



HAL
open science

Externalisation en restitution binaurale non-individualisée avec head-tracking : aspects objectifs et perceptifs

Tom Colas

► **To cite this version:**

Tom Colas. Externalisation en restitution binaurale non-individualisée avec head-tracking : aspects objectifs et perceptifs. Acoustique [physics.class-ph]. Université de Bretagne occidentale - Brest, 2024. Français. NNT : 2024BRES0053 . tel-04817050

HAL Id: tel-04817050

<https://theses.hal.science/tel-04817050v1>

Submitted on 3 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT DE

L'UNIVERSITÉ DE BRETAGNE OCCIDENTALE

ÉCOLE DOCTORALE N° 644

Mathématiques et Sciences et Technologies

de l'Information et de la Communication en Bretagne Océane

Spécialité : *Signal, Image, Vision, Son*

Par

Tom COLAS

Externalisation en restitution binaurale non-individualisée avec head-tracking : aspects objectifs et perceptifs

Thèse présentée et soutenue à Brest, le 08 Octobre 2024

Unité de recherche : Lab-STICC UMR 6285

Rapporteurs avant soutenance :

Mathieu LAVANDIER Directeur de recherche, ENTPE Lyon
Mitsuko ARAMAKI Directrice de recherche, PRISM Marseille

Composition du Jury :

Présidente :	Isabelle VIAUD-DELMON	Directrice de recherche, IRCAM Paris
Examineurs :	Mathieu LAVANDIER	Directeur de recherche, ENTPE Lyon
	Mitsuko ARAMAKI	Directrice de recherche, PRISM Marseille
	Isabelle VIAUD-DELMON	Directrice de recherche, IRCAM Paris
	Clara SUJED	Chargée de recherche, IRBA Brétigny-sur-Orge
Dir. de thèse :	Mathieu PAQUIER	Professeur des Universités, UBO, Brest
Co-encadrant :	Étienne HENDRICKX	Maître de Conférences, UBO, Brest
Co-encadrant :	Nicolas FARRUGIA	Maître de Conférences, IMT Atlantique, Plouzané

REMERCIEMENTS

Avant toute chose, je tiens à adresser un énorme remerciement à mon directeur Mathieu Paquier, ainsi qu'à mes encadrants Étienne Hendrickx et Nicolas Farrugia. Merci pour la confiance que vous m'avez accordé en me proposant ce projet de thèse. Merci à tous les trois pour votre accompagnement, votre soutien et votre patience. Le partage de vos connaissances scientifiques et techniques m'a permis d'apprendre tant de choses et de développer mes propres compétences.

Je remercie Mathieu Lavandier et Mitsuko Aramaki d'avoir accepté d'évaluer mon travail de thèse en étant les rapporteurs de ce manuscrit ainsi que Isabelle Viaud-Delmon et à Clara Sued d'avoir accepté de participer au jury. Je remercie également Luc Arnal et Olivier Machery pour leur participation à mon comité de suivi individuel.

Je tiens à remercier toute l'équipe de l'ISB. Merci à Luc Pennamen et François Barvec d'avoir partagé avec moi leurs connaissances techniques et de m'avoir aidé dans les installations des tests perceptifs. Merci à Vincent Koehl et à Gauthier Berthomieu pour leurs conseils avisés et pour leur bonne humeur inaltérable. J'adresse un remerciement particulier à Clément Rappin pour avoir partagé avec moi ces trois années en tant que doctorant. J'adresse également mes remerciements à l'équipe BRAIn de l'IMT Atlantique de m'avoir accueilli dans leurs locaux et d'avoir répondu à mes questions de programmeur novice. Je tiens à remercier François Salmon de m'avoir aidé dans la création des stimuli sonores et Mathilde De Quarti d'avoir prêté sa voix pour les enregistrements. Enfin, j'adresse toute ma gratitude aux nombreuses personnes qui ont rendu mon travail de thèse possible en participant aux tests perceptifs.

J'adresse un énorme remerciement à mes parents, qui m'ont soutenu et encouragé tout au long de mes études, et plus généralement pour toutes les valeurs qu'ils m'ont transmises, pour leur accompagnement et pour leur amour.

Pour finir, je remercie du fond du coeur Emily qui a su rendre ma vie plus belle en partageant mon quotidien pendant ces trois années.

TABLE DES MATIÈRES

Introduction	9
1 Neurosciences et perception auditive binaurale	13
1.1 Signaux acoustiques et activité cérébrale	14
1.1.1 Traitements périphériques et centraux de l'audition	14
1.1.2 L'électrophysiologie	22
1.1.3 Les potentiels évoqués	27
1.2 La localisation auditive	35
1.2.1 Localisation en azimut	36
1.2.2 Localisation en élévation	41
1.2.3 Localisation en distance	41
1.3 La restitution binaurale	45
1.3.1 Définition et principe	45
1.3.2 Synthèse binaurale dynamique	51
1.3.3 Binaural individualisé ou non-individualisé	56
1.4 L'externalisation	58
1.4.1 Indices d'externalisation	58
1.4.2 Mesure et problématique	63
1.4.3 Externalisation et activité cérébrale	67
1.4.4 Objectifs de la thèse	73
2 Étude comparative comportementale vs. EEG de l'externalisation en restitution binaurale dynamique	75
2.1 Introduction	75
2.2 Matériel et méthode	76
2.2.1 Participants	76
2.2.2 Stimulus	76
2.2.3 Enregistrement et traitement du stimulus sonore	77
2.2.4 Système de reproduction audio	79

TABLE DES MATIÈRES

2.2.5	Enregistrement EEG	79
2.2.6	Procédure	82
2.2.7	Pré-traitement de l'EEG et création des epochs	85
2.2.8	Analyses statistiques des ERPs	87
2.3	Résultats	89
2.3.1	Résultats comportementaux	89
2.3.2	Résultats EEG	93
2.4	Discussion	99
2.4.1	Constats comportementaux	99
2.4.2	Constats EEG	100
2.5	Conclusion et perspectives	105
3	Étude de la persistance de l'effet des mouvements de tête sur l'externalisation en restitution binaurale dynamique	107
3.1	Introduction	107
3.2	Expérience 1	109
3.2.1	Matériel et méthode	109
3.2.2	Résultats	114
3.2.3	Discussion	116
3.3	Expérience 2	117
3.3.1	Matériel et méthode	118
3.3.2	Résultats	120
3.3.3	Discussion	122
3.4	Expérience 3	123
3.4.1	Matériel et méthode	123
3.4.2	Résultats	125
3.4.3	Discussion	125
3.5	Discussion générale	127
3.6	Conclusion	129
	Conclusion générale	131
	Annexes	135
A	Annexe A : données des mouvements de tête	135

B	Annexe B : statistiques bayésiennes	140
C	Annexe C : analyses supplémentaires des données EEG	142
	C.1 Méthode	142
	C.2 Résultats	143
	Bibliographie	145

INTRODUCTION

La restitution binaurale est une méthode de restitution du son au casque, qui cherche à se rapprocher le plus possible d'une écoute naturelle.

La technologie binaurale connaît un développement croissant dans des milieux variés tels que la réalité virtuelle, le jeu vidéo, la radio et la muséographie. Les productions dans ces domaines ont souvent pour objectif d'immerger le spectateur dans un univers audiovisuel et le réalisme de l'espace sonore est primordial pour que l'immersion soit convaincante. La restitution binaurale semble particulièrement efficace pour répondre à cette problématique, néanmoins les paramètres qui influent sur le réalisme et sur la qualité perçue sont encore mal maîtrisés et peuvent varier substantiellement d'un individu à l'autre.

Lors d'une écoute naturelle, l'auditeur perçoit les sources sonores comme provenant de l'environnement extérieur, hors de sa tête. Les sources sont alors *externalisées*. En revanche, lors d'une écoute au casque, les sons sont souvent perçus comme provenant de l'intérieur de la tête de l'auditeur. Les sources sont alors *internalisées*. Pour qu'une reproduction binaurale soit convaincante et réaliste, il est crucial que les sources sonores virtuelles soient au contraire externalisées, comme dans la "vraie vie". Malheureusement, il est courant que certaines sources, en particulier celles situées à l'avant et à l'arrière, ne soient pas correctement externalisées, ce qui nuit à la qualité de la restitution binaurale.

L'utilisation d'un dispositif de suivi des mouvements de tête, ou "head-tracking", peut permettre d'améliorer l'externalisation lors de la restitution binaurale. Ce dispositif permet de modifier la synthèse binaurale en fonction des mouvements de tête de l'auditeur, ajustant ainsi la position des sources virtuelles pour qu'elles restent fixes par rapport à l'environnement extérieur. Cela permet d'améliorer l'externalisation pendant et après les mouvements de tête, l'effet persistant même une fois les mouvements terminés.

Jusqu'à présent, les études se sont plutôt intéressées à l'externalisation d'un point de vue perceptif (subjectif). Néanmoins, l'externalisation est un percept complexe dont la définition peut porter à débat et qui varie beaucoup d'un individu à l'autre. Cette subjectivité peut être une contrainte importante dans la recherche et le développement des technologies binaurales.

L'électroencéphalographie (EEG) est une méthode d'exploration de l'activité cérébrale qui mesure l'activité électrique du cerveau grâce à des électrodes placées sur le cuir chevelu. Les

potentiels évoqués auditifs correspondent à une activité EEG particulière induite par une stimulation auditive ponctuelle. Ces potentiels peuvent être enregistrés pour étudier la relation entre les caractéristiques d'une stimulation auditive et l'activité électrique cérébrale résultante. En associant les potentiels évoqués auditifs à des mesures perceptives d'externalisation, il serait possible de surmonter les contraintes de subjectivité des études perceptives.

Les travaux présentés dans cette thèse sont centrés autour de l'externalisation en écoute binaurale avec head-tracking. Les objectifs sont :

- de trouver un corrélat neuronal de l'externalisation en restitution binaurale. Pour cela, une étude composée d'une expérience EEG et d'une expérience comportementale a été réalisée.
- de caractériser la persistance de l'amélioration d'externalisation obtenue après des mouvements de tête avec head-tracking en restitution binaurale. Pour cela, une étude composée de trois expériences comportementales a été réalisée.

Le chapitre 1 est dédiée à l'exposition des concepts et des méthodes de neurophysiologie et de psychoacoustique indispensables à la compréhension de la contribution de la thèse. On y explore le traitement des signaux acoustiques, depuis leur réception par les oreilles jusqu'à leur décodage par le cerveau, pour introduire le fonctionnement de l'électroencéphalographie et des potentiels évoqués. Les bases de la localisation auditive, notamment les différences interaurales et les HRTFs, sont ensuite présentées, des notions essentielles pour appréhender la technologie binaurale qui est discutée par la suite. Les principes de la synthèse binaurale dynamique et du head-tracking sont détaillés, mettant en lumière leur intérêt pour améliorer la restitution binaurale, notamment en optimisant l'externalisation. La notion d'externalisation est donc expliquée, avec les paramètres influençant sa qualité et les défis associés à sa mesure.

Le chapitre 2 présente la première étude réalisée dans le cadre de la thèse. Cette étude a exploré la relation entre externalisation et activité électrique cérébrale en comparant les données issues d'une expérience comportementale et d'une expérience électroencéphalographique de sujets soumis à des stimuli plus ou moins externalisés. Une méthode innovante, s'appuyant sur la persistance de l'amélioration d'externalisation due aux mouvements de tête avec head-tracking, a été développée afin de comparer des stimuli "acoustiquement identiques" mais présentant des différences d'externalisation.

Les résultats d'externalisation obtenus dans la première étude ont soulevé plusieurs problématiques concernant la robustesse de cette amélioration d'externalisation. Le chapitre 3 présente

une étude composée de trois expériences réalisées afin d'explorer ces problématiques. La persistance de l'effet des mouvements de tête a été étudiée avec différentes sources sonores dont la longueur variait.

Enfin, les principaux résultats obtenus dans les études sont résumés et des pistes de réflexion pour de futurs travaux sont discutées dans la conclusion générale.

NEUROSCIENCES ET PERCEPTION

AUDITIVE BINAURALE

Ce premier chapitre aborde les notions et méthodes de neurosciences et de psychoacoustique nécessaires à la compréhension des travaux présentés dans les chapitres suivants. Tout d'abord, la section 1.1 expose les fondements de l'audition en détaillant le trajet de l'information sonore depuis l'entrée des oreilles jusqu'à son arrivée au cerveau. En effet, l'oreille externe joue un rôle de filtrage fréquentiel primordial dans le fonctionnement de la technologie binaurale et l'oreille interne effectue la transduction des signaux acoustiques en signaux électriques qui peuvent être étudiés pour caractériser les fonctions auditives. Cette section présente également le principe et les méthodes de l'électroencéphalographie, notamment le fonctionnement des potentiels évoqués auditifs, qui sont étudiés dans le deuxième chapitre de la thèse.

La section 1.2 présente les fondements de la localisation auditive, notamment les indices interauraux et monauraux utilisés par le système auditif pour localiser les sons dans l'espace. La technologie binaurale étant une technique de spatialisation des sons, son fonctionnement s'appuie sur ces différents indices de localisation.

Une fois ces bases établies, la section 1.3 expose le fonctionnement de la technologie binaurale et en particulier de la synthèse binaurale. Pour cela, l'acquisition des fonctions de transfert de la tête (les HRTFs), les problématiques liées à l'individualisation des HRTFs et leurs impacts sur la restitution binaurale sont présentés en détail. Cette section introduit également le principe de la synthèse binaurale dynamique, qui permet de prendre en compte les mouvements de l'auditeur pour obtenir une synthèse binaurale plus réaliste et une meilleure externalisation.

L'externalisation constitue le coeur du travail de recherche présenté dans cette thèse. La section 1.4 définit la notion d'externalisation et les indices qui influent sur cette dernière. Les problèmes liés aux méthodes de mesure perceptive de l'externalisation sont abordés ainsi que l'apport que pourraient constituer les méthodes objectives pour contourner ces limitations et apporter de nouvelles connaissances sur le fonctionnement de l'externalisation en restitution binaurale.

1.1 Signaux acoustiques et activité cérébrale

Lorsqu'une onde sonore parvient aux oreilles d'un auditeur, elle subit une série de modifications et de transductions afin d'être transmise aux neurones des voies auditives sous la forme de signaux électrophysiologiques puis interprétée par le cerveau. L'étude de ces signaux peut permettre de comprendre quand et comment les différents paramètres du son sont traités par le cerveau.

1.1.1 Traitements périphériques et centraux de l'audition

Oreille externe et moyenne

L'oreille externe est constituée du pavillon et du conduit auditif. Son rôle principal est de concentrer les ondes sonores et de les amplifier afin de les conduire jusqu'au tympan. D'abord, le pavillon dirige les ondes sonores environnantes vers l'entrée du conduit auditif externe (CAE). Ces ondes sonores progressent ensuite dans le CAE, qui, associé à la conque, crée une résonance acoustique. Le tympan est une membrane située à l'extrémité du CAE, entre l'oreille externe et l'oreille moyenne. Il est mis en vibration par les ondes acoustiques et transmet ces vibrations aux osselets de l'oreille moyenne. L'onde sonore subit plusieurs modifications lors de son déplacement entre le pavillon et le tympan. La diffraction du pavillon provoque un filtrage dépendant largement de l'angle d'incidence de l'onde sonore. Ce filtrage pavillonnaire permet notamment de faciliter la localisation des sources et est inclus dans la fonction de transfert liée à la tête (HRTF) qui est détaillée dans la section 1.3.1. Le CAE permet une amplification relative de certaines fréquences de l'onde sonore [1]. Cette amplification, fondée en partie sur des phénomènes de résonance acoustique, dépend de la fréquence, mais reste indépendante de l'angle d'incidence du son. La figure 1.1 illustre le gain acoustique apporté par l'oreille externe.

La vibration du tympan est transmise à l'oreille interne via la chaîne ossiculaire, composée de trois osselets : le marteau, l'enclume et l'étrier. Le marteau est directement attaché au centre du tympan et vibre avec ce dernier, transmettant ses vibrations à l'enclume, qui à son tour les transmet à l'étrier. Cette chaîne crée un effet de levier qui permet une amplification. La platine de l'étrier est fixée à la fenêtre ovale, qui fait office d'interface entre l'oreille moyenne et l'oreille interne (voir figure 1.2).

