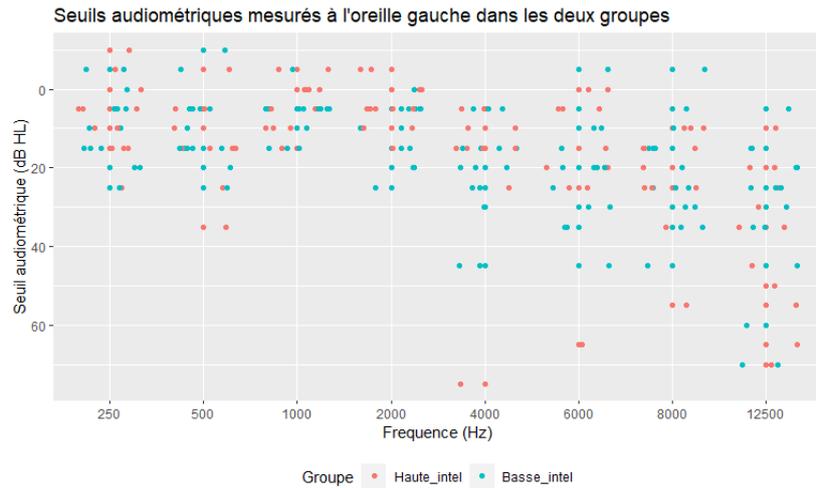


Figure 67 : Les seuils audiométriques mesurés à l'oreille gauche en fonction de deux groupes à haute intelligibilité et à basse intelligibilité. Chaque point indique un participant et chaque couleur l'appartenance au groupe (rouge : haute intelligibilité ; vert : basse intelligibilité).

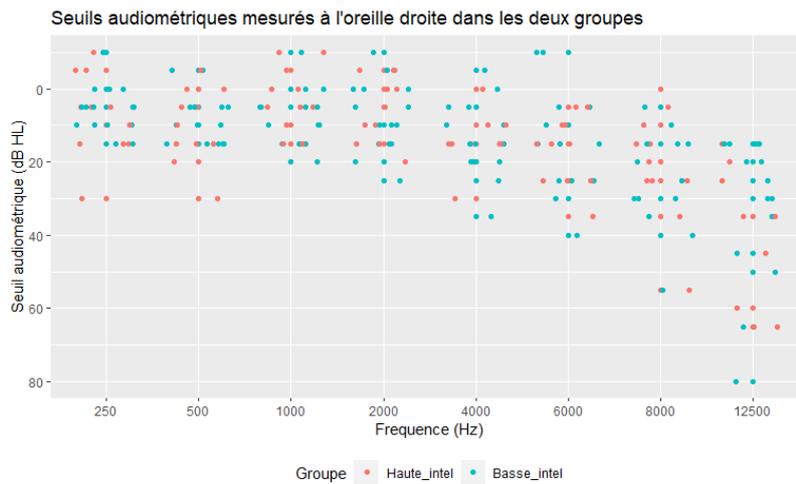


Figure 68 : Les seuils audiométriques mesurés à l'oreille droite en fonction de deux groupes à haute intelligibilité et à basse intelligibilité. Chaque point indique un participant et chaque couleur l'appartenance au groupe (rouge : haute intelligibilité ; vert : basse intelligibilité).

V. Discussion

La présente étude concernait une population militaire et visait à évaluer les effets de la séparation spatiale en situations multilocuteurs à la fois sur l'intelligibilité de la voix cible et sur l'effort d'écoute ressenti par le participant lors de la réalisation de la tâche. Les effets obtenus ont été ensuite étudiés en fonction de l'état auditif des participants. La charge cognitive des opérateurs des systèmes de communication militaire peut avoir des conséquences sur leur sécurité et sur celles des individus travaillant avec eux, aussi il est important d'identifier des pistes pour la réduire, par exemple via une diminution de l'effort d'écoute. Or il est crucial de savoir si l'efficacité de potentielles solutions dépend de l'état auditif de l'opérateur, qui peut être très variable d'un opérateur militaire à l'autre. En effet, leurs antécédents d'exposition au bruit varient selon la survenue d'accidents ou les postes occupés durant leur carrière. Par ailleurs, il est maintenant admis que l'audiométrie tonale classique ne suffit pas à décrire les capacités auditives d'un individu. Par conséquent, une première partie de l'étude a été dédiée à la récolte d'informations concernant les antécédents d'exposition au bruit et les difficultés ressenties au quotidien lors des situations d'écoute complexe (p. ex., compréhension lors de situations multi-locuteurs ou dans le bruit) à travers l'utilisation de questionnaires subjectifs (c.-à-d., questionnaire démographique et SSQ). Parmi ces évaluations réalisées, un test de compréhension de la parole dans le bruit a été aussi mené, ainsi que la mesure des produits de distorsion des otoémissions acoustiques.

Contrairement à la deuxième étude, l'effort d'écoute ressenti et la motivation à réaliser la tâche ont été mesurés dans cette étude pour permettre de mieux cerner les effets de la séparation spatiale des locuteurs.

Les résultats ont montré que l'intelligibilité la plus élevée était atteinte dans la condition avec le plus grand écartement entre les locuteurs. Effectivement la condition CRM.20 ne semblait pas suffisante pour garantir une bonne intelligibilité. Xia et al. (2015) dans leur étude ont révélé une grande variabilité parmi les participants avec des troubles de l'audition dans l'utilisation des indices spatiaux (c.-à-d., trois voix étaient placées à un écartement de 15° l'une de l'autre). Cette grande variabilité pourrait être le reflet de différences individuelles dans la capacité à encoder les indices spatiaux. Cette variabilité dans l'intelligibilité a été retrouvée également chez des auditeurs normo-entendants (Ruggles & Shinn-Cunningham, 2011). En revanche, l'augmentation de la séparation spatiale des locuteurs de 20° à 60° augmentait significativement l'intelligibilité des participants de la présente étude. Par ailleurs, la différence d'intelligibilité

liée à l'écartement des voix masques de 20° à 60° était corrélée à l'intelligibilité à 20°. Ainsi l'écartement profite le plus aux participants qui ont déjà les meilleures performances lorsque les voix masques sont proches de la voix cible. Ce résultat n'est pas en faveur d'un effet nivelant de l'écartement comme cela a pu être observé entre des auditeurs jeunes et âgés dans une précédente étude (Füllgrabe et al., 2015). Au contraire l'écartement a tendance à rendre les participants bons encore meilleurs sans améliorer la performance des auditeurs les moins bons, ce qui conduit à creuser l'écart entre ces deux populations. En effet, dans la condition CRM.60, deux groupes de participants semblaient ainsi s'opposer en termes d'intelligibilité atteinte : un groupe *haute intelligibilité* et un groupe à *basse intelligibilité*. Il semble donc que les participants à haute intelligibilité étaient avantagés par la présence de la séparation spatiale des locuteurs tandis que les autres ne l'étaient pas. Une analyse exploratoire de la présence de ces deux groupes a conduit à observer que parmi les variables mesurées, celle qui prédisait le mieux cette distinction était l'effort d'écoute ressenti par les participants dans la condition de compréhension de la parole dans le bruit. En d'autres termes, les participants qui déclaraient ressentir un effort d'écoute majeur dans la condition avec la voix cible dans le bruit étaient ceux appartenant au groupe avec une intelligibilité basse dans la condition CRM.60. Un mécanisme commun sous-jacent pourrait être les capacités de codage temporel, engagées à la fois dans la compréhension de la parole dans le bruit (Lorenzi et al., 2006) que dans l'intelligibilité en situation multilocuteurs (Ruggles & Shinn-Cunningham, 2011). Le rôle particulier joué par l'effort d'écoute dans le bruit correspond effectivement aux témoignages des personnes qui déclarent avoir des difficultés à comprendre la parole dans le bruit sans avoir de problème à l'entendre. Toutefois, ce lien n'a pas été confirmé avec les résultats obtenus dans la présente étude en utilisant le SSQ. Ce questionnaire ne montrait aucune corrélation avec la mesure comportementale de l'intelligibilité de la parole dans le bruit ou les mesures subjectives de l'effort d'écoute. Il est probable que la tâche d'écoute proposée n'était pas suffisamment proche des situations de la vie quotidienne auxquelles font référence les questions du SSQ. La structure rigide de la tâche expérimentale et la perception subjective de la séparation spatiale des locuteurs ont pu influencer ce résultat. Par exemple, au quotidien, l'auditeur peut s'appuyer pour la compréhension sur des indices qui lui parviennent de la situation (p. ex., lecture labiale, compréhension intuitive grâce au contexte) tandis que dans la tâche proposée ici il est pratiquement impossible de prévoir le couple couleur-chiffre qui va être donné par la voix cible. Dans les prochaines études, une idée pourrait être par exemple d'utiliser des tâches plus *écologiques*, c'est-à-dire plus proches des situations du quotidien.

Une autre hypothèse testée concernait la possibilité d'étudier l'effet de la séparation spatiale des locuteurs afin de réduire l'effort d'écoute ressenti chez le participant. Les résultats ont montré que l'effort ressenti dans la condition CRM.60 était plus bas par rapport à toutes les autres conditions. Passer d'un écartement des voix masques de 20° à un écartement de 60° a effectivement conduit à une réduction de l'effort d'écoute ressenti. Sachant que la réduction de l'effort ressenti du fait de l'augmentation de l'écartement des voix masques était corrélée à l'amélioration de l'intelligibilité, nous mettons ici en évidence le fait que la capacité à tirer profit de la séparation spatiale entraîne à la fois une réduction de l'effort d'écoute ressenti et une amélioration de l'intelligibilité, et cela chez les auditeurs présentant les performances les meilleures lorsque l'écartement est faible.

De façon intéressante, dans la phase initiale de la première étude et dans la deuxième étude, la meilleure intelligibilité était atteinte avec un engagement plus important de ressources cognitives par rapport aux autres conditions (respectivement : phase initiale de la première étude engagement plus important de l'IFG dans la condition favorable et dans la deuxième étude engagement plus important de l'IFG, DLPFC et STG dans la condition 60). Cela pourrait suggérer un lien entre engagement important de ressources cognitives, meilleure intelligibilité et moindre effort. Il semblerait donc que le déploiement des ressources cognitives soit susceptible d'être optimal et plus fonctionnel et qu'il n'est donc pas nécessairement lié à un effort plus important. Bien que cela reste une déduction, cela semble pointer vers le résultat de la première étude en ce qui concerne une réduction de l'activité cérébrale avec une meilleure intelligibilité (en lien avec l'efficacité neuronale).

Une dernière hypothèse concernait le rôle joué par la motivation dans l'effort d'écoute afin d'atteindre une meilleure intelligibilité. Les résultats n'ont montré aucune corrélation significative entre les différentes variables. Ce résultat pourrait être expliqué de plusieurs manières. Il est probable que le manque de relation entre ces variables vienne du choix de l'échelle administrée. De fait, au contraire de l'échelle de l'effort d'écoute ressenti, cette échelle était une échelle à dix points allant de 1 : '*Je n'ai jamais renoncé à écouter*' à 10 : '*Pour toutes les questions, j'ai abandonné*'. Le reste des labels étaient laissés vides et à remplir à la discrétion du participant. Chaque participant pouvait donc juger lui-même quelles nuances considérer entre les points 1 et 10. Cela a pu laisser la place à une grande variabilité d'attribution de signification motivationnelle. Il se peut donc que le manque de catégories précises ait joué sur la difficulté à quantifier la motivation d'une façon plus homogène parmi tous les participants. Il faut aussi souligner que les participants ne remplissaient pas de façon *secrète* leur niveau de

motivation, mais il le communiquait à haut voix à l'expérimentateur. Il est possible que le fait de l'énoncer publiquement poussait les participants à adapter leurs réponses en fonction de la présence de l'expérimentateur

Par ailleurs, il est important de souligner que l'importance attribuée par la littérature sur l'effort d'écoute à la variable de la motivation n'est pas associée à une disponibilité d'échelles ou de questionnaires pour l'évaluer dans ce domaine d'étude. Généralement la plupart des études essaient de déduire le niveau de motivation en modulant le comportement du participant à travers les récompenses monétaires (Bianco et al., 2021) ou des *matrix pay off* (Thompson et al., 2015) ou encore en déduisant des circuits cérébraux impliqués pour réaliser une tâche une plus ou moins grande motivation (c.-à-d., circuits de la récompense et dopamine), mais il n'existe pas de méthode pour l'évaluer directement.

À la lumière de ces résultats, nous suggérons qu'une partie des efforts de la communauté scientifique dans l'étude de l'effort d'écoute pourrait être dévouée à la conception d'échelles ou d'autres méthodes plus précises pour évaluer la motivation.

Une autre partie de cette étude a concerné l'évaluation des relations entre les mesures de l'état auditif (c.-à-d., audiométrie tonale liminaire, PDA et intelligibilité de la parole dans le bruit) et les différences d'intelligibilité et d'effort d'écoute ressenti liées à l'écartement des voix masques de 20° à 60°. Les résultats concernant l'intelligibilité ont montré que les mesures de l'état auditif n'étaient pas corrélées significativement avec la différence d'intelligibilité. Les seuils audiométriques ne semblent pas expliquer la capacité à bénéficier de la présence de la séparation spatiale des locuteurs, y compris pour le seuil très haute fréquence (12500 Hz). Concernant les PDA, seule la mesure à 2000 Hz a montré une corrélation significative avec la différence d'intelligibilité liée à l'écartement des voix masques. Des amplitudes plus élevées sont supposées refléter une meilleure fonctionnalité des CCE. Ce résultat serait en faveur du fait que ce sont les participants ayant la meilleure audition qui tirent parti de la séparation spatiale, confirmant l'absence d'effet nivelant lié à la séparation spatiale.

Néanmoins, la vision globale des résultats liés à l'évaluation de l'état auditif est plutôt en faveur d'un lien faible entre l'état auditif et la différence de performance induite par l'écartement des voix masques de 20° à 60°.

Concernant la différence d'effort d'écoute ressenti, la seule corrélation significative est retrouvée avec certaines fréquences mesurées avec l'audiométrie (c.-à-d., 6000 et 8000 Hz). Globalement, les participants avec un seuil audiométrique plus bas (c.-à-d., une audition

théoriquement moins altérée) présentaient une réduction de l'effort plus importante. Cependant, ce résultat était limité à une seule oreille.

Les résultats concernant les relations entre état auditif et intelligibilité ou effort d'écoute n'ont ainsi pas mis en évidence de liens notables. Globalement, les participants qui arrivent à mieux bénéficier de la présence de la séparation spatiale des locuteurs, et pour l'intelligibilité et pour l'effort d'écoute ressenti, sont les participants avec les meilleures performances lorsque la séparation spatiale est faible, et ce indépendamment de leur état auditif.

Une explication possible à la variabilité interindividuelle observée dans cette étude pourrait être liée à la technique de séparation spatiale des locuteurs et au positionnement des différentes voix. En effet, lors de l'utilisation de techniques de séparation spatiale des locuteurs, il est fréquent que l'auditeur éprouve des difficultés à percevoir la position « droit devant » qui était ici la position de la voix cible. La voix a tendance à être perçue sur-élevée et non au niveau du plan horizontal ou derrière l'auditeur et non devant (Wightman et Kistler, 1999 ; Andéol et al., 2015) provoquant une dissonance cognitive gênant le participant. Lors d'une prochaine étude, ce point pourrait être contrôlé en demandant à l'auditeur où la voix cible était perçue. De plus, il est possible que de meilleure performance soient obtenues avec le même degré de séparation spatiale, mais en évitant de positionner la voix cible droit devant (par exemple en la plaçant sur les côtés).

Ainsi, la séparation spatiale des locuteurs telle qu'elle est proposée dans cette étude ne semble pas être une solution permettant de réduire la variabilité interindividuelle dans l'intelligibilité en amenant les moins bons au niveau des meilleurs. Au contraire, elle a tendance à accentuer cette ségrégation. Néanmoins cette variabilité n'est pas expliquée par l'état auditif, ce dernier ne serait donc pas un facteur limitant l'intelligibilité dans ce type de situation multilocuteurs. Il reste à déterminer les facteurs sous tendant cette variabilité pour la réduire et améliorer l'intelligibilité pour le maximum d'opérateurs.

Synthèse

L'intérêt principal de ces travaux de thèse était porté sur l'étude de l'intelligibilité en situations multilocuteurs. Ces dernières sont définies comme des situations d'écoute complexe, dans lesquelles un auditeur doit focaliser son attention sur une voix cible dans un contexte où plusieurs locuteurs sont présents (voix-masques).

Plus précisément, les trois études présentées dans ce manuscrit de thèse ont permis d'étudier la problématique de l'effort d'écoute lié à ce type de situations. En tenant compte de la nature multidimensionnelle du concept de l'effort d'écoute, diverses variables dont il est composé ont été étudiées (c.-à-d., engagement de ressources cognitives, motivation et vécu subjectif) avec différents outils (c.-à-d., tâches comportementales, enregistrements de l'activité corticale, échelles subjectives, mesures audiométriques). Chaque étude s'est concentrée sur l'investigation d'une ou plusieurs de ces variables. En manipulant expérimentalement deux paramètres acoustique - les différences d'intensité entre les voix (c.-à-d., TMR) dans la première étude et la séparation spatiale des locuteurs dans la deuxième et dans la troisième étude - nous avons cherché à étudier les variables liées à l'effort d'écoute pour mieux comprendre ce concept.

Ainsi, l'étude de l'effort d'écoute s'est déclinée de différentes manières dans les trois études et chaque étude a répondu à des questions différentes. Le premier objectif, d'ordre fondamental, visait à observer l'engagement des ressources cognitives via l'activité corticale, et ainsi les mécanismes cérébraux sous-jacents à l'écoute en situation multilocuteurs chez les personnes normo-entendantes au sein d'un large réseau cortical/cérébral fonctionnel (première et deuxième études).

Le deuxième objectif, d'ordre principalement clinique, s'est adressé aux militaires suspectés d'avoir développé, du fait de leur carrière, des lésions auditives. Cette étude s'intéressait aux effets de la séparation spatiale des locuteurs sur l'intelligibilité de la voix cible et sur l'effort d'écoute ressenti en fonction de l'état auditif tel qu'il peut être mesuré par les tests audiologiques actuels (troisième étude).

Enfin, le troisième objectif, d'ordre opérationnel, visait à étudier comment les indices sonores tels quels le TMR, l'entraînement à utiliser ce dernier et la séparation spatiale des locuteurs pouvaient aider à améliorer l'intelligibilité de la voix cible et réduire l'effort d'écoute en vue de la conception de systèmes de communication plus efficaces et plus adaptés à l'opérateur (première, deuxième et troisième étude).

Première étude : l'entraînement à l'utilisation du TMR conduit à une amélioration de l'intelligibilité et à un moindre engagement de ressources cognitives

Dans cette première étude, nous avons étudié les différences interindividuelles dans l'utilisation du TMR (*phase initiale*) et les changements intra-individuels de l'intelligibilité dans l'utilisation du TMR suite à des sessions d'entraînement (*phase d'entraînement*). L'étude s'adressait à des auditeurs normo-entendants qui réalisaient une tâche d'écoute multilocuteurs pendant que l'activité corticale était enregistrée à l'aide de la fNIRS.

La tâche utilisée pour les trois études était la tâche du Coordinate Response Measure – CRM (Bolia et al., 2000). Le CRM est une des tâches les plus utilisées lorsqu'il s'agit d'étudier la perception de la parole en situation multilocuteurs. La structure rigide dont elle est composée permet d'observer de façon contrôlée l'interaction de différents facteurs, par exemple : le genre des locuteurs, le nombre de locuteurs, le TMR entre les locuteurs (Andéol et al., 2017 ; Brungart, 2001 ; Humes et al., 2017).

Dans la première étude, la tâche consistait en la détection de mots-clés énoncés par une voix cible en présence d'une voix masque. Les deux voix étaient présentées à trois niveaux différents de TMR, respectivement : favorable, intermédiaire et défavorable.

Dans la phase initiale, nous avons validé l'efficacité du TMR pour une meilleure compréhension de la voix cible en situation multilocuteurs. Cette efficacité était clairement évidente dans la condition favorable, qui était la condition avec le plus haut taux de bonnes réponses et plus discutable dans les conditions défavorable et intermédiaire dans lesquelles les performances étaient plus variables. De fait, dans ces deux conditions, il semble que tous les participants n'aient pas pu bénéficier de la présence du TMR, jusqu'à en être désavantagés, tandis que dans la condition défavorable certains participants ont, de manière contrintuitive, conservé une bonne intelligibilité en utilisant le TMR (Andéol et al., 2017). Au niveau cortical, les résultats ont montré une sollicitation majeure du gyrus frontal inférieur (IFG) gauche dans la condition favorable comparée à la condition défavorable. Dans le but d'étudier ce lien au sein de chaque condition, des analyses en régression robuste ont été ensuite réalisées. Les résultats ont montré une corrélation positive entre l'activité corticale et l'intelligibilité au sein des régions frontales (IFGs bilatéraux, cortex préfrontal bilatéral) et temporales (cortex auditif gauche), et ce, uniquement dans la condition défavorable. En d'autres termes, dans la condition censée être la plus difficile, les participants qui engageaient le plus de ressources cognitives

étaient ceux qui avaient une meilleure compréhension de la voix cible. Ces deux résultats montrent le rôle de l'IFG dans les situations multilocuteurs et confirment ainsi son rôle dans les situations d'écoute difficile en général (Abrams et al., 2013 ; Alain et al., 2018 ; Peelle, 2018). Sur la base de cette observation, nous avons souhaité étudier les processus cérébraux associés à l'entraînement à utiliser le TMR chez les participants qui semblaient ne pas en tirer parti initialement.

En partant des résultats de la littérature qui montrent la possibilité d'apprendre à utiliser le TMR (Brungart & Simpson, 2004 ; Thompson et al., 2015), nous avons focalisé notre intérêt sur la condition défavorable qui présentait la plus grande une marge de progression. Ainsi, les auditeurs ayant obtenu les moins bons scores en présence du TMR défavorable dans la phase initiale ont été recrutés pour participer à la phase d'entraînement. Les hypothèses des effets de cet entraînement portaient sur une probable amélioration de l'intelligibilité associée soit 1) à un engagement plus grand de ressources cognitives initialement non déployées, soit 2) de manière moins intuitive, avec un moindre engagement de ressources cognitives. Dans ce cas, l'hypothèse était que l'apprentissage et l'utilisation plus efficace de stratégies afin de réaliser la tâche (c.-à-d., inhibition de la voix masque avec focalisation facilitée sur la voix cible) auraient conduit à une telle diminution.

Les participants étaient divisés en deux groupes selon l'entraînement qui leur était proposé. D'un côté le *driven group* recevait des instructions explicites sur le TMR suivi d'un feedback auditif pour renforcer la bonne réponse (p. ex., l'expérimentateur déclarer explicitement au participant le niveau d'intensité sonore de la voix cible) et de l'autre côté le *not driven group* devait tout simplement réaliser la même tâche que celle faite pendant la première partie de l'étude (aucune instruction ni aucun feedback).

Le premier résultat intéressant était que les scores d'intelligibilité au terme de cette phase d'entraînement ne différaient pas entre les deux groupes : une nette amélioration de l'intelligibilité a été mise en évidence dans les deux groupes. Ainsi, que les participants aient reçu une consigne claire sur la manière d'utiliser le TMR à leur avantage, ou qu'ils aient simplement pratiqué la tâche pendant plusieurs sessions, la plupart ont pu réussir à s'améliorer. L'absence d'effet du type d'entraînement peut s'expliquer de plusieurs manières. Il est probable que 3 jours d'entraînement aient été suffisants pour réaliser la tâche de façon optimale conduisant à l'atteinte d'un plateau. Le fait que les conditions étaient regroupées en blocs homogènes a peut-être aidé les participants à comprendre la structure de la tâche et la présence du TMR. Les participants peuvent avoir développé leur propre stratégie pour réaliser la tâche (Thompson et al., 2015), probablement basée sur le contrôle inhibiteur. Par ailleurs, les

améliorations observées peuvent être dues, du moins en partie, à un apprentissage de la différenciation des caractéristiques des voix des locuteurs par les participants (Johnsrude et al., 2013). Ce résultat montre donc que le TMR en tant qu'indice sonore en situation multilocuteurs, s'il n'est pas forcément accessible initialement, peut être utilisé de manière relativement intuitive à mesure que l'on pratique ce genre de tâche auditive et conduire à une nette amélioration de l'intelligibilité du langage.

Cette amélioration était associée à une diminution de l'activité corticale au sein des IFGs bilatéraux qui étaient les régions parmi les plus activées en concomitance avec une meilleure intelligibilité dans la phase initiale. Ces résultats sont donc en faveur d'une association entre apprentissage et activité corticale plus faible, validant de fait notre seconde hypothèse. Ce résultat s'aligne avec les hypothèses de la théorie de l'efficacité neuronale, de l'anglais *neural efficiency theory* (Bless et al., 2014 ; Curtin et al., 2019 ; Fairclough et al., 2018 ; Putkinen & Saarikivi, 2018) qui associe l'apprentissage avec une moindre activité cérébrale. L'entraînement au cours des trois jours a pu conduire à un développement de stratégies individuelles qui ont permis de réaliser la tâche en améliorant par conséquent l'intelligibilité. Ces stratégies apprises soit automatiquement (*not driven group*) soit par instructions externes (*driven group*), auraient donc permis aux auditeurs, qui initialement avaient des difficultés, de réussir la tâche même si le TMR était défavorable. Parmi ces stratégies, nous proposons que figure une meilleure capacité à inhiber la voix masque et à se focaliser sur la voix cible. Cette explication a pu être proposée du fait de la réduction de l'activité au sein des IFGs supposés être impliqués dans les processus liés au contrôle inhibiteur (Chavan et al., 2015).

La première étude posait ainsi les bases pour la deuxième étude dans laquelle les questions concernaient la possibilité d'utiliser un autre indice sonore et notamment la séparation spatiale des locuteurs afin d'évaluer la relation entre l'intelligibilité et l'engagement de ressources cognitives (c.-à-d., diminution/augmentation de l'engagement de ressources cognitives).

Deuxième étude : la séparation spatiale des locuteurs améliore l'intelligibilité au prix d'un engagement plus grand de ressources cognitives

La deuxième étude se positionnait donc dans la suite naturelle de la première. L'effort d'écoute y a été étudié en situation multilocuteurs à travers la manipulation de la séparation spatiale des locuteurs, qui a déjà montré son efficacité sur l'intelligibilité dans certaines situations (Andéol et al., 2017).

La tâche consistait en la détection de mots-clés (tâche du CRM) énoncés par une voix cible qui pouvait être présentée avec deux ou quatre voix masques placées à différents degrés d'écartement de la voix cible positionnée droit devant (0°). Une première condition avec un écartement faible (c.-à-d. deux voix masques positionnées à $\pm 20^\circ$) ; une deuxième condition avec un grand écartement (c.-à-d. deux voix masques positionnées à $\pm 60^\circ$) ; et une troisième condition qui comprenait les deux types d'écartements (c.-à-d. quatre voix masques positionnées à $\pm 20^\circ$ et $\pm 60^\circ$) ont ainsi été utilisées. Ces différentes configurations ont été testées en faisant varier le nombre de locuteurs présents dans la scène auditive et leur localisation. Pendant le déroulement de la tâche, l'activité corticale des auditeurs était enregistrée à l'aide de la fNIRS. Comme dans la première étude, ces enregistrements visaient à évaluer l'engagement des ressources cognitives des participants, avec une éventuelle implication des régions corticales mises en évidence lors de la première étude, et en particulier l'IFG. Comme nous l'avons déjà souligné, un grand nombre d'études ont mis en évidence la contribution importante de cette région en situation d'écoute difficile (Abrams et al., 2013 ; Alain et al., 2018 ; Lawrence et al., 2018 ; Chapitre I. Première étude). Ce type de situation d'écoute induit une certaine demande cognitive (p. ex., mémoire, attention, contrôle inhibiteur) qui peut dépasser le domaine spécifique lié au traitement du langage (*general domain*) ; c'est-à-dire des processus transversaux à différentes tâches et indépendants de la modalité du stimulus (Peelle, 2018 ; Zekveld et al., 2007). Un rôle de type *general domain* a été plusieurs fois avancé pour l'IFG (Lawrence et al., 2018 ; Tettamanti et al., 2009) en particulier dans ces situations d'écoute complexe. Ce rôle serait également lié à une sollicitation majeure de ressources en mémoire, attention, prise de décision et contrôle inhibiteur (Eisner et al., 2010 ; Lawrence et al., 2018 ; Wild et al., 2012). Du fait des résultats de la première étude, qui ont attribué un rôle spécifique de l'IFG dans le contrôle inhibiteur, nous avons opté dans la présente étude pour une observation plus approfondie de cette région. Notre objectif était de dissocier les rôles de l'IFG en subdivisant l'analyse selon les trois sous-régions qui le composent : la *pars triangularis*, la *pars opercularis* et la *pars orbitalis*. De plus, les mesures

d'activité corticale comprenaient également les régions frontales avec le DLPFC, temporales avec le STG et pariétales avec l'IPC, ceci dans un souci de reproductibilité de la première étude. Les hypothèses principales de la deuxième étude portaient a) sur l'amélioration de l'intelligibilité de la voix cible grâce à la séparation spatiale des locuteurs b) sur un engagement de ressources cognitives plus important associé à une meilleure intelligibilité en présence de la séparation spatiale des locuteurs (depuis les Résultats – Phase initiale de la première étude) et c) sur la possibilité d'obtenir, grâce à la séparation spatiale des locuteurs, une bonne compréhension de la voix cible y compris en présence d'un nombre élevé de locuteurs.

Les résultats comportementaux ont montré à quel point la séparation spatiale des locuteurs peut être un indice utile pour une ségrégation efficace de la voix cible parmi les autres voix, mais sous certaines conditions. Ainsi, la meilleure intelligibilité de la voix cible était atteinte avec le degré de séparation entre les locuteurs le plus élevé (condition 60). La réduction de l'écart à 20° entre la voix cible et les deux voix masques conduisait à une réduction de l'intelligibilité et la présence de 4 voix masques conduisait à une intelligibilité dégradée, bien que non nulle. En effet, la littérature a mis en évidence une large variabilité interindividuelle dans la condition avec un écartement de voix masques d'environ 20° (Ruggles & Shinn-Cunningham, 2011 ; Xia et al., 2015). Il semble donc que la séparation spatiale des locuteurs soit de plus en plus bénéfique à mesure que l'écart grandit entre les locuteurs et que les bénéfices qu'elle apporte sont réduits par le nombre croissant de locuteurs.

Les résultats de l'activité corticale ont montré l'implication d'un large réseau cérébral fronto-temporo-pariétal (Alain et al., 2018 ; Peelle, 2018 ; Chapitre I. Première étude). De même que lors de la première étude, nous avons observé que la condition avec la plus haute intelligibilité (condition 60) correspondait à une activation majeure de plusieurs régions corticales parmi lesquelles l'IFG *pars triangularis*, le DLPFC bilatéral et le STG gauche. Le rôle joué par l'IFG dans le contrôle inhibiteur, impliqué dans l'inhibition des voix masques, semble être ici à nouveau confirmé. En particulier, une étude récente a montré l'implication de la *pars triangularis* de l'IFG gauche dans une tâche avec séparation spatiale des locuteurs en situation multilocuteurs (Wang et al., 2021). En soutien de l'hypothèse du contrôle inhibiteur, dans cette étude, l'implication des DLPFC bilatéraux est mise en évidence (voir Scannella, 2011 pour une description de leur rôle).

Ainsi l'ensemble des résultats des deux premières études suggère un rôle des mécanismes d'inhibition dans les situations multilocuteurs. L'IFG confirme avoir, entre la première et deuxième étude, un rôle dans le contrôle inhibiteur. Une meilleure intelligibilité est atteinte

grâce à une plus grande séparation spatiale des locuteurs et ceci se ferait donc en permettant un meilleur contrôle inhibiteur des voix masques au sein de l'IFG.

Enfin, il est important de rappeler que les effets bénéfiques de la séparation spatiale des locuteurs s'amointrissent lorsque le nombre de locuteurs présents augmente. Dans la condition 20-60 – avec quatre voix masques—l'intelligibilité de la voix cible s'est réduite malgré la séparation spatiale. Les meilleures performances restent donc celles avec un nombre de locuteurs limité, dans notre cas trois, et présentés à un degré important de séparation spatiale (Pinto et al., 2020).

Enfin, pour la condition plus difficile (20-60), les résultats d'imagerie en fNIRS n'ont pas révélé d'activation particulière dans les régions fronto-temporales ; comme ce fût le cas pour la condition 60. En revanche, dans cette condition, une activité corticale plus élevée, par rapport aux autres deux conditions, a été relevée dans le lobule pariétal inférieur droit (IPC). Plusieurs études chez des sujets sains et chez des patients avec des lésions cérébrales ont montré que la localisation d'un son dans l'espace sollicite l'activation de régions postérieures parmi lesquelles est compris l'IPC bilatéral (Lewald et al., 2018 ; Tissieres et al., 2019). Il est probable que les participants aient essayé d'engager des ressources cognitives dans cette condition pour exploiter le plus possible le seul indice sonore présent pour ségréger les voix afin de comprendre, sans pourtant y parvenir. Il est possible que l'engagement de certaines ressources cognitives, ou tout du moins, l'effort se soit concentré sur cet aspect du stimulus uniquement.

Enfin, le manque de mesures subjectives telles que l'effort ressenti et/ou de la motivation à continuer à faire face à un défi acoustique élevé ne permet malheureusement pas d'approfondir ce résultat. Cette lacune révèle l'importance d'avoir, dans la compréhension de la problématique de l'effort d'écoute, des évaluations subjectives de la part de l'auditeur pendant la réalisation de ces tâches d'écoute.

Troisième étude : la séparation spatiale améliore l'intelligibilité et réduit l'effort d'écoute ressenti

La troisième étude était directement inspirée des résultats de la deuxième et s'adressait à une population de militaires de carrière, qui du fait de leur métier sont susceptibles d'avoir été exposés à des niveaux sonores intenses. L'objectif ici était d'évaluer chez cette population, l'effet de la séparation spatiale des locuteurs sur l'intelligibilité et l'effort d'écoute ressenti. Les effets obtenus ont été ensuite étudiés en fonction de l'état auditif des participants.

Étant donné que l'activité professionnelle de ces opérateurs est souvent demandeuse en ressources cognitives (p. ex., planification ou gestion de tâches multiples en situations complexes), il est important d'élaborer des interfaces de communication qui puissent à la fois limiter l'effort d'écoute ressenti lors de la réalisation des tâches et améliorer l'intelligibilité des messages échangés. De plus, ces militaires peuvent avoir une audition dégradée qui peut leur causer des difficultés de compréhension des messages auditifs tout en augmentant l'effort d'écoute associé.

Il est important de rappeler que certaines lésions auditives présentes chez ces personnes restent difficiles à détecter, comme c'est le cas notamment des lésions concernant les synapses présentes entre les CCI et le nerf auditif (synaptopathie cochléaire, Liberman & Kujawa, 2017), ou des dysfonctions fines au niveau de l'amplificateur cochléaire que sont les CCE (Parker, 2020). À cet égard, et afin d'évaluer de façon plus fine l'état auditif des participants, des mesures complémentaires à l'audiométrie tonale liminaire « classique » ont été effectuées : l'audiométrie très haute fréquence (12500 Hz) et l'enregistrement des produits de distorsion des otoémissions acoustiques. De plus, parmi les conditions expérimentales testées, une condition expérimentale supplémentaire a été ajoutée pour évaluer la compréhension de la voix cible dans le bruit, cette capacité étant peu corrélée à l'audiométrie tonale (Phatak et al., 2019). Ce dernier point est en lien avec l'existence de témoignages de difficultés à comprendre la parole dans le bruit au sein de cette population, quantifié dans cette étude par la version courte du questionnaire SSQ version française (Moulin et al., 2015).

En utilisant la même tâche que celle utilisée dans la deuxième étude (c.-à-d. tâche du CRM), les hypothèses de la troisième étude concernaient la possibilité, à travers l'utilisation de la séparation spatiale des locuteurs, d'améliorer l'intelligibilité des participants et de réduire l'effort d'écoute ressenti. La mesure subjective de l'effort d'écoute ressenti était complétée par l'évaluation, toujours subjective, de la motivation à réaliser la tâche d'écoute. L'hypothèse, ici, concernait la possibilité, pour les participants les plus motivés, d'atteindre l'intelligibilité la

meilleure du fait d'un effort plus grand. Cette hypothèse est tirée des résultats de la première étude (phase initiale) et de la deuxième étude, qui ont montré qu'une meilleure intelligibilité était atteinte avec un plus grand engagement de ressources cognitives. Ici, la question concernait la possibilité de mieux réussir la tâche grâce à une séparation spatiale des locuteurs plus importante et d'être plus motivés à atteindre l'objectif de la compréhension.

Premièrement, les résultats obtenus ont montré que la séparation spatiale des locuteurs conduisait bien à une réduction de l'effort d'écoute ressenti ainsi qu'à une amélioration de l'intelligibilité du message. Ce dernier résultat obtenu dans la deuxième étude chez de jeunes normo-entendants était ici étendu à une population militaire avec une audition dégradée. Les résultats concernant l'intelligibilité montraient qu'un écartement faible parmi les locuteurs (c.-à-d. condition 20) n'était pas suffisant afin d'obtenir une bonne intelligibilité et qu'en revanche la meilleure intelligibilité était atteinte avec un degré d'écartement plus important (c.-à-d. condition 60). Donc, écarter les voix masques de 20° à 60° augmentait significativement l'intelligibilité. En détail, la différence d'intelligibilité entre ces deux écartements était corrélée à l'intelligibilité obtenue dans la condition 20. Ce résultat montre que les participants qui avaient déjà une bonne intelligibilité avec un écartement faible étaient les mêmes qui arrivaient à profiter le plus de la différence d'écartement, tandis que les participants qui n'ont pas pu bénéficier de l'écartement plus faible n'ont pas pu bénéficier d'un écartement plus grand.

De plus, nous avons retrouvé une large variabilité interindividuelle dans ces deux conditions (c.-à-d., condition 20 et condition 60). En effet, certains participants semblaient bénéficier de la séparation spatiale des locuteurs et d'autres non. Il est probable que ces personnes à cause de leurs troubles d'audition aient développé des stratégies personnelles et différentes afin de comprendre la parole au quotidien, ce qui serait la source de cette variabilité interindividuelle. Des séances d'entraînement avec des tâches d'écoute proposant différentes configurations spatiales des locuteurs pourraient être intéressantes afin d'évaluer la présence de cette variabilité. Il n'est pas exclu que cette variabilité soit également due aux techniques de séparation spatiale des locuteurs utilisées dans cette étude et à la tâche elle-même. Il se peut que certains auditeurs n'aient pas été vraiment capables d'utiliser leurs propres stratégies en raison de la tâche qui n'était pas assez écologique. En raison de cela, il serait aussi utile de considérer des tâches d'écoute plus écologiques et proches du quotidien, où l'auditeur arrive à s'impliquer.

Nous nous sommes donc concentrés sur la condition avec la meilleure intelligibilité et la plus grande variabilité interindividuelle (c.-à-d. condition 60), et nous avons observé que, parmi toutes les variables possibles, celle qui pourrait expliquer la variabilité de la performance dans

cette condition était l'effort d'écoute ressenti par les participants dans la condition de compréhension de la parole dans le bruit. En d'autres termes, les participants ayant obtenu les moins bonnes performances dans la condition avec la plus grande séparation spatiale (condition de masquage informationnel), étaient ceux qui percevaient l'effort de compréhension de la parole dans le bruit comme étant le plus élevé (condition de masquage énergétique). Ce lien pourrait renvoyer aux témoignages des personnes déclarant avoir des difficultés à comprendre le contenu de la parole dans le bruit. Toutefois, ce résultat n'a pas été corroboré par les résultats du SSQ. La faible relation entre les résultats inhérents à la tâche d'écoute et les réponses au SSQ pourrait être expliquée par la divergence entre les conditions expérimentales proposées et les situations du quotidien. Il est possible que les participants n'aient pas été capables d'appliquer à la tâche expérimentale les mêmes stratégies utilisées pour faire face aux défis acoustiques de tous les jours.

Une autre hypothèse testée visait à étudier les effets de la séparation spatiale des locuteurs sur la réduction de l'effort d'écoute ressenti chez les participants.

La condition avec le plus grand écartement des locuteurs était la condition dans laquelle les participants reportaient ressentir un effort moindre. L'écartement des voix masques de 20° à 60° conduisait aussi à une réduction de l'effort d'écoute ressenti. Les participants qui ont le plus diminué leur effort d'écoute ressenti avec un écartement des voix masques sont également ceux dont l'intelligibilité a été la plus améliorée avec un écartement des voix masques de 20° à 60°.

Il semblait donc que la capacité à tirer profit de la séparation spatiale entraînait à la fois une réduction de l'effort ressenti et une amélioration de l'intelligibilité, et cela chez les auditeurs présentant les performances les meilleures lorsque l'écartement est faible. L'utilisation de cet indice sonore pourrait donc être une solution intéressante à prendre en compte dans le perfectionnement des outils de communications en milieu opérationnel (p. ex., pour les radiocommunications). Cela permettrait de s'orienter de plus en plus vers une conception intégrée de l'outil de communication et des besoins et/ou difficultés de la personne qui l'utilise. Cela pourrait être particulièrement important compte tenu de la variabilité interindividuelle constatée dans l'étude réalisée et donc des caractéristiques de chaque opérateur.

Une dernière hypothèse concernait le rôle joué par la motivation dans l'effort d'écoute afin d'atteindre une meilleure intelligibilité. Les résultats n'ont montré aucune corrélation significative entre les différentes variables. Probablement, le manque de résultats peut être dû en partie à l'échelle même qu'a été utilisée afin d'évaluer la motivation. La structure de cette échelle ne permettait pas d'homogénéiser les réponses des participants (c.-à-d. l'échelle était

constituée de 10 points dont seulement deux avec des labels) et donc d'évaluer de façon précise leur niveau de motivation. De plus, les participants reportaient directement à l'investigateur leur niveau de motivation, ce qui a pu rendre l'évaluation *altérée* du fait de la présence de l'investigateur.

Une autre partie de cette étude a concerné l'évaluation des relations entre les mesures de l'état auditif (c.-à-d. audiométrie tonale liminaire, PDA et intelligibilité de la parole dans le bruit) et la différence d'intelligibilité et d'effort d'écoute lié à l'écartement des voix masque de 20° à 60°. Globalement, les résultats concernant les liens entre intelligibilité, effort d'écoute et état auditif n'ont pas mis en évidence de lien notable. L'état auditif ne semblait pas être un facteur limitant l'intelligibilité dans ce type de situation multilocuteurs.

Il semblait plutôt que les participants qui arrivaient à mieux bénéficier de la présence de la séparation spatiale des locuteurs étaient les participants avec les meilleures performances lorsque la séparation spatiale est faible.

Nous avons suggéré que la présence de cette variabilité interindividuelle ait pu être liée à la structure de la tâche (c.-à-d. technique de séparation spatiale des locuteurs). Il est probable que les participants n'aient pas perçu de la même façon la position où était censée être la voix cible (c.-à-d., droit devant) ; (Wightman et Kistler, 1999 ; Andéol et al, 2014). Par conséquent, pour certains participants cet indice n'a pas pu être complètement accessible.

Discussion générale

Ce travail de thèse regroupe trois études qui ont chacune contribué à apporter une caractérisation multifactorielle de l'effort d'écoute. L'objectif principal était d'améliorer la compréhension et la définition de cette problématique pour pouvoir proposer des pistes d'amélioration de systèmes de communication et de prise en charge des personnes ayant des troubles de la compréhension en situations complexes. Ceci a été fait par une approche originale regroupant la mesure comportementale associée à l'activité corticale sous-jacente ainsi que le report subjectif de la motivation et de l'effort ressenti en situation d'écoute complexe.

Pourquoi l'effort d'écoute ?

La réponse à cette question réside dans l'intérêt multiple suscité par ce concept. D'une part parce qu'il reste intéressant de comprendre cette capacité extraordinaire de notre cerveau à se concentrer sur une seule voix en présence d'autres voix interférentes même si elles sont plus fortes. D'autre part, parce qu'il existe un réel besoin d'améliorer la compréhension de la parole dans ce contexte, parfois difficile et conduisant à générer de l'effort. Nous pouvons en faire l'expérience directe lorsque nous essayons de communiquer avec une personne au milieu d'autres, ou lorsque nous sommes dans la rue et que le bruit d'une moto qui passe dégrade la qualité de la conversation en cours. Cet effort peut se traduire par une implication supplémentaire de ressources qui ne seraient pas déployées en condition d'écoute optimale. Cette implication supplémentaire risque de consommer des ressources qui pourraient être destinées à la réalisation d'autres tâches concurrentes et impacter par exemple la qualité de leur réalisation.

Chaque étude a donc été élaborée pour permettre d'apporter des réponses à ces questions, certaines de nature plus fondamentale, d'autres plus clinique ou encore opérationnelle. Ainsi cette thèse a eu vocation à mieux définir le concept d'effort d'écoute en situation multilocuteurs pour une meilleure prise en charge des patients ayant des difficultés de compréhension et de quantifier l'apport de l'utilisation des indices sonores pour la conception d'outils de communication plus efficaces.

Concernant les intérêts de nature fondamentale, nos résultats, soutenus par des observations en neuro-imagerie, s'insèrent dans un fil d'études où il devient de plus en plus clair que le concept d'effort d'écoute doit être pensé davantage en termes de coopération entre des régions cérébrales qui créent de véritables réseaux cérébraux. La complexité de la situation exige un

engagement plus important de ressources cognitives afin de bien réussir la tâche d'écoute. Il est intéressant de considérer que cet engagement de ressources cognitives est probablement soutenu et maintenu par une certaine motivation de l'auditeur à faire face à cette complexité. Cette assomption reste une hypothèse à vérifier et, pour l'instant, n'est qu'une déduction du travail réalisé au cours de cette thèse.

Un autre aspect fondamental est que les situations multilocuteurs représentent un grand défi dans la mesure où, pour se concentrer sur la voix cible, il faut être capable d'inhiber la/les voix masque(s). Il semblerait donc que les processus liés au contrôle inhibiteur soient cruciaux en situation multilocuteurs afin de focaliser sur la voix cible et la comprendre. À cet égard, à partir des résultats d'imagerie de la première et de la deuxième étude, qui confirment une partie des résultats de la littérature concernant le rôle de l'IFG dans le contrôle inhibiteur (du type *general domain*) nous proposons un rôle de l'IFG dans le contrôle inhibiteur qui s'étend aux situations multilocuteurs. Ce dernier point est renforcé par les résultats obtenus suite à un entraînement à l'utilisation d'indices sonores (première étude). En effet, si dans un premier temps une meilleure compréhension est associée à un engagement plus important de ressources cognitives impliquant aussi l'IFG (phase initiale dans la première étude) cet engagement se réduit au sein de l'IFG à mesure de l'amélioration de l'efficacité de ces mécanismes (phase d'entraînement). La réduction de l'engagement des ressources cognitives projette une lumière intéressante sur la question de la familiarisation de l'auditeur (par l'entraînement) à une tâche d'écoute complexe. Il est intéressant de voir que la maîtrise de la complexité de la tâche permet de mettre en place des stratégies efficaces qui, à leur tour, permettront la diminution des ressources cognitives engagées. Une fois que l'auditeur maîtrise la gestion de ses propres ressources et connaît la complexité de la tâche, il optimise l'utilisation de ces ressources certainement via une amélioration des stratégies concomitante à une réduction de l'activité corticale (Bless et al., 2014). Dans la deuxième étude, un engagement plus important de l'IFG était aussi présent dans les situations où une meilleure ségrégation des voix était obtenue grâce à la séparation spatiale des locuteurs, qui elle-même aurait permis une meilleure inhibition de la voix masque. Pour conclure, il semblerait qu'il n'y ait pas seulement une réduction quantitative de l'engagement, mais aussi une amélioration qualitative de celui-ci.

Dans ces deux premières études, le manque d'évaluations subjectives ne permet pas d'estimer quel est le coût personnel derrière l'engagement de ces ressources cognitives. Dans la première étude par exemple, il est envisageable qu'un effort d'écoute perçu initialement comme plus important puisse avoir été compensé par une diminution de ce dernier grâce à l'entraînement.

Avec la troisième étude, nous avons donc tenté de combler cette lacune avec l'utilisation d'évaluations subjectives de l'effort d'écoute. Et en effet, dans cette étude, les résultats ont montré que l'amélioration de l'intelligibilité de la voix cible atteinte avec le plus grand degré de séparation spatiale des locuteurs était associée à un vécu d'effort subjectif réduit.

L'effort d'écoute, un concept multifactoriel

La synthèse bibliographique, associée aux résultats obtenus dans chaque étude du présent travail de thèse, nous conduit à souligner l'importance de prendre en compte un modèle d'étude de l'effort d'écoute multidimensionnel (Herrmann & Johnsrude, 2020 ; Peelle, 2018 ; Pichora-Fuller et al., 2016). En effet, à travers ces observations, nous voyons que la compréhension d'une voix cible en situation multilocuteurs est influencée par l'intervention de plusieurs variables composant l'effort d'écoute (p. ex., les caractéristiques du contexte d'échange des messages, du message échangé, de l'auditeur même...). Finalement, c'est l'interaction dynamique entre ces variables qui génère chez l'auditeur le sentiment d'effort d'écoute. Au centre de ce modèle se trouve l'auditeur avec ses caractéristiques personnelles, ses ressources, la motivation et le vécu ressenti pour réaliser la tâche d'écoute. Le même auditeur est immergé dans un contexte spécifique qui peut soulever des défis acoustiques plus ou moins gérables selon leur niveau d'interférence. Le défi pour la compréhension de la voix cible se joue alors via l'interaction entre ces deux facteurs : la demande soulevée par le contexte et la réponse que l'auditeur peut apporter face à cette demande.