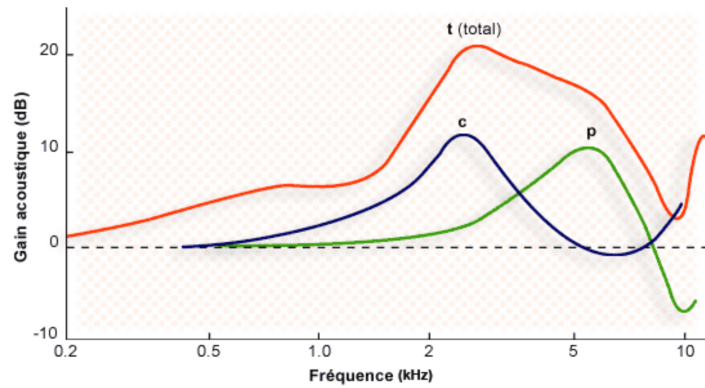


FIGURE 1.1 – Gain acoustique induit par l'oreille externe pour une source située à 45° dans le plan horizontal. Courbe verte : gain induit par le pavillon, courbe bleue : gain induit par le CAE, courbe rouge ; gain total. Image extraite de [2].

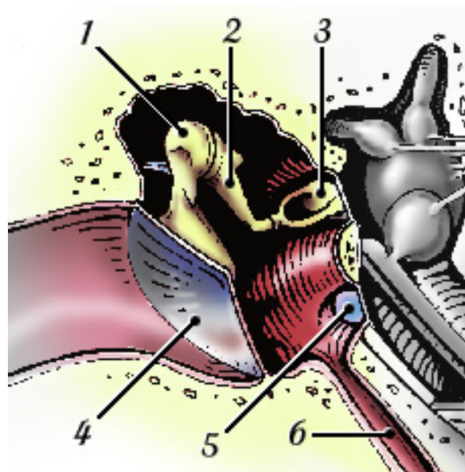


FIGURE 1.2 – Schéma de l'oreille moyenne représentant : (4) le tympan, (1) le marteau, (2) l'enclume et (3) l'étrier, (5) la fenêtre ronde, (6) la trompe d'Eustache. Image extraite de [2].

Le rapport de surfaces entre le tympan et la fenêtre ovale, environ 20 pour 1, engendre une deuxième amplification. L'amplification mécanique de l'oreille moyenne varie selon la fréquence des ondes à transmettre et peut atteindre 20 dB aux fréquences les plus amplifiées. La figure 1.3 montre le gain total apporté par l'oreille moyenne en fonction de la fréquence.

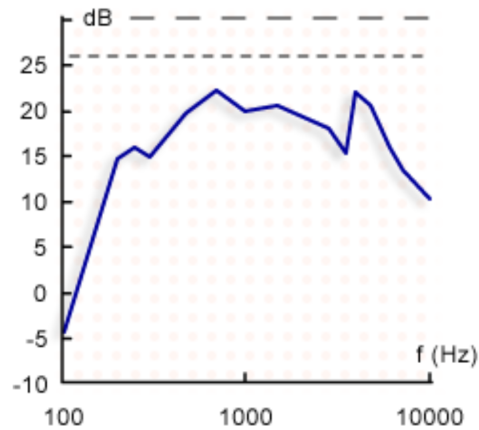


FIGURE 1.3 – Gain apporté par l'oreille moyenne en fonction de la fréquence. Image extraite de [2].

Oreille interne et cochlée

L'oreille interne contient la cochlée, un organe creux, enroulé sur lui-même qui permet une décomposition fréquentielle des sons et une transduction des ondes mécaniques en influx nerveux pouvant circuler dans les neurones auditifs. Trois canaux s'étendent le long de la cochlée : le canal cochléaire, rempli d'endolymphe, et les rampes vestibulaire et tympanique, contenant toutes deux de la périlymphe (voir figure 1.4). Lorsque la fenêtre ovale bouge sous l'effet de l'étrier, elle transmet ses vibrations à la périlymphe dans la rampe vestibulaire à la base de la cochlée. Ces vibrations se propagent le long de la cochlée et font vibrer la membrane basilaire. La zone qui vibre avec l'amplitude maximale dépend de la fréquence de l'onde, les basses fréquences affectant l'apex et les hautes fréquences affectant la base de la cochlée (voir figure 1.5). Cette organisation de la membrane basilaire selon la fréquence est appelée tonotopie cochléaire. Une fois la zone d'amplitude maximale atteinte, l'onde passe dans la rampe tympanique, puis se dirige vers la fenêtre ronde.

Cette membrane basilaire est recouverte de cellules sensorielles appelées cellules ciliées. Les cellules ciliées internes (CCI) sont les éléments transducteurs de la cochlée qui transforment l'information mécanique en information chimique afin d'exciter les neurones auxquels elles

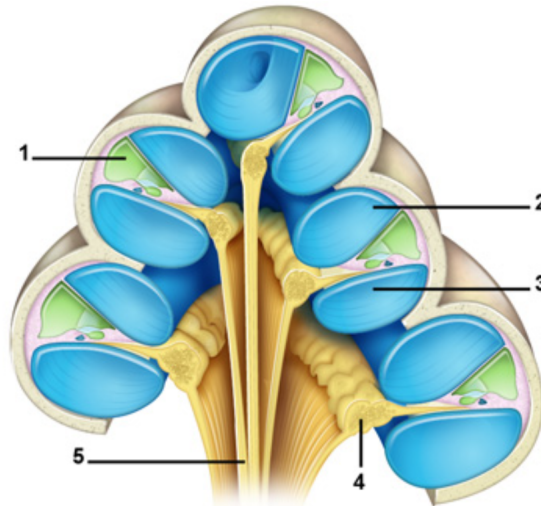


FIGURE 1.4 – Coupe de la cochlée. (1) canal cochléaire, (2) rampe vestibulaire, (3) rampe tympanique, (4) ganglion cochléaire, (5) nerf auditif. Image extraite de [2].

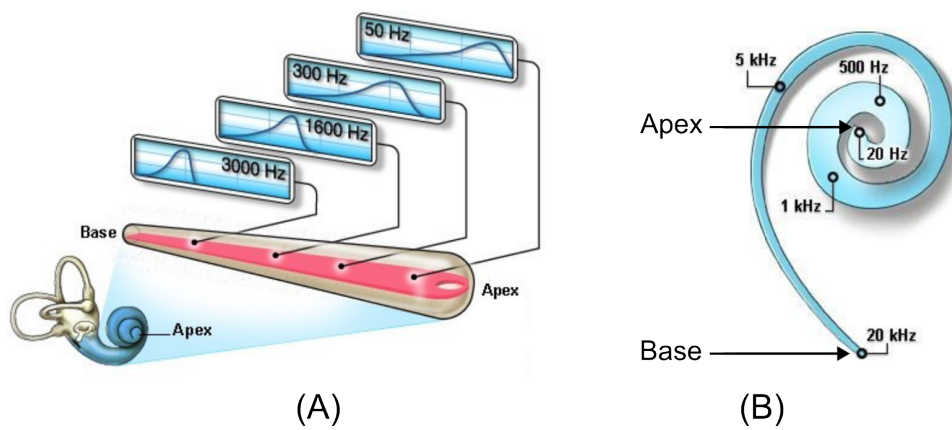


FIGURE 1.5 – (A) Vibration de la membrane basilaire en fonction de la fréquence de l'onde sonore. (B) Tonotopie cochléaire de 20 Hz à 20 KHz. Images extraites de [2].

sont connectées. Le neurone est la plus petite partie du système nerveux qui assure la transmission de l'influx nerveux au travers du corps humain (figure 1.6). Les dendrites sont de petites ramifications autour du corps cellulaire qui conduisent l'influx nerveux vers ce dernier. Le corps cellulaire est la partie centrale du neurone. L'axone est la partie par laquelle l'influx est propagé hors du corps cellulaire vers les neurones suivants. Tout au bout de l'axone, on retrouve les terminaisons nerveuses. Les connexions entre les cellules ciliées et les neurones ou d'un neurone à l'autre s'établissent au niveau des dendrites. Ces connexions sont nommées des synapses (figure 1.6). Les CCI possèdent des vésicules de neuromédiateur à la base de leur corps cellulaire qui est connecté aux dendrites des premiers neurones du nerf auditif. Lorsque la membrane basilaire se déforme sous l'effet de la vibration d'un son, les CCI entrent en contact avec la membrane tectoriale ce qui a pour effet d'ouvrir des canaux ioniques, situés au voisinage du sommet des cellules. Un courant de cations s'établit vers l'intérieur de la cellule et provoque la dépolarisation de la cellule. Lorsque la dépolarisation est suffisante, l'émission de vésicules de neuromédiateurs dans la fente synaptique a lieu et aboutit à l'apparition de potentiels d'action dans les neurones connectés à la cellule qui a été dépolarisée.

Un autre type de cellules ciliées est présent dans la cochlée. Il s'agit des cellules ciliées externes (CCE) qui n'ont pas pour fonction de transmettre des signaux aux neurones. De la même manière que pour les CCI, les vibrations acoustiques induisent un mouvement des cils des CCE, ce qui modifie le potentiel de ces cellules. Cependant, contrairement aux CCI, ces variations de potentiel entraînent un changement de forme des CCE. En fonction de leur polarisation, elles se raccourcissent ou s'allongent, agissant ainsi comme des oscillateurs qui amplifient localement la vibration de la membrane basilaire. Cette activité des CCE augmente la sensibilité de la cochlée d'environ 50 dB et améliore sa capacité à discriminer des fréquences très proches. [4]. De plus, les cellules ciliées sont en contact avec des fibres nerveuses efférentes qui transmettent des potentiels d'action depuis le cerveau vers l'oreille interne permettant un contrôle du cerveau sur l'activité de la cochlée. Ce contrôle permet notamment d'adapter sélectivement l'amplification effectuée par les CCE [5-7]. La figure 1.7 schématise l'organisation des cellules ciliées par rapport aux membranes basilaire et tectoriale à l'intérieur de la cochlée.

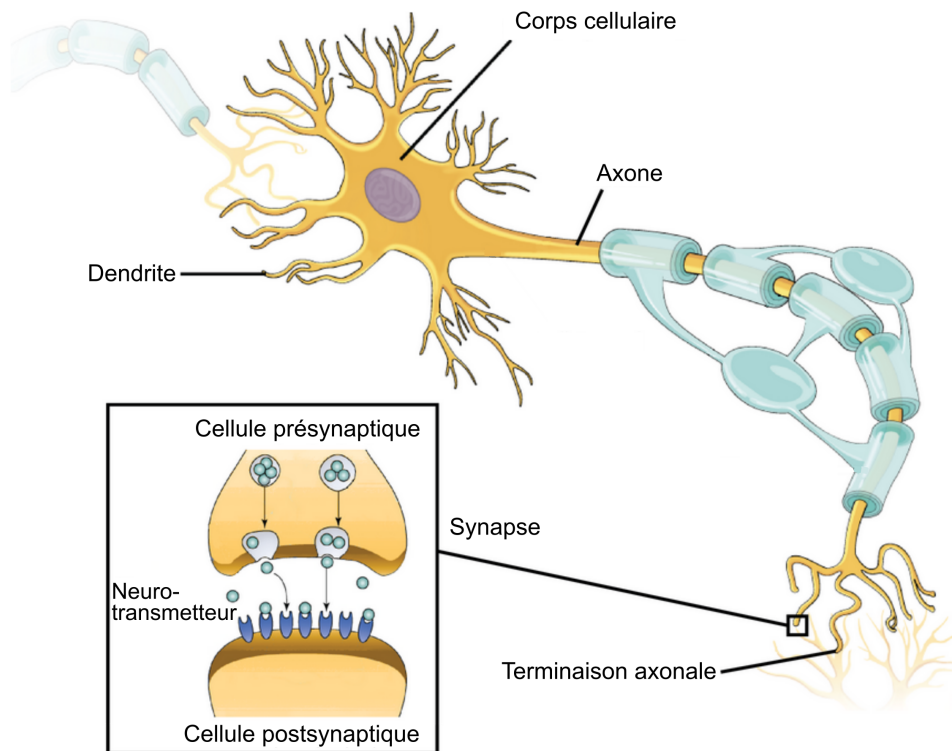


FIGURE 1.6 – Schéma d'un neurone et d'une connexion synaptique. Image adaptée de [3]

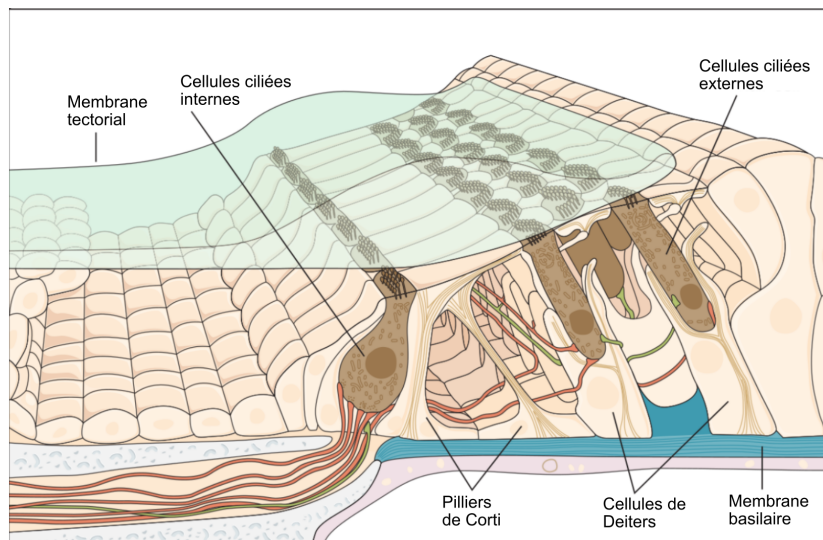


FIGURE 1.7 – Schéma de l'organisation des cellules ciliées par rapport aux membranes basillaires et tectoriales à l'intérieur de l'organe de Corti. Image adaptée de [8]

Codage dans les voix nerveuses

Les CCI effectuent la transduction des signaux mécaniques en potentiels d'action électriques qu'elles transmettent au nerf auditif auquel elles sont connectées. Le nerf auditif est constitué d'un grand nombre de neurones qui assurent la transmission des potentiels d'action au cerveau.

La tonotopie fréquentielle obtenue au niveau de la cochlée est conservée dans le nerf auditif puisque chaque neurone n'est connecté qu'à une seule CCI. Les potentiels d'action ainsi créés sont transmis au cerveau par les voies auditives ascendantes. Ces voies sont constituées d'un certain nombre de relais neuraux effectuant chacun un travail spécifique de décodage et d'interprétation qui est ensuite transmis aux relais supérieurs, jusqu'au cortex auditif. Il existe la voie auditive primaire, dédiée exclusivement à la fonction auditive, et des voies non primaires où convergent l'ensemble des modalités sensorielles. La figure 1.8 représente la voie auditive primaire. Les relais de cette voie sont :

1. Les noyaux cochléaires, au niveau desquels un premier décodage de la durée, de la fréquence et de l'intensité du message sonore est effectué.
2. Le complexe olivaire supérieur qui joue un rôle dans la localisation des sons.
3. Le colliculus inférieur qui joue également un rôle dans la localisation des sons.
4. Le corps genouillé médian (thalamus), au niveau duquel s'effectue un travail d'intégration et de préparation d'une réponse motrice, par exemple vocale.

Après le thalamus, le message sonore est transmis au cortex auditif (figure 1.9) où peuvent être exécutées des fonctions de plus haut niveau telles que la reconnaissance du message, sa mémorisation, ou encore son intégration dans une réponse volontaire.

Il est important de noter que l'analyse fréquentielle réalisée au niveau cochléaire est conservée au travers de chaque centre auditif. Les neurones répondant à une fréquence de manière préférentielle sont regroupés dans une zone donnée, et les différentes zones spécifiques d'une fréquence sont rangées topographiquement de manière ordonnée, par fréquences croissantes ou décroissantes. Une deuxième caractéristique remarquable est que tous les noyaux à partir du complexe olivaire supérieur reçoivent des informations en provenance des deux oreilles. C'est en confrontant ces informations que les indices interauraux peuvent être extraits, permettant notamment une première étape de l'analyse de la localisation des sources sonores dans le complexe olivaire supérieur et le colliculus inférieur.