À partir de ce postulat, il est clair qu'atteindre l'objectif de comprendre n'est pas toujours une mission facile. Par exemple, si le défi acoustique soulevé par la situation est élevé et que l'auditeur n'est pas capable de lui faire face, la réalisation de l'objectif de compréhension peut être compromise. Cet exemple nous montre à quel point il est important de prendre en compte la variable motivationnelle dans ce modèle. Un auditeur motivé et qui considère que comprendre *en vaut la peine*, est aussi un auditeur qui est plus prédisposé à engager des ressources cognitives afin de réaliser la tâche. En revanche, un auditeur qui n'est pas assez motivé peut se désengager facilement de la tâche et renoncer à la compréhension, même si celle-ci peut représenter un défi objectif réalisable. D'autre part, la motivation peut également être alimentée par le vécu subjectif de l'auditeur du moment où l'objectif a été atteint. Ainsi, l'auditeur peut associer ses vécus subjectifs de valence positive (plaisir, joie), de la curiosité à comprendre le contenu de la conversation et cet aspect peut le motiver encore plus à rester engagé. À l'inverse, il peut associer ses vécus subjectifs de valence négative (frustration, ennui)

à comprendre le contenu de la conversation qui peuvent le démotiver et le pousser à renoncer à la tâche d'écoute. Dans la Figure 69 un schéma représente la combinaison des deux modèles de la littérature de Peelle, 2018 et Herrmann et Johnsrude, 2020, la manière dont cette thèse a abordé les éléments qui le composent et comment les résultats obtenus permettent de le discuter. Parler d'effort d'écoute est donc une question complexe, car cela signifie qu'il faut prendre en compte de nombreuses variables liées à l'auditeur (incluant la motivation à atteindre un objectif), au contexte et à la dynamique qui existe entre les deux.

Garder à l'esprit toutes ces variables et la manière dont elles interagissent devient nécessaire à la fois pour une meilleure compréhension des mécanismes qui sous-tendent la compréhension dans les situations multilocuteurs, mais aussi pour prendre en compte, dans la conception d'outils de communication (p. ex., les radiocommunications ou les appareils auditifs), les exigences et les besoins de la personne en question.

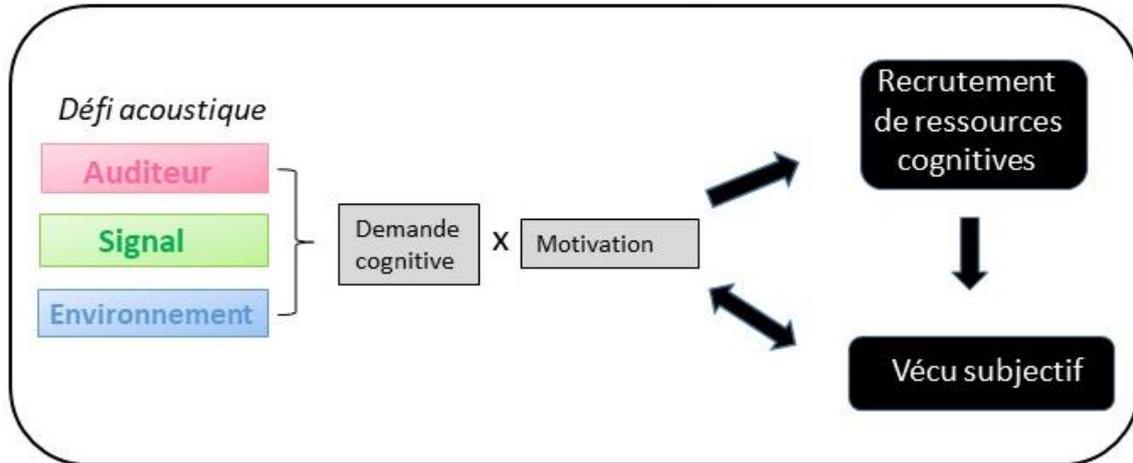
Motivation

Les modèles théoriques qui expliquent l'effort d'écoute réservent à la variable de la motivation une place importante, elle est souvent définie comme ayant un rôle *modérateur* dans l'effort d'écoute (Herrmann & Johnsrude, 2020; Peelle, 2018; Pichora-Fuller et al., 2016). Le fait pour un auditeur d'être plus ou moins motivé va avoir un impact important sur la compréhension. Un auditeur plus motivé a plus de chances de réussir la tâche qu'un auditeur non motivé. Il semble donc essentiel pour les chercheurs qui étudient cette problématique de prendre en compte cette variable dans leurs protocoles. En revanche, cette nécessité n'est pas soutenue par une disponibilité d'outils pour l'évaluer de façon directe. En fait, afin d'évaluer le niveau de motivation, certains chercheurs essaient de provoquer des changements dans la réponse des participants à travers l'offre de récompenses monétaires (Bianco et al., 2021) ou par des systèmes de récompenses et de punitions (p. ex., à travers des *matrix pay off*) ; (Thompson et al., 2015a). Généralement dans le domaine des neurosciences, la motivation est aussi déduite en enregistrant l'implication des régions qui sont censées faire partie du *circuit de la récompense* (notamment le striatum, l'amygdale, le cortex préfrontal ventro-médian et le cortex orbito-frontal). Ce circuit est impliqué dans la motivation, dans les émotions positives et en particulier celles dont le plaisir est une composante importante (p. ex., la joie, l'euphorie). Pour cette raison, un rôle important est joué par la dopamine, le neurotransmetteur *du plaisir* qui est donc impliqué lors d'une prise de décision, de l'évaluation des options à prendre en compte, du calcul de probabilités et donc de la motivation à agir (Haber & Knutson, 2010).

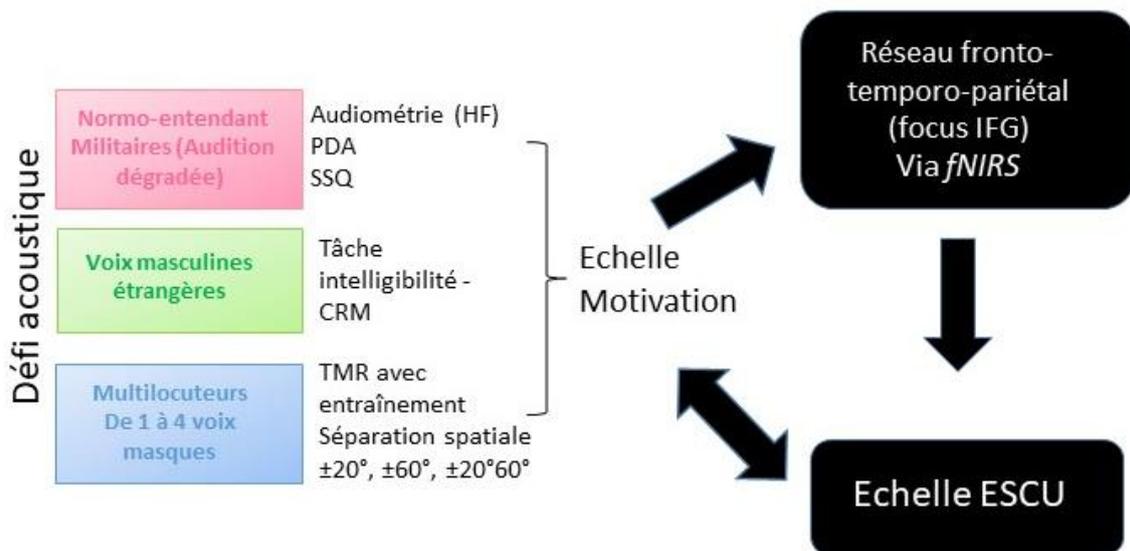
Les recherches cliniques qui étudient les pathologies et syndromes liés à la composante motivationnelle (p. ex., troubles dépressifs, addiction au jeu) ajoutent donc des éléments importants afin de comprendre les mécanismes sous-jacents à la motivation et la méthode pour l'évaluer. Le Bouc & Pessiglione, 2018, en partant de la théorie de la décision pour laquelle la motivation peut être vue comme la résultante du compromis entre l'effort fait et la récompense obtenue, proposent des modèles computationnels afin d'évaluer la motivation. Ces modèles computationnels se proposent de *phénotyper* le comportement des patients dans des tests de motivation afin de caractériser l'état du patient pour adapter des traitements personnalisés.

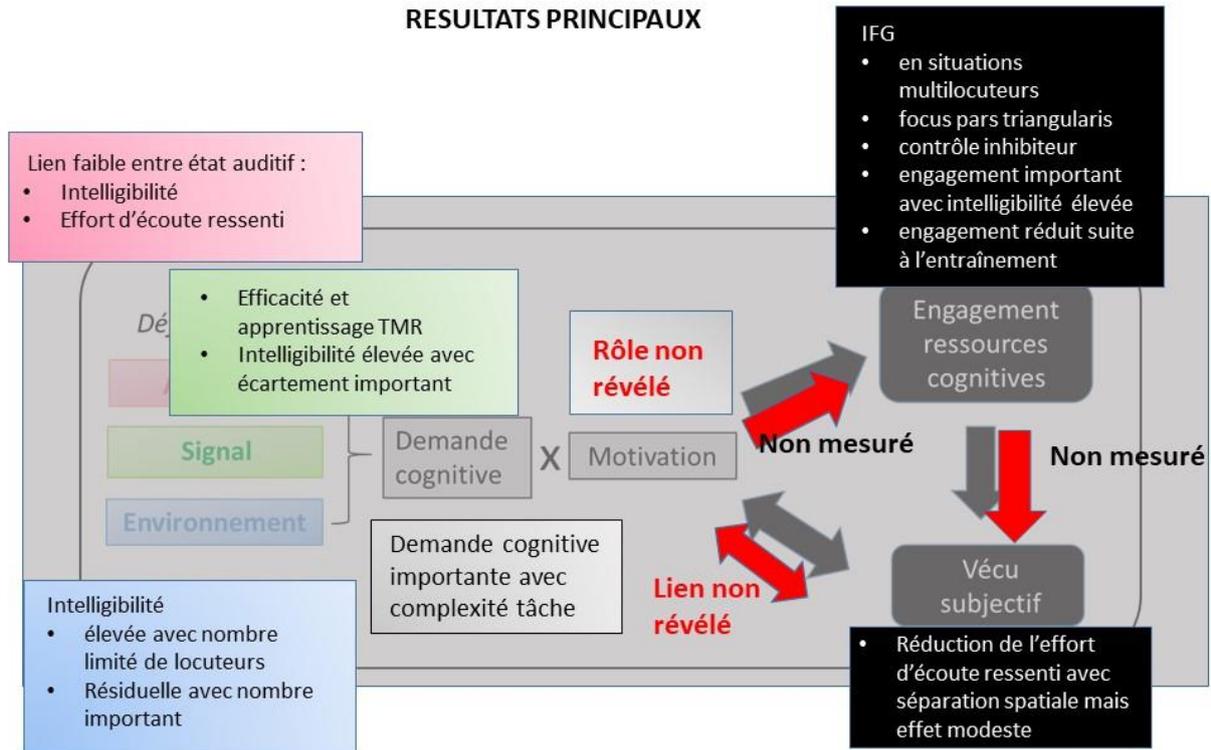
La recherche sur les stratégies pour évaluer la motivation en clinique est en cours. Il sera envisageable aussi que, dans le domaine des études sur l'effort d'écoute, des stratégies pour évaluer l'implication de cette variable soient prises en compte pour rendre l'étude sur l'effort d'écoute encore plus complète et précise. Par ailleurs, l'étude de l'implication de cette variable est importante aussi afin de mieux comprendre les difficultés des opérateurs qui travaillent en situation d'écoute complexe ou pour les personnes avec des troubles de l'audition. Ces personnes, du fait de la complexité de la tâche qu'ils doivent réaliser, peuvent facilement se désengager et se démotiver dans l'atteinte de la compréhension. Cette demande semble alors encore plus urgente compte tenu de la *transversalité* de cette problématique de l'effort d'écoute qui peut comprendre différents domaines de travail (p. ex., conducteurs de trains, professeurs, musiciens) et qui peut avoir plusieurs conséquences sur le bien-être de la personne.

MODELES THEORIQUES



ETUDES DE LA THESE





58

Figure 69 : Schéma qui représente respectivement : les *Modèles* proposés dans la littérature sur l'effort d'écoute (Peelle, 2018 : en haut à gauche et Herrmann et Johsrude, 2020 : en haut à droite) les *Études de la thèse* et les *Résultats principaux*.

Les études de la thèse ont été le résultat de la combinaison entre les deux modèles de la littérature. Les résultats obtenus ouvrent la voie à différentes perspectives qui s'orientent principalement vers l'amélioration des techniques de séparation spatiale avec des protocoles plus *écologiques* et vers l'étude de nouvelles méthodes d'évaluation des composantes subjectives (motivation, vécu subjectif).

Conclusion

Les situations multilocuteurs sont fréquentes dans la vie quotidienne et engendrent souvent un effort d'écoute que chacun a pu expérimenter. Dans certaines populations, cet effort d'écoute peut avoir des conséquences sérieuses : parmi les personnes âgées, il serait à l'origine de l'évitement des situations multilocuteurs et pourrait conduire à l'isolement social et au déclin cognitif. Parmi les professionnels des environnements à risque (personnel navigant, militaire), l'effort d'écoute en situations multilocuteurs peut contribuer à l'apparition d'une surcharge cognitive compromettant leur sécurité. Mieux comprendre l'effort d'écoute, les dimensions qui le composent comme les facteurs qui le modulent, constituent un réel enjeu pour aider ces populations.

Les modèles récents suggèrent que l'effort d'écoute peut être appréhendé en croisant le défi acoustique avec la motivation de l'individu : cette combinaison conduit au déploiement de ressources cognitives provoquant un vécu subjectif (positif ou négatif) influençant à son tour la motivation. Dans notre travail, ces différentes dimensions ont été abordées par la fNIRS pour décrire les ressources cognitives mises en jeu via le pattern d'activation corticale, la motivation et le vécu subjectif via des questionnaires et échelles, et le défi acoustique via des mesures de l'état auditif ou la manipulation de l'environnement expérimental.

En gardant constant le défi acoustique tout en manipulant les ressources cognitives de l'auditeur par un entraînement, une première étude a permis de mettre en évidence le rôle central joué par le contrôle inhibiteur sur l'intelligibilité en situation multilocuteurs. Grâce à l'efficacité neurale mise en jeu par l'entraînement, une diminution des ressources cognitives employées semble possible, ouvrant la voie à une diminution de l'effort d'écoute.

Dans une seconde étude, le défi acoustique a été manipulé en variant la position spatiale des locuteurs, compliquant (pour des locuteurs proches) ou facilitant (pour des locuteurs éloignés), la ségrégation de leur flux de parole. Cette seconde étude a confirmé le rôle joué par le contrôle inhibiteur dans les processus cognitifs mis en œuvre dans les situations multilocuteurs. Il serait ainsi une cible de choix pour tester de futures méthodes de réduction de l'effort d'écoute en améliorant son efficacité chez l'auditeur, par exemple par des procédures d'entraînement spécifiques.

Le défi acoustique dépend également de l'état auditif de l'auditeur : chez les militaires, comme chez les personnes âgées, il est attendu qu'une altération du capteur cochléaire ait des conséquences en termes d'effort d'écoute. Cependant, en termes de vécu subjectif, une

troisième étude a mis en évidence que faciliter le contrôle inhibiteur en éloignant les locuteurs produit une diminution de l'effort sans lien évident avec l'état auditif, tout en améliorant l'intelligibilité. Combiner les résultats de la deuxième et de la troisième étude conduirait à penser qu'un engagement du contrôle inhibiteur objectivé par l'activation corticale enregistrée en fNIRS serait concomitant à une diminution de l'effort ressenti. Cependant l'absence de mesures conjointes de ces deux dimensions de l'effort ne permet pas de l'affirmer. Disposer d'une réelle évaluation multidimensionnelle de l'effort dans une même expérience semble crucial pour des travaux ultérieurs.

Parmi les dimensions de l'effort, la motivation est probablement la plus délicate à évaluer et à manipuler. Notre tentative de la quantifier par une autoévaluation avec une échelle dédiée n'a pas donné satisfaction, la mesure ne variant que très peu entre les participants. S'inspirer d'autres domaines des neurosciences où elle est explorée depuis plus longtemps pourrait aider la communauté de la recherche en audition à mieux l'appréhender. Ainsi les travaux réalisés en neurosciences de l'addiction pourraient être d'une aide précieuse.

Enfin, aider efficacement les populations concernées par l'effort d'écoute en situation multilocuteurs passe par la confirmation des effets observés dans des situations plus écologiques. Pour les militaires, l'usage de moyens de simulations impliquant simultanément un grand nombre de joueurs, comme le jeu sérieux VBS 3, permettrait de valider la séparation spatiale des locuteurs comme un moyen de diminution de l'effort d'écoute et de lutte contre les dangers de la surcharge cognitive.

Bibliographie

- Aasted, C. M., Yücel, M. A., Cooper, R. J., Dubb, J., Tsuzuki, D., Becerra, L., Petkov, M. P., Borsook, D., Dan, I., & Boas, D. A. (2015). Anatomical guidance for functional near-infrared spectroscopy: AtlasViewer tutorial. *Neurophotonics*, 2(2), 020801–020801. <https://doi.org/10.1117/1.NPh.2.2.020801>
- Abrams, D. A., Ryali, S., Chen, T., Balaban, E., Levitin, D. J., & Menon, V. (2013). Multivariate Activation and Connectivity Patterns Discriminate Speech Intelligibility in Wernicke's, Broca's, and Geschwind's Areas. *Cerebral Cortex*, 23(7), 1703–1714. <https://doi.org/10.1093/cercor/bhs165>
- Alain, C., Du, Y., Bernstein, L. J., Barten, T., & Banai, K. (2018). Listening under difficult conditions: An activation likelihood estimation meta-analysis. *Human Brain Mapping*, 39(7), 2695–2709. <https://doi.org/10.1002/hbm.24031>
- Albinet, C. T., Mandrick, K., Bernard, P. L., Perrey, S., & Blain, H. (2014). Improved cerebral oxygenation response and executive performance as a function of cardiorespiratory fitness in older women: A fNIRS study. *Frontiers in Aging Neuroscience*, 6. <https://doi.org/10.3389/fnagi.2014.00272>
- Andéol, G., Suied, C., Scannella, S., & Dehais, F. (2017). The Spatial Release of Cognitive Load in Cocktail Party Is Determined by the Relative Levels of the Talkers. *JARO - Journal of the Association for Research in Otolaryngology*, 18(3), 457–464. <https://doi.org/10.1007/s10162-016-0611-7>
- Arbogast, T. L., Mason, C. R., & Kidd, G. (2002). The effect of spatial separation on informational and energetic masking of speech. *The Journal of the Acoustical Society of America*, 112(5), 2086–2098. <https://doi.org/10.1121/1.1510141>
- Arbogast, T. L., Mason, C. R., & Kidd, G. (2005). The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 117(4), 2169–2180. <https://doi.org/10.1121/1.1861598>
- Arlinger, S. (2003). Negative consequences of uncorrected hearing loss—A review. *International journal of audiology*, 42 Suppl 2, 2S17-20. <https://doi.org/10.3109/14992020309074639>
- Ashmore, J. (2008). Cochlear outer hair cell motility. *Physiological Reviews*, 88(1), 173–210. <https://doi.org/10.1152/physrev.00044.2006>
- Baddeley, A. (1996). Exploring the Central Executive. *The Quarterly Journal of Experimental Psychology Section A*, 49(1), 5–28. <https://doi.org/10.1080/713755608>
- Badri, R., Siegel, J. H., & Wright, B. A. (2011). Auditory filter shapes and high-frequency hearing in adults who have impaired speech in noise performance despite clinically normal audiograms. *The Journal of the Acoustical Society of America*, 129(2), 852–863. <https://doi.org/10.1121/1.3523476>

- Barker, J. W., Aarabi, A., & Huppert, T. J. (2013). Autoregressive model based algorithm for correcting motion and serially correlated errors in fNIRS. *Biomedical Optics Express*, 4(8), 1366–1366. <https://doi.org/10.1364/boe.4.001366>
- Belyk, M., Brown, S., Lim, J., & Kotz, S. A. (2017). Convergence of semantics and emotional expression within the IFG pars orbitalis. *NeuroImage*, 156, 240–248. <https://doi.org/10.1016/j.neuroimage.2017.04.020>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Ben-Shachar, M., Palti, D., & Grodzinsky, Y. (2004). Neural correlates of syntactic movement: Converging evidence from two fMRI experiments. *NeuroImage*, 21(4), 1320–1336. <https://doi.org/10.1016/j.neuroimage.2003.11.027>
- Best, V., Marrone, N., Mason, C. R., & Kidd, G. (2012). The influence of non-spatial factors on measures of spatial release from masking. *The Journal of the Acoustical Society of America*, 131(4), 3103–3110. <https://doi.org/10.1121/1.3693656>
- Best, V., Ozmeral, E. J., Kopco, N., & Shinn-Cunningham, B. G. (2008). Object continuity enhances selective auditory attention. *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.0803718105>
- Bianco, R., Mills, G., Kerangal, M. de, Rosen, S., & Chait, M. (2021). Reward enhances online participants' engagement with a demanding auditory task. *BioRxiv*, 2021.02.09.430467. <https://doi.org/10.1101/2021.02.09.430467>
- Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience*, 7(3), 295–301. <https://doi.org/10.1038/nn1198>
- Bless, J. J., Westerhausen, R., Kompus, K., Gudmundsen, M., & Hugdahl, K. (2014). Self-supervised, mobile-application based cognitive training of auditory attention: A behavioral and fMRI evaluation. *Internet Interventions*, 1(3), 102–110. <https://doi.org/10.1016/j.invent.2014.06.001>
- Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America*, 107(2), 1065–1066. <https://doi.org/10.1121/1.428288>
- Botvinick, M. M., Niv, Y., & Barto, A. G. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*, 113(3), 262–280. <https://doi.org/10.1016/j.cognition.2008.08.011>
- Bramhall, N., Beach, E. F., Epp, B., Le Prell, C. G., Lopez-Poveda, E. A., Plack, C. J., Schaette, R., Verhulst, S., & Canlon, B. (2019). The search for noise-induced cochlear synaptopathy in humans: Mission impossible? *Hearing Research*, 377, 88–103. <https://doi.org/10.1016/j.heares.2019.02.016>
- Bregman, A. S. (2009). Auditory Scene Analysis. *Encyclopedia of Neuroscience*, 729–736. <https://doi.org/10.1016/B978-008045046-9.00246-1>

- Bronkhorst, A. W. (2015). The cocktail-party problem revisited: Early processing and selection of multi-talker speech. In *Attention, Perception, and Psychophysics*. <https://doi.org/10.3758/s13414-015-0882-9>
- Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America*, *109*(3), 1101–1109. <https://doi.org/10.1121/1.1345696>
- Brungart, D. S., & Simpson, B. D. (2004). Within-ear and across-ear interference in a dichotic cocktail party listening task: Effects of masker uncertainty. *The Journal of the Acoustical Society of America*, *115*(1), 301–310. <https://doi.org/10.1121/1.1628683>
- Brungart, D. S., & Simpson, B. D. (2005). Improving Multitalker Speech Communication with Advanced Audio Displays. *New Directions for Improving Audio Effectiveness*, *937*, 30-1-30–18.
- Brungart, D. S., & Simpson, B. D. (2007). Effect of target-masker similarity on across-ear interference in a dichotic cocktail-party listening task. *The Journal of the Acoustical Society of America*, *122*(3), 1724–1734. <https://doi.org/10.1121/1.2756797>
- Buchsbaum, B. R., & D'Esposito, M. (2008). The search for the phonological store: From loop to convolution. *Journal of Cognitive Neuroscience*, *20*(5), 762–778. <https://doi.org/10.1162/jocn.2008.20501>
- Caldwell, J. (2005). Fatigue in aviation. *Travel Medicine and Infectious Disease*, *3*, 85–96. <https://doi.org/10.1016/j.tmaid.2004.07.008>
- Catani, M., Dell'Acqua, F., Vergani, F., Malik, F., Hodge, H., Roy, P., Valabregue, R., & Thiebaut de Schotten, M. (2012). Short frontal lobe connections of the human brain. *Cortex*, *48*(2), 273–291. <https://doi.org/10.1016/j.cortex.2011.12.001>
- Chang, E. F., Raygor, K. P., & Berger, M. S. (2015). Contemporary model of language organization: An overview for neurosurgeons. *Journal of Neurosurgery*, *122*(2), 250–261. <https://doi.org/10.3171/2014.10.JNS132647>
- Chanquoy, L., Tricot, A., & Sweller, J. (2007). *La charge cognitive: Théorie et applications*. Armand Colin.
- Chavan, C. F., Mouthon, M., Draganski, B., van der Zwaag, W., & Spierer, L. (2015). Differential patterns of functional and structural plasticity within and between inferior frontal gyri support training-induced improvements in inhibitory control proficiency. *Human Brain Mapping*, *36*(7), 2527–2543. <https://doi.org/10.1002/hbm.22789>
- Chern, A., & Golub, J. S. (2019). Age-related hearing loss and dementia. *Alzheimer disease and associated disorders*, *33*(3), 285–290. <https://doi.org/10.1097/WAD.0000000000000325>
- Cherry, E. C. (1953). Cocktail Party Effect Cherry 1953.pdf. *The Journal of the Acoustical Society of America*, *25*(5), 975–979.
- Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America*, *123*(1), 414–427. <https://doi.org/10.1121/1.2804952>

- Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., & Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature Neuroscience*, 3(3), 292–297. <https://doi.org/10.1038/73009>
- Curtin, A., Ayaz, H., Tang, Y., Sun, J., Wang, J., & Tong, S. (2019). Enhancing neural efficiency of cognitive processing speed via training and neurostimulation: An fNIRS and TMS study. *NeuroImage*, 198(January), 73–82. <https://doi.org/10.1016/j.neuroimage.2019.05.020>
- Cushing, S. (1994). *Fatal Words: Communication Clashes and Aircraft Crashes*. University of Chicago Press.
- Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does Semantic Context Benefit Speech Understanding through “Top–Down” Processes? Evidence from Time-resolved Sparse fMRI. *Journal of Cognitive Neuroscience*, 23(12), 3914–3932. https://doi.org/10.1162/jocn_a_00084
- Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, 23(8), 3423–3431. <https://doi.org/10.1523/jneurosci.23-08-03423.2003>
- Deal, J. A., Betz, J., Yaffe, K., Harris, T., Purchase-Helzner, E., Satterfield, S., Pratt, S., Govil, N., Simonsick, E. M., Lin, F. R., & for the Health ABC Study Group. (2017). Hearing Impairment and Incident Dementia and Cognitive Decline in Older Adults: The Health ABC Study. *The Journals of Gerontology: Series A*, 72(5), 703–709. <https://doi.org/10.1093/gerona/glw069>
- Defenderfer, J., Kerr-German, A., Hedrick, M., & Buss, A. T. (2017). Investigating the role of temporal lobe activation in speech perception accuracy with normal hearing adults: An event-related fNIRS study. *Neuropsychologia*, 106(August), 31–41. <https://doi.org/10.1016/j.neuropsychologia.2017.09.004>
- Dehais, F., Duprès, A., Blum, S., Drougard, N., Scannella, S., Roy, R. N., & Lotte, F. (2019). Monitoring pilot’s mental workload using erps and spectral power with a six-dry-electrode EEG system in real flight conditions. *Sensors (Switzerland)*, 19(6). <https://doi.org/10.3390/s19061324>
- Dehais, F., Dupres, A., Di Flumeri, G., Verdiere, K., Borghini, G., Babiloni, F., & Roy, R. (2018). Monitoring Pilot’s Cognitive Fatigue with Engagement Features in Simulated and Actual Flight Conditions Using an Hybrid fNIRS-EEG Passive BCI. *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 544–549. <https://doi.org/10.1109/SMC.2018.00102>
- Dehais, F., Tessier, C., Christophe, L., & Reuzeau, F. (2010). The Perseveration Syndrome in the Pilot’s Activity: Guidelines and Cognitive Countermeasures. In P. Palanque, J. Vanderdonckt, & M. Winckler (A c. Di), *Human Error, Safety and Systems Development* (pagg. 68–80). Springer. https://doi.org/10.1007/978-3-642-11750-3_6
- Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, 18(1), 193–222. <https://doi.org/10.1146/annurev.ne.18.030195.001205>

- D'Esposito, M., & Postle, B. R. (2015). The Cognitive Neuroscience of Working Memory. *Annual Review of Psychology*, *66*(1), 115–142. <https://doi.org/10.1146/annurev-psych-010814-015031>
- Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*(1), 80–90. <https://doi.org/10.1037/h0039515>
- Dobie, R. A., & Humes, L. E. (2017). Commentary on the regulatory implications of noise-induced cochlear neuropathy. *International Journal of Audiology*, *56*(sup1), 74–78. <https://doi.org/10.1080/14992027.2016.1255359>
- Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in cognitive sciences*, *12*(3), 99–105. <https://doi.org/10.1016/j.tics.2008.01.001>
- Duffau, H., Moritz-Gasser, S., & Mandonnet, E. (2014). A re-examination of neural basis of language processing: Proposal of a dynamic hodotopical model from data provided by brain stimulation mapping during picture naming. *Brain and Language*, *131*, 1–10. <https://doi.org/10.1016/j.bandl.2013.05.011>
- Eckert, M. A., Teubner-Rhodes, S., & Vaden, K. I. (2016). Is Listening in Noise Worth It? The Neurobiology of Speech Recognition in Challenging Listening Conditions. *Ear and hearing*, *37*(Suppl 1), 101S-110S. <https://doi.org/10.1097/AUD.0000000000000300>
- Eddins, D. A., & Liu, C. (2012). Psychometric properties of the coordinate response measure corpus with various types of background interference. *The Journal of the Acoustical Society of America*, *131*(2), EL177–EL183. <https://doi.org/10.1121/1.3678680>
- Eickhoff, S. B., Heim, S., Zilles, K., & Amunts, K. (2009). A systems perspective on the effective connectivity of overt speech production. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *367*(1896), 2399–2421. <https://doi.org/10.1098/rsta.2008.0287>
- Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., & Scott, S. K. (2010). Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *Journal of Neuroscience*, *30*(21), 7179–7186. <https://doi.org/10.1523/JNEUROSCI.4040-09.2010>
- Ekkekakis, P. (2009). Illuminating the Black Box: Investigating Prefrontal Cortical Hemodynamics during Exercise with Near-Infrared Spectroscopy. *Journal of Sport and Exercise Psychology*, *31*(4), 505–553. <https://doi.org/10.1123/jsep.31.4.505>
- Evans, S., McGettigan, C., Agnew, Z. K., Rosen, S., & Scott, S. K. (2016). Getting the Cocktail Party Started: Masking Effects in Speech Perception. *Journal of Cognitive Neuroscience*, *28*(3), 483–500. https://doi.org/10.1162/jocn_a_00913
- Fairclough, S., Ewing, K., Burns, C., & Kreplin, U. (2018). Neural efficiency and mental workload: Locating the red line. *Neuroergonomics: The Brain at Work and in Everyday Life*, 73–77. <https://doi.org/10.1016/B978-0-12-811926-6.00012-9>
- Ferbinteanu, J. (2019). Memory systems 2018 – Towards a new paradigm. *Neurobiology of Learning and Memory*, *157*, 61–78. <https://doi.org/10.1016/j.nlm.2018.11.005>

- Ferrari, M., & Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *NeuroImage*, *63*(2), 921–935. <https://doi.org/10.1016/j.neuroimage.2012.03.049>
- Ferreri, L., Bigand, E., Perrey, S., Muthalib, M., Bard, P., & Bugajska, A. (2014). Less Effort, Better Results: How Does Music Act on Prefrontal Cortex in Older Adults during Verbal Encoding? An fNIRS Study. *Frontiers in Human Neuroscience*, *8*. <https://doi.org/10.3389/fnhum.2014.00301>
- Fiedler, L., Wöstmann, M., Herbst, S. K., & Obleser, J. (2019). Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions. *NeuroImage*, *186*(August 2018), 33–42. <https://doi.org/10.1016/j.neuroimage.2018.10.057>
- Fiez, J. A., Raichle, M. E., Miezin, F. M., Petersen, S. E., Tallal, P., & Katz, W. F. (1995). PET Studies of Auditory and Phonological Processing: Effects of Stimulus Characteristics and Task Demands. *Journal of Cognitive Neuroscience*, *7*(3), 357–375. <https://doi.org/10.1162/jocn.1995.7.3.357>
- Fishburn, F. A., Hlutkowsky, C. O., Bemis, L. M., Huppert, T. J., Wakschlag, L. S., & Perlman, S. B. (2019). Irritability uniquely predicts prefrontal cortex activation during preschool inhibitory control among all temperament domains: A LASSO approach. *NeuroImage*, *184*(July 2018), 68–77. <https://doi.org/10.1016/j.neuroimage.2018.09.023>
- Foxe, J. J., & Snyder, A. C. (2011). The Role of Alpha-Band Brain Oscillations as a Sensory Suppression Mechanism during Selective Attention. *Frontiers in Psychology*, *0*. <https://doi.org/10.3389/fpsyg.2011.00154>
- Francis, A. L., & Love, J. (2020). Listening effort: Are we measuring cognition or affect, or both? *WIREs Cognitive Science*, *11*(1). <https://doi.org/10.1002/wcs.1514>
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *The Journal of the Acoustical Society of America*, *109*(5), 2112–2122. <https://doi.org/10.1121/1.1354984>
- Freyman, R. L., Helfer, K. S., McCall, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech Spatial release from informational masking in speech recognition The effect of spatial separation on informational and energetic masking of speech Some Experiments on the Recognition of Speech, with One and with Two Ears. *Citation: The Journal of the Acoustical Society of America*, *106*, 1101–1101. <https://doi.org/10.1121/1.428211>
- Friederici, A. D. (2011). The Brain Basis of Language Processing: From Structure to Function. *Physiological Reviews*, *91*(4), 1357–1392. <https://doi.org/10.1152/physrev.00006.2011>
- Friederici, A. D., & Gierhan, S. M. (2013). The language network. *Current Opinion in Neurobiology*, *23*(2), 250–254. <https://doi.org/10.1016/j.conb.2012.10.002>
- Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention—Focusing the searchlight on sound. *Current Opinion in Neurobiology*. <https://doi.org/10.1016/j.conb.2007.07.011>

- Furman, A. C., Kujawa, S. G., & Liberman, M. C. (2013). Noise-induced cochlear neuropathy is selective for fibers with low spontaneous rates. *Journal of Neurophysiology*, *110*(3), 577–586. <https://doi.org/10.1152/jn.00164.2013>
- Gagné, J.-P., Besser, J., & Lemke, U. (2017). Behavioral Assessment of Listening Effort Using a Dual-Task Paradigm: A Review. *Trends in Hearing*, *21*, 2331216516687287. <https://doi.org/10.1177/2331216516687287>
- Gatehouse, S., & Noble, W. (2004). The Speech, Spatial and Qualities of Hearing Scale (SSQ). *International Journal of Audiology*, *43*(2), 85–99. <https://doi.org/10.1080/14992020400050014>
- Giraudet, L., St-Louis, M.-E., Scannella, S., & Causse, M. (2015). *P300 Event-Related Potential as an Indicator of Inattentive Deafness?* <https://doi.org/10.1371/journal.pone.0118556>
- Goldstein, E. B. (2008). *Blackwell Handbook of Sensation and Perception*. John Wiley & Sons.
- Golestani, N., & Zatorre, R. J. (2004). Learning new sounds of speech: Reallocation of neural substrates. *NeuroImage*, *21*(2), 494–506. <https://doi.org/10.1016/j.neuroimage.2003.09.071>
- Gondusky, J. S., & Reiter, M. P. (2005). Protecting Military Convoys in Iraq: An Examination of Battle Injuries Sustained by a Mechanized Battalion during Operation Iraqi Freedom II. *Military Medicine*, *170*(6), 546–549. <https://doi.org/10.7205/MILMED.170.6.546>
- Haber, S. N., & Knutson, B. (2010). The Reward Circuit: Linking Primate Anatomy and Human Imaging. *Neuropsychopharmacology*, *35*(1), 4–26. <https://doi.org/10.1038/npp.2009.129>
- Hartmann, L., Sallard, E., & Spierer, L. (2016). Enhancing frontal top-down inhibitory control with Go/NoGo training. *Brain Structure and Function*, *221*(7), 3835–3842. <https://doi.org/10.1007/s00429-015-1131-7>
- Heald, S., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience*, *0*. <https://doi.org/10.3389/fnsys.2014.00035>
- Hendriks, D., Smits, A., Lavanga, M., De Wel, O., Thewissen, L., Jansen, K., Caicedo, A., Van Huffel, S., & Naulaers, G. (2019). Measurement of Neurovascular Coupling in Neonates. *Frontiers in Physiology*, *10*, 65. <https://doi.org/10.3389/fphys.2019.00065>
- Herrmann, B., & Johnsrude, I. S. (2020). A model of listening engagement (MoLE). *Hearing Research*, *397*, 108016–108016. <https://doi.org/10.1016/j.heares.2020.108016>
- Hertrich, I., Dietrich, S., Blum, C., & Ackermann, H. (2021). The Role of the Dorsolateral Prefrontal Cortex for Speech and Language Processing. *Frontiers in Human Neuroscience*, *0*. <https://doi.org/10.3389/fnhum.2021.645209>
- Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1), 67–99. <https://doi.org/10.1016/j.cognition.2003.10.011>
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. <https://doi.org/10.1038/nrn2113>
- Hill, K. T., & Miller, L. M. (2010). Auditory attentional control and selection during cocktail party listening. *Cerebral Cortex*, *20*(3), 583–590. <https://doi.org/10.1093/cercor/bhp124>

- Hoshi, Y. (2003). Functional near-infrared optical imaging: Utility and limitations in human brain mapping. *Psychophysiology*, *40*(4), 511–520. <https://doi.org/10.1111/1469-8986.00053>
- Humes, L. E., Kidd, G. R., & Fogerty, D. (2017). Exploring Use of the Coordinate Response Measure in a Multitalker Babble Paradigm. *Journal of Speech, Language, and Hearing Research*, *60*(3), 741–754. https://doi.org/10.1044/2016_JSLHR-H-16-0042
- Huppert, T. J., Hoge, R. D., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2006). A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *NeuroImage*, *29*(2), 368–382. <https://doi.org/10.1016/j.neuroimage.2005.08.065>
- Ihlefeld, A., & Shinn-Cunningham, B. (2008). Spatial release from energetic and informational masking in a selective speech identification task. *The Journal of the Acoustical Society of America*, *123*(6), 4369–4379. <https://doi.org/10.1121/1.2904826>
- Irani, F., Platek, S. M., Bunce, S., Ruocco, A. C., & Chute, D. (2007). Functional Near Infrared Spectroscopy (fNIRS): An Emerging Neuroimaging Technology with Important Applications for the Study of Brain Disorders. *The Clinical Neuropsychologist*, *21*(1), 9–37. <https://doi.org/10.1080/13854040600910018>
- James, W. (2007). *The Principles of Psychology*. Cosimo, Inc.
- Jenson, D., Harkrider, A. W., Thornton, D., Bowers, A. L., & Saltuklaroglu, T. (2015). Auditory cortical deactivation during speech production and following speech perception: An EEG investigation of the temporal dynamics of the auditory alpha rhythm. *Frontiers in Human Neuroscience*, *0*. <https://doi.org/10.3389/fnhum.2015.00534>
- Jöbsis, F. F. (1977). Noninvasive, Infrared Monitoring of Cerebral and Myocardial Oxygen Sufficiency and Circulatory Parameters. *Science*, *198*(4323), 1264–1267. <https://doi.org/10.1126/science.929199>
- Johnson, J., Xu, J., Cox, R., & Pendergraft, P. (2015). A Comparison of Two Methods for Measuring Listening Effort As Part of an Audiologic Test Battery. *American Journal of Audiology*, *24*(3), 419–431. https://doi.org/10.1044/2015_AJA-14-0058
- Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (2013). Swinging at a Cocktail Party: Voice Familiarity Aids Speech Perception in the Presence of a Competing Voice. *Psychological Science*, *24*(10), 1995–2004. <https://doi.org/10.1177/0956797613482467>
- Johnsrude, I. S., & Rodd, J. M. (2015). Factors That Increase Processing Demands When Listening to Speech. In *Neurobiology of Language* (pagg. 491–502). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-407794-2.00040-7>
- Johnston, W. A., & Heinz, S. P. (1978). Flexibility and capacity demands of attention. *Journal of Experimental Psychology: General*, *107*(4), 420–435. <https://doi.org/10.1037/0096-3445.107.4.420>
- Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.
- Kerns, J. G. (2006). Anterior cingulate and prefrontal cortex activity in an fMRI study of trial-to-trial adjustments on the Simon task. *NeuroImage*, *33*(1), 399–405. <https://doi.org/10.1016/j.neuroimage.2006.06.012>

- Kidd, G., Arbogast, T. L., Mason, C. R., & Gallun, F. J. (2005). The advantage of knowing where to listen. *The Journal of the Acoustical Society of America*. <https://doi.org/10.1121/1.2109187>
- Kidd, G., Mason, C. R., Rohtla, T. L., & Deliwala, P. S. (1998). Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. *The Journal of the Acoustical Society of America*, *104*(1), 422–431. <https://doi.org/10.1121/1.423246>
- Kocsis, L., Herman, P., & Eke, A. (2006). The modified Beer–Lambert law revisited. *Physics in Medicine and Biology*, *51*(5), N91–N98. <https://doi.org/10.1088/0031-9155/51/5/N02>
- Kramer, S. E., Teunissen, C. E., & Zekveld, A. A. (2016). Cortisol, chromogranin A, and pupillary responses evoked by speech recognition tasks in normally hearing and hard-of-hearing listeners: A pilot study. *Ear and Hearing*, *37*(March 2018), 126S–135S. <https://doi.org/10.1097/AUD.0000000000000311>
- Krause, M., & Richards, S. (2014). Prevalence of traumatic brain injury and access to services in an undergraduate population: A pilot study. *Brain Injury*, *28*(10), 1301–1310. <https://doi.org/10.3109/02699052.2014.916416>
- Krueger, M., Schulte, M., Zokoll, M. A., Wagener, K. C., Meis, M., Brand, T., & Holube, I. (2017). Relation Between Listening Effort and Speech Intelligibility in Noise. *American Journal of Audiology*, *26*(3S), 378–392. https://doi.org/10.1044/2017_AJA-16-0136
- Kujawa, S. G., & Liberman, M. C. (2009). Adding insult to injury: Cochlear nerve degeneration after «temporary» noise-induced hearing loss. *Journal of Neuroscience*, *29*(45), 14077–14085. <https://doi.org/10.1523/JNEUROSCI.2845-09.2009>
- Lawrence, R. J., Wiggins, I. M., Anderson, C. A., Davies-Thompson, J., & Hartley, D. E. H. (2018). Cortical correlates of speech intelligibility measured using functional near-infrared spectroscopy (fNIRS). *Hearing Research*, *370*, 53–64. <https://doi.org/10.1016/j.heares.2018.09.005>
- Le Bouc, R., & Pessiglione, M. (2018). La motivation dans tous ses K: Approche computationnelle des pathologies neuropsychiatriques. *médecine/sciences*, *34*(3), 238–246. <https://doi.org/10.1051/medsci/20183403012>
- Lee, A. K. C., Larson, E., Maddox, R. K., & Shinn-Cunningham, B. G. (2014). Using neuroimaging to understand the cortical mechanisms of auditory selective attention. *Hearing Research*, *307*, 111–120. <https://doi.org/10.1016/j.heares.2013.06.010>
- Lemke, U., & Besser, J. (2016). Cognitive Load and Listening Effort: Concepts and Age-Related Considerations. *Ear and Hearing*, *37* Suppl 1, 77S–84S. <https://doi.org/10.1097/AUD.0000000000000304>
- Leung, H.-C. (2000). An Event-related Functional MRI Study of the Stroop Color Word Interference Task. *Cerebral Cortex*, *10*(6), 552–560. <https://doi.org/10.1093/cercor/10.6.552>
- Lewald, J., Schlüter, M. C., & Getzmann, S. (2018). Cortical processing of location changes in a “cocktail-party” situation: Spatial oddball effects on electrophysiological correlates of auditory selective attention. *Hearing Research*. <https://doi.org/10.1016/j.heares.2018.04.009>

- Liakakis, G., Nickel, J., & Seitz, R. J. (2011). Diversity of the inferior frontal gyrus—A meta-analysis of neuroimaging studies. *Behavioural Brain Research*, 225(1), 341–347. <https://doi.org/10.1016/j.bbr.2011.06.022>
- Lieberman, M. C., Epstein, M. J., Cleveland, S. S., Wang, H., & Maison, S. F. (2016). Toward a Differential Diagnosis of Hidden Hearing Loss in Humans. *PloS One*, 11(9), e0162726. <https://doi.org/10.1371/journal.pone.0162726>
- Lieberman, M. C., & Kujawa, S. G. (2017). Cochlear synaptopathy in acquired sensorineural hearing loss: Manifestations and mechanisms. *Hearing Research*. <https://doi.org/10.1016/j.heares.2017.01.003>
- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. J. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences*, 103(49), 18866–18869. <https://doi.org/10.1073/pnas.0607364103>
- Mackersie, C. L., & Calderon-Moultrie, N. (2016). Autonomic Nervous System Reactivity During Speech Repetition Tasks: Heart Rate Variability and Skin Conductance. *Ear and Hearing*, 37, 118S. <https://doi.org/10.1097/AUD.0000000000000305>
- Maikala, R. V. (2010). Modified Beer's Law – historical perspectives and relevance in near-infrared monitoring of optical properties of human tissue. *International Journal of Industrial Ergonomics*, 40(2), 125–134. <https://doi.org/10.1016/j.ergon.2009.02.011>
- Mandrick, K. (s.d.). *Application de la spectroscopie proche infrarouge dans la discrimination de la charge de travail*. 212.
- Manuel, A. L., Bernasconi, F., & Spierer, L. (2013). Plastic modifications within inhibitory control networks induced by practicing a stop-signal task: An electrical neuroimaging study. *Cortex*, 49(4), 1141–1147. <https://doi.org/10.1016/j.cortex.2012.12.009>
- Matthen, M. (2016). Effort and Displeasure in People Who Are Hard of Hearing. *Ear and Hearing*, 37, 28S. <https://doi.org/10.1097/AUD.0000000000000292>
- McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group «white paper». *International Journal of Audiology*, 53(7), 433–440. <https://doi.org/10.3109/14992027.2014.890296>
- Medina-Garin, D. R., Dia, A., Bedubourg, G., Deparis, X., Berger, F., & Michel, R. (2016). Acute Acoustic Trauma in the French Armed Forces During 2007–2014. *Noise & Health*, 18(85), 297–302. <https://doi.org/10.4103/1463-1741.195802>
- Meek, J. H., Firbank, M., Elwell, C. E., Atkinson, J., Braddick, O., & Wyatt, J. S. (1998). Regional Hemodynamic Responses to Visual Stimulation in Awake Infants. *Pediatric Research*, 43(6), 840–843. <https://doi.org/10.1203/00006450-199806000-00019>
- Melara, R., Rao, A., & Tong, Y. (2002). The duality of selection: Excitatory and inhibitory processes in auditory selective attention. *Journal of experimental psychology. Human perception and performance*, 28, 279–306. <https://doi.org/10.1037/0096-1523.28.2.279>

- Moser, T., & Starr, A. (2016). Auditory neuropathy—Neural and synaptic mechanisms. *Nature Reviews Neurology*, *12*(3), 135–149. <https://doi.org/10.1038/nrneurol.2016.10>
- Mosnier, I., Bebear, J.-P., Marx, M., Fraysse, B., Truy, E., Lina-Granade, G., Mondain, M., Sterkers-Artières, F., Bordure, P., Robier, A., Godey, B., Meyer, B., Frachet, B., Poncet, C., Bouccara, D., & Sterkers, O. (2014). Predictive Factors of Cochlear Implant Outcomes in the Elderly. *Audiology and Neurotology*, *19*(Suppl. 1), 15–20. <https://doi.org/10.1159/000371599>
- Moulin, A., Pauzie, A., & Richard, C. (2015). Validation of a French translation of the speech, spatial, and qualities of hearing scale (SSQ) and comparison with other language versions. *International Journal of Audiology*, *54*(12), 889–898. <https://doi.org/10.3109/14992027.2015.1054040>
- Nakai, T., Kato, C., & Matsuo, K. (2005). An fMRI Study to Investigate Auditory Attention: A Model of the Cocktail Party Phenomenon. In *Magnetic Resonance in Medical Sciences* (Vol. 4, Numero 2, pagg. 75–82). <http://www.mrc-cbu.cam.ac.uk>
- Neubauer, A. C., & Fink, A. (2009). Intelligence and neural efficiency. *Neuroscience & Biobehavioral Reviews*, *33*(7), 1004–1023. <https://doi.org/10.1016/j.neubiorev.2009.04.001>
- Nixon, P., Lazarova, J., Hodinott-Hill, I., Gough, P., & Passingham, R. (2004). The Inferior Frontal Gyrus and Phonological Processing: An Investigation using rTMS. *Journal of Cognitive Neuroscience*, *16*(2), 289–300. <https://doi.org/10.1162/089892904322984571>
- Obleser, J., & Kotz, S. A. (2010). Expectancy Constraints in Degraded Speech Modulate the Language Comprehension Network. *Cerebral Cortex*, *20*(3), 633–640. <https://doi.org/10.1093/cercor/bhp128>
- Obleser, J., Wise, R. J. S., Dresner, M. A., & Scott, S. K. (2007). Functional Integration across Brain Regions Improves Speech Perception under Adverse Listening Conditions. *Journal of Neuroscience*, *27*(9), 2283–2289. <https://doi.org/10.1523/JNEUROSCI.4663-06.2007>
- Obrig, H., Wenzel, R., Kohl, M., Horst, S., Wobst, P., Steinbrink, J., Thomas, F., & Villringer, A. (2000). Near-infrared spectroscopy: Does it function in functional activation studies of the adult brain? *International Journal of Psychophysiology*, *35*(2), 125–142. [https://doi.org/10.1016/S0167-8760\(99\)00048-3](https://doi.org/10.1016/S0167-8760(99)00048-3)
- Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., & Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences*, *89*(13), 5951–5955. <https://doi.org/10.1073/pnas.89.13.5951>
- Panikratova, Y. R., Vlasova, R. M., Akhutina, T. V., Korneev, A. A., Sinitsyn, V. E., & Pechenkova, E. V. (2020). Functional connectivity of the dorsolateral prefrontal cortex contributes to different components of executive functions. *International Journal of Psychophysiology*, *151*, 70–79. <https://doi.org/10.1016/j.ijpsycho.2020.02.013>
- Parasuraman, R. (2003). Neuroergonomics: Research and practice. *Theoretical Issues in Ergonomics Science*, *4*(1–2), 5–20. <https://doi.org/10.1080/14639220210199753>
- Parker, M. A. (2020). Identifying three otopathologies in humans. *Hearing Research*, *398*, 108079. <https://doi.org/10.1016/j.heares.2020.108079>

- Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing, 39*(2). <https://doi.org/10.1097/AUD.0000000000000494>
- Peelle, J. E., Troiani, V., & Grossman, M. (2009). Interaction between process and content in semantic memory: An fMRI study of noun feature knowledge. *Neuropsychologia, 47*(4), 995–1003. <https://doi.org/10.1016/j.neuropsychologia.2008.10.027>
- Peng, C., & Hou, X. (2021). Applications of functional near-infrared spectroscopy (fNIRS) in neonates. *Neuroscience Research, 170*, 18–23. <https://doi.org/10.1016/j.neures.2020.11.003>
- Perrey, S. (2008). Non-invasive NIR spectroscopy of human brain function during exercise. *Methods, 45*(4), 289–299. <https://doi.org/10.1016/j.ymeth.2008.04.005>
- Peterson, B. S., Skudlarski, P., Gatenby, J. C., Zhang, H., Anderson, A. W., & Gore, J. C. (1999). An fMRI study of stroop word-color interference: Evidence for cingulate subregions subserving multiple distributed attentional systems. *Biological Psychiatry, 45*(10), 1237–1258. [https://doi.org/10.1016/S0006-3223\(99\)00056-6](https://doi.org/10.1016/S0006-3223(99)00056-6)
- Phatak, S. A., Brungart, D. S., Zion, D. J., & Grant, K. W. (2019). Clinical Assessment of Functional Hearing Deficits: Speech-in-Noise Performance. *Ear and Hearing, 40*(2), 426–436. <https://doi.org/10.1097/AUD.0000000000000635>
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing, 37*, 5S-27S. <https://doi.org/10.1097/AUD.0000000000000312>
- Pichora-Fuller, M. K., Mick, P., & Reed, M. (2015). Hearing, Cognition, and Healthy Aging: Social and Public Health Implications of the Links between Age-Related Declines in Hearing and Cognition. *Seminars in Hearing, 36*(3), 122–139. <https://doi.org/10.1055/s-0035-1555116>
- Pichora-Fuller, M. K., & Souza, P. E. (2003). Effects of aging on auditory processing of speech. *International Journal of Audiology, 42*(Suppl2), 2S11-2S16.
- Pinti, P., Scholkmann, F., Hamilton, A., Burgess, P., & Tachtsidis, I. (2019). Current Status and Issues Regarding Pre-processing of fNIRS Neuroimaging Data: An Investigation of Diverse Signal Filtering Methods Within a General Linear Model Framework. *Frontiers in Human Neuroscience, 12*(January), 1–21. <https://doi.org/10.3389/fnhum.2018.00505>
- Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., & Burgess, P. W. (2020). The present and future use of functional near- infrared spectroscopy (fNIRS) for cognitive neuroscience. *Annals of the New York Academy of Sciences, 1464*(1), 5–29. <https://doi.org/10.1111/nyas.13948>
- Pinto, D., Agmon, G., & Golumbic, E. Z. (2020). The Role of Spatial Separation on Selective and Distributed Attention to Speech. *BioRxiv, 2020.01.27.920785*. <https://doi.org/10.1101/2020.01.27.920785>

- Piquado, T., Benichov, J. I., Brownell, H., & Wingfield, A. (2012). The hidden effect of hearing acuity on speech recall, and compensatory effects of self-paced listening. *International Journal of Audiology*, *51*(8), 576–583. <https://doi.org/10.3109/14992027.2012.684403>
- Plack, C. J., Léger, A., Prendergast, G., Kluk, K., Guest, H., & Munro, K. J. (2016). Toward a Diagnostic Test for Hidden Hearing Loss. *Trends in Hearing*, *20*. <https://doi.org/10.1177/2331216516657466>
- Prendergast, G., Guest, H., Munro, K. J., Kluk, K., Léger, A., Hall, D. A., Heinz, M. G., & Plack, C. J. (2016). Effects of noise exposure on young adults with normal audiograms I: Electrophysiology. *Hearing Research*. <https://doi.org/10.1016/j.heares.2016.10.028>
- Pryce, H., Metcalfe, C., Claire, L. St., & Hall, A. (2010). Causal attributions in King-Kopetzky syndrome. *International Journal of Audiology*, *49*(7), 482–487. <https://doi.org/10.3109/14992021003682319>
- Putkinen, V., & Saarikivi, K. (2018). Neural correlates of enhanced executive functions: Is less more? *Annals of the New York Academy of Sciences*, *1423*(1), 117–125. <https://doi.org/10.1111/nyas.13645>
- Qi, Z., Han, M., Wang, Y., de los Angeles, C., Liu, Q., Garel, K., Chen, E. S., Whitfield-Gabrieli, S., Gabrieli, J. D. E., & Perrachione, T. K. (2019). Speech processing and plasticity in the right hemisphere predict variation in adult foreign language learning. *NeuroImage*, *192*(February), 76–87. <https://doi.org/10.1016/j.neuroimage.2019.03.008>
- Quaranta, N., Coppola, F., Casulli, M., Barulli, O., Lanza, F., Tortelli, R., Capozzo, R., Leo, A., Tursi, M., Grasso, A., Solfrizzi, V., Sobbà, C., & Logroscino, G. (2014). The Prevalence of Peripheral and Central Hearing Impairment and Its Relation to Cognition in Older Adults. *Audiology and Neurotology*, *19*(Suppl. 1), 10–14. <https://doi.org/10.1159/000371597>
- Quaresima, V., & Ferrari, M. (2019). Functional Near-Infrared Spectroscopy (fNIRS) for Assessing Cerebral Cortex Function During Human Behavior in Natural/Social Situations: A Concise Review. *Organizational Research Methods*, *22*(1), 46–68. <https://doi.org/10.1177/1094428116658959>
- Rahne, T., Fröhlich, L., Plontke, S., & Wagner, L. (2021). Influence of surgical and N95 face masks on speech perception and listening effort in noise. *PLOS ONE*, *16*(7), e0253874. <https://doi.org/10.1371/journal.pone.0253874>
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature neuroscience*, *12*(6), 718–724. <https://doi.org/10.1038/nn.2331>
- Rennies, J., Best, V., Roverud, E., & Kidd, G. (2019). Energetic and Informational Components of Speech-on-Speech Masking in Binaural Speech Intelligibility and Perceived Listening Effort. *Trends in Hearing*, *23*, 1–21. <https://doi.org/10.1177/2331216519854597>
- Rennies, J., & Kidd, G. (2018). Benefit of binaural listening as revealed by speech intelligibility and listening effort. *The Journal of the Acoustical Society of America*, *144*(4), 2147–2159. <https://doi.org/10.1121/1.5057114>

- Renz, T. (2020). Open-plan offices with room-high sound screens and small amount on collaborative work—Effects of sound masking. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 261(5), 1831–1842.
- Rodd, J. M., Longe, O. A., Randall, B., & Tyler, L. K. (2010). The functional organisation of the fronto-temporal language system: Evidence from syntactic and semantic ambiguity. *Neuropsychologia*, 48(5), 1324–1335. <https://doi.org/10.1016/j.neuropsychologia.2009.12.035>
- Roelofsen, P. (2008). Performance loss in open-plan offices due to noise by speech. *Journal of Facilities Management*, 6(3), 202–211. <https://doi.org/10.1108/14725960810885970>
- Roldán, M., & Kyriacou, P. A. (2021). Near-Infrared Spectroscopy (NIRS) in Traumatic Brain Injury (TBI). *Sensors*, 21(5), 1586.
- Rubia, K., Smith, A. B., Brammer, M. J., & Taylor, E. (2003). Right inferior prefrontal cortex mediates response inhibition while mesial prefrontal cortex is responsible for error detection. *NeuroImage*, 20(1), 351–358. [https://doi.org/10.1016/S1053-8119\(03\)00275-1](https://doi.org/10.1016/S1053-8119(03)00275-1)
- Rudner, M., Lyberg-Åhlander, V., Brännström, J., Nirme, J., Pichora-Fuller, M. K., & Sahlén, B. (2018). Listening Comprehension and Listening Effort in the Primary School Classroom. *Frontiers in Psychology*, 0. <https://doi.org/10.3389/fpsyg.2018.01193>
- Ruggles, D., & Shinn-Cunningham, B. (2011). Spatial selective auditory attention in the presence of reverberant energy: Individual differences in normal-hearing listeners. *JARO - Journal of the Association for Research in Otolaryngology*, 12(3), 395–405. <https://doi.org/10.1007/s10162-010-0254-z>
- Santosa, H., Zhai, X., Fishburn, F., & Huppert, T. (2018). The NIRS Brain AnalyzIR toolbox. *Algorithms*, 11(5). <https://doi.org/10.3390/A11050073>
- Scannella, S. (2011). *Bases cérébrales du conflit visuo-auditif spatial et sémantique: Études en IRM fonctionnelle et EEG* [These de doctorat, Toulouse 3]. <https://www.theses.fr/2011TOU30192>
- Scannella, S., Causse, M., Chauveau, N., Pastor, J., & Dehais, F. (2013). Effects of the audiovisual conflict on auditory early processes. *International Journal of Psychophysiology*, 89(1), 115–122. <https://doi.org/10.1016/J.IJPSYCHO.2013.06.009>
- Schroeter, M. L., Kupka, T., Mildner, T., Uludağ, K., & von Cramon, D. Y. (2006). Investigating the post-stimulus undershoot of the BOLD signal—A simultaneous fMRI and fNIRS study. *NeuroImage*, 30(2), 349–358. <https://doi.org/10.1016/j.neuroimage.2005.09.048>
- Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, 123(12), 2400–2406. <https://doi.org/10.1093/brain/123.12.2400>
- Scott, S. K., & McGettigan, C. (2013). The neural processing of masked speech. *Hearing Research*, 303, 58–66. <https://doi.org/10.1016/j.heares.2013.05.001>
- Scott, S. K., Rosen, S., Wickham, L., & Wise, R. J. S. (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception.

The Journal of the Acoustical Society of America, 115(2), 813–821.
<https://doi.org/10.1121/1.1639336>

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The Expected Value of Control: An Integrative Theory of Anterior Cingulate Cortex Function. *Neuron*, 79(2), 217–240.
<https://doi.org/10.1016/j.neuron.2013.07.007>

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, 12(5), 182–186. <https://doi.org/10.1016/j.tics.2008.02.003>

Shomstein, S., & Yantis, S. (2006). Parietal Cortex Mediates Voluntary Control of Spatial and Nonspatial Auditory Attention. *The Journal of Neuroscience*, 26(2), 435–439.
<https://doi.org/10.1523/JNEUROSCI.4408-05.2006>

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2007). Speech-associated gestures, Broca's area, and the human mirror system. *Brain and Language*, 101(3), 260–277.
<https://doi.org/10.1016/j.bandl.2007.02.008>

Smith, G. N. L., Pisoni, D. B., & Kronenberger, W. G. (2019). High-Variability Sentence Recognition in Long-Term Cochlear Implant Users: Associations With Rapid Phonological Coding and Executive Functioning. *Ear and hearing*, 40(5), 1149–1161.
<https://doi.org/10.1097/AUD.0000000000000691>

Sohoglu, E., & Davis, M. H. (2016). *Perceptual learning of degraded speech by minimizing prediction error*. <https://doi.org/10.1073/pnas.1523266113>

Sommers, M., & Danielson, S. (1999). Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context. *Psychology and aging*, 14, 458–472. <https://doi.org/10.1037/0882-7974.14.3.458>

Stephens, D., & Zhao, F. (2000). The role of a family history in King Kopetzky Syndrome (obscure auditory dysfunction). *Acta Oto-Laryngologica*, 120(2), 197–200.
<https://doi.org/10.1080/000164800750000900>

Strauß, A., Wöstmann, M., & Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Frontiers in Human Neuroscience*, 8(MAY), 1–7.
<https://doi.org/10.3389/fnhum.2014.00350>

Studebaker, G. A. (1985). A «Rationalized» Arcsine Transform. *Journal of Speech, Language, and Hearing Research*, 28(3), 455–462. <https://doi.org/10.1044/jshr.2803.455>

Swinehart, D. F. (1962, luglio 1). *The Beer-Lambert Law* (world) [Research-article]. ACS Publications; Division of Chemical Education. <https://doi.org/10.1021/ed039p333>

Szalárdy, O., Tóth, B., Farkas, D., György, E., & Winkler, I. (2019). Neuronal correlates of informational and energetic masking in the human brain in a multi-talker situation. *Frontiers in Psychology*, 10(APR), 1–16. <https://doi.org/10.3389/fpsyg.2019.00786>

Taren, A. A., Venkatraman, V., & Huettel, S. A. (2011). A Parallel Functional Topography between Medial and Lateral Prefrontal Cortex: Evidence and Implications for Cognitive Control. *Journal of Neuroscience*, 31(13), 5026–5031.
<https://doi.org/10.1523/JNEUROSCI.5762-10.2011>

- Tepe, V., Smalt, C., Nelson, J., Quatieri, T., & Pitts, K. (2017). Hidden Hearing Injury: The Emerging Science and Military Relevance of Cochlear Synaptopathy. *Military Medicine*, 182(9–10), e1785–e1795. <https://doi.org/10.7205/MILMED-D-17-00025>
- Tettamanti, M., Rotondi, I., Perani, D., Scotti, G., Fazio, F., Cappa, S. F., & Moro, A. (2009). Syntax without language: Neurobiological evidence for cross-domain syntactic computations. *Cortex*, 45(7), 825–838. <https://doi.org/10.1016/j.cortex.2008.11.014>
- Thompson, E. R., Iyer, N., Simpson, B. D., Wakefield, G. H., Kieras, D. E., & Brungart, D. S. (2015a). *In Speech-on-Speech Masking Experiments a*). 138(3), 1297–1304.
- Tissieres, I., Crottaz-Herbette, S., & Clarke, S. (2019). Implicit representation of the auditory space: Contribution of the left and right hemispheres. *Brain Structure and Function*, 224(4), 1569–1582. <https://doi.org/10.1007/s00429-019-01853-5>
- Tóth, B., Farkas, D., Urbán, G., Szalárdy, O., Orosz, G., Hunyadi, L., Hajdu, B., Kovács, A., Szabó, B. T., Shestopalova, L. B., & Winkler, I. (2019). Attention and speech-processing related functional brain networks activated in a multi-speaker environment. In *PLoS ONE* (Vol. 14, Numero 2, pag. 31). <https://doi.org/10.1371/JOURNAL.PONE.0212754>
- Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, 12(4), 242–248. <https://doi.org/10.1080/17470216008416732>
- Vaden, K. I., Kuchinsky, S. E., Cute, S. L., Ahlstrom, J. B., Dubno, J. R., & Eckert, M. A. (2013). The Cingulo-Opercular Network Provides Word-Recognition Benefit. *Journal of Neuroscience*, 33(48), 18979–18986. <https://doi.org/10.1523/JNEUROSCI.1417-13.2013>
- van Esch, T. E. M., Kollmeier, B., Vormann, M., Lyzenga, J., Houtgast, T., Hällgren, M., Larsby, B., Athalye, S. P., Lutman, M. E., & Dreschler, W. A. (2013). Evaluation of the preliminary auditory profile test battery in an international multi-centre study. *International Journal of Audiology*, 52(5), 305–321. <https://doi.org/10.3109/14992027.2012.759665>
- Verney, S. P., Granholm, E., & Marshall, S. P. (2004). Pupillary responses on the visual backward masking task reflect general cognitive ability. *International Journal of Psychophysiology*, 52(1), 23–36. <https://doi.org/10.1016/j.ijpsycho.2003.12.003>
- Villringer, A., & Chance, B. (1997). Non-invasive optical spectroscopy and imaging of human brain function. *Trends in Neurosciences*, 20(10), 435–442. [https://doi.org/10.1016/S0166-2236\(97\)01132-6](https://doi.org/10.1016/S0166-2236(97)01132-6)
- Visentin, C., Prodi, N., Cappelletti, F., Torresin, S., & Gasparella, A. (2018). Using listening effort assessment in the acoustical design of rooms for speech. *Building and Environment*, 136, 38–53. <https://doi.org/10.1016/j.buildenv.2018.03.020>
- Walsh, B. J., Buonocore, M. H., Carter, C. S., & Mangun, G. R. (2011). Integrating Conflict Detection and Attentional Control Mechanisms. *Journal of Cognitive Neuroscience*, 23(9), 2211–2221. <https://doi.org/10.1162/jocn.2010.21595>
- Wang, J., Chen, J., Yang, X., Liu, L., Wu, C., Lu, L., Li, L., & Wu, Y. (2021). Common Brain Substrates Underlying Auditory Speech Priming and Perceived Spatial Separation. *Frontiers in Neuroscience*, 15. <https://doi.org/10.3389/fnins.2021.664985>

- Weeks, S. R., McAuliffe, C. L., DuRussel, D., & Pasquina, P. F. (2010). Physiological and Psychological Fatigue in Extreme Conditions: The Military Example. *PM&R*, 2(5), 438–441. <https://doi.org/10.1016/J.PMRJ.2010.03.023>
- White, B. E., & Langdon, C. (2021). The cortical organization of listening effort: New insight from functional near-infrared spectroscopy. *NeuroImage*, 240, 118324. <https://doi.org/10.1016/j.neuroimage.2021.118324>
- Wightman, F. L., & Kistler, D. J. (1989). Headphone simulation of free- field listening. I: Stimulus synthesis. *The Journal of the Acoustical Society of America*, 85(2), 858–867. <https://doi.org/10.1121/1.397557>
- Wijayasiri, P., Hartley, D. E. H., & Wiggins, I. M. (2017). Brain activity underlying the recovery of meaning from degraded speech: A functional near-infrared spectroscopy (fNIRS) study. *Hearing Research*, 351, 55–67. <https://doi.org/10.1016/j.heares.2017.05.010>
- Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful Listening: The Processing of Degraded Speech Depends Critically on Attention. *Journal of Neuroscience*, 32(40), 14010–14021. <https://doi.org/10.1523/JNEUROSCI.1528-12.2012>
- Wingfield, A. (2016). Evolution of Models of Working Memory and Cognitive Resources. *Ear and Hearing*, 37, 35S. <https://doi.org/10.1097/AUD.0000000000000310>
- Xia, J., Nooraei, N., Kalluri, S., & Edwards, B. (2015). Spatial release of cognitive load measured in a dual-task paradigm in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 137(4), 1888–1898. <https://doi.org/10.1121/1.4916599>
- Xiong, Y., & Newman, S. (2021). Both activation and deactivation of functional networks support increased sentence processing costs. *NeuroImage*, 225, 117475. <https://doi.org/10.1016/j.neuroimage.2020.117475>
- Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top–down and bottom–up processes in speech comprehension. *NeuroImage*, 32(4), 1826–1836. <https://doi.org/10.1016/j.neuroimage.2006.04.199>
- Zekveld, A. A., Heslenfeld, D. J., Johnsrude, I. S., Versfeld, N. J., & Kramer, S. E. (2014). The eye as a window to the listening brain: Neural correlates of pupil size as a measure of cognitive listening load. *NeuroImage*, 101, 76–86. <https://doi.org/10.1016/j.neuroimage.2014.06.069>
- Zekveld, A. A., Rudner, M., Johnsrude, I. S., Heslenfeld, D. J., & Rönnerberg, J. (2012). Behavioral and fMRI evidence that cognitive ability modulates the effect of semantic context on speech intelligibility. *Brain and Language*, 122(2), 103–113. <https://doi.org/10.1016/j.bandl.2012.05.006>
- Zekveld, A. A., Rudner, M., Kramer, S. E., Lyzenga, J., & Rönnerberg, J. (2014). Cognitive processing load during listening is reduced more by decreasing voice similarity than by increasing spatial separation between target and masker speech. *Frontiers in Neuroscience*, 0. <https://doi.org/10.3389/fnins.2014.00088>

Zekveld, A. A., van Scheepen, J. A. M., Versfeld, N. J., Veerman, E. C. I., & Kramer, S. E. (2019). Please try harder! The influence of hearing status and evaluative feedback during listening on the pupil dilation response, saliva-cortisol and saliva alpha-amylase levels. *Hearing Research*, 381, 107768. <https://doi.org/10.1016/j.heares.2019.07.005>

Zhang, M., Mary Ying, Y. L., & Ihlefeld, A. (2018). Spatial Release From Informational Masking: Evidence From Functional Near Infrared Spectroscopy. *Trends in Hearing*, 22, 1–12. <https://doi.org/10.1177/2331216518817464>

Zhao, D. S., Fei. (2000). The Role of a Family History in King Kopetzky Syndrome (Obscure Auditory Dysfunction). *Acta Oto-Laryngologica*, 120(2), 197–200. <https://doi.org/10.1080/000164800750000900>

Annexe 1. Article

Frontiers in Neuroergonomics, section Augmented and Synthetic Neuroergonomics

Cocktail party training induces decreased activity in bilateral IFGs. A fNIRS study

Cosima Lanzilotti^{1,2,3}, Guillaume Andéol², Christophe Micheyl⁴, Sébastien Scannella^{2*}

¹ Département Conception et Conduite des Véhicules Aéronautiques et Spatiaux, Institut Supérieur de l'Aéronautique et de l'Espace (ISAE-Supaéro), Toulouse, France

² Département Neurosciences et sciences cognitives, Institut de Recherche Biomédicale des Armées, 1 place du général Valérie André, BP 73, 91223, Brétigny sur Orge, France

³ Thales SIX GTS France S.A.S, 4 Avenue des Louvresses, 92622, Gennevilliers, France

⁴ Starkey Hearing Technologies, 23 Rue Claude Nicolas Ledoux 94000 Créteil, France

*Correspondence: sebastien.scannella@isae-superaero.fr

Keywords: near infra-red spectroscopy¹, left inferior frontal gyrus activity², cocktail party training³, target-to-masker ratio⁴, speech-in-speech maskings⁵.

Abstract

The brain networks responsible for human listeners' ability to selectively attend to, and decode a voice amid other talkers remain incompletely elucidated to date. The present study aimed to investigate relationships between cortical activity and performance in a concurrent-speech listening task, before and after training-induced improvements. In a first experiment, cortical activity was measured using functional near infrared spectroscopy (fNIRS) in 74 participants while they performed a speech-intelligibility task that required selectively listening to target sentences played concurrently with a masker speech sentence. Three different target-to-masker level ratios (TMR)—low, high, and intermediate—were tested, eliciting different levels of listening challenges as well as different listening strategies. Results showed large inter-individual differences in task performance and cortical activity: whereas performance increased monotonically with TMR in some participants, in others performance failed to decrease, or even improved, in the lowest-TMR condition—an effect reflecting different listening strategies, with some participants taking advantage of sound-level differences between the target and masker streams to more easily identify the target voice. On the neuronal level, fronto-parieto-temporal and left inferior frontal gyrus activities were found to be correlated with behavioral

performance. In Experiment II, 27 participants with poor performance in the lowest-TMR condition of Experiment I received perceptual training aiming at improving their performance in this condition. Results show significant performance improvements and decreased activity in bilateral inferior frontal gyri after training—an effect that may reflect lower neural activity linked to higher efficiency in processes involved in masker inhibition after training. Overall, these findings add to the evidence for an involvement of the inferior frontal gyrus in speech listening in adverse conditions.

INTRODUCTION

The ability of humans and other animals to selectively attend to a specific sound source in the presence of other sounds has been a research topic of interest to psychologists and neuroscientists for several decades (Bee and Micheyl, 2008; Bronkhorst, 2015; Cherry, 1953; McDermott, 2009; Micheyl and Oxenham, 2010; Shinn-Cunningham and Best, 2008). For modern humans, an important everyday-life example of this ability can be found in multi-talker conversations.

Several previous studies have started to elucidate psychoacoustical, cognitive, and neurobiological processes involved in the ability of normal-hearing or hearing-impaired listeners, to selectively attend and decode a speech stream in the presence of concurrent streams, including competing voices. Broadly, auditory stream segregation and selective auditory attention likely play a crucial role as pre-requisites for good speech intelligibility in such listening situations (Alain, 2000; Bronkhorst, 2015; Cherry, 1953; Fritz et al., 2007; McDermott, 2009; Shinn-Cunningham, 2008). Psychoacoustic studies have outlined various acoustic cues, such as fundamental-frequency, sound-intensity, or source-location differences between the target and competing voices (Best et al., 2010; Oxenham, 2008; Popham et al., 2018), which the auditory system can use to separate concurrent voices, and to group or ‘track’ over time, sounds from a given voice.

In parallel to these behavioral investigations, brain-imaging studies have started to shed light on brain networks involved in speech comprehension in multi-talker listening situations. In particular, results indicate that this ability requires the involvement of a large fronto-temporo-parietal network (Evans et al., 2016; Fiedler et al., 2019; Hill and Miller, 2010; Szalárdy et al., 2019; Tóth et al., 2019; Zhang et al., 2018). Within this network, intelligible sentences appear to be processed mainly within the bilateral temporal cortices and the inferior frontal gyrus (IFG)

(Davis and Johnsrude, 2003; Okada et al., 2010; Peelle et al., 2010; Peelle, 2018; Rodd et al., 2005). In particular, activity in the left IFG has been found to be related to speech intelligibility, especially, in challenging listening tasks involving distorted speech or speech presented concurrently with other sounds, such as noise (Alain et al., 2018; Lawrence et al., 2018; Peelle, 2018; Wijayasiri et al., 2017; Wild et al., 2012). Greater prefrontal activity has also been found to be associated with higher speech intelligibility (Andéol et al., 2017), perhaps, reflecting greater attentional engagement (c.-à-d., target selection and/or masker inhibition). Consistent with the view that inhibition of attentional processes associated with non-target sounds plays an important role in task performance (Johnsrude and Rodd, 2016), errors in speech-intelligibility tasks with concurrent speech streams often stem from participants reporting masker words instead of target words (Brungart, 2001).

One difficulty, in interpreting the results of some of these previous behavioral and brain-imaging studies, stems from the fact that even in the simplest multi-talker situation, with only two talkers involved, speech intelligibility can be influenced by multiple factors, including physical stimulus (acoustic) differences between the target and the masker voices, inter-individual differences in the perception of these stimulus differences, as well as more-difficult-to-control cognitive variables, such as attentional and memory abilities. Any or all of these factors have the potential to influence the extent to which different brain regions are involved in task performance. Further advances in our understanding of the brain underpinnings of this important behavioral ability for normal- and hearing-impaired listeners, hinge critically on being able to disentangle those factors (Lee et al., 2014).

The present study takes advantage of an approach based on inter-individual differences in, and intra-individual changes of, perceptual strategies to address brain networks involved in performance of a speech-intelligibility task with two concurrent speech streams. Specifically, the approach builds upon earlier observations that, for some listeners, speech intelligibility improves as the difference in sound level between the target and masker voices (aka, the target-to-masker ratio, TMR) increases monotonically from negative (adverse) to positive (favorable). For other listeners, intelligibility is actually highest for negative and positive TMRs, and intermediate for TMRs close to zero, that is to say, when the target and masker voices have approximately the same sound level (Andéol et al., 2017; Brungart, 2001; Brungart and Simpson, 2007; Cooke et al., 2008; Thompson et al., 2015). Although the latter finding may initially seem counter-intuitive, it can easily be accounted for by considering that, when the target and masker voices have different sound levels, listeners can in principle take advantage of these differences to aid in the identification of target and masker sounds. However, this

perceptual strategy is only available to listeners who perceive the sound-level difference and who are actually able to use it as an advantage. Here, we posited that listeners' ability to exploit this perceptual strategy could be enhanced via the provision of specific instructions, practice (perceptual training) and feedback to participants. The fact that speech intelligibility can improve with perceptual training is well established (Eisner et al., 2010; Golestani and Zatorre, 2004; Sehm et al., 2013; Sohoglu and Davis, 2016). Moreover, it is known that changes in perceptual strategies and improvements in performance in multi-talker speech-intelligibility tasks can be achieved, also, by providing participants with specific instructions and/or feedback (Brungart and Simpson, 2004; Thompson et al., 2015). In the present study, we used a combination of instructions, feedback, and task practice to influence participants' listening strategies. This way, we were able to study changes in brain activity related specifically to changes in task performance, unconfounded by concomitant physical-stimulus changes.

Using this behavioral paradigm, we tested the hypothesis that fronto-temporo-parietal cortical activity is related to speech intelligibility in a multi-talker situation. To this aim, we used functional near infrared spectroscopy (fNIRS). fNIRS has been used successfully to study brain processes involved in language processing before (Quaresima and Ferrari, 2019).

The study consisted of two experiments. In a first experiment (observational phase), initially naïve and untrained participants performed a speech-intelligibility task while their cortical activity was recorded. We predicted that participants' speech intelligibility scores in an adverse (negative) TMR condition would co-vary with prefrontal cortex activity (Andéol et al., 2017) and with a larger fronto-temporo-parietal network activity (Alain et al., 2018; Peelle, 2018; Wild et al., 2012). In this second experiment (interventional phase), participants who had the poorest speech-intelligibility scores in experiment I were recruited and trained (through instructions, task practice, and feedback) to take advantage of sound-level differences between the target and masker voices. We predicted that, following this training, speech intelligibility would improve and that cortical activity would vary concomitantly. Based on earlier findings (Golestani and Zatorre, 2004; Qi et al., 2019; Sohoglu and Davis, 2016), one hypothesis was that IFG activity would decrease bilaterally, reflecting improved efficiency in inhibition of the masker voice. A second hypothesis was that, for conditions where cognitive engagement was initially low (i.e. adverse TMR), cortical activity in left IFG would increase after training, reflecting greater engagement of cognitive resources for task performance.

MATERIALS AND METHODS

Participants

Seventy-four adult volunteers (20 females, mean age = 23.6 ± 3.2 years) were recruited to take part in the study. All were students from ISAE-SUPAERO and Toulouse University with normal hearing, defined as pure-tone hearing thresholds better than 20 dB hearing level (HL) at octave frequencies from 125 Hz to 8 kHz (Audio Console Oscilla® headphones with built-in software).

The study was cleared by the local ethics committee (Comité de Protection des Personnes sud-ouest et Outre Mer II, IDRCB 2017-A00859–44). Before starting the study, all participants completed a written consent form. They were paid 15 euros for their participation in the first experiment; participants who additionally completed a second experiment received 60 euros in total.

Speech Stimuli

Stimuli used in the speech-intelligibility task were sentences from the Coordinate Response Measure (CRM) corpus (Bolia et al., 2001). CRM sentences are all constructed using the same pattern: “Ready call sign go to color number now”, where call sign, color, and number are placeholders. The actual call sign, color, and number played during a sentence are drawn at random, and independently, from a list of 8 call signs (Baron, Charlie, Ringo, Eagle, Arrow, Hopper, Tiger, Laker), 4 colors (red, blue, green, white) and 8 numbers (from ‘one’ to ‘eight’) resulting in 256 possible sentences. The sentences used in this study were uttered by four male speakers.

For the purpose of the present study, the sentences were always presented in pairs, with one target sentence indicated by the call sign, Baron, and a concurrent (simultaneously presented) masker sentence indicated by one of the seven remaining call signs. The call sign used for the masker sentence was drawn at random from the seven available call signs.

Three conditions, corresponding to three different TMRs, were tested: an ‘adverse’ condition (TMR = -12 dB), an ‘intermediate’ condition (TMR = -4 dB), and a ‘favorable’ condition (TMR = 4 dB). Stimuli corresponding to one of the three conditions were presented consecutively to the participant in a block of 32 pairs of sentences, before the next condition was tested, also as a block. The testing order of the three conditions was randomized across participants. Each stimulus (a pair of simultaneous sentences) lasted two seconds. After each stimulus, participants had to report the color and the number corresponding to the target sentence. They provided

answers using a home-made response box. To allow sufficient time for measurement of the hemodynamic response, a minimum inter-stimulus-interval (ISI) of 18 s was enforced.

Stimuli were generated using custom software written in Matlab® (2018b). The audio files (.wav) were transmitted from the computer (Dell Optiflex 990) to a digital signal processor (RX8; Tucker Davis Technologies) to be filtered in real time with HRTFs measured using an artificial head (Neuman KU-100). The signal was preamplified (HB7, TDT) and presented binaurally through headphones (Beyer Dynamics DT770) at an overall level of approximately 70 dB SPL (rms). The target and masker sentences were filtered by the same, frontal HRTF, such that they appeared to the listener to both be coming from straight ahead (0 degree). On each trial, the SPL of combined target-plus-masker signal was drawn at random from a 6-dB range (from 67 to 73 dB SPL, in 1-dB steps). Behavioral responses were recorded using a homemade response box with three rows of buttons: a first row for the colors, a second row for numbers one to four, and a third row for numbers five to eight. The response box was linked to the real-time processor and the responses were recorded with Matlab.

fNIRS data collection

Cortical activity was assessed with a whole-head fNIRS device NirScout (Nirx <https://nirx.net/nirscout/>). A 15-by-23 optode array, including 16 emitters ($\lambda_1=760$ nm; $\lambda_2=850$ nm) and 23 detectors, yielding 44 measurement channels in total (Figure 1.A), was used. The source-detector spacing was maintained at approximately 30 mm using plastic spacers. The frequency sampling was 7.81 Hz. Data were acquired using NIRSTAR v14.2 software. Channel positioning was made using the international 10-20 EEG system. The 15 sources were positioned at Fpz, AF7, AF8, F3, F4, FC5, FC6, T7, T8, TPP7h, TPP8h, P3, Pz, P4, Oz; the 23 detectors were positioned at Fp1, Fp2, AF3, AF4, F5, F6, FT7, FC3, FC4, FT8, C5, C6, TP7, TP8, P5, P1, P2, P6, PO1, POZ, PO2, O1, O2.

Cortical sensitivity maps were computed using a Monte-Carlo simulation algorithm within Atlas Viewer (V2.8) (Figure 1.C). Default parameters were used, except for the number of photons, which was set to $1E7$ (Aasted et al., 2015). Thirteen brain region of interests (ROIs) were defined (Figure 1.B and Table 1), corresponding to: Prefrontal Cortex (PFC), Dorsolateral Prefrontal Cortex (DLPFC), Inferior Frontal Gyrus (IFG), Auditory Cortex (AC), Superior Parietal Lobe (SPL), Inferior Parietal Lobe (IPL), and Visual Cortex (V1). The spatial resolution of fNIRS provided a brain observation area of roughly 1 cm^2 around the center of the Montreal Neurological Institute (MNI) coordinate system (Evans et al., 1993).

ROI	Left Hemisphere					Right Hemisphere				
	Channels		MNI Coordinates			Channels		MNI Coordinates		
	Source	Detector	x	y	z	Source	Detector	x	y	z
PFC	S1	D1	-12	72	1	S1	D2	15	70	0
	S3	D1	-25	58	2	S2	D2	36	67	5
	S5	D3	-23	49	24	S4	D4	23	44	26
DLPFC	S5	D9	-40	32	31	S4	D10	46	23	32
	S5	D5	-36	35	18	S4	D6	50	-40	-25
IFG	S7	D7	-42	9	16	S6	D8	48	9	10
	S7	D9	-57	24	34	S6	D10	46	13	25
	S7	D5	-55	34	22	S6	D6	46	24	19
AC	S9	D11	-50	-12	9	S8	D12	64	-10	12
	S9	D13	-58	-20	8	S8	D14	51	-20	12
SPL	S13	D17	-33	-69	54	S12	D18	39	-70	53
	S13	D19	-36	-76	47	S12	D20	39	-75	44
IPL	S13	D15	-50	-12	9	S12	D16	43	-64	33
V1	S15	D23	-10	-97	6	S15	D22	20	-99	7
						S15	D21*	3	-96	24

Table 1. fNIRS ROIs and measurement channels MNI coordinates (x, y, z of the centroids) extracted from Atlas viewer (V2.8). *The S15-D21 channel is actually centered between the two hemispheres, but is presented here as part of the right hemisphere to simplify the table layout.

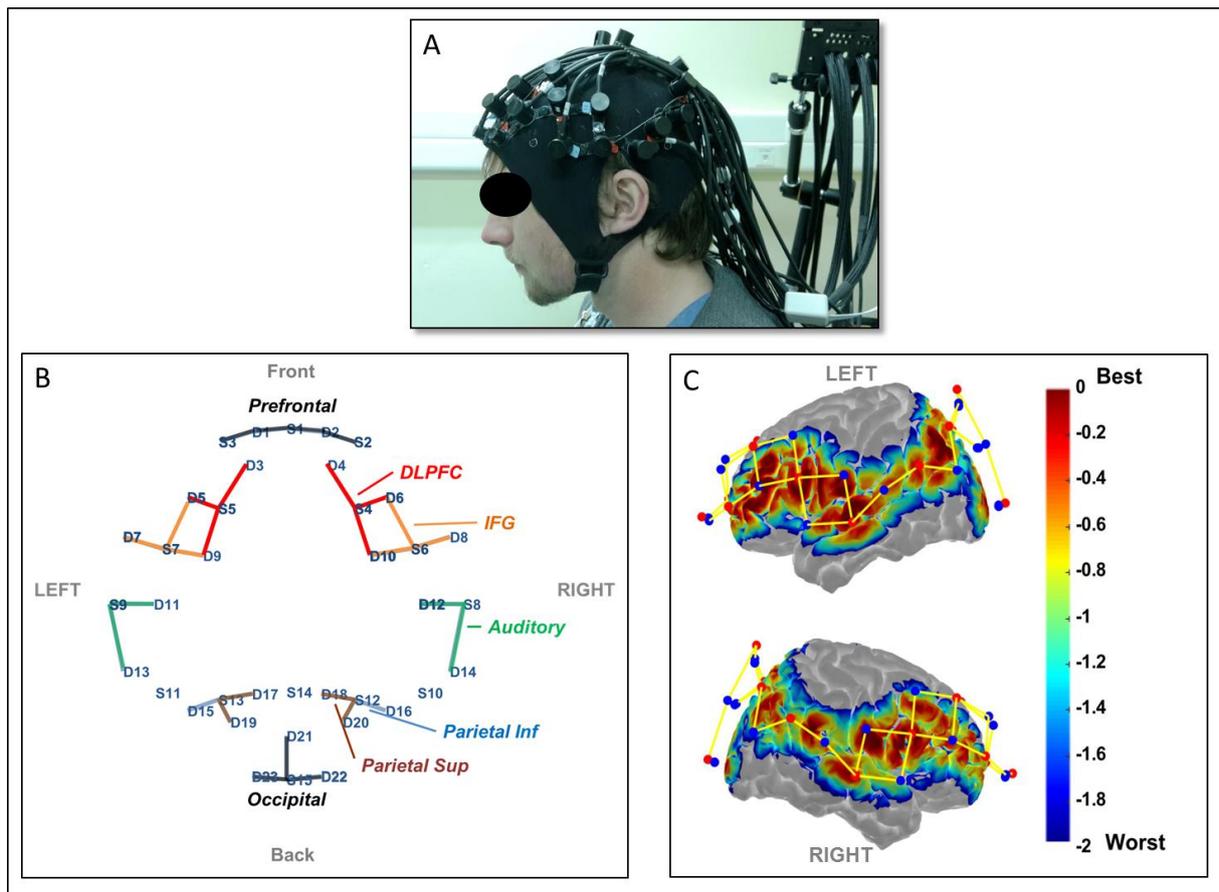


Figure 1. A) Photography of the optode-array holder placed on the head of a participant for demonstration purposes. (B) 2D representation of ROI location; each line represents a measurement channel. (C) Left (up) and right (down) views of the cortical sensitivity (in mm⁻¹) maps obtained using Monte-Carlo simulation within AtlasViewer (V2.8; default parameters except for the number of photons that has been set to 107).

fNIRS preprocessing

Data were processed using the Matlab-based NIRS AnalyzIR toolbox (Santosa et al., 2018) V.615, with Matlab R2018b. First, raw fNIRS data were imported in Matlab, and a demographic table was created based on the NIRstar demographic inputs and on task-performance results. Stimulus duration was set to 2 s, corresponding to the per-sentence duration of CRM sentences. Raw data were then converted into optical density and submitted to a Temporal Derivative Distribution Repair (TDDR) for motion correction (Fishburn et al., 2019). These authors have shown that this correction is highly effective in correcting movement artifacts in fNIRS data. Concentration changes for oxygenated and deoxygenated hemoglobin were computed using the modified Beer-Lambert law with a partial path length factor (ppf) of 0.1 for both wavelengths. Following Pinti et al. (2019) recommendations, to minimize systemic component effects, data were bandpass corrected using a fourth-order bandpass Butterworth filter with a high-pass frequency of 0.01Hz and a low-pass frequency of 0.09Hz. A final pre-processing step involved principal-component-analysis (PCA) filtering that was used to filter-out the two main systemic components (c.-à-d., cardiac activity and Mayer waves).

Experimental design

This study included two experiments. A first experiment, Experiment I, involved all 74 participants. This was followed by a second experiment, Experiment II, which involved only a subset of participants from Experiment I, selected based on their performance in the adverse condition of Experiment I. Experiment II involved a training program over three consecutive days (T1, T2, T3), followed by one final test session, which was similar to the single test session of Experiment I (Figure 2).

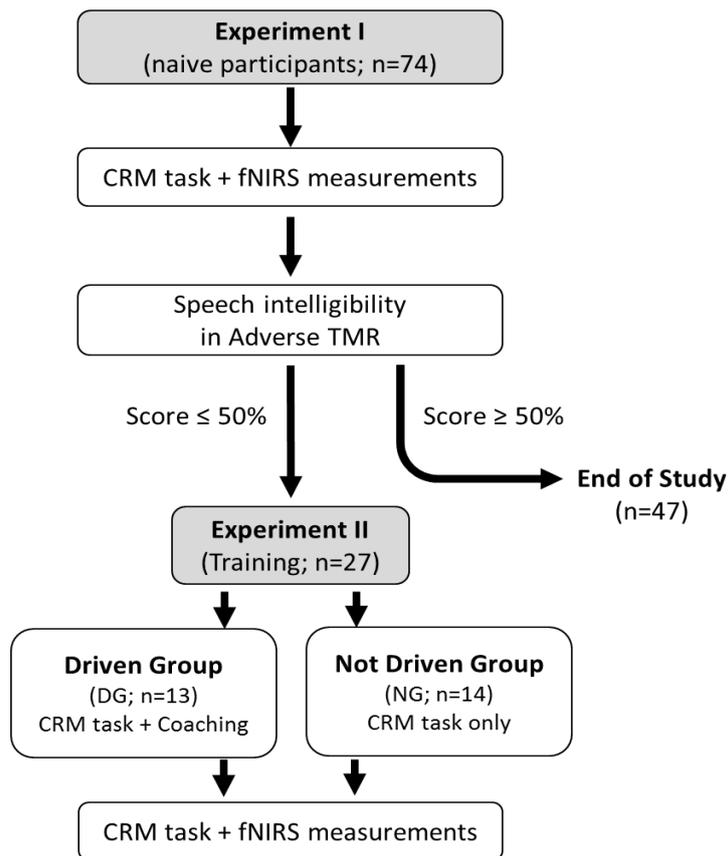


Figure 2. Schematic of the experimental design for experiments I and II.

EXPERIMENT I: NAIVE PARTICIPANTS

METHODS

Participants in Experiment I performed the concurrent-speech intelligibility task described above while their brain activity was recorded with the 16-by-23 optode NIRSCOUT fNIRS device (Nirx <https://nirx.net/nirscout/>). Before fNIRS data acquisition, all participants performed one short training block of 16 speech stimuli at 0-dB TMR to verify correct understanding of the instructions and provide basic familiarization with the task. This experiment lasted approximately one hour.

STATISTICAL ANALYSIS

Behavior

A response was considered as correct when both the color and the number were correctly reported. The correct response rate for each condition was transformed using the Rationalized Arcsin Unit (RAU) transform (Studebaker, 1985); this transform yields results more suitable for parametric statistical analyses than untransformed percent-correct data.

A preliminary Mauchly's Sphericity test came up significant for the TMR factor ($\chi^2=11.24$; $p<0.01$), hence a multivariate analysis of variance (MANOVA) was used to analyze RAU scores, with TMRs (Adverse vs. Intermediate vs. Favorable) as a within-subject factor. All behavioral statistical analyses were performed using the software STATISTICA© (V10) from StatSoft.

fNIRS

Task activation

Task activation was quantified by convolving the fNIRS signal with a canonical hemodynamic response function (HRF) for each condition (adverse, intermediate and favorable TMRs). Following Barker et al. (2013), beta coefficients were then estimated using an autoregressive iteratively reweighted least squares (AR-IRLS) algorithm; this algorithm was used to account for the presence of serial correlations in fNIRS data and to control for Type-1 errors. The estimated betas for the three conditions were submitted to a robust mixed-effects model, with condition as a fixed effect and subject as a random effect. Lastly, statistical differences in the weighted betas across conditions were assessed using multiple two-sample t-tests. Multiple comparisons across ROIs were controlled for using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995), and p-values were corrected using the false discovery rate (FDR).

Relationship between cortical activity and speech intelligibility scores

The relationship between speech intelligibility scores and brain activations was assessed using second-level robust mixed-effects models for each condition separately, with condition as a fixed effect and speech-intelligibility scores as regressors.

RESULTS

Behavior

Behavioral results showed a significant main effect of TMR [$F(2, 146)=90.63$, $p< 0.001$; see Figure 3]. Post-hoc comparisons with Tukey's HSD test showed higher performance ($M = 76.03$; $SD = 14.93$; $p<0.001$) with the favorable TMR compared to the intermediate ($M = 31.56$;

SD = 26.85) and adverse (M = 35.37; SD = 28.95) TMRs. No statistically significant difference was observed between the intermediate and adverse TMRs ($p = 0.5$).

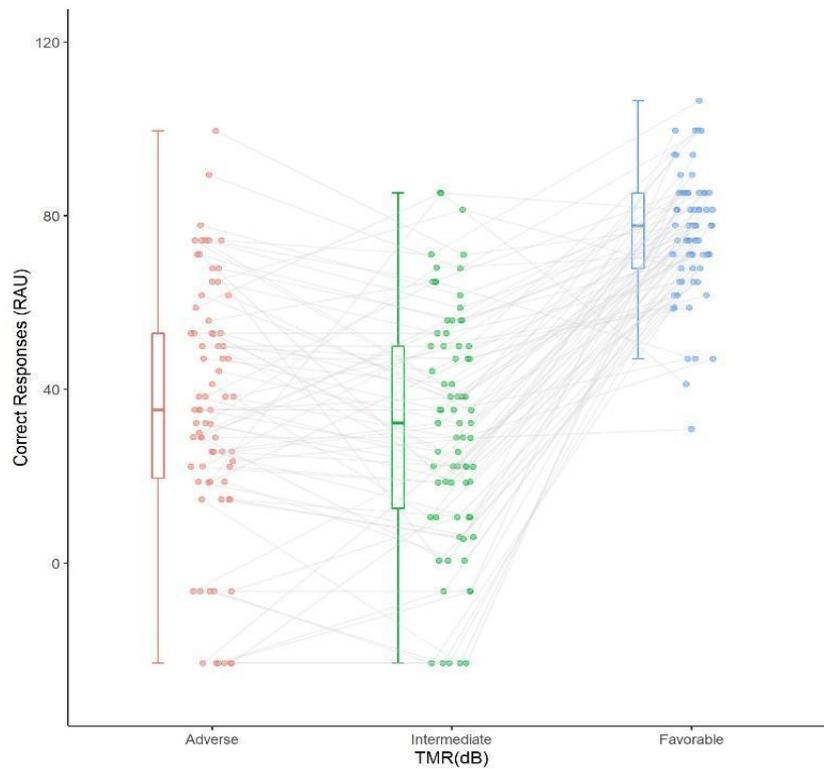


Figure 3. Correct responses (Rationalized Arcsin Units) as a function of TMR in Experiment I. Boxplots show the median (horizontal bar) and the interquartile range (boxes). The whiskers reach from the lowest to the highest observed value within 1.5 times the interquartile range. Dots show individual data points. Individual data points corresponding to the same listener across different conditions are connected with grey lines.

fNIRS

Task activation

Greater activation was observed for the favorable-TMR condition than for the adverse-TMR condition in the left IFG ($p < 0,05$; FDR-corrected). By contrast, in the left IPL, significantly weaker activation was observed for the favorable-TMR condition compared to the adverse-TMR condition ($p < 0,05$; FDR-corrected). No other contrast in any region showed significant results. The significant t-test contrasts between the adverse and favorable TMR conditions are shown in Figure 4.

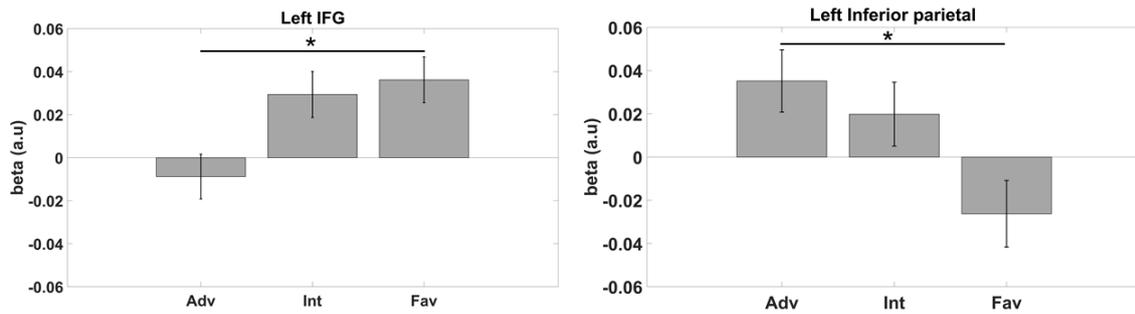


Figure 4. Hbo activations for the left IFG and the left inferior parietal ROIs. *: $p < 0,05$ (FDR corrected) for the number of ROI areas. Vertical bars denote the standard error of the mean.