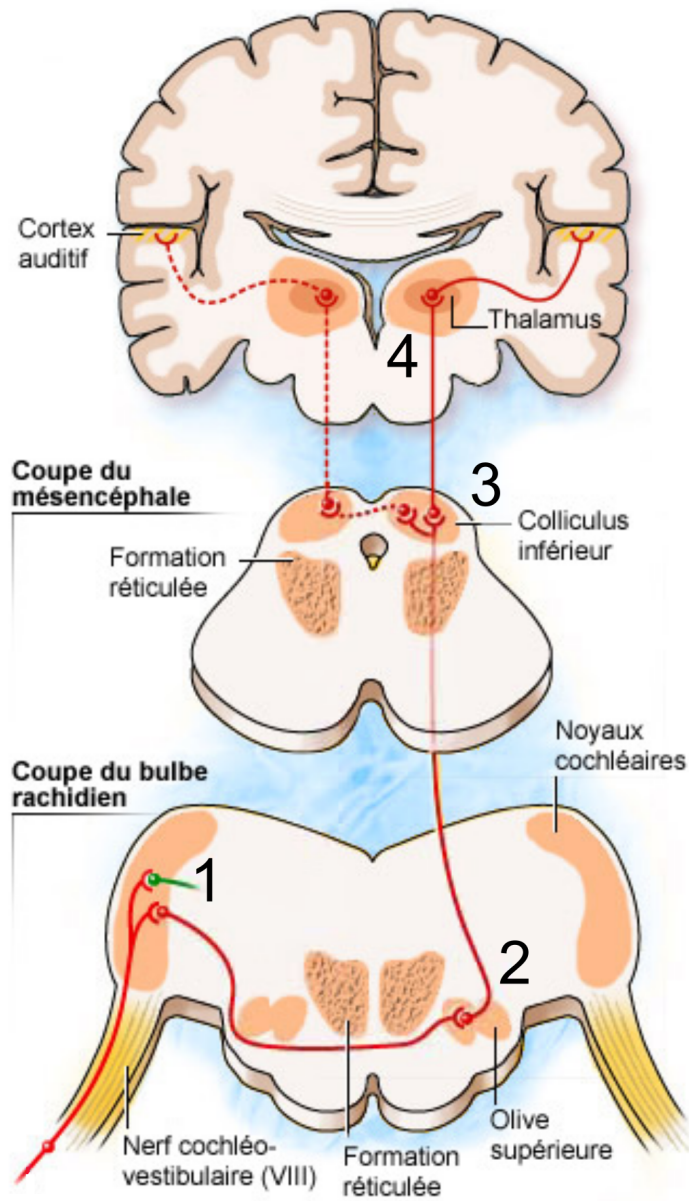


FIGURE 1.8 – Schéma des relais de la voie auditive primaire. Image adaptée de [2]

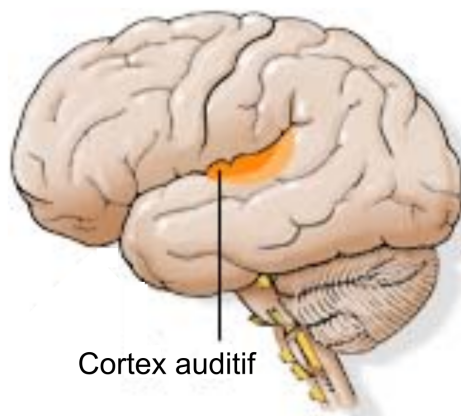


FIGURE 1.9 – Représentation de l'emplacement du cortex auditif dans le cerveau. Image adaptée de [2]

1.1.2 L'électrophysiologie

Les mesures de l'activité cérébrale permettent d'étudier les fonctions perceptives, cognitives et motrices de façon objective. Il existe une variété de méthodes pour mesurer l'activité cérébrale de façon non invasive, notamment :

- L'électroencéphalographie (EEG), qui permet de mesurer l'activité du cerveau de façon directe en enregistrant le champ électrique généré par l'activation des neurones.
- La magnétoencéphalographie (MEG), qui consiste à enregistrer le champ magnétique généré par l'activité neurale.
- L'imagerie par résonance magnétique fonctionnelle (IRMf), qui permet de visualiser, de manière indirecte, l'activité du cerveau en enregistrant des variations hémodynamiques (variation des propriétés du flux sanguin) cérébrales.

Les travaux présentés dans cette thèse ont été effectués avec des mesures EEG, car le dispositif EEG laisse davantage de liberté de mouvement au sujet et peut être compatible avec des protocoles incluant de mouvements de tête des sujets. De plus, la résolution temporelle précise de l'EEG permet de mesurer des potentiels évoqués qui sont utiles pour l'exploration de la perception auditive.

Fonctionnement de l'électroencéphalographie

Les synapses électriques se trouvent à travers tout le système nerveux et permettent au courant électrique de circuler passivement d'un neurone à l'autre. L'électroencéphalographie (EEG)

consiste à enregistrer les variations électriques résultant des activités synaptiques des neurones. Cette activité électrique est générée par l'activation simultanée d'un grand nombre de neurones et nécessite la synchronisation de centaines de milliers de neurones pour devenir visible à la surface du cuir chevelu. La surface corticale minimale nécessaire pour détecter un potentiel est estimée à 6 cm² [9].

L'EEG mesure la différence de potentiels entre deux électrodes placées sur le cuir chevelu. Cette différence de potentiels est mesurée entre une électrode "utile", qui recueille l'activité cérébrale que l'on souhaite étudier, et une électrode de référence, la plus inactive possible. Le choix du placement de l'électrode de référence est un problème qui a beaucoup été débattu et aucun consensus n'a été établi [9]. Dans le cas des enregistrements d'EEG auditifs, plusieurs références ont été proposées : sur le nez [10, 11], sur le sommet du crâne [12, 13], ou encore sur les mastoïdes [14, 15]. Une autre solution, souvent utilisée en clinique, est d'utiliser comme référence la moyenne de l'ensemble des électrodes [16, 17].

Le signal recueilli par les électrodes est transmis à un amplificateur puis enregistré par un ordinateur sous la forme de *tracé EEG*, dont un exemple est donné dans la figure 1.10. La morphologie des tracés dépend de l'état d'activation des neurones enregistrés et cet état d'activation va se traduire par des variations de fréquence et d'amplitude des ondes enregistrées. Comparé à d'autres techniques d'imagerie comme l'IRMf, l'EEG fournit des informations en temps réel avec une résolution temporelle très fine, de l'ordre de la milliseconde. [18].

Les électrodes

Les électrodes peuvent être passives ou actives. Avec des électrodes passives, le signal recueilli à la surface du cuir chevelu est directement transmis, au travers d'un câble, à l'amplificateur. Ces électrodes sont très simples et peu coûteuses, mais comme le signal recueilli est très faible, il est sujet à de nombreuses perturbations électromagnétiques lors du transfert jusqu'à l'amplificateur. Les électrodes actives rassemblent deux composants : l'électrode qui capte le signal électrique et le préamplificateur qui procède à une première amplification du signal pour pouvoir le transmettre avec moins de perturbations. Cependant, ces électrodes sont plus coûteuses et plus lourdes à mettre en place, car elles incluent les câbles nécessaires au fonctionnement du préamplificateur. Lors d'enregistrements dans des conditions de laboratoire, les électrodes passives sont généralement plus pertinentes puisque les perturbations électromagnétiques peuvent être correctement contrôlées. Il existe trois types d'électrodes passives :

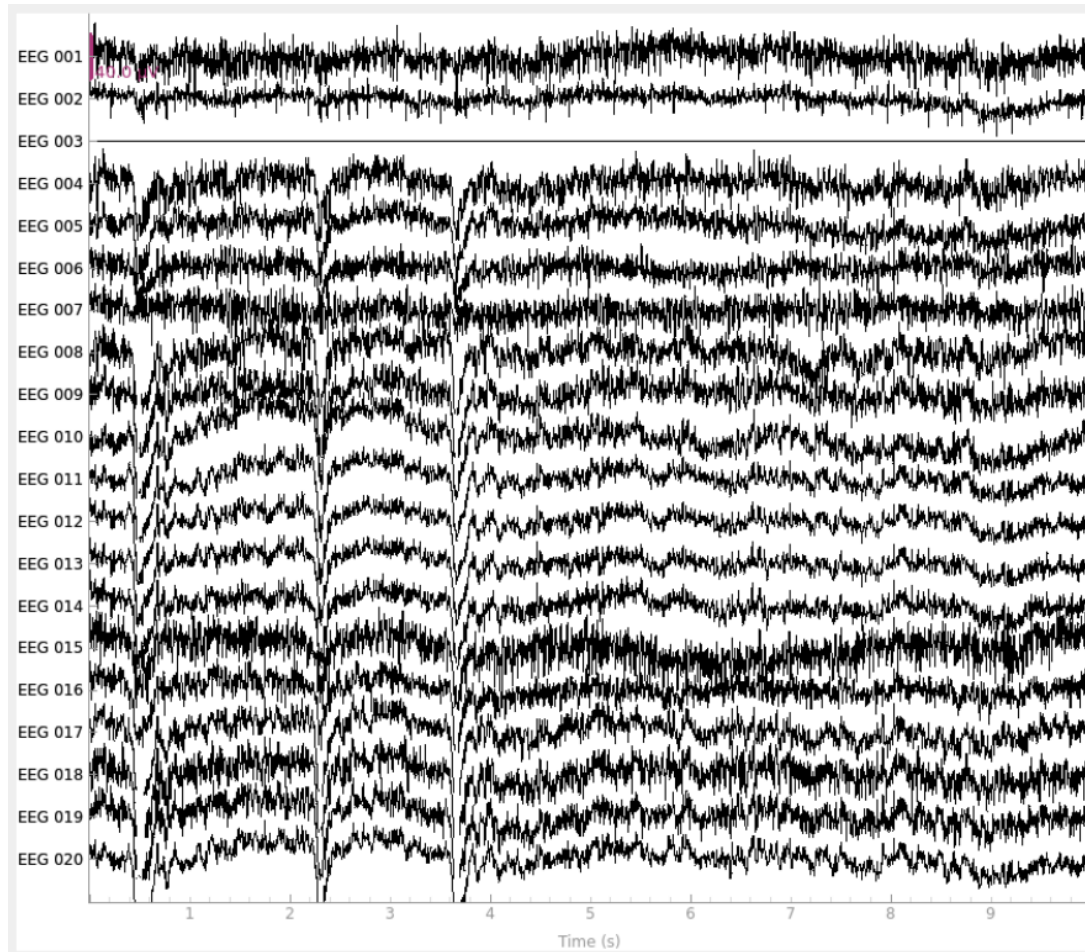


FIGURE 1.10 – Exemple de tracé EEG. Chaque ligne (EEG 001, EEG 002...) correspond au signal enregistré par une électrode. Image extraite de [19].

- L'électrode "tampon" se présente sous la forme d'un petit dôme d'environ 2 cm^2 à la base. Elle est fabriquée en chlorure d'argent et est recouverte d'un tampon de tissu imbibé d'eau saline (figure 1.11 A). Elle est simplement placée sur le cuir chevelu après dégraissage et application d'une pâte conductrice. Un bonnet spécial peut être utilisé pour la maintenir en place sur le cuir chevelu.
- L'électrode "cupule" se présente sous la forme d'un disque en argent de 5 mm à 10 mm de diamètre, avec une cupule centrale permettant l'application d'une pâte conductrice (figure 1.11 B). Elle peut également être maintenue en place à l'aide d'un bonnet. Ces électrodes sont recommandées pour les enregistrements de longue durée.
- L'électrode "aiguille" est une aiguille en acier inoxydable ou argenté, insérée juste sous la peau de façon parallèle à la surface (figure 1.11 C). Elle convient rarement aux tests en laboratoire, car elle est à usage unique et elle est plus invasive que les autres types d'électrodes.

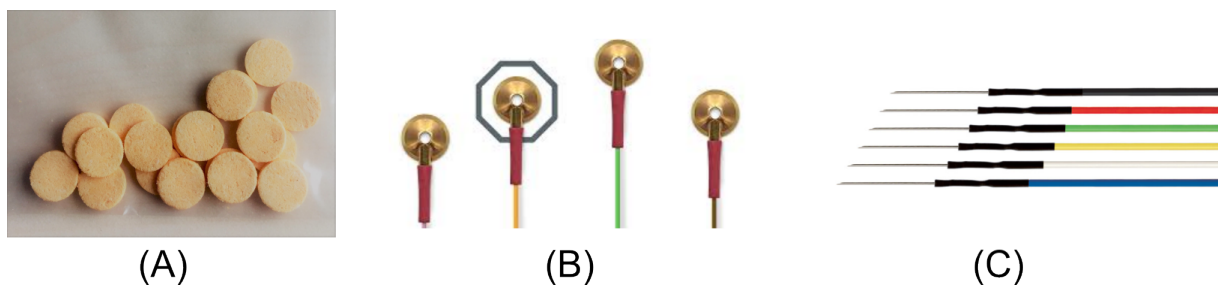


FIGURE 1.11 – (A) électrodes tampons [20], (B) électrodes cupules [21], (C) électrodes aiguilles [22].

Pour garantir un positionnement spatial précis des électrodes en fonction des repères anatomiques propres à chaque individu, des normes ont été développées. Ces normes, déterminées en fonction de la résolution spatiale souhaitée, dictent le nombre d'électrodes nécessaires. L'emplacement des électrodes est standardisé grâce à une nomenclature internationale appelée système 10-20 [23, 24] (voir figure 1.12). Ce système fixe de manière proportionnelle la position des électrodes par rapport à des repères osseux immuables. Les positions sont définies en divisant les lignes crâniennes des plans médian et transversal en segments correspondant à 10 % et 20 % de leur longueur. Chaque électrode est identifiée par un nom spécifique : une lettre majuscule indique une zone du crâne (A = le lobe de l'oreille, C = central, P = pariétal, F = frontal, Fp = frontal polaire, O = occipital, T = temporal) suivi soit d'un chiffre pair pour désigner le côté droit, soit d'un chiffre impair pour indiquer le côté gauche ou soit de la lettre "z" pour indiquer l'axe central.

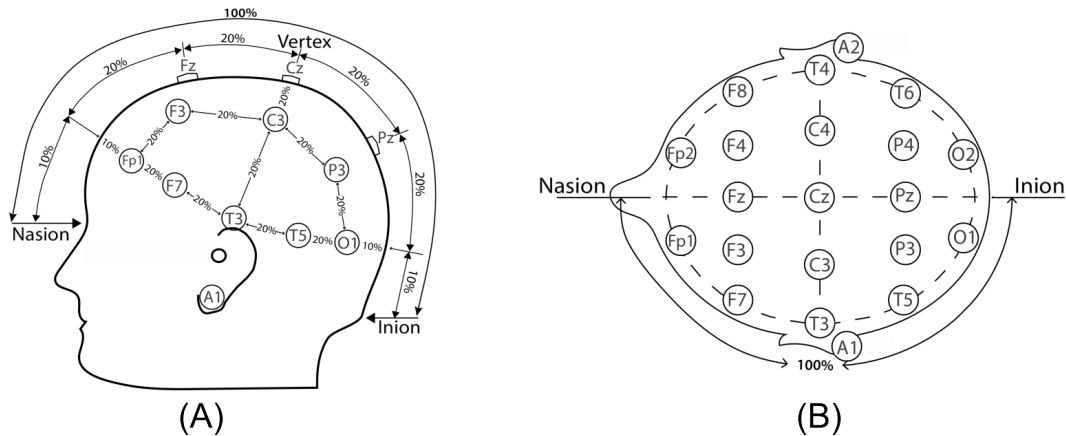


FIGURE 1.12 – Représentation du système 10 - 20 : (A) vue de côté, (B) vue du dessus. Image adaptée de [25]

Les artefacts de mesure

En raison de l’amplitude faible du signal électrique capté sur le cuir chevelu, l’électroencéphalogramme est sujet à de nombreuses perturbations qui peuvent potentiellement masquer l’activité cérébrale d’intérêt (figure 1.13 A). Ces perturbations peuvent être dues à une activité électrophysiologique non cérébrale d’origine corporelle (on parle alors d’artefact électrophysiologique propre au sujet) ou à du bruit extérieur au sujet. Les artefacts électrophysiologiques les plus communs sont :

- Les mouvements oculaires : il s’agit de déplacements volontaires ou involontaires des yeux, créant des variations lentes de potentiels sur les électrodes frontales. Ces mouvements induisent souvent des distorsions significatives du signal EEG, ce qui complique son interprétation (figure 1.13 B).
- Les clignements des yeux : il s’agit de mouvements mécaniques des paupières. Chaque clignement ou fermeture des paupières induit une déflexion importante du signal. Ces artefacts ont leur amplitude maximale au niveau des électrodes situées dans la région frontale (figure 1.13 C).
- Les artefacts musculaires : ils sont dus à l’activité des muscles de la tête ou du visage (par exemple la contraction de la mâchoire ou des sourcils) qui se traduisent par des phénomènes électriques rapides. Les artefacts musculaires produisent des signaux à haute fréquence, souvent d’une amplitude supérieure à celle de l’activité EEG mesurable (figure 1.13 D).
- L’activité cardiaque : chaque battement cardiaque induit un léger mouvement de la tête,

habituellement non perceptible. Cependant, au niveau local, les électrodes proches d'une artère peuvent être affectées par des mouvements additionnels, générant ainsi des artefacts. Ces artefacts sont généralement faciles à repérer puisqu'ils suivent le rythme cardiaque (figure 1.13 E).

Les bruits extérieurs peuvent être dus à des perturbations mécaniques telles que le mouvement d'une électrode mal fixée qui bouge sur le cuir chevelu, ou à des perturbations électromagnétiques dans l'environnement proche des électrodes telles que l'alimentation mal isolée d'un appareil électrique. Certaines perturbations électromagnétiques constantes sont faciles à identifier et à supprimer. Il est par exemple fréquent que l'activité électrique du secteur se retrouve dans l'enregistrement EEG sous la forme d'une fréquence constante à 50 Hz et à ses harmoniques (100 Hz, 150 Hz). Il suffit alors d'appliquer un filtre coupe-bande à ces fréquences. En revanche, les perturbations non continues sont plus difficiles à supprimer et peuvent masquer le signal d'intérêt.

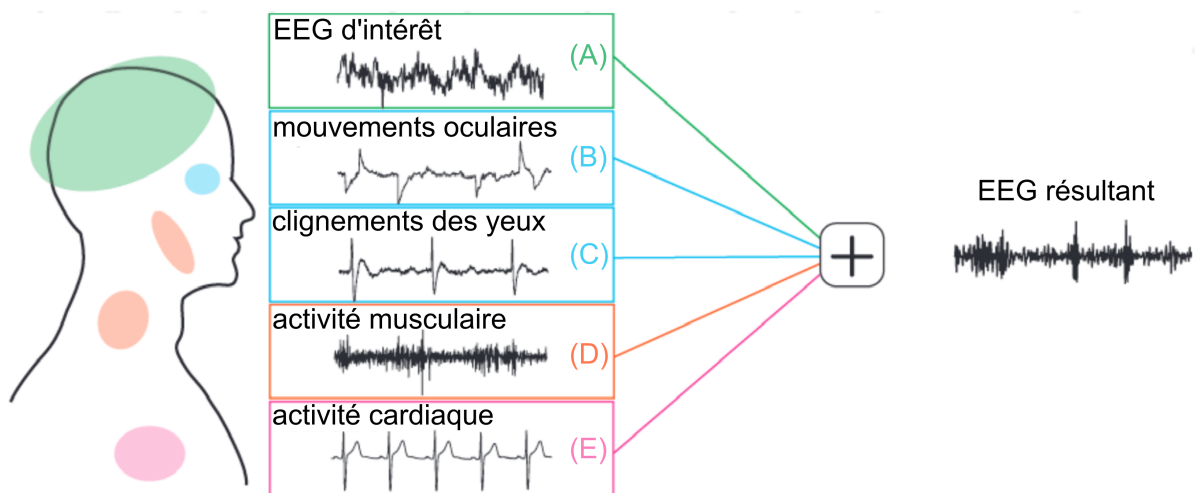


FIGURE 1.13 – Exemple des différents types d'artefacts les plus communs en EEG et leur impact sur l'EEG résultant. (A) signal EEG d'intérêt, (B) artefacts de mouvements oculaires, (C) artefacts de clignements des yeux, (D) artefacts liés à l'activité musculaire, (E) artefacts liés à l'activité cardiaque. Image adaptée de [26].

1.1.3 Les potentiels évoqués

L'exposition d'un individu à n'importe quelle stimulation entraîne une activation spécifique des neurones. Les fluctuations de tension résultantes de cette activation, enregistrables à la surface du cuir chevelu, s'appellent les potentiels évoqués (ERP pour *event-related potential*) et

sont caractéristiques de la stimulation utilisée. L'utilisation des ERP est un moyen efficace pour étudier le comportement du cerveau en fonction de la nature des stimulations utilisées. L'activité électrique des neurones évolue rapidement et couvre des zones spatiales étendues, c'est pourquoi les ERPs doivent être enregistrés avec une résolution temporelle de l'ordre de la milliseconde et à différents emplacements du cuir chevelu, à l'aide de techniques d'exploration cérébrale telles que l'EEG ou la MEG. Les travaux présentés dans cette thèse explorent les potentiels évoqués auditifs (aERP pour *auditory event-related potential*) enregistrés à l'aide d'EEG. On distingue un ERP *single trial*, correspondant à l'activité électrique générée par un seul stimulus, de l'ERP moyen (appelé simplement ERP), correspondant à la moyenne d'un grand nombre d'ERPs *single trial*. Un ERP *single trial* n'est pas exploitable individuellement, car beaucoup trop bruité. C'est pourquoi les études utilisant les ERPs se concentrent principalement sur les ERPs moyens plutôt que sur les ERPs *single trial*.

Obtention d'ERPs

La forme d'onde de l'ERP est obtenue à partir d'un procédé de répétition d'un stimulus et du calcul de la moyenne des réponses électriques de ces répétitions. Lors de l'enregistrement de l'EEG, il faut répéter chaque stimulus un grand nombre de fois. Chaque répétition génère un ERP *single trial*. Ensuite, il faut extraire les parties des tracés EEG correspondants à une fenêtre temporelle synchronisée avec chaque répétition du stimulus. On obtient ainsi un grand nombre d'extraits d'EEG, tous de la même longueur et pouvant être considérés équivalents puisqu'ils correspondent tous à l'activité électrique des neurones en réponse à la même stimulation. Ces extraits sont appelés *epochs*. Chaque *epochs* contient un ERP *single trial*. Tous ces *epochs* sont moyennés pour obtenir la forme d'onde correspondant à l'ERP d'un stimulus. Ce procédé est schématisé dans la figure 1.14. Le fait de moyennner un grand nombre d'*epochs* permet d'améliorer le rapport signal / bruit en limitant l'influence des fluctuations d'amplitude générées par l'activité aléatoire des neurones, mettant ainsi en valeur l'activité générée par le stimulus [27].

Les EEG enregistrés sur des périodes relativement longues, de plusieurs minutes, présentent souvent de faibles augmentations ou diminutions globales de l'amplitude moyenne qu'on appelle les dérives lentes. Ces dérives peuvent poser problème lors de la création des ERPs. En effet, des *epochs* enregistrés avec plusieurs minutes d'intervalle peuvent avoir une amplitude moyenne très différente à cause des dérives lentes. Si ces dérives ne sont pas compensées lors de la création des *epochs*, les ERPs résultants ne sont pas exploitables. Pour compenser la dérive lente du signal EEG dans le temps, chaque *epoch* est corrigé par rapport à sa ligne de base [29].

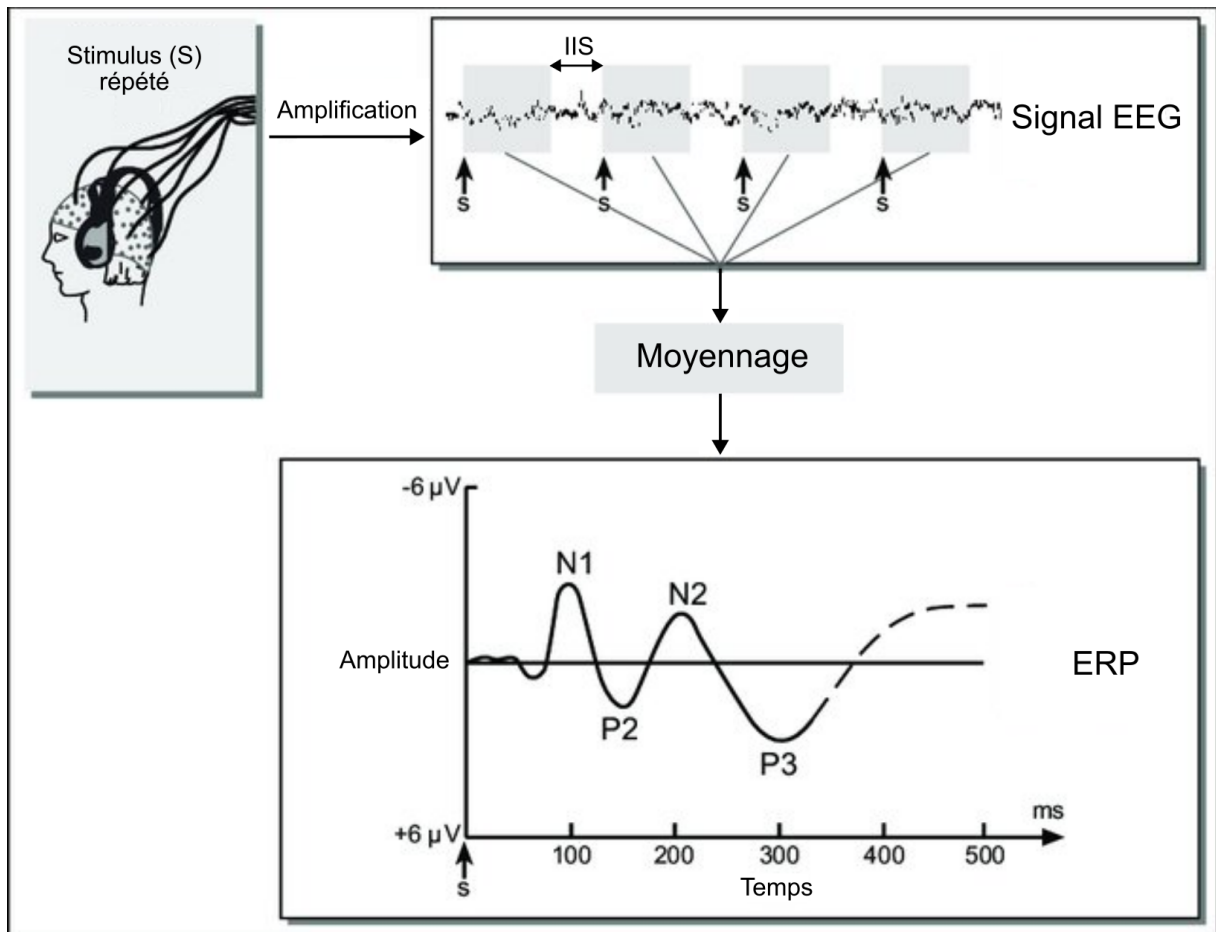


FIGURE 1.14 – Différentes étapes nécessaires à l'obtention d'ERP. le symbole "S" indique l'apparition du stimulus. IIS = Interval Inter-Stimulus. On note que l'axe de l'amplitude est "inversé" (amplitude positive vers le bas et négative vers le haut), comme c'est souvent le cas dans la littérature. Image adaptée de [28].

La ligne de base est un intervalle de base (couvrant généralement les 100 à 200 ms précédant l'apparition du stimulus) pour lequel l'activité neuronale est minimale. En effet, le décalage d'amplitude globale, dû à la dérive lente, affecte la ligne de base de la même façon que le signal après l'apparition du stimulus, mais l'activité liée à la stimulation n'est pas présente dans la ligne de base. Ainsi, en soustrayant la valeur d'amplitude moyenne de la ligne de base au signal après l'apparition du stimulus, la dérive lente est supprimée tout en conservant l'activité liée à la stimulation. Un filtre passe-haut avec une fréquence de coupure très basse peut également être utilisé pour compenser la dérive lente du signal avant la création des epochs.

Composantes ERP

Une forme d'onde de l'ERP se présente comme une succession de déflexions d'amplitudes positives et négatives que l'on appelle composantes (partie basse de la figure 1.14). Les composantes sont caractérisées par leur polarité (positive ou négative), par leur latence par rapport au début du stimulus et par leur emplacement topographique sur le cuir chevelu. Dans le cas des ERP auditifs, on distingue trois catégories de composantes classifiées en fonction de leur latence : les composantes précoces de latences inférieures à 10 ms, les composantes semi-précoces dont les latences sont comprises entre 10 ms et 80 ms et les composantes tardives dont les latences sont supérieures à 80 ms. L'origine des composantes précoces est bien connue, elles sont générées par l'activité synaptique des neurones depuis le nerf auditif et jusqu'au colliculus inférieur (*cf.* section 1.1.1). En revanche, l'origine des composantes semi-précoces et tardives n'est pas établie de façon claire. Les composantes semi-précoces pourraient être générées par l'activité synaptique des neurones du thalamus et du cortex auditif. Les composantes tardives sont probablement générées par l'activité des neurones du cortex auditif et d'autres cortex impliqués dans certains traitements cognitifs liés à la stimulation sonore.

Les composantes précoces sont dites exogènes, car elles reflètent des réponses directes du système auditif indépendamment des changements d'état psychologique du sujet. Elles ne dépendent que de la nature du stimulus et de l'état des voies auditives. C'est également le cas, dans une certaine mesure, des composantes semi-précoces. Pour ces raisons, l'étude des ERP précoces et semi-précoces est couramment effectuée lors de diagnostics médicaux, pour vérifier le bon fonctionnement des fonctions auditives. Les composantes tardives sont générées et modulées par des mécanismes cognitifs de plus haut niveau. Ces composantes sont dites endogènes, car elles dépendent du stimulus ainsi que de la tâche effectuée par le sujet, son attention et ses attentes. Les ERP tardifs sont fréquemment étudiés pour explorer des processus auditifs de haut niveau tels que la perception de la parole, la perception des rythmes dans la musique ou

la localisation des sons.

ERP tardif

La forme d'onde de la figure 1.14 montre certaines des composantes tardives les plus couramment étudiées. On les distingue par leur polarité (P pour les composantes d'amplitude positive et N pour les négatives) et leur latence. Par exemple, la N1 désigne une composante négative qui apparaît autour de 100 ms après le début du stimulus et la P2 désigne une composante positive qui apparaît autour de 200 ms après le début du stimulus [30]. La liste suivante décrit les composantes pertinentes par rapport aux travaux présentés dans cette thèse :

- **N1 / P2** : L'apparition d'un stimulus auditif ou visuel génère une réponse ERP caractéristique connue sous le nom de complexe N1-P2, qui consiste en une déflexion négative autour de 100 ms suivie d'une déflexion positive autour de 200 ms. Bien que fortement corrélées aux propriétés physiques du stimulus, ces composantes sont parmi les premières à présenter des variations en fonction de variables cognitives. Par exemple, l'augmentation d'amplitude observée lorsqu'on compare la N1 en réponse à des stimuli attendus par rapport à la N1 en réponse à des stimuli non attendus est considérée comme reflétant l'activité cognitive liée à l'attention sélective, qui facilite la détection du stimulus [31-33]. Il est important de noter que certaines études montrent également des variations liées à l'état cognitif plus précoces comme la P50 [34].
- **N2 / MMN** : Une deuxième déflexion négative de la réponse ERP associée au traitement cognitif est la composante N2 qui se produit environ 200 ms après le début du stimulus. Contrairement à la N1 qui est systématiquement générée lors d'une stimulation, la composante N2 est générée par la détection d'un stimulus déviant des stimuli précédents [35]. La N2 peut être divisée en deux sous-composantes, la N2a et la N2b, sur la base de distributions topographiques et de caractéristiques fonctionnelles différentes [36]. La N2b n'est déclenchée que si le stimulus déviant est pertinent pour la tâche effectuée par le sujet et est donc associée à des processus de détection consciente de la déviance. En revanche, l'apparition de la N2a est indépendante de l'attention. Cette dernière sous-composante est équivalente à la *mismatch negativity* (MMN) décrite pour la première fois par Näätänen *et al.* en 1978 [37]. La MMN est supposée refléter un "mécanisme automatique et préconscient de détection des changements" [38]. Le fait que la MMN soit déclenchée de manière fiable même en l'absence d'attention focalisée en a fait un objet d'étude pertinent dans l'exploration des processus cognitifs automatiques.

- **P300** : La composante P300 (ou P3) est également générée dans le contexte de paradigmes "standard - déviant" (paradigmes *oddball*). Il s'agit d'une déflexion positive prédominante de l'ERP qui atteint généralement son maximum environ 300 ms après le début du stimulus. Découverte pour la première fois par Sutton *et al.* en 1965 [39], cette onde est probablement la composante de l'ERP la plus étudiée et la plus fiable (pour une revue détaillée, voir [40, 41]). Comme le N2b, son apparition est liée à la détection consciente d'un stimulus inattendu. Elle peut être observée en réponse à des stimuli particuliers qui ont été définis a priori comme des cibles au sein d'une tâche de détection de la déviance ou qui sont suffisamment saillants pour déclencher un changement d'attention involontaire. L'apparition de la P300 peut donc être considérée comme un indicateur de la classification consciente d'un stimulus comme étant "différent". Inversement, l'absence de P300 en réponse à un stimulus déviant est généralement considérée comme un signe que la déviance a échappé à la conscience [42-44]. On distingue fréquemment deux sous-composantes de la P300 : la P3a, qui est provoquée par un nouveau stimulus et est associée à un effet de surprise, et la P3b, que l'on observe lorsque le sujet se trouve face à un stimulus imprévisible qui demande une réponse (donc en lien avec la mémoire, l'évaluation d'un stimulus et la prise de décision). Pour une revue complète de ces sous-composantes, voir [45].
- **ACC** : Le complexe de changement acoustique (ACC pour *acoustic change complex*) apparaît à la suite d'un changement au sein d'un stimulus auditif continu en cours. Le changement à l'origine de l'apparition d'un ACC peut être de différentes natures : changement d'enveloppe spectrale [46-49], changement de période [46], changement d'intensité sonore [48-51] ou changement de localisation [52]. Ses composantes sont organisées de façon temporellement similaire aux composantes N1 / P2 et ont été nommées N1' et P2' [52] ou cN1 et cP2 [47]. La cN1 apparaît environ 100 ms et la cP2 environ 200 ms après que le changement s'est produit dans le stimulus. Certains auteurs ont également défini une composante P1', qui peut apparaître 50 ms après le changement, comme faisant partie de l'ACC [52]. Ainsi, le complexe N1 / P2 est le marqueur de l'apparition d'un stimulus et le complexe ACC est le marqueur d'un changement au sein d'un stimulus. L'ACC peut servir de mesure objective des processus neuronaux qui sous-tendent les changements au sein d'un stimulus acoustique.

L'apparition et la modulation de ces ERPs dépendent donc de la nature des stimuli ainsi que de l'état psychologique de l'auditeur. Ainsi, la mise en relation des ERPs tardifs avec la perception subjective des stimuli peut permettre de caractériser objectivement des aspects de

la perception tels que la localisation des sources, le niveau sonore ou l'externalisation. Les corrélats neuronaux de l'externalisation sont discutés dans la section 1.4.3.

ERPs et protocoles expérimentaux

La mise en place de protocoles expérimentaux pour étudier les ERPs repose sur un certain nombre de compromis entre la durée des stimuli, nombre de répétitions de chaque stimulus, la durée entre deux stimuli et la longueur du test.

Les tests utilisant des EEG peuvent être relativement pénibles pour les sujets, car ceux-ci doivent limiter au maximum leurs mouvements afin de ne pas contaminer les enregistrements EEG tout en restant concentrés et alertes. Même si les sujets n'ont pas de tâche particulière à effectuer, il est nécessaire qu'ils restent dans le même état d'éveil tout au long du test. Par exemple, si un sujet passe d'un état complètement éveillé à un état de somnolence au cours du test, l'enregistrement EEG résultant sera différent d'un état à l'autre. De plus, l'impédance des électrodes peut varier au cours du temps. Les électrodes humides (*e.g.* les électrodes imbibées d'eau ou celles utilisant du gel) en particulier ont tendance à sécher avec le temps. Pour toutes ces raisons, il est nécessaire que les sessions de test EEG ne soient pas trop longues et il est généralement admis qu'une session ne doit pas dépasser 60 minutes. Si un protocole donne lieu à des sessions trop longues, il peut être divisé en plusieurs sessions, mais cela est à éviter, car l'état psychologique du sujet peut varier de façon importante d'un jour à l'autre, ce qui peut avoir une répercussion sur les enregistrements EEG.

Un certain nombre de paramètres dépendent des composantes ERPs qu'on souhaite étudier et ont une influence sur la durée du protocole :

- Le choix de la fenêtre temporelle de création des epochs. Plus les composantes qu'on souhaite étudier sont tardives, plus cette fenêtre devra être grande. Donc, plus l'intervalle inter-stimulus (IIS) devra être grand. l'IIS correspond au temps entre la fin d'un stimulus et le début du stimulus suivant.
- Le nombre de répétitions du stimulus. Les composantes ERPs les plus larges ont besoin de moins de répétitions que les composantes les plus petites. Par exemple, une composante large comme la P300 nécessite entre 30 et 60 répétitions (*e.g.* [53]), une composante de taille moyenne comme la N1 nécessite entre 100 et 200 répétitions (*e.g.* [54]), tandis qu'une composante de petite taille comme la P1 nécessite plus de 200 répétitions (*e.g.* [55]).

- Les stimuli utilisés. Les stimuli les plus utilisés dans les premières études des ERPs et encore largement utilisés aujourd'hui sont des sons très courts et simples tels que des clics ou des impulsions de bruit rose (des *bursts*, *e.g.* [15, 56]). Ce genre de stimulus court peut être répété un grand nombre de fois et permet donc de générer facilement des ERPs. Néanmoins, dans certains cas comme l'étude de la parole ou quand on souhaite un cadre d'étude plus écologique, il est nécessaire d'utiliser des stimuli plus complexes tels que de la voix (*e.g.* [57, 58]) ou de la musique (*e.g.* [59, 60]). Dans ces cas-là, les réponses ERP associées peuvent être plus complexes et contenir des chevauchements de composantes, rendant l'interprétation potentiellement plus difficile.

1.2 La localisation auditive

La technologie binaurale repose sur différentes techniques de spatialisation du son. Ces techniques exploitent les mécanismes sollicités par le système auditif pour localiser les sons dans l'environnement de l'auditeur. Les oreilles sont positionnées à des emplacements distincts, à droite et à gauche de la tête de l'auditeur. La comparaison par le système auditif des signaux captés par chacune des oreilles est un premier mécanisme permettant de localiser les sources. De plus, lors de son déplacement de la source aux oreilles d'un auditeur, les paramètres physiques de l'onde sonore sont modifiés par l'environnement de propagation ainsi que par le corps de l'auditeur, notamment par les réflexions sur la tête, le buste et le pavillon des oreilles. Ces modifications dépendent de la position de la source sonore et constituent un indice que le système auditif est capable d'interpréter pour localiser la source. La description de la position des sources dans l'espace s'appuie sur un référentiel propre à l'auditeur, dont la tête constitue le centre (figure 1.15).

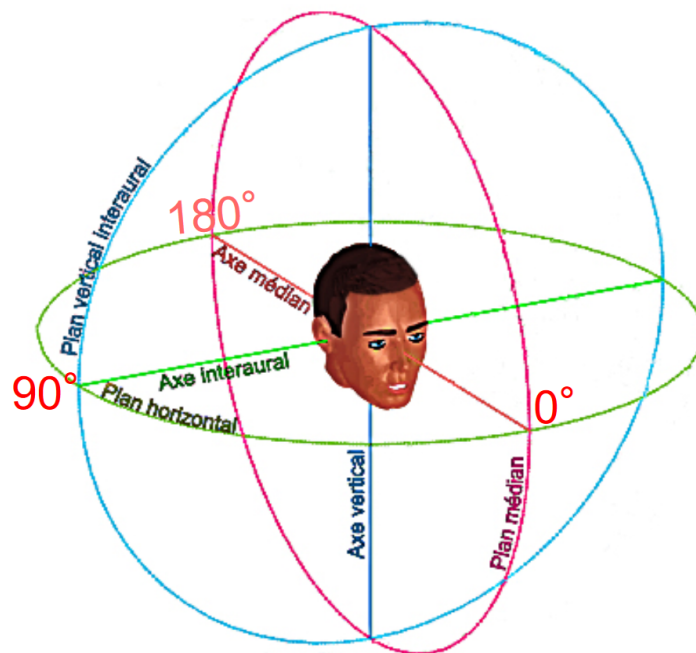


FIGURE 1.15 – Représentation des axes et plans du référentiel propre à l'auditeur. Image adaptée de [61].

1.2.1 Localisation en azimut

Un son est capté par les deux oreilles de l'auditeur qui sont situées à deux endroits différents. Ainsi, l'onde sonore issue d'une source située ailleurs que dans le plan médian n'arrive pas en même temps aux deux tympans ce qui engendre une différence interaurale de temps (ITD : *Interaural Time Difference*). De plus, l'onde sonore arrive sur les deux tympans avec une différence de niveau (ILD : *Interaural Level Difference* ou IID : *Interaural Intensity Difference*). Ces différences constituent les indices interauraux qui jouent un rôle important pour la localisation des sons en azimut [62].

Différence interaurale de temps

Les Différences Interaurales de Temps (ITD) se produisent lorsqu'un son atteint chaque oreille à des instants différents en raison de la différence de temps qu'il faut aux ondes sonores pour parcourir les distances entre la source sonore et chaque oreille. Lorsqu'une source est située à un emplacement latéral par rapport à l'auditeur, le son atteint d'abord l'oreille la plus proche (l'oreille ipsilatérale), provoquant ainsi un délai temporel avant d'atteindre l'oreille opposée (l'oreille controlatérale).

Woodworth [63] a proposé un modèle de tête sphérique pour estimer les ITD en fonction de la position de la source et du rayon de la tête. La figure 1.16 donne une représentation visuelle de ce modèle. La source sonore parcourt la même distance et arrive donc en même temps à l'oreille ipsilatérale et au point A. Pour arriver à l'oreille controlatérale, la source devra parcourir une distance supplémentaire D

Tant que la longueur d'onde est suffisamment grande par rapport à D , l'ITD est assimilable à une différence de phase (IPD : *Interaural Phase Difference*). Brian Moore [65] a montré que l'IPD constitue un indice de localisation efficace jusqu'à 1500 Hz (voir également [66]).

Différence interaurale d'intensité

Un autre indice que le système auditif est capable d'exploiter pour localiser les sources en azimut est la différence entre les niveaux sonore capté à chaque oreille (ILD, parfois appelé IID). Pour une source latéralisée, la tête de l'auditeur se trouve sur le chemin acoustique pour atteindre l'oreille controlatérale et agit comme un obstacle diffractant et réfléchissant les ondes sonores en direction de cette dernière. Plus une source est située à un angle grand par rapport au plan médian, plus la tête constitue un obstacle important sur le chemin des ondes en direction de l'oreille controlatérale. Ainsi, l'ILD varie en fonction de l'azimut de la source mais également

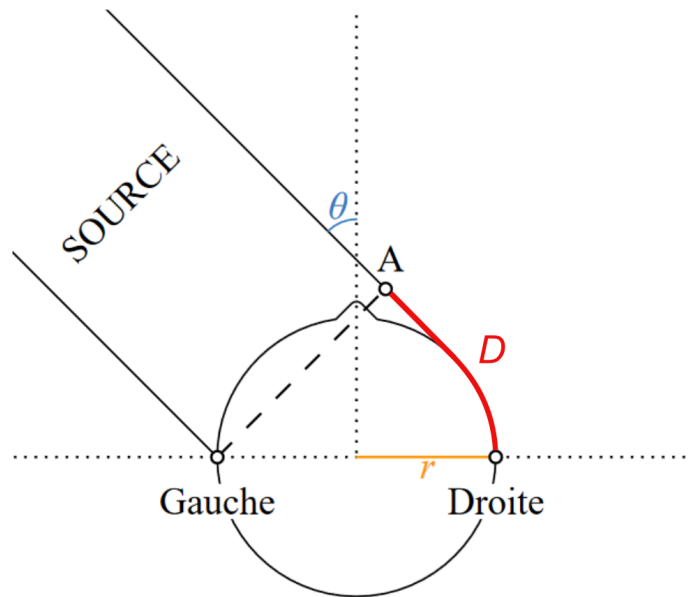


FIGURE 1.16 – Modélisation de ITD pour une source d'azimut θ atteignant les oreilles d'une tête de rayon r . Image adaptée de [64]

de la fréquence du son. En effet, pour un son de basse fréquence ($\lambda \geq r$ avec λ la longueur d'onde et r le rayon de la tête) la tête ne constitue pas un obstacle significatif et l'ILD résultant sera négligeable. En revanche, plus la fréquence augmente ($\lambda < r$) plus les phénomènes de diffraction et de réflexion sont importants et plus l'ILD augmente [67]. Plusieurs études ont déterminé que l'ILD devenait un indice de localisation efficace à partir de 1000 Hz environ [67-69]. La figure 1.17 montre l'évolution de l'ILD, mesurée sur un sujet, en fonction de l'angle d'incidence de la source et de la fréquence du son. On peut voir que pour une fréquence basse (272 Hz) l'ILD reste faible (< 3 dB) quel que soit l'angle d'incidence. En revanche, plus la fréquence augmente, plus un angle d'incidence élevé engendrera une ILD importante.

Limitation des indices interauraux

Les ITD et ILD permettent de localiser les sources sonores en azimut. Cependant, ces indices ne sont pas suffisants puisqu'il existe différentes positions de l'espace pour lesquelles les ITD et les ILD sont identiques, notamment les ITD et les ILD estimées à des positions symétriques par rapport au plan frontal. L'existence de ces ambiguïtés peut conduire à des inversions avant-arrière telles que schématisées dans la figure 1.18 : l'auditeur perçoit une source avec un angle d'incidence de 45° alors que la source a un angle d'incidence de 135° [71].

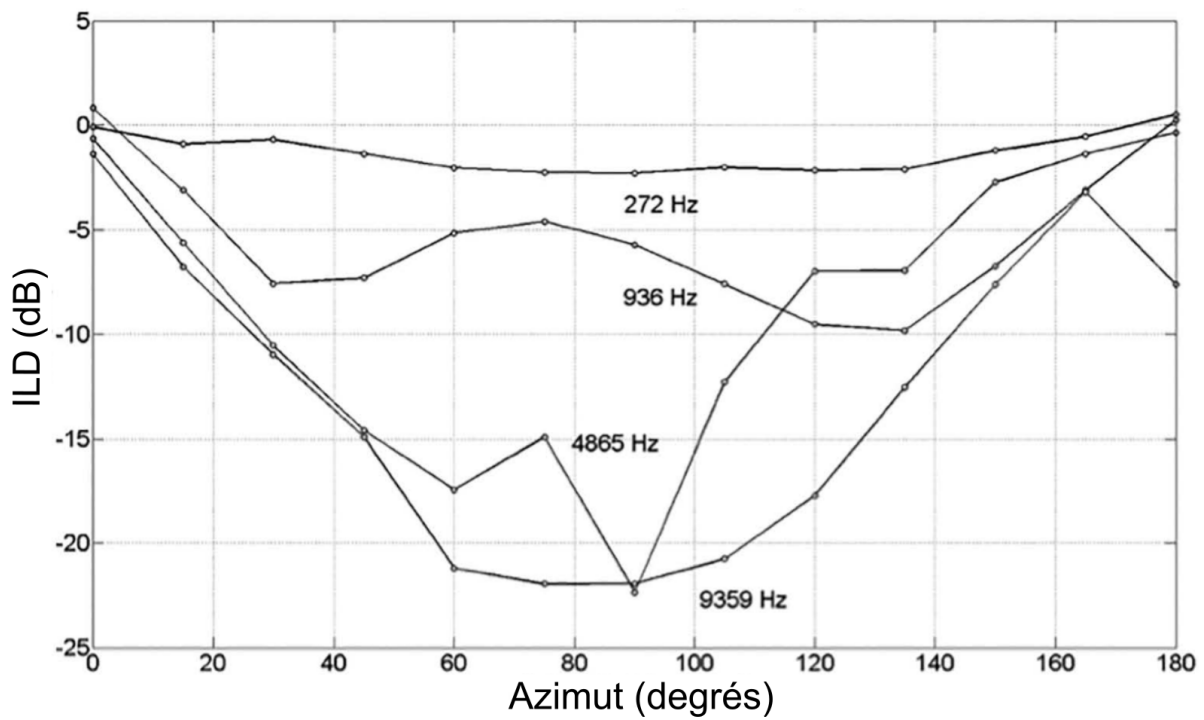


FIGURE 1.17 – Évolution de l'ILD en fonction de l'angle d'incidence de la source et de la fréquence du son. Mesurée sur un sujet à 71 dB SPL. Image adaptée de [70].

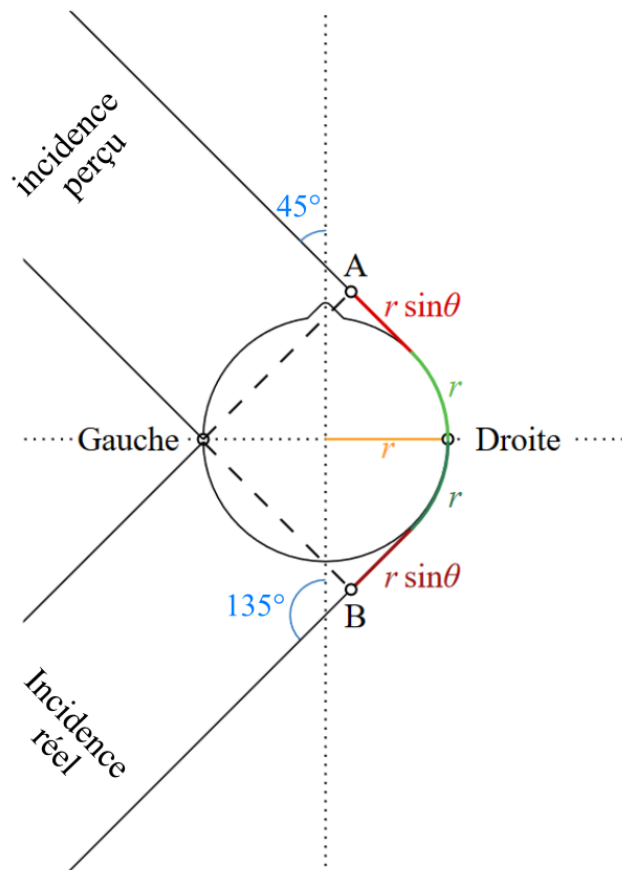


FIGURE 1.18 – Exemple d’inversion avant-arrière : une source avec les même ITD et ILD peut être perçu à 45° ou à 135° . Image adaptée de [64]

Il est cependant possible de lever cette ambiguïté au moyen de mouvement de tête de l'auditeur. En effet, lors d'un mouvement, les ITD et ILD évoluent en fonction du mouvement et cette évolution sera différente pour des sources à 45° et 135° . Le système auditif est capable d'interpréter ces modifications dynamiques d'ITD et d'ILD afin d'éviter de potentielles inversions [72-74].

Un type similaire d'ambiguïté peut apparaître en élévation. De la même façon que dans le plan horizontal, il existe plusieurs points pour lesquelles les ITD et ILD sont identiques dans le plan vertical. Par exemple, toutes les sources localisées à 0° en azimut engendrent des différences interaurales identiques, quelle que soit leur élévation. Dans ce cas précis, des mouvements de tête de l'auditeur engendrent des modifications d'ITD et d'ILD plus importants si la source est localisé à 0° que si la source est localisée à 90° en élévation et permettent ainsi de lever l'ambiguïté. Toutefois, pour des sources situées hors du plan médian, les différences de modification d'ITD et d'ILD engendrées par des mouvements de tête entre des sources plus ou moins élevées deviennent moins importantes et ne permettent plus de localiser les sources en élévation. En fait la localisation en élévation est plutôt permise par les indices spectraux (cf. section 1.2.2).

En considérant l'azimut et l'élévation, on remarque que des ITD ou ILD constantes peuvent renvoyer à un grand nombre de positions possibles pour la source sonore. Pour une différence interaurale donnée, l'ensemble des positions possibles se présente comme des hyperboloïdes ayant pour foyer les oreilles de l'auditeur. Elles ont une taille qui dépend de l'ITD ou ILD considérée et, en champ lointain, peuvent être assimilées à des cônes. C'est pourquoi ces zones d'ambiguïtés se nomment "cônes de confusion" [75]. la figure 1.19 représente deux cônes de confusion obtenu pour deux ITD différents. Dans cet exemple, l'ITD du cône A est inférieur à l'ITD du cône B.

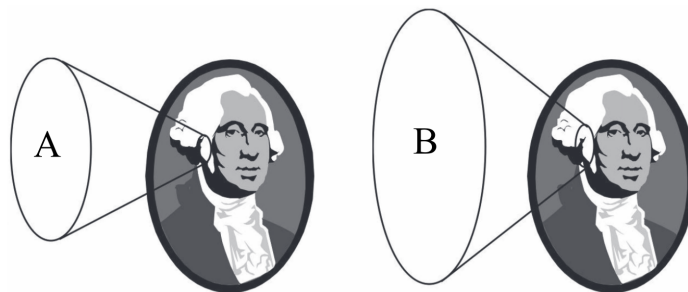


FIGURE 1.19 – Exemple de cônes de confusion : $ITD_A < ITD_B$. Image adaptée de [76]

1.2.2 Localisation en élévation

Les modèles descriptifs d'ITD et d'ILD exposés précédemment considèrent la tête comme une sphère rigide sans prendre en compte les détails de la morphologie humaine. Le signal arrivant aux oreilles est modifié, lors de son trajet, par des micro réflexions sur le buste, la tête et les pavillons des oreilles de l'auditeur ce qui engendre une modification du spectre sonore, autrement dit une coloration spectrale. Le pavillon des oreilles fonctionne comme un système de plusieurs résonateurs dont les modes sont plus ou moins excités en fonction de l'angle d'incidence de la source. Ces modes peuvent apporter une coloration spectrale sur des fréquences s'étendant de 3 kHz à 17 kHz [77, 78]. L'influence du torse est principalement liée à l'angle d'élévation des sources. Pour une source proche du plan horizontal, il n'y a que peu de réflexions sur le torse et son effet est négligeable. En revanche, il induit des réflexions importantes pour des sources en hauteur et produit un effet d'ombre acoustique pour des sources en dessous de la tête de l'auditeur [79].

Le système auditif est capable d'interpréter ces modifications du signal sonore afin de localiser la source en élévation [80] et d'éviter les inversions avant-arrière évoquées précédemment [81]. On parle alors d'indices spectraux de localisation. Ces indices spectraux sont monauraux puisque leur décodage repose sur une interprétation du signal à chaque oreille, indépendamment du signal à l'autre oreille.

Pour une position de source donnée, l'ensemble des modifications de temps, d'intensité et spectrales que subit le signal entre la source et le tympan de l'auditeur peut être exprimé par une fonction de transfert unique pour chaque oreille. Ces fonctions de transfert, dépendant largement de l'angle de la source, s'appellent les HRTFs et sont décrites dans la section 1.3.1

1.2.3 Localisation en distance

Niveau de la source

Le niveau sonore global d'un son constitue l'indice le plus important de localisation en distance. Plus une source s'éloigne de l'auditeur, plus son niveau sonore décroît. En champ libre, cette décroissance est de 6 dB par doublement de la distance et constitue un indice efficace de localisation en distance. Cependant, le niveau capté par l'auditeur dépend du niveau de diffusion de la source en plus de sa distance. Ainsi, des artefacts de localisation en distance peuvent apparaître dans le cas où le système auditif confond distance et niveau de diffusion [82]. Cet indice est plus efficace pour des sources connues telles que la voix ou les sons naturels que pour des sources moins fréquentes [83].

Rapport champ direct / champ réverbéré

Le rapport entre le niveau sonore du champ direct et le niveau sonore du champ réverbéré capté par les oreilles d'un auditeur est un indice efficace pour déterminer la distance d'une source dans un milieu réverbérant. On parle de DRR (*Direct-to-Reverberant energy Ratio*). Le champ direct est défini comme le son allant directement de la source aux oreilles de l'auditeur et le champ réverbéré est défini comme l'ensemble de toutes les ondes réfléchies sur les différentes surfaces de l'environnement avant d'arriver aux oreilles de l'auditeur. Pour qu'un champ réverbéré existe, il faut que l'auditeur se trouve dans un lieu comportant des surfaces sur lesquelles le son peut être réfléchi. C'est le cas de n'importe quel lieu intérieur, mais également de certains lieux extérieurs comportant des surfaces de réflexion suffisamment proches telles que des lieux urbains. Le niveau du champ direct diminue de 6 dB par doublement de la distance, en revanche le niveau du champ réverbéré reste globalement constant en tout point du lieu dans lequel diffuse la source. Ainsi, une source proche de l'auditeur aura un DRR plus important qu'une source plus éloignée et le système auditif est capable d'interpréter ce ratio pour localiser une source en distance [84]. Le DRR constitue un indice de localisation en distance absolu qui n'a pas besoin de se baser sur des connaissances à priori du son et est donc très efficace [85]. Cependant, cet indice ne peut pas être utilisé dans des lieux non réverbérants tels que des lieux extérieurs sans paroi proche et il ne peut pas être utilisé au-delà d'une certaine distance. En effet, dans les lieux réverbérants suffisamment grands, il existe une distance entre la source et l'auditeur à partir de laquelle le niveau du champ direct est tellement faible que l'auditeur ne capte plus que le champ réverbéré. Au-delà de cette distance, appelé *distance critique*, le DRR reste égal à 0 et ne constitue plus un indice de localisation fiable puisque deux positions de sources à des distances différentes auront le même DRR aux oreilles de l'auditeur [86].

Basses fréquences en champ proche

La plupart des études cherchant à caractériser l'ILD considèrent des sources sonores en champ lointain pour lesquelles les ondes peuvent être considérées comme étant planes. En champ proche, la nature sphérique des ondes peut avoir une influence sur la perception du niveau. En effet, pour des distances courtes (< 1 m), d'importantes différences de niveau apparaissent, notamment pour les basses fréquences (jusqu'à 20 dB à 0.12 m d'après [87]). Ces différences de niveau induites par la distance peuvent être interprétées pour localiser des sources sonores [88].

Parallaxe acoustique

Quand une source est éloignée, la différence entre l'angle de la source par rapport au centre de la tête et l'angle de la source par rapport à l'emplacement des oreilles est négligeable (figure 1.20 A). En revanche, lorsque la source est proche, cette différence d'angle peut devenir importante (figure 1.20 B). Cette disparité entre l'angle relatif au centre de la tête et l'angle relatif à l'emplacement de l'oreille s'appelle "la parallaxe acoustique" et peut être interprétée par le système auditif pour localiser les sources en distance [89].

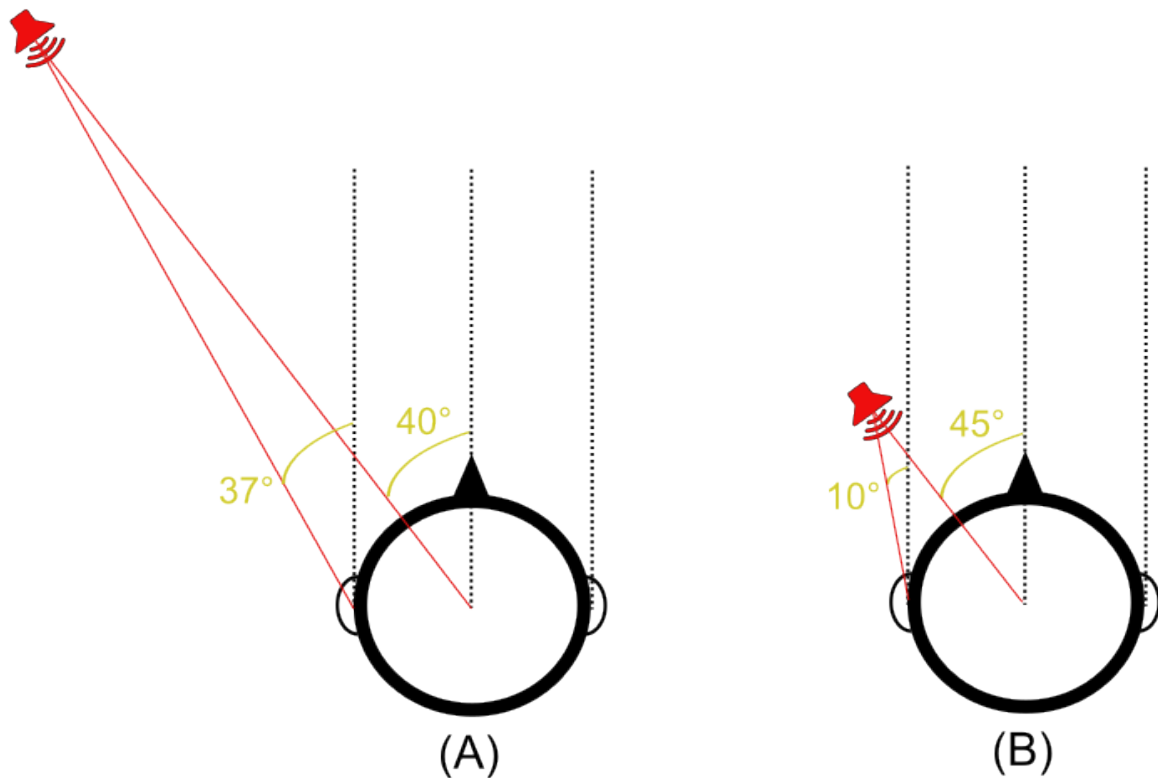


FIGURE 1.20 – Exemple de l'effet de parallaxe acoustique. (A) : source lointaine, parallaxe acoustique négligeable. (B) : source proche, parallaxe acoustique importante.

Coloration spectrale

L'énergie d'une onde sonore est progressivement absorbée par le milieu dans lequel elle se propage. Cette absorption dépend de la nature du milieu et de la distance parcourue. Plus la distance est grande, plus l'absorption est importante. Dans l'air, l'absorption est relativement faible [90] et concerne principalement les hautes fréquences. Pour un son large bande, la perte

d'énergie dans le haut de spectre peut néanmoins devenir un indice de perception de la distance au-delà de 15 m [91]. Cet indice de localisation est limité, parce qu'il n'est utilisable que pour des sources très éloignées et parce qu'il constitue un indice relatif. En effet, pour pouvoir interpréter l'atténuation des hautes fréquences comme étant dû à la distance, le système auditif doit se référer à une connaissance à priori du spectre fréquentiel de la source captée [92]. L'interprétation de cet indice peut donc être biaisée par une mauvaise connaissance à priori de la source et donner lieu à des artefacts de localisation.

1.3 La restitution binaurale

Les casques (et écouteurs) audio constituent les techniques de diffusion du son les plus directes. Les signaux audio passent des haut-parleurs droit et gauche aux conduits auditifs droit et gauche en s'affranchissant des interactions avec l'environnement et d'une partie des interactions avec la morphologie de l'auditeur. Il est ainsi possible de délivrer à l'auditeur les signaux audio les plus contrôlés possibles, le seul facteur pouvant varier d'un auditeur à l'autre étant les fonctions de transfert relatives au casque audio, les HPTFs (*HeadPhone Transfer Function*). Il devrait alors, en théorie, être possible pour les ingénieurs du son d'obtenir les contenus audio les plus réalistes possibles. Dans la pratique, lors d'une écoute au casque, le son est souvent perçu comme provenant de l'intérieur de la tête de l'auditeur et donc pas parfaitement réaliste. La restitution binaurale est une technique de reproduction du son, au casque, qui cherche à pallier les problèmes liés à l'écoute stéréophonique au casque classique pour se rapprocher davantage d'une écoute naturelle.

1.3.1 Définition et principe

Au sens strict, le son binaural désigne n'importe quel son constitué de deux canaux distincts qui pénètre chacun dans l'oreille de l'auditeur pour laquelle il est destiné. D'après cette définition, n'importe quel contenu audio stéréophonique est binaural. En pratique, les termes *restitution binaurale*, *contenu binaural* et *audio binaural* sont utilisés pour désigner les contenus audio restitués au casque après avoir été filtrés par des HRTFs. À l'inverse, les termes *contenu stéréophonique* et *audio stéréophonique* sont utilisés pour désigner le son à deux canaux, diffusés au casque sans filtrage de HRTFs.

HRTF

La coloration spectrale induite par le pavillon des oreilles et les autres parties du corps constitue un filtrage particulier qui modifie le signal sonore arrivant aux oreilles de l'auditeur de différentes manières en fonction de l'angle d'incidence de la source. Pour une position de source donnée, l'ensemble de ces effets peut être exprimé par une combinaison de deux fonctions de transfert, une pour chaque oreille, qui considérées ensemble contiennent également les différences interaurales relatives à la position de la source. Pour chaque position de source, il existe donc une combinaison de deux fonctions de transfert qui contient l'ensemble des informations de localisation. Elles sont nommées les HRTFs (*Head-Related Transfer Functions*). Les équiva-

lents temporels des HRTFs sont des réponses impulsionnelles appelées HRIRs (*Head-Related Impulse Responses*).

En filtrant n'importe quel signal sonore avec la HRTF de chaque oreille, il est possible d'imposer au signal d'origine les caractéristiques spatiales de la position correspondant aux HRTFs utilisés. Lors d'une restitution au casque des deux signaux ainsi créés, la source sonore devrait donner l'illusion de provenir de la position relative aux HRTFs. C'est le principe du son binaural.

Acquisition de HRTFs

Il est possible d'acquérir des HRTFs, qui peuvent ensuite être utilisées pour réaliser une synthèse binaurale (discuté dans la section 1.3.1). Pour chaque position de l'espace, il existe une paire de HRTFs unique. On définit l'ensemble des HRTFs connues d'une personne (ou d'un mannequin) comme un set de HRTFs. Celui-ci est défini par sa précision d'échantillonnage qui dépend du nombre de positions de l'espace pour lesquelles il existe une paire de HRTFs connue. Un set de HRTFs comportant une paire de HRTFs pour des positions espacées de 30° est considéré imprécis, alors qu'un set comportant une paire de HRTFs pour chaque position tous les 2° est considéré beaucoup plus précis, mais est plus lourd et sera plus long et complexe à acquérir. Généralement, les sets de HRTFs sont plus précis en azimut qu'en élévation, car le système auditif est moins précis pour localiser des sources en élévation [93] qu'en azimut [94] et donc les potentiels artefacts de localisation en élévation dus à l'utilisation d'un set de HRTFs imprécis sont moins perceptibles. Il est également fréquent que les sets de HRTFs ne couvrent pas les positions les plus basses ($< -45^\circ$ d'élévation), car il est très rare que des sources sonores soient situées aussi bas par rapport à l'auditeur. Par exemple, les sets de HRTFs de la base de données LISTEN [95] comportent des paires de HRTFs couvrant les positions entre -45° et $+90^\circ$ en élévations, avec une paire de HRTFs tous les 15° pour un total de 10 positions en élévation. En azimut, les sets de HRTFs comportent une paire tous les 15° pour un total de 24 mesures réparties tout autour de la tête.

La méthode la plus couramment utilisée pour acquérir des HRTFs est la mesure acoustique. Cette mesure est généralement effectuée dans une chambre anéchoïque, afin de ne capter que les effets acoustiques propres au corps de l'auditeur. Si l'enregistrement des HRTFs se fait dans un environnement non anéchoïque, les fonctions de transfert enregistrées contiendront également les modifications fréquentielles liées à la propagation acoustique dans l'environnement. Les réponses impulsionnelles associées à de telles HRTFs s'appellent des BRIR (*Binaural Room*

Impulse Response).

Un microphone est disposé au niveau de chaque oreille, généralement à l'entrée du conduit auditif bouché. Le positionnement du microphone au plus proche du tympan, avec le conduit auditif ouvert, est possible, mais plus compliqué puisqu'il faut utiliser des microphones adaptés et qu'il faut prendre en compte l'existence de noeuds et de ventres de pression dans le conduit auditif. La tête, équipée des microphones, est positionnée en un point précis et une enceinte émet un signal de mesure. Après une opération de déconvolution, on obtient un signal par oreille, qui correspond à la fonction de transfert entre n'importe quelle source à cette position et le tympan. Pour avoir un set de HRTFs le plus complet possible, il faut utiliser un système permettant de déplacer l'enceinte par pas réguliers les plus petits possibles et répéter la mesure à chaque position (voir figure 1.21). Plus le pas est petit, plus le set de HRTFs mesuré est précis, mais plus la mesure est longue. Si la tête mesurée est celle d'un humain, la longueur peut devenir contraignante, car il faut que la tête reste exactement dans la même position d'une mesure à l'autre. Il est également possible d'effectuer la mesure d'un set de HRTFs sur un mannequin tel que ceux utilisés pour l'enregistrement de binaural natif. Les HRTFs sont généralement mesurées en condition de champ lointain (> 1 m) afin de pouvoir négliger la dépendance radiale des HRTFs.

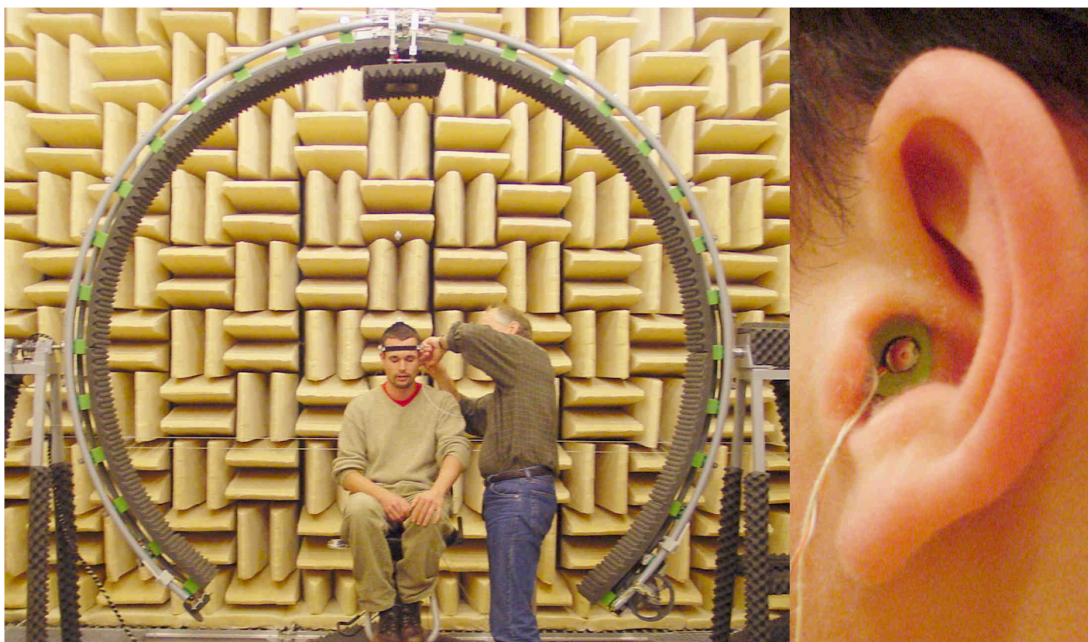


FIGURE 1.21 – Exemple de mesure de HRTFs. Image extraite de [96]

La mesure acoustique de HRTFs étant un procédé particulièrement long et complexe, différentes méthodes de modélisation de HRTFs ont été proposées. Ces méthodes peuvent permettre de calculer des HRTFs de façon précise [97] mais requièrent une puissance de calcul numérique très importante, ce qui constitue un frein à leur développement [98].

HPTF

Il est important que noter que lors d'une écoute au casque le filtrage du signal par une paire de HRTFs ne suffit pas complètement à obtenir un signal identique à un signal naturel. En effet, tous les casques audio possèdent une fonction de transfert propre, due à la réponse acoustique des transducteurs, qui déforme légèrement le signal lors de la restitution. De plus, le couplage entre le casque et le pavillon des oreilles induit également une modification du signal qui dépend à la fois de la forme du casque, de son positionnement sur la tête et de la morphologie des oreilles de l'auditeur [99]. Les modifications fréquentielles liées au casque audio et à son couplage avec les oreilles sont définies par les fonctions de transfert HPTFs (*HeadPhone Transfer Function*).

La fonction de transfert propre au casque est généralement donnée par le constructeur et est donc facile à compenser. En revanche, la coloration spectrale due au couplage entre le casque et les oreilles est beaucoup plus compliquée à évaluer et à compenser puisqu'elle est propre à chaque auditeur et varie d'un placement du casque à l'autre [100]. Cette variation de HPTF d'une position à l'autre est perceptible [101] et pourrait donc impacter négativement la qualité de la restitution binaurale. Certains auteurs indiquent cependant que la coloration spectrale des HPTFs n'est pas problématique, car elle reste bien moins importante que celle des HRTFs [99]. Dans le cas de test perceptif, les HPTFs sont rarement prises en compte et compensées, car il faudrait les mesurer individuellement pour chaque auditeur et de manière répétée pour prendre en compte les différents positionnements possibles du casque audio.

Prise de son binaurale

La manière la plus directe d'obtenir des signaux binauraux est de prendre en compte le filtrage des HRTFs dès l'enregistrement : c'est le principe de la prise de son binaurale (ou "binaural natif"). Une première méthode d'enregistrement consiste à placer un micro à l'entrée du conduit auditif de chaque oreille d'un humain (figure 1.22 A). Il convient d'utiliser des microphones adaptés et positionnés de manière à correctement capturer les colorations spectrales induites par le pavillon des oreilles. Cette méthode est relativement simple à mettre en place,

mais nécessite que la personne portant les micros reste le plus immobile possible et limite au maximum ses bruits de présence afin de ne pas parasiter l'enregistrement. Une autre méthode consiste à utiliser un mannequin avec des microphones intégrés dans les oreilles. Ce mannequin peut n'être constitué que d'une paire de pavillons d'oreille (figure 1.22 B). Ce dispositif léger est néanmoins très limité puisqu'il ne prend en compte qu'une partie des ITDs et des colorations spectrales induites par les pavillons en omettant la tête et le torse. Il peut s'agir d'une tête artificielle seule ou même d'une tête avec un torse (figure 1.22 C et D).

Pour la restitution, il suffit que chaque haut-parleur du casque diffuse le signal enregistré à l'oreille correspondante. La prise de son binaurale est considérée comme permettant une reproduction audio immersive et convaincante, mais comporte plusieurs limitations importantes. Tout d'abord, comme le filtrage des HRTFs est inclus dans le signal sonore dès l'enregistrement, il n'est pas possible d'utiliser un autre set de HRTFs après enregistrement. De plus, l'effet de salle est contenu dans les signaux binauraux et ne peut pas être modifié par la suite. Enfin, la position des sources sonores est figée et ne peut pas être modifiée après enregistrement, ce qui exclut notamment l'utilisation d'un système de suivi des mouvements de tête de l'auditeur (présenté dans la section 1.3.2). Une solution pour pallier ces problèmes est d'utiliser la synthèse binaurale.

Synthèse binaurale

Le principe de la synthèse binaurale est de générer des signaux binauraux à partir d'un signal sonore monophonique anéchoïque, ou issu d'une prise de son de proximité. Il faut sélectionner la paire de HRTFs correspondant à la position virtuelle souhaitée et effectuer une convolution entre le signal monophonique et la HRTF de chaque oreille. On obtient deux signaux que l'on peut diffuser dans chacun des haut-parleurs du casque. Il est possible de réaliser cette synthèse avec plusieurs signaux monophoniques correspondant à plusieurs objets localisés à différentes positions de l'espace en utilisant les HRTFs correspondants à chacune de ces positions afin de créer une scène sonore complexe. En plus de permettre la création de scènes sonores complexes à partir d'éléments simples, la synthèse binaurale offre la possibilité de contrôler l'ensemble des paramètres de la convolution et ainsi d'utiliser n'importe quelle réverbération et d'avoir le choix parmi un grand nombre de sets de HRTFs différents. Il est possible de choisir la position des sources virtuelles et d'obtenir des mouvements de sources en passant de la paire de HRTFs d'une position à la paire de HRTFs d'une autre position voisine. Ces mouvements peuvent être effectués en temps réel, en prenant en compte les mouvements de tête de l'auditeur pour avoir une restitution binaurale dynamique.



(A)



(B)



(C)



(D)

FIGURE 1.22 – Différentes techniques de prise de son binaurale : (A) microphone 4101-B placé à l'entrée du conduit auditif [102], (B) paire de pavillons d'oreille FS 3Dio [103], (C) tête KU-100 [104], (D) : tête et buste KEMAR [105].

1.3.2 Synthèse binaurale dynamique

La synthèse binaurale offre la possibilité de modifier le rendu audio en temps réel en prenant en compte l'évolution de n'importe quel paramètre. Cela peut être utile dans des applications interactives telles que le jeu vidéo ou la réalité virtuelle pour lesquelles l'auditeur contrôle, par le biais d'actions motrices, l'évolution du son. Les paramètres qui peuvent être modifiés sont le niveau sonore, la réverbération, le ratio champ direct / champ réverbéré et la position en azimut et en élévation des sources. Le contrôle en temps réel des trois premiers paramètres n'est utile principalement que dans des situations où l'auditeur se déplace (ou déplace un personnage qu'il incarne) dans un environnement virtuel. Par exemple dans des simulations de réalité virtuelle, des jeux vidéo ou des installations de muséographies. En effet, dans ces cas-là, il est judicieux d'augmenter le niveau sonore d'une source quand l'auditeur s'en approche ou encore de modifier le temps d'arrivée des réflexions quand l'auditeur passe d'une salle à une autre. Un contrôle correct de ces paramètres permet de rendre la scène sonore plus réaliste et d'améliorer l'immersion de l'auditeur. Le contrôle de la position des sources est, pour sa part, utile dans un grand nombre de situation d'écoutes au casque et peut être effectué à l'aide d'un système de *head-tracking*.

Principe du head-tracking

Lors d'une écoute naturelle, si une source est située devant l'auditeur à l'angle d'azimut 0° (figure 1.23 A), et que celui-ci tourne la tête de 90° vers la gauche, la source se trouvera à 90° à droite de l'auditeur (figure 1.23 B). La position de la source restera fixe dans le référentiel terrestre, mais changera de position dans le référentiel propre à l'auditeur. En revanche, lors d'une écoute au casque sans prise en compte des mouvements de l'auditeur, si celui-ci tourne la tête de 90° vers la gauche, la source "tournera" en même temps et restera toujours face à l'auditeur, à 0° (figure 1.23 C). Dans ce cas, la source restera fixe dans le référentiel propre à l'auditeur et changera de position dans le référentiel terrestre. Cette spécificité de l'écoute au casque peut poser problème du point de vue du réalisme et de l'immersion et diminuer les performances de localisation de l'auditeur.

Il est possible d'utiliser un système de *head-tracking* pour capter les mouvements de l'auditeur et envoyer ces informations au moteur de synthèse binaural afin de modifier la position des sources rendues. Avec un tel système actif, si l'auditeur tourne la tête de 90° vers la gauche, la position de la source passera de 0° à 90° vers la droite en balayant toutes les positions intermédiaires (figure 1.23 D). Ainsi, à la fin du mouvement, la source se trouvera à 90° à droite

de l'auditeur et sera donc restée fixe dans le référentiel terrestre tout au long du mouvement, comme lors d'une écoute naturelle.

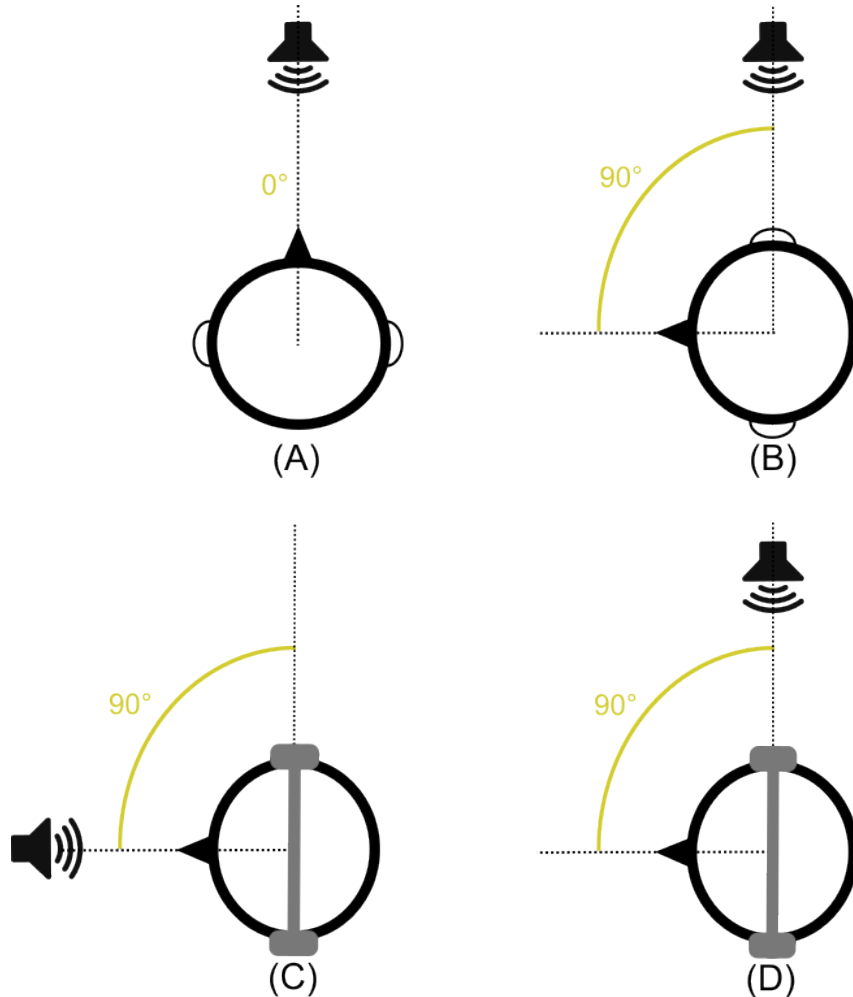


FIGURE 1.23 – (A) : source localisé à 0° , devant l'auditeur, en écoute naturelle. (B) : mouvement de 90° vers la gauche en écoute naturelle. (C) : mouvement de 90° vers la gauche en écoute au casque sans head-tracking. (D) : mouvement de 90° vers la gauche en écoute au casque avec head-tracking.

Pour modifier la position de la source, il suffit de remplacer la paire de filtres HRTFs de la position d'origine par la paire de filtres de la nouvelle position. Lors d'un déplacement continu de source correspondant à un mouvement de tête, il est nécessaire d'appliquer successivement toutes les paires de HRTFs de toutes les positions par lesquelles doit passer la source. L'obtention de HRTFs se fait par la mesure acoustique ou la modélisation pour une position donnée (*cf.* section 1.3.1). Comme l'acquisition des HRTFs correspondant à l'ensemble des positions de l'espace est impossible, les moteurs de synthèse binaurale utilisent des sets de HRTFs cor-

respondant à un nombre limité de positions (par exemple tous les 5°). Pour les positions intermédiaires, le moteur de synthèse effectue une interpolation. L'interpolation consiste en une estimation d'un filtre de HRTF à partir des HRTFs mesurées pour des positions voisines. Plusieurs méthodes d'interpolation existent [106-108]. Pour une revue complète, voir [109].

Effet de la latence

La latence du système est un paramètre important dans la synthèse binaurale dynamique. Elle représente le laps de temps entre le moment où l'auditeur effectue un mouvement de tête et le moment où les filtres correspondant à la nouvelle position sont actualisés dans l'écoute. Elle dépend du temps nécessaire au head-tracker pour transmettre les informations de mouvement et du temps que prend le moteur de synthèse binaural pour mettre à jour les filtres. La latence du moteur de lecture audio fait également partie de la latence totale du système, mais est généralement extrêmement courte et donc négligeable. Quand elle est trop élevée, la latence totale constitue un facteur important de détérioration des performances de localisation en synthèse binaurale dynamique [110]. Le seuil de détection de la latence a été estimé à 60 ms pour les sujets les plus performants [111]. La valeur de latence à partir de laquelle elle devient problématique varie en fonction de la durée du son. Cette valeur a été estimée à 75 ms pour des sons courts [111] et à 250 ms pour des sons longs [112].

Différents systèmes de head-trackers

Il existe de nombreux systèmes pour suivre les mouvements et la position de la tête de l'auditeur à partir d'un point de référence de l'espace. Les systèmes à 3 degrés de liberté permettent de suivre l'orientation de la tête de l'auditeur selon trois directions appelées *yaw*, *pitch* et *roll* (voir figure 1.24). Les systèmes à 6 degrés de liberté permettent, en plus de suivre l'orientation de la tête, de suivre la position de l'auditeur dans l'espace en 3 dimensions (profondeur, largeur et hauteur, ou x , y et z). Les systèmes à 6 degrés de liberté sont plus complets, mais également plus complexes à mettre en place, et ne sont utiles que dans des contextes bien spécifiques, par exemple lorsque l'auditeur doit se déplacer dans des installations de muséographe ou de réalité augmentée.

Les head-tracker existants peuvent être divisés en quatre catégories :

- **Acoustiques** : Les head-trackers acoustiques fonctionnent en analysant le temps nécessaire à un signal acoustique pour atteindre un récepteur. Généralement, plusieurs émet-

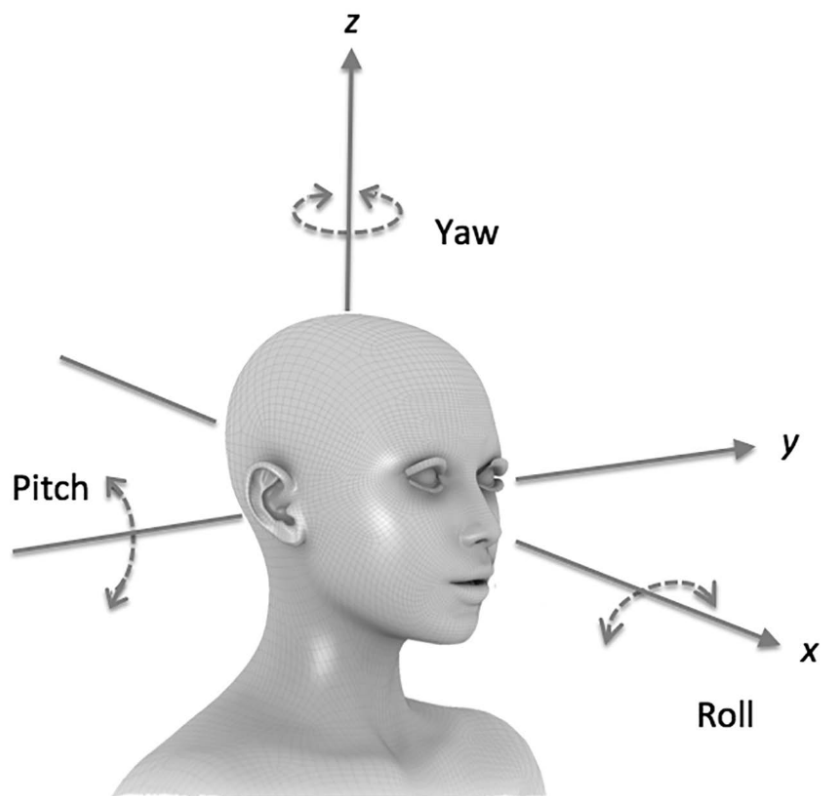


FIGURE 1.24 – Représentation des différents axes de rotation de la tête et des différentes directions de déplacement possibles. Image extraite de [113]).

teurs émettent des ultrasons dans un environnement et plusieurs récepteurs sont placés sur la tête de l'auditeur pour suivre sa position et son orientation. Ces systèmes sont limités en portée du fait de l'absorption de l'air.

- **Inertiels** : Les capteurs inertiels utilisent des accéléromètres et des gyroscopes pour extraire des informations de position. Le gyroscope indique l'orientation de la tête. L'accéléromètre permet de calculer la position dans laquelle se trouve l'auditeur en effectuant une double intégration à partir des informations d'accélération de la tête. Une première intégration permet d'obtenir la vitesse, puis une seconde intégration permet d'obtenir la position après un mouvement par rapport à la position initiale. Les capteurs inertiels sont très courants dans les téléphones mobiles et autres appareils grand public en raison de leur faible coût, de leur taux de rafraîchissement élevé et de leur faible latence, mais ils ont tendance à subir une dérive de position à faible vitesse, qui peut être prise en compte et compensée dans une certaine mesure, mais peut devenir problématique à très faible vitesse.
- **Optiques** : La technologie des head-trackers optiques s'appuie sur une variété de méthodes d'analyse basées sur l'image pour extraire des informations de position à partir d'une caméra. Les méthodes de suivi optique utilisent deux techniques : 1) le suivi d'objets équipés de marqueurs passifs ou actifs disposés selon un motif connu, 2) le suivi d'objets sans marqueurs ayant une géométrie connue (par exemple, une tête ou un visage). Il est possible d'extraire les informations d'orientation et de position de l'auditeur à partir de ces suivis d'objets. Ces systèmes sont peu utilisés, car leur portée est limitée à la zone captée par les caméras et ils nécessitent une puissance de calcul informatique importante.
- **Magnétiques** : Les head-trackers magnétiques utilisent un émetteur fixe pour créer un champ magnétique. Un récepteur magnétique est placé sur le dessus de la tête de l'auditeur et détecte les variations du champ magnétique causées par les mouvements de ce dernier. Il détermine l'intensité et l'angle du champ magnétique, et donc la position et l'orientation du capteur par rapport à l'émetteur. Les head-trackers magnétiques ont une très grande précision, mais sont sensibles aux perturbations électromagnétiques externes et ont généralement une courte portée de fonctionnement.

Le choix du système pour une application donnée dépend principalement du lieu d'utilisation (*e.g.* intérieur ou extérieur, isolé ou sujet à des perturbations électromagnétiques), de la précision souhaitée et des contraintes d'encombrement.

1.3.3 Binaural individualisé ou non-individualisé

Binaural individualisé

Qu'il s'agisse de prises de son ou de synthèse binaurale, l'auditeur écoute un signal qui a été filtré par les HRTFs d'une personne (ou d'un mannequin) contenant des caractéristiques propres à la morphologie de la personne. Si l'auditeur écoute un signal filtré avec ses propres HRTFs, la restitution binaurale est dite *individualisée*. Cela est possible si l'auditeur a effectué lui-même l'enregistrement avec des microphones dans ses oreilles ou si la synthèse binaurale a été effectuée avec des HRTFs mesurées directement sur lui. Cette restitution binaurale individualisée est, en théorie, la plus fidèle possible, la plus proche possible d'une écoute naturelle, puisque, en écoute naturelle, les signaux aux tympons d'un auditeur sont toujours filtrés par ses propres HRTFs. Cependant, la restitution binaurale individualisée est rare, car l'acquisition de HRTFs est relativement longue et complexe et la plupart des auditeurs ne possèdent pas d'enregistrement de leur propre set de HRTFs.

Binaural non-individualisé

Les contraintes liées à la restitution binaurale individualisée ont incité les diffuseurs à favoriser la restitution binaurale non-individualisée pour la diffusion de leurs contenus. En écoute non-individualisée, l'auditeur reçoit des signaux qui ont été filtrés par des HRTFs qui ne lui sont pas propres : d'une autre personne ou d'un mannequin. La restitution binaurale non-individualisée est, en général, considéré moins convaincante que la restitution individualisée, car le filtrage HRTF utilisé n'est pas connu de l'auditeur. En effet, tous les individus ont des morphologies différentes et les filtrages résultants peuvent varier de façon importante [114]. Plusieurs études perceptives ont montré que l'utilisation de HRTFs non-individuelles a un impact sur les performances de localisation [115-118], l'externalisation [118-121] et d'autres paramètres tels que la clarté, la richesse et la préférence globale [114]. Certains résultats suggèrent que cet impact est négatif et qu'il faudrait donc privilégier l'utilisation de HRTFs individuelles, d'autres résultats suggèrent que cet impact est négligeable ou même, de façon surprenante, bénéfique. Il semble que l'ajout de modalités tels que la vision ou le suivi des mouvements de tête de l'auditeur permettent de s'affranchir de l'impact négatif de l'utilisation de HRTFs non-individualisées [118, 120]. Ces problématiques sont discutées en détail dans la section 1.4.1.

Apprentissage de HRTFs

Plusieurs études ont montré qu'il est possible pour le système auditif "d'apprendre" des HRTFs non-individuelles [122]. Après une période d'apprentissage plus ou moins longue, les performances de localisation d'un sujet pour un set de HRTFs non-individuelles s'améliorent. Blum *et. al.* [123] ont observé une amélioration modeste, mais significative après un apprentissage de seulement 12 minutes. Parseihian et Katz [124] ont observé une amélioration graduelle corrélée avec la durée d'apprentissage. Bien qu'en 36 minutes d'apprentissage, les performances de localisation n'aient pas atteint une précision aussi élevée qu'en écoute individualisée, les résultats suggèrent que l'amélioration pourrait être encore plus importante avec une période d'apprentissage plus longue. Les résultats d'autres études [125, 126] suggèrent que l'utilisation de modalités visuelles et motrices en plus du son dans la période d'apprentissage favorise la rapidité et la qualité de l'apprentissage. Ces techniques d'apprentissage, bien que prometteuses, demanderaient la mise en place d'une période d'apprentissage pour tous les auditeurs et ne sont pas développées auprès du grand public.

1.4 L'externalisation

Lors d'une écoute en environnement naturel, l'auditeur perçoit les sons *externalisés* : les sources semblant provenir de l'extérieur de sa tête. À l'inverse, lors d'une écoute au casque, l'auditeur perçoit souvent les sources sonores comme provenant de l'intérieur de sa tête : *internalisées*. En fonction de la nature du contenu audio diffusé, cet artefact de perception peut nuire au réalisme de la reproduction audio. Un exemple commun de source internalisée est la voix d'un locuteur perçue par lui-même avec les oreilles bouchées. En dehors de ce cas précis, le phénomène d'internalisation des sons se produit souvent lors d'une écoute stéréophonique classique au casque [127] et parfois lors d'une restitution binaurale. Bien que ces phénomènes d'internalisation et d'externalisation aient été étudiés depuis la fin du 19e siècle [128-131], le développement accru depuis le début du 21e siècle des technologies d'écoute au casque, en particulier de la technologie binaurale, a accéléré la recherche sur ces phénomènes. La revue de Best *et al.* [132] discute de nombreux aspects de l'externalisation et constitue une référence importante de cette section.

1.4.1 Indices d'externalisation

Plenge [133] a défini que le prérequis principal pour avoir une externalisation convaincante est d'utiliser des signaux sonores "adéquats à l'oreille". La probabilité qu'un signal audio restitué au casque soit correctement externalisé augmente à mesure qu'il se rapproche d'un signal naturel, en tenant compte des propriétés de l'environnement et de la morphologie de l'auditeur. De plus, l'externalisation semble dépendre fortement du réalisme de la scène sonore et de l'écoute [134, 135]. Ainsi, la présence du casque audio sur la tête de l'auditeur, l'écoute d'une scène sonore incohérente par rapport à l'environnement dans lequel se trouve l'auditeur ou l'utilisation de sources sonores connues sont des paramètres qui peuvent favoriser l'internalisation du son.

Angle d'incidence

L'angle d'incidence de la source, en particulier l'azimut, a une influence importante sur l'externalisation. Plus une source est éloignée du plan médian, plus elle paraît externalisée. Ainsi, une source située à droite ou à gauche de l'auditeur ($\pm 90^\circ$) paraît plus externalisée que la même source située devant ou derrière l'auditeur (0° ou 180°) [119, 121, 136-138]. Une explication de cet effet pourrait venir de la présence de différences interaurales. En effet, à $\pm 90^\circ$

les ITD et ILD sont importantes et il n'existe pas d'ambiguïté possible entre la gauche et la droite puisque une source à $+90^\circ$ a des ITD et ILD inverse d'une source à -90° alors que, à 0° et 180° , les ITD et ILD sont identiques [75]. Le système auditif dispose de moins d'indices pour localiser une source proche du plan médian qu'une source située sur le côté et il est possible que cette ambiguïté favorise l'internalisation de la source. De plus, pour des sources proches du plan médian, les sources situées derrière l'auditeur sont mieux externalisées que les mêmes sources situées devant. Cet effet pourrait venir de l'influence de la modalité visuelle sur l'externalisation, discutée dans la suite de cette section.

Contenu spectral

Le fait de filtrer chaque source de la scène sonore par la paire de HRTFs correspondant à sa position permet de grandement améliorer l'externalisation [139, 140]. Ainsi, une même scène sonore est plus externalisée lors d'une restitution binaurale que lors d'une restitution stéréophonique classique.

De plus, la qualité des HRTFs utilisées est importante. En effet, plusieurs études ont montré que la dégradation des HRTFs entraînait une baisse d'externalisation [141-143]. Le problème de l'individualisation des HRTFs (discuté dans la section 1.3.3) dépend également du contenu spectral. En effet, pour un auditeur donné, un signal filtré avec des HRTFs non-individualisées est moins adéquat qu'un signal filtré avec ses propres HRTFs. Cette différence spectrale pourrait mener à une baisse d'externalisation avec des HRTFs non individualisées, cependant les études récentes n'ont pas mis en évidence cet effet de façon claire. Certaines études trouvent une baisse d'externalisation importante quand les sujets écoutent avec des HRTFs non individualisées [119, 120] alors que d'autres trouvent des différences plus modérées et qui dépendent