Relationship between cortical activation and intelligibility scores

To test for relationships between the cortical activation and the behavioral performance, a robust regression was carried-out using the weighted betas (c.-à-d., cortical activation) as the dependent variable and the RAU score (c.-à-d., behavioral performance) as the independent variable. For the adverse-TRM condition, statistically significant positive correlations between these two variables were observed within a large fronto-temporo-parietal network including: prefrontal cortex (left: $r=0,18$; $p<0,05$; right: $r=0,17$; $p<0,05$), left auditory cortex ($r=0,18$; $p<0,05$), right IFG ($r=0,17$; $p<0,05$), and left IFG ($r=0,17$; $p<0,05$). A negative correlation was also found between the activation of the left SPL and RAU scores ($r = -0,17$; $p<0,05$); (see Figure 5). For the intermediate and favorable conditions, the same robust-regression analyses failed to show statistically significant linear associations between RAU scores and cortical activation.

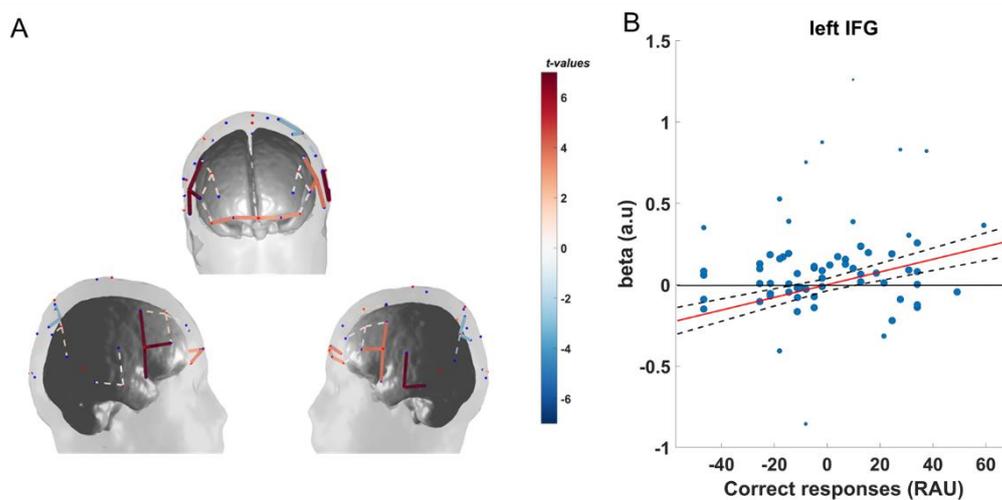


Figure 5. Illustrations of the relationship between cortical activity (beta coefficients for Hbo) and speech intelligibility (RAU scores) for the adverse condition. Left, 3D representation of the channels significantly correlated to the RAU scores (solid lines; $p<0,05$ FDR corrected): left and right IFGs ($r=0,17$ and $r=0,19$ respectively), left auditory cortex ($r=0,18$), prefrontal

cortex ($r=0,16$) and the left SPL ($r=-0,17$). Right, scatter plot of the robust linear regression results for the left IFG. Size of points represents the robust regression weights (c.-à-d., outliers).

DISCUSSION I

In Experiment I, a speech-in-speech listening test was used to explore the relationship between speech intelligibility and cortical activity. As previously shown, intelligibility at the group level was higher in the favorable TMR but did not decrease from intermediate to adverse TMR. Higher masker level provided indeed larger level cues, facilitating the segregation of the voices; allowing the inhibition of the masker and the selection of the target talker (Andéol et al., 2017; Brungart, 2001). However, large individual differences revealed that many listeners were not able to achieve such behavior (Figure 3).

Task activation

Significant effects of TMR over brain activity have been found over two main brain areas: the left IFG and the left IPL. Greater activation was observed for the favorable-TMR condition compared to the adverse-TMR condition in the left IFG while lower activation was observed in the left IPL for the same comparison.

The task used in this study likely engages auditory working memory (i.e. phonological short term memory) which is necessary to manage colors and numbers under challenging situations. The role of left IPL in auditory working memory has been already assumed (Baldo and Cronkers, 2006; Buchsbaum and D'Esposito, 2008; Ravizza et al., 2004; Yue et al., 2019) and even considered as part of the phonological loop (Buchsbaum and D'Esposito, 2008). Indeed, the phonological loop would be responsible for the storage and maintenance of information in a verbal form. It is considered as the central component underlying working memory for verbal material (Buchsbaum, 2015). The increased activation in the adverse-TMR condition can be thought as a higher demand of working memory resources due to the more challenging situation. This could explain the larger parietal activation in the adverse condition compared to the favorable condition.

On the other hand, the significant higher activation of the left IFG in the favorable-TMR condition compared to the adverse-TMR condition could be interpreted in different ways depending on the literature: the difference of the acoustic stimulation like the increase in TMR

(Alain et al., 2018; Bauernfeind et al., 2018); the increase in intelligibility per se (Davis and Johnsruide, 2003), the increase of activity of neural processes involved in the inhibition of the masker talker (Johnsruide and Rodd, 2016), or the increase in cognitive engagement to reach better speech intelligibility (Lawrence et al., 2018). Since the relationship between speech intelligibility and cortical activation is unclear when considering differences between TMRs, we assessed it at the individual level in a regression analyses.

Relationship between speech intelligibility and brain activation

The relationship between speech intelligibility scores and brain activation within each condition led to the involvement of a large fronto-temporo-parietal network for the adverse TMR condition only. We found a positive correlation between the left IFG activation and speech intelligibility scores in the adverse condition (Figure 5). This result is consistent with the TMR effect over the left IFG activation (Figure 4): higher activation goes with higher speech intelligibility. The literature has already established the role of the left IFG in speech-in-noise paradigms (Alain et al., 2018). We suggest now that the left IFG is involved in speech intelligibility whatever the masking type: energetic (speech-in-noise) or informational (speech-in-speech). In addition, we also found a significant relationship between right IFG activation and speech-intelligibility scores in the adverse condition that was not expected according to the previous neuroimaging speech studies. We propose here two plausible explanations of this result.

First, in our informational masking task, performance may hinge critically on the separation of the target and masker signals based on the level cues that differentiate these two signals. Previous studies have shown pronounced changes of activation in bilateral temporal cortex and IFG with increasing sound level (Bauernfeind et al., 2018; Bilecen et al., 2002; Jäncke et al., 1998; Neuner et al., 2014).

Second, once the separation occurs, speech intelligibility in informational masking requires the inhibition of the masker talker to focus on the target talker. Previous study also found that the right IFG plays a central role in inhibition processing in a Stop Signal or Go-No Go task for instance (Hampshire et al., 2010). Indeed, inhibitory control is necessary to suppress irrelevant cognitive processes or motor plans (Aron et al., 2004) and maintaining and adapting goal directed behaviors in changing situations (Aron, 2007; Dillon and Pizzagalli, 2007). This ability relies on a right-lateralized fronto-basal brain network that includes the right IFG (Aron, 2007; Aron et al., 2004; Hampshire et al., 2010).

Furthermore, the literature about challenging listening conditions highlights the implication of a larger fronto-temporo-parietal network not limited to the bilateral IFG for optimal speech intelligibility (Fiedler et al., 2019; Hill and Miller, 2010; Szalárdy et al., 2019; Tóth et al., 2019). While the higher activation of the auditory cortex is more linked to the acoustic features of the stimulus, the prefrontal and parietal regions show a stimulus-independent activation (Peelle, 2018; Peelle et al., 2010). This suggests that the involvement of these brain areas can be attributed to higher order processes of attention, monitoring and memory (Andéol et al., 2017; Peelle, 2018). Nevertheless, the reported explained variances of these correlations are modest. To provide stronger evidence of the left and right IFGs role in speech-in-speech processing and assess the four previously mentioned hypotheses (c.-à-d., increase in TMR, increase in speech intelligibility, increase in inhibition of the masker, or increase in cognitive engagement), we hence designed a training paradigm in Experiment II using an approach based on perceptual training and/or coaching. We have determined that task performance in the adverse-TMR condition could be modified (specifically improve), in participants who initially had not performed well in this condition, by (a) drawing their attention to the voice level differences (c.-à-d., the target voice had a lower sound level than the masker voice: level cues coaching), and (b) letting the participants to further practice the task (training). Importantly, any improvement in behavioral task performance and/or cognitive engagement, which might result from such coaching or training, could not be ascribed to a change in the physical characteristics of the signals since TMRs remained the same before and after training/coaching. Therefore, any change in cortical activity between before and after this intervention would likely be indicative of a true association between cortical activity and behavioral performance.

EXPERIMENT II – TRAINED PARTICIPANTS

METHOD

Twenty-seven participants from Experiment I who achieved less than 50% of correct responses in the adverse condition were recruited to take part in Experiment II. Two types of training were tested with two subgroups of participants. The first group, hereafter referred to as the ‘*not-driven group*’, simply repeated the same task as in Experiment I. The second group hereafter referred to as the ‘*driven group*’, received explicit verbal information that was meant to facilitate task performance by making it easier for the participants to distinguish the target from the masker. Specifically, prior to a block of trials, participants from the *driven group* were

informed whether the target voice would be softer (negative TMR condition) or louder (positive TMR condition) than the masker voice. Moreover, auditory feedback, consisting in the target sentence without the masker, was systematically presented after each trial and participants were asked to validate their answer one more time. The presentation of the target sentence in isolation was meant to facilitate learning of relevant perceptual cues and strategies for more easily identifying and hearing-out the target sentence from the target-plus-masker mixture.

All participants were randomly assigned to the *not-driven* or *driven group*. The training program was made of three consecutive days of training session (T1, T2 and T3). Each session lasted about 20 minutes and no fNIRS recordings were made during them. At the end of the training program, all 27 participants performed the same task as in the Experiment I with fNIRS recordings. This session, corresponding to the post-training session, lasted 30 minutes.

STATISTICAL ANALYSIS

Behavior

Mauchly's sphericity tests on the RAU scores yielded significant results ($\chi^2=17,43$; $p<0,001$ for TMR, and $\chi^2=8,56$; $p<0,05$ for the TMR-by-session interaction). Accordingly, two-by-three MANOVAs were used to analyze the RAU scores, with Session (Pre vs. Post training) and TMRs (Adverse vs. Intermediate vs. Favorable) as within-subject factors and Group (*not-driven group* vs *driven group*) as a between-subject factor. All post-hoc comparisons involved Tukey's honestly significant difference (HSD) tests. All behavioral statistical analyses were performed using the software STATISTICA© (V10) from Stat Soft and a significant threshold was set to 0,05.

fNIRS

fNIRS analyses were performed identically to Experiment I. Since behavioral analyses showed no group differences, all fNIRS results presented hereafter refer to the whole population (*driven + not-driven groups*).

RESULTS

Behavior

No significant effect of group or interaction between group and other factors was found ($p \geq 0.09$ in best case; Figure 7). These results indicate that the two types of training did not lead to any post-training behavioral differences. However, a significant interaction between TMR and session was found ($F(2,52)=12.14$; $p<0.001$, see Figure 6). Tukey's HSD post-hoc tests showed a significant training effect for the three TMR conditions ($p<0.001$ for all TMRs). Therefore, the interaction reflects a smaller session effect within the favorable condition compared to the two other conditions.

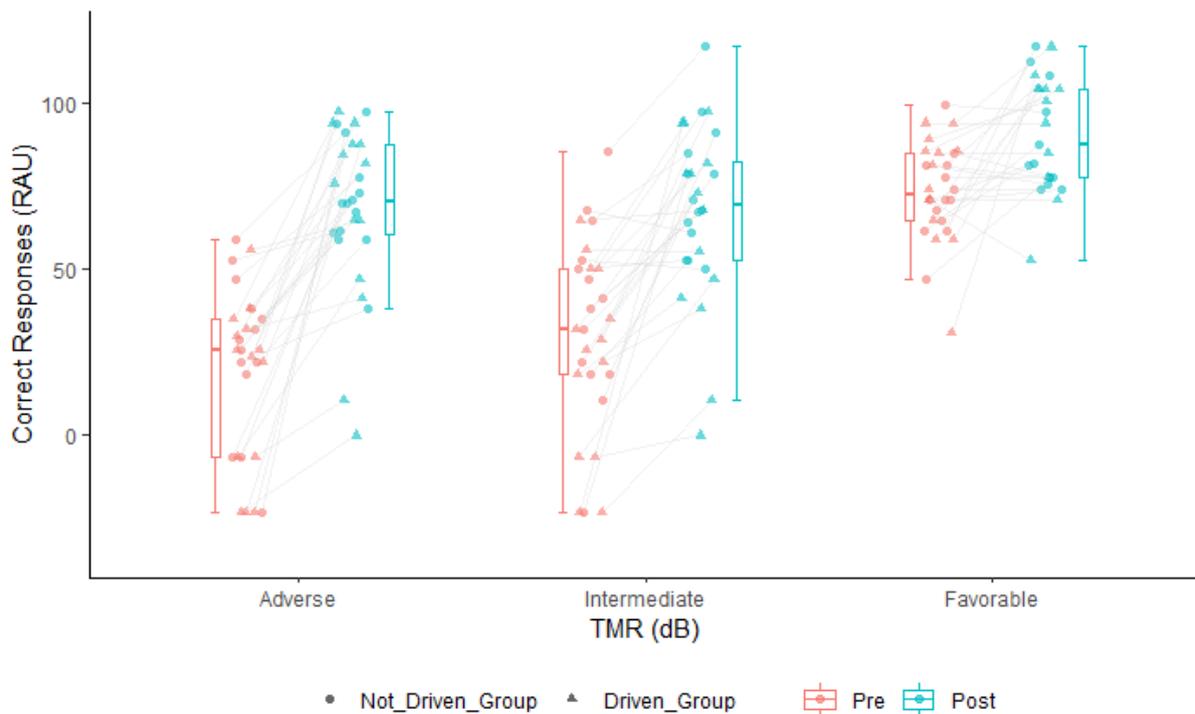


Figure 6. Speech intelligibility scores (RAUs) for each TMR (Adverse, Intermediate, Favorable), training session (Pre, Post), training group (Driven, Not-Driven). Boxplots show medians (horizontal bars) and interquartile ranges (boxes). Whiskers reach from the lowest to the highest observed value within the 1.5 interquartile range.

A main effect of session was found, consistent with a significant increase in speech intelligibility after the training ($M_{pre}=41.25$, $SD=3.18$; $M_{post}=77.20$, $SD=3.80$; $F(1,26)=72.2$; $p<0,001$).

fNIRS

T-contrasts comparing pre-training with post-training results showed a significant decrease in bilateral IFG activation after training ($p < 0,05$; FDR corrected) for the adverse-TMR condition (see Figure 7). These contrasts also revealed a decrease of right IFG activation following training for the intermediate-TMR condition. No ROI showed a significant change with training for the favorable condition.

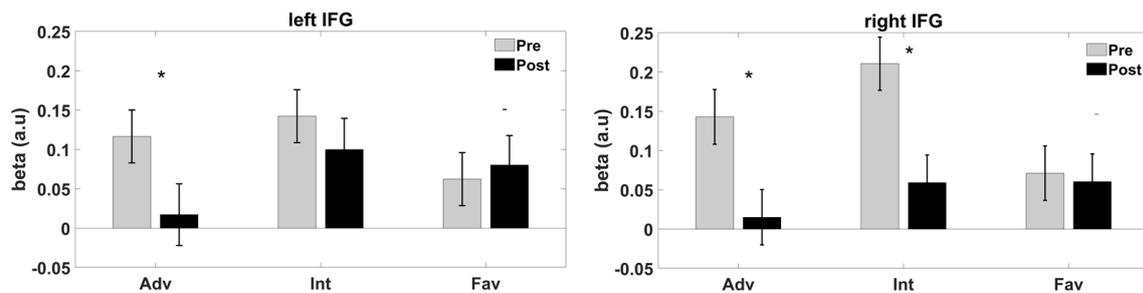


Figure 7. Effects of training on left and the right IFG activations for each TMR condition (Adv: adverse, Int: intermediate and Fav: favorable). *: $p < 0,05$ (FDR corrected).

DISCUSSION II

The findings of similar, statistically significant, improvements in behavioral performance in the *not-driven* and the *driven groups* indicate that the provision of explicit instructions and feedback regarding response correctness did not yield greater performance improvements than those afforded by supplementary task practice alone. The improvements observed may have been due, at least in part, to further learning of differentiating the talkers' voices characteristics by the participants (Johnsrude et al., 2013). In addition, or instead, participants may have developed their own strategy to achieve the task (Thompson et al., 2015), probably based on inhibitory control.

In the present study, once the two voices are segregated, the listener has to select the target voice and inhibit the masker voice. Besides, in the adverse condition, the masker voice is louder and therefore easier to select. However, the listener has to inhibit this natural trend in order to focus on the weaker target voice. Thus, inhibitory control seems to be crucial for good speech intelligibility in our informational masking task. The development of an efficient inhibitory control has already been demonstrated during a short period of training without feedback or explicit instructions for well-known inhibition tests (p. ex., Go/NoGo, Stop Signal, Stroop task), in EEG (Manuel et al., 2013) and fMRI (Chavan et al., 2015) experiments.

In the present study, the training led to a decrease of bilateral IFGs activity within the adverse-TMR condition, as well as a decrease of right IFG activity in the intermediate-TMR condition. This contrasts with previous studies and our first experiment results that showed greater cortical activation associated with higher speech intelligibility (Eisner et al., 2010; Sehm et al., 2013). We can argue that measurement of cortical activity during the first training sessions could have also shown greater activation of IFGs that would have subsequently decrease due to automatization mechanisms. Previous studies have found such initial cortical activity increase, followed by a progressive decrease during training (Hempel et al., 2004; Qi et al., 2019). In Qi et al.'s study, the authors measured the cortical activity during a discrimination tasks with speech (Mandarin) tones and sinewave tones, before and after intensive Mandarin course (c.-à-d., listening, reading and writing) in native English-speakers. They showed that participants exhibiting a larger decrease of right IFG activity showed better long term retention of Mandarin language. A similar pattern, with an increase followed by a decrease of IFG activity during training, has also been observed in an fMRI study using a visuo-spatial working memory task (Hempel et al., 2004). Finally, decreased IFG activity has also been found in studies involving inhibitory-control tasks (Hartmann et al., 2016; Manuel et al., 2013).

The role of inhibitory control in the task is the second major difference between our study and previous studies that reported better speech intelligibility associated with a higher IFG activation across conditions. Indeed, we used a speech-in-speech task with two talkers producing intelligible sentences whereas most of the previous studies used degraded speech with only one talker (Eisner et al., 2010; Sehm et al., 2013). As a consequence, the inhibition of the loud masker talker, that plays a major role in our task achievement, is not required in the former studies since it doesn't exist.

At the neural level, inhibitory control has been associated with a consistent network of brain areas involving mainly the frontal cortex and more specifically, the bilateral IFG (Aron et al., 2014; Chavan et al., 2015; Jonides et al., 1998; Swick et al., 2011; Xue et al., 2008). Inhibitory control belongs to a domain-general system that manages the ability to inhibit interference from task-irrelevant information (Aron et al., 2004; Miyake et al., 2000). Chavan et al. (2015) for instance, used a Go/NoGo proficiency task to study the effects of a two-week training on frontal top-down inhibitory control mechanisms. They found that behavioral improvements were associated with decreased activities in bilateral IFG, consistent with our training results. They argued that participants presented a minor engagement of neural populations involved in the inhibitory process to achieve the task.

As an explanation, some authors suggest that neural modifications (p. ex., synaptic weight updates), leading to more efficient and lower neural activity, would occur during the training. We suggest now that similar mechanisms were involved in the present study, and that it may explain the decreased activity of the IFG following the training.

GENERAL DISCUSSION

The present study aimed to investigate the neural mechanisms underlying the ability of human listeners to selectively attend to, and understand, a talker's voice in the presence of a competing voice—a particular case of a more general ability, sometimes referred to as 'cocktail party' listening. To address this question, two complementary approaches were used. In a first experiment, correlations between behavioral performance and cortical activity were assessed for three conditions, corresponding to different TMRs, yielding different levels of task difficulty. For the adverse-TMR condition only, correlations were observed, indicating greater cortical activity with higher performance. Importantly, for some participants, performance was lower in this adverse-TMR condition than in the two others, more favorable TMR conditions (as expected based on the level of the masker). For other participants, performance was higher in the adverse-TMR condition than in the intermediate-TMR condition. The latter arguably took advantage of the sound level difference between the target and the masker sentences to identify the target. Consequently, a second experiment used a different approach to further investigate the relationships between task performance and cortical activity. In this approach, training, specific task instructions, and feedback were used to improve the performance of participants who had initially performed poorly in the adverse-TMR condition in the first experiment. Changes in cortical activity were observed, related to changes in behavioral performance with training, providing further evidence for the existence of relationships between these two variables. Together, these two approaches, observational (Experiment I) and interventional (Experiment II), provide additional support for the hypothesis proposed in earlier studies (Alain et al., 2018), that left IFG is involved in challenging listening situations.

The increasing activation of left IFG across the three TMR conditions observed in Experiment I could be related to increasing TMR (a physical sound parameter), to increasing speech intelligibility (a measure of perceptual performance), the increasing activity of the neural processes related to the inhibition of masker talker, or to increasing cognitive engagement (c.-à-d., a cognitive effect). The findings in Experiment II of increasing speech intelligibility with training despite unchanged TMRs, rule out the first above explanation. Furthermore, the

decreased activity within the left IFG in the adverse-TMR condition after the training argues against a direct positive relationship between speech intelligibility and left-IFG activation.

Moreover, if the cognitive engagement effect was the main explanation, one would expect that correlation between left IFG activity and speech intelligibility found in the Experiment I would be interpreted as a result of low engagement of cognitive resources in conditions where performance was lowest: in those conditions those participants just ‘stopped trying’. As a consequence, training would have conducted to increase the cortical activity by increasing the engagement in poor listeners, but instead a decrease of cortical activity has been observed.

This leaves the neural activity related to masker inhibition as the most plausible explanation for the combined results of Experiment I and II. By associating learning with synaptic weighting updates and lower neural activity (Bless et al., 2014; Curtin et al., 2019; Fairclough et al., 2018; Putkinen and Saarikivi, 2018), the neural efficiency theory provides a comprehensive framework for the dual observation of a better performance and a lower brain activity in terms of a lower cognitive engagement. In line with the assumptions of this theory, Bless et al., (2014) found that training in a dichotic task led to better performance and lower cortical activity in brain regions associated with selective auditory processing. A decrease of neural activity resulting from training likely results in cognitive-resource savings that can be used subsequently to perform concurrent tasks. This could be assessed, for instance, in future studies using a dual-task paradigm.

Finally, the multi-talker task used in this study involves inhibition of the masker’s voice to focus on the talker’s voice, suggesting that training benefits may stem from better inhibitory control. Indeed, the inhibitory control is particularly relevant here since the louder masker in the adverse TMR condition is much easier to report than the target talker. Hence, the listener has to struggle against this natural trend to perform correctly. Nevertheless, a broader role of inhibitory control in generic speech-in-speech task has to be confirmed. The crucial role of inhibitory control may be limited to multi-talker situations with large level differences between talkers. For instance, it is expected that no natural trend would be in favor of a given talker if talkers differ only by their gender. However, inhibitory control would help, at least, to limit the interference of the masker talker in any multi talker situations (Strauß et al., 2014). Future studies could target the role of inhibitory control thanks to specific protocols. For instance, an experiment could reinforce the inhibitory control by training with a non-auditory task, and then test whether that training can improve the speech-in-speech intelligibility. Reciprocally, other experiment could test the effect of speech-in-speech training on non-auditory measurement of inhibitory control.

The present findings argue that the improvement in speech-in-speech intelligibility can be mediated by a cognitive ability, namely the inhibitory control. Since that ability is not auditory specific, it could be trained by non-auditory tasks (p. ex., serious games) and then benefits patients with clinical or subclinical auditory lesions. Moreover, as suggested by the decrease of the left IFG activity following the training, such training should spare cognitive resources and therefore be more beneficial for the elderly or the workers submitted to cognitive constraints.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

All the authors contributed to the experiment design, results discussion, and paper redaction. Data collection and statistical analysis were made by CL, SS, and GA.

Funding

This work was supported by THALES SIX GTS and Starkey Hearing Technologies.

Acknowledgments

The authors warmly thank Bruno Sourdillat and François Capman for their helpful comments and Jean Christophe Bouy for software development.

Contributions to the field Statement

The theme of our research was to investigate the neural mechanisms underlying language comprehension in challenging listening situations, notably multi-talker situations.

We focused on better understanding the contribution of the inferior frontal gyrus (IFG) in speech intelligibility. To do this, after confirming the already documented correlation between speech intelligibility and IFG activity, we asked ourselves how this might change after a period of training to increase intelligibility.

To address this question, we used two approaches, one observational resulting in Experiment I and the other interventional resulting in experiment II.

These two experiments together not only confirm the role of the IFG in challenging listening situations but also add an important piece to our understanding of the nature of its role. We argue that the role of the IFG in speech intelligibility may be related to a better control in the inhibition of the message to be ignored following the training. We consider this an innovative and original point of our research.

We consider that these results may be useful to a wide audience of people interested in the domain of auditory neuroscience.

Furthermore, considering that we used training in our research, the results could also be of inspiration for topics including learning in its broader sense.

Reference styles

Aasted, C. M., Yücel, M. A., Cooper, R. J., Dubb, J., Tsuzuki, D., Becerra, L., Petkov, M. P., Borsook, D., Dan, I., and Boas, D. A. (2015). Anatomical guidance for functional near-infrared spectroscopy: AtlasViewer tutorial. *Neurophotonics*, 2(2), 020801. <https://doi.org/10.1117/1.NPh.2.2.020801>

Alain, C. (2000). Selectively attending to auditory objects. *Frontiers in Bioscience*, 5(1), d202. <https://doi.org/10.2741/Alain>

Alain, C., Du, Y., Bernstein, L. J., Barten, T., and Banai, K. (2018). Listening under difficult conditions: An activation likelihood estimation meta-analysis. *Human Brain Mapping*, 39(7), 2695–2709. <https://doi.org/10.1002/hbm.24031>

Andéol, G., Suied, C., Scannella, S., and Dehais, F. (2017). The Spatial Release of Cognitive Load in Cocktail Party Is Determined by the Relative Levels of the Talkers. *JARO - Journal of the Association for Research in Otolaryngology*, 18(3), 457–464. <https://doi.org/10.1007/s10162-016-0611-7>

Aron, A. R. (2007). The neural basis of inhibition in cognitive control. *Neuroscientist*, 13(3), 214–228. <https://doi.org/10.1177/1073858407299288>

Aron, A. R., Robbins, T. W., and Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, 8(4), 170–177. <https://doi.org/10.1016/j.tics.2004.02.010>

Aron, A. R., Robbins, T. W., and Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: One decade on. *Trends in Cognitive Sciences*, 18(4), 177–185. <https://doi.org/10.1016/j.tics.2013.12.003>

- Baldo, J. V., and Cronkers, N. F. (2006). The role of inferior parietal and inferior frontal cortex in working memory. *Neuropsychology*, 20(5), 529–538. <https://doi.org/10.1037/0894-4105.20.5.529>
- Barker, J. W., Aarabi, A., and Huppert, T. J. (2013). Autoregressive model based algorithm for correcting motion and serially correlated errors in fNIRS. *Biomedical optics express*, 4(8), 1366-1379. <https://doi.org/10.1364/BOE.4.001366>
- Bauernfeind, G., Wriessnegger, S. C., Haumann, S., and Lenarz, T. (2018). Cortical activation patterns to spatially presented pure tone stimuli with different intensities measured by functional near-infrared spectroscopy. *Human Brain Mapping*, 39(7), 2710–2724. <https://doi.org/10.1002/hbm.24034>
- Bee, M. A., and Micheyl, C. (2008). *The “Cocktail Party Problem”: What Is It? How Can It Be Solved? And Why Should Animal Behaviorists Study It?* <https://doi.org/10.1037/0735-7036.122.3.235>
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1), 289-300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Best, V., Shinn-Cunningham, B. G., Ozmeral, E. J., and Kopčo, N. (2010). Exploring the benefit of auditory spatial continuity Spatial release from informational masking in speech recognition Exploring the benefit of auditory spatial continuity. *Citation: The Journal of the Acoustical Society of America*, 127, 2112. <https://doi.org/10.1121/1.3431093>
- Bilecen, D., Seifritz, E., Scheffler, K., Henning, J., and Schulte, A.-C. (2002). Amplitude of the Human Auditory Cortex: An fMRI Study. *NeuroImage*, 17(2), 710–718. <https://doi.org/10.1006/nimg.2002.1133>
- Bless, J. J., Westerhausen, R., Kompus, K., Gudmundsen, M., and Hugdahl, K. (2014). Self-supervised, mobile-application based cognitive training of auditory attention: A behavioral and fMRI evaluation. *Internet Interventions*, 1(3), 102–110. <https://doi.org/10.1016/j.invent.2014.06.001>

- Bolia, R. S., Nelson, W. T., and Morley, R. M. (2001). Asymmetric performance in the cocktail party effect: Implications for the design of spatial audio displays. *Human Factors*, 43(2), 208–216. <https://doi.org/10.1518/001872001775900887>
- Bronkhorst, A. W. (2015). The cocktail-party problem revisited: early processing and selection of multi-talker speech. In *Attention, Perception, and Psychophysics*. <https://doi.org/10.3758/s13414-015-0882-9>
- Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America*, 109(3), 1101–1109. <https://doi.org/10.1121/1.1345696>
- Brungart, D. S., and Simpson, B. D. (2004). Within-ear and across-ear interference in a dichotic cocktail party listening task: Effects of masker uncertainty. *The Journal of the Acoustical Society of America*, 115(1), 301–310. <https://doi.org/10.1121/1.1628683>
- Brungart, D. S., and Simpson, B. D. (2007). Cocktail party listening in a dynamic multitalker environment. *Perception and Psychophysics*. <https://doi.org/10.3758/BF03194455>
- Buchsbaum, B. R. (2015). Working Memory and Language. In *Neurobiology of Language* (pp. 863–875). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-407794-2.00069-9>
- Buchsbaum, B. R., and D’Esposito, M. (2008). The search for the phonological store: From loop to convolution. *Journal of Cognitive Neuroscience*, 20(5), 762–778. <https://doi.org/10.1162/jocn.2008.20501>
- Chavan, C. F., Mouthon, M., Draganski, B., van der Zwaag, W., and Spierer, L. (2015). Differential patterns of functional and structural plasticity within and between inferior frontal gyri support training-induced improvements in inhibitory control proficiency. *Human Brain Mapping*, 36(7), 2527–2543. <https://doi.org/10.1002/hbm.22789>
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America*, 25(5), 975-979.

<https://doi.org/10.1121/1.1907229>

Cooke, M., Garcia Lecumberri, M. L., and Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America*, 123(1), 414–427. <https://doi.org/10.1121/1.2804952>

Curtin, A., Ayaz, H., Tang, Y., Sun, J., Wang, J., and Tong, S. (2019). Enhancing neural efficiency of cognitive processing speed via training and neurostimulation: An fNIRS and TMS study. *NeuroImage*, 198(January), 73–82. <https://doi.org/10.1016/j.neuroimage.2019.05.020>

Davis, M. H., and Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, 23(8), 3423–3431. <https://doi.org/10.1523/jneurosci.23-08-03423.2003>

Dillon, D. G., and Pizzagalli, D. A. (2007). Inhibition of action, thought, and emotion: a selective neurobiological review. *Applied and Preventive Psychology*, 12(3), 99-114.

Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., and Scott, S. K. (2010). Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *Journal of Neuroscience*, 30(21), 7179–7186. <https://doi.org/10.1523/JNEUROSCI.4040-09.2010>

Evans AC, Collins DL, Mills SR, Brown ED, Kelly RL, P. T. (193 C.E.). 3D Mri. *Nucl. Sci. Symp. Med. Imaging Conf. (IEEE Conf. Rec., 1813–1817.*

Evans, S., Mcgettigan, C., Agnew, Z. K., Rosen, S., and Scott, S. K. (2016). *Getting the cocktail party started: masking effects in speech perception.* https://doi.org/10.1162/jocn_a_00913

Fairclough, S., Ewing, K., Burns, C., and Kreplin, U. (2018). Neural efficiency and mental workload: Locating the red line. *Neuroergonomics: The Brain at Work and in Everyday Life*, 73–77. <https://doi.org/10.1016/B978-0-12-811926-6.00012-9>

- Fiedler, L., Wöstmann, M., Herbst, S. K., and Obleser, J. (2019). Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions. *NeuroImage*, 186(August 2018), 33–42. <https://doi.org/10.1016/j.neuroimage.2018.10.057>
- Fishburn, F. A., Hlutkowsky, C. O., Bemis, L. M., Huppert, T. J., Wakschlag, L. S., and Perlman, S. B. (2019). Irritability uniquely predicts prefrontal cortex activation during preschool inhibitory control among all temperament domains: A LASSO approach. *NeuroImage*, 184(July 2018), 68–77. <https://doi.org/10.1016/j.neuroimage.2018.09.023>
- Fritz, J. B., Elhilali, M., David, S. V., and Shamma, S. A. (2007). Auditory attention - focusing the searchlight on sound. In *Current Opinion in Neurobiology*. <https://doi.org/10.1016/j.conb.2007.07.011>
- Golestani, N., and Zatorre, R. J. (2004). Learning new sounds of speech: Reallocation of neural substrates. *NeuroImage*, 21(2), 494–506. <https://doi.org/10.1016/j.neuroimage.2003.09.071>
- Hampshire, A., Chamberlain, S. R., Monti, M. M., Duncan, J., and Owen, A. M. (2010). The role of the right inferior frontal gyrus: inhibition and attentional control. *NeuroImage*, 50(3), 1313–1319. <https://doi.org/10.1016/j.neuroimage.2009.12.109>
- Hartmann, L., Sallard, E., and Spierer, L. (2016). Enhancing frontal top-down inhibitory control with Go/NoGo training. *Brain Structure and Function*, 221(7), 3835–3842. <https://doi.org/10.1007/s00429-015-1131-7>
- Hempel, A., Giesel, F. L., Garcia Caraballo, N. M., Amann, M., Meyer, H., Wüstenberg, T., Essig, M., and Schröder, J. (2004). Plasticity of Cortical Activation Related to Working Memory during Training. *American Journal of Psychiatry*, 161(4), 745–747. <https://doi.org/10.1176/appi.ajp.161.4.745>
- Hill, K. T., and Miller, L. M. (2010). Auditory attentional control and selection during cocktail party listening. *Cerebral Cortex*, 20(3), 583–590. <https://doi.org/10.1093/cercor/bhp124>

- Jäncke, L., Shah, N. J., Posse, S., Grosse-Ryuken, M., and Müller-Gärtner, H. W. (1998). Intensity coding of auditory stimuli: An fMRI study. *Neuropsychologia*, 36(9), 875–883. [https://doi.org/10.1016/S0028-3932\(98\)00019-0](https://doi.org/10.1016/S0028-3932(98)00019-0)
- Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., and Carlyon, R. P. (2013). Swinging at a Cocktail Party: Voice Familiarity Aids Speech Perception in the Presence of a Competing Voice. *Psychological Science*, 24(10), 1995–2004. <https://doi.org/10.1177/0956797613482467>
- Johnsrude, I. S., and Rodd, J. M. (2016). Factors that increase processing demands when listening to speech. *Neurobiology of language*, 491-502. <https://doi.org/10.1016/B978-0-12-407794-2.00040-7>
- Jonides, J., Smith, E. E., Marshuetz, C., Koeppe, R. A., and Reuter-Lorenz, P. A. (1998). Inhibition in verbal working memory revealed by brain activation. *Proceedings of the National Academy of Sciences of the United States of America*, 95(14), 8410–8413. <https://doi.org/10.1073/pnas.95.14.8410>
- Lawrence, R. J., Wiggins, I. M., Anderson, C. A., Davies-Thompson, J., and Hartley, D. E. H. (2018). Cortical correlates of speech intelligibility measured using functional near-infrared spectroscopy (fNIRS). *Hearing Research*, 370, 53–64. <https://doi.org/10.1016/j.heares.2018.09.005>
- Lee, A. K. C., Larson, E., Maddox, R. K., and Shinn-Cunningham, B. G. (2014). Using neuroimaging to understand the cortical mechanisms of auditory selective attention. In *Hearing Research* (Vol. 307, pp. 111–120). Elsevier. <https://doi.org/10.1016/j.heares.2013.06.010>
- Manuel, A. L., Bernasconi, F., and Spierer, L. (2013). Plastic modifications within inhibitory control networks induced by practicing a stop-signal task: An electrical neuroimaging study. *Cortex*, 49(4), 1141–1147. <https://doi.org/10.1016/j.cortex.2012.12.009>
- McDermott, J. H. (2009). The cocktail party problem. *Current Biology*, 19(22), 1024–1027. <https://doi.org/10.1016/j.cub.2009.09.005>

- Micheyl, C., and Oxenham, A. J. (2010). Pitch, harmonicity and concurrent sound segregation: Psychoacoustical and neurophysiological findings. In *Hearing Research* (Vol. 266, Issues 1–2, pp. 36–51). Elsevier. <https://doi.org/10.1016/j.heares.2009.09.012>
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex “Frontal Lobe” Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*(1), 49–100. <https://doi.org/10.1006/cogp.1999.0734>
- Neuner, I., Kawohl, W., Arrubla, J., Warbrick, T., Hitz, K., Wyss, C., Boers, F., and Shah, N. J. (2014). Cortical Response Variation with Different Sound Pressure Levels: A Combined Event-Related Potentials and fMRI Study. *PLoS ONE*, *9*(10). <https://doi.org/10.1371/journal.pone.0109216>
- Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I. H., Saberi, K., Serences, J. T., and Hickok, G. (2010). Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex*, *20*(10), 2486–2495. <https://doi.org/10.1093/cercor/bhp318>
- Oxenham, A. J. (2008). *Pitch Perception and Auditory Stream Segregation: Implications for Hearing Loss and Cochlear Implants*. <https://doi.org/10.1177/1084713808325881>
- Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing*, *39*(2). <https://doi.org/10.1097/AUD.0000000000000494>
- Peelle, J. E., Eason, R. J., Schmitter, S., Schwarzbauer, C., and Davis, M. H. (2010). Evaluating an acoustically quiet EPI sequence for use in fMRI studies of speech and auditory processing. *NeuroImage*, *52*(4), 1410–1419. <https://doi.org/10.1016/j.neuroimage.2010.05.015>
- Pinti, P., Scholkmann, F., Hamilton, A., Burgess, P., and Tachtsidis, I. (2019). Current status and issues regarding pre-processing of fNIRS neuroimaging data: an investigation of

- diverse signal filtering methods within a general linear model framework. *Frontiers in human neuroscience*, 12, 505. <https://doi.org/10.3389/fnhum.2018.00505>
- Popham, S., Boebinger, D., Ellis, D. P. W., Kawahara, H., and Mcdermott, J. H. (n.d.). *Inharmonic speech reveals the role of harmonicity in the cocktail party problem*. <https://doi.org/10.1038/s41467-018-04551-8>
- Putkinen, V., and Saarikivi, K. (2018). Neural correlates of enhanced executive functions: Is less more? *Annals of the New York Academy of Sciences*, 1423(1), 117–125. <https://doi.org/10.1111/nyas.13645>
- Qi, Z., Han, M., Wang, Y., de los Angeles, C., Liu, Q., Garel, K., Chen, E. S., Whitfield-Gabrieli, S., Gabrieli, J. D. E., and Perrachione, T. K. (2019). Speech processing and plasticity in the right hemisphere predict variation in adult foreign language learning. *NeuroImage*, 192(February), 76–87. <https://doi.org/10.1016/j.neuroimage.2019.03.008>
- Quaresima, V., and Ferrari, M. (2019). Functional Near-Infrared Spectroscopy (fNIRS) for Assessing Cerebral Cortex Function During Human Behavior in Natural/Social Situations: A Concise Review. *Organizational Research Methods*, 22(1), 46–68. <https://doi.org/10.1177/1094428116658959>
- Ravizza, S. M., Delgado, M. R., Chein, J. M., Becker, J. T., and Fiez, J. A. (2004). Functional dissociations within the inferior parietal cortex in verbal working memory. *NeuroImage*, 22(2), 562–573. <https://doi.org/10.1016/j.neuroimage.2004.01.039>
- Rodd, J. M., Davis, M. H., and Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex*, 15(8), 1261–1269. <https://doi.org/10.1093/cercor/bhi009>
- Santosa, H., Zhai, X., Fishburn, F., and Huppert, T. (2018). The NIRS Brain AnalyzIR toolbox. *Algorithms*, 11(5). <https://doi.org/10.3390/A11050073>
- Sehm, B., Schnitzler, T., Obleser, J., Groba, A., Ragert, P., Villringer, A., and Obrig, H. (2013). Facilitation of inferior frontal cortex by transcranial direct current stimulation induces

- perceptual learning of severely degraded speech. *Journal of Neuroscience*, 33(40), 15868–15878. <https://doi.org/10.1523/JNEUROSCI.5466-12.2013>
- Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, 12(5), 182–186. <https://doi.org/10.1016/j.tics.2008.02.003>
- Shinn-Cunningham, B. G., and Best, V. (2008). Selective Attention in Normal and Impaired Hearing. *Trends in Amplification*. <https://doi.org/10.1177/1084713808325306>
- Sohoglu, E., and Davis, M. H. (2016). Perceptual learning of degraded speech by minimizing prediction error. *Proceedings of the National Academy of Sciences*, 113(12), E1747–E1756. <https://doi.org/10.1073/PNAS.1523266113>
- Strauß, A., Wöstmann, M., and Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Frontiers in Human Neuroscience*, 8(MAY), 1–7. <https://doi.org/10.3389/fnhum.2014.00350>
- Studebaker, G. A. (1985). A "rationalized" arcsine transform. *Journal of Speech, Language, and Hearing Research*, 28(3), 455–462. <https://doi.org/10.1044/jshr.2803.455>
- Swick, D., Ashley, V., and Turken, U. (2011). Are the neural correlates of stopping and not going identical? Quantitative meta-analysis of two response inhibition tasks. *NeuroImage*, 56(3), 1655–1665. <https://doi.org/10.1016/j.neuroimage.2011.02.070>
- Szalárdy, O., Tóth, B., Farkas, D., György, E., and Winkler, I. (2019). Neuronal correlates of informational and energetic masking in the human brain in a multi-talker situation. *Frontiers in Psychology*, 10(APR), 1–16. <https://doi.org/10.3389/fpsyg.2019.00786>
- Thompson, E. R., Iyer, N., Simpson, B. D., Wakefield, G. H., Kieras, D. E., and Brungart, D. S. (2015). *in Speech-on-Speech Masking Experiments a*). 138(3), 1297–1304.
- Tóth, B., Farkas, D., Urbán, G., Szalárdy, O., Orosz, G., Hunyadi, L., Hajdu, B., Kovács, A., Szabó, B. T., Shestopalova, L. B., and Winkler, I. (2019). Attention and speech-processing related functional brain networks activated in a multi-speaker environment. In *PLoS ONE*

(Vol. 14, Issue 2). <https://doi.org/10.1371/JOURNAL.PONE.0212754>

Wijayasiri, P., Hartley, D. E. H., and Wiggins, I. M. (2017). Brain activity underlying the recovery of meaning from degraded speech: A functional near-infrared spectroscopy (fNIRS) study. *Hearing Research*, 351, 55–67. <https://doi.org/10.1016/j.heares.2017.05.010>

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., and Johnsrude, I. S. (2012). Effortful Listening: The Processing of Degraded Speech Depends Critically on Attention. *Journal of Neuroscience*, 32(40), 14010–14021. <https://doi.org/10.1523/JNEUROSCI.1528-12.2012>

Xue, G., Aron, A. R., and Poldrack, R. A. (2008). Common neural substrates for inhibition of spoken and manual responses. *Cerebral Cortex*, 18(8), 1923–1932. <https://doi.org/10.1093/cercor/bhm220>

Yue, Q., Martin, R. C., Cris Hamilton, A., and Rose, N. S. (2019). Non-perceptual Regions in the Left Inferior Parietal Lobe Support Phonological Short-term Memory: Evidence for a Buffer Account? *Cerebral Cortex*, 29(4), 1398–1413. <https://doi.org/10.1093/cercor/bhy037>

Zhang, M., Mary Ying, Y. L., Ihlefeld, A., Ying, Y.-L. M., and Ihlefeld, A. (2018). Spatial Release From Informational Masking: Evidence From Functional Near Infrared Spectroscopy. *Trends in Hearing*, 22, 1–12. <https://doi.org/10.1177/2331216518817464>

Et à quelle occasion :.....
.....
.....

- Êtes-vous dyslexique (oui/non) ? :
- Avez-vous déjà travaillé dans un environnement bruyant (oui/non) ?
Si oui,
Pouvez-vous détailler quand, le lieu et pendant combien de temps ?

.....
.....
...
Portez-vous des protections auditives ?

- Avez-vous des loisirs bruyants, par exemple la course automobile ou le tir ?
Si oui, pendant combien de temps et à quelle fréquence ?

-
...
• Est-ce que vous jouez d'un instrument de musique ?
Si oui,
Quel instrument ?

Depuis quel âge ?.....

Combien de temps par semaine (heures) ?.....

Quel genre de musique ?

- Combien de temps passez-vous à écouter de la musique par semaine ?

Sous casque ? Quelle sorte de casque ?.....

À quel volume ?

Type de lecteur ?

- Allez-vous en boîte de nuit ? Aux concerts ?

À quel âge avez-vous commencé ?

À quelle fréquence ?

Portez-vous des protections auditives ? De quelle sorte ?

.....
...

- Pouvez-vous quantifier approximativement votre exposition au bruit durant les 5 dernières années ? (le nombre d'heures par semaine)

.....
...

HV :

.....
...

Fait à..... , le

Résumé :

Suivre une conversation en situations multilocuteurs peut être une tâche difficile. Dans ces situations d'écoute complexe, l'intelligibilité de la parole est souvent dégradée et la compréhension peut être compromise. L'effort d'écoute en situation multilocuteurs peut être élevé du fait des ressources cognitives nécessaires à obtenir une intelligibilité satisfaisante. Un effort d'écoute élevé peut engendrer une surcharge cognitive compromettant la sécurité de certains travailleurs, comme les militaires qui doivent gérer plusieurs tâches simultanément et sous une forte pression temporelle. Par ailleurs les militaires, du fait de leur travail, peuvent avoir des troubles de l'audition dont l'impact sur l'effort d'écoute est mal connu.

Mieux comprendre l'effort d'écoute, les dimensions qui le composent comme les facteurs qui le modulent, constituent un réel enjeu pour aider ces populations

L'effort d'écoute est un concept multidimensionnel qui comprend les dimensions liées aux caractéristiques de l'auditeur (ressources cognitives, vécu subjectif, motivation, audition) ; les dimensions externes liées aux caractéristiques du message à comprendre (clair ou dégradé) et du contexte où l'échange se déroule (différence d'intensités sonores entre les locuteurs et séparation spatiale de ceux-ci).

Les trois études présentées dans cette thèse ont étudié ces différentes dimensions de l'effort d'écoute à travers des mesures objectives (performance, engagement de ressources cognitives via fNIRS, mesures audiologiques) et subjectives (échelles et questionnaires).

La première étude, à travers des enregistrements de l'activité corticale, a quantifié l'engagement cognitif de jeunes normo-entendants avant et après un entraînement à une tâche d'écoute multilocuteurs. Dans cette étude, les locuteurs étaient chacun placés à des différents niveaux d'intensité sonore. Cette étude a permis de mettre en évidence d'une part le rôle central du contrôle inhibiteur dans la compréhension d'un locuteur cible et d'autre part l'efficacité de l'entraînement dans la réduction des ressources cognitives employées présupposant une possible diminution de l'effort d'écoute.

La seconde étude, réalisée également avec des jeunes normo-entendants, a utilisé la séparation spatiale des locuteurs confirmant le rôle joué par le contrôle inhibiteur dans les processus cognitifs permettant la compréhension en situations multilocuteurs.

La troisième étude a été réalisée avec une population de militaires illustrant une diminution de l'effort d'écoute ressenti avec un meilleur contrôle inhibiteur. Ces effets n'étaient pas liés ni à l'état auditif dégradé ni à la motivation investie dans la tâche d'écoute.

Ainsi, la séparation spatiale et l'entraînement à l'utilisation des indices de différence d'intensité sonore entre les voix des locuteurs, pourrait permettre une réduction de l'effort d'écoute qui reste à confirmer en situation écologique.

Mots clés : situations multilocuteurs, effort d'écoute, fNIRS, contrôle inhibiteur.

Abstract :

Following a conversation in multitalker situations can be a difficult task. In these complex listening situations, speech intelligibility is often degraded and comprehension may be compromised. Listening effort in multitalker situations can be high because of the cognitive resources required to achieve satisfactory intelligibility. High listening effort can lead to cognitive overload compromising the safety of some workers, such as military personnel who have to manage several tasks simultaneously and under high time pressure. Furthermore, military personnel, because of their work, may have hearing disorders whose impact on listening effort is not well known.

A better understanding of listening effort, including the dimensions that make it up and the factors that modulate it, is a real challenge in helping these populations.

Listening effort is a multidimensional concept that includes dimensions related to the characteristics of the listener (cognitive resources, subjective experience, motivation, hearing); external dimensions related to the characteristics of the message to be understood (clear or degraded) and the context in which the exchange takes place (difference in sound intensity between talkers and spatial separation of talkers).

The three studies presented in this thesis investigated these different dimensions of listening effort through objective (performance, cognitive resource engagement via fNIRS, audiological measures) and subjective (scales and questionnaires) measures.

The first study, using cortical activity recordings, quantified the cognitive engagement of young normal-hearing individuals before and after training in a multitalker listening task. In this study, the talkers were each placed at different levels of intensity. This study demonstrated the central role of inhibitory control in the comprehension of a target talker and the effectiveness of training in reducing the cognitive resources used, presupposing a possible reduction in listening effort.

The second study, also conducted with young normal-hearing, used spatial separation of talkers, confirming the role played by inhibitory control in the cognitive processes that enable comprehension in multitalker situations.

The third study was conducted with a military population showing a decrease in listening effort experienced with better inhibitory control. These effects were not related to either the impaired auditory state or the motivation invested in the listening task.

Thus, spatial separation and training in the use of cues to the difference in intensity between talkers' voices could lead to a reduction in listening effort that remains to be confirmed in an ecological situation.

Keywords: multitalker situations, listening effort, fNIRS, inhibitory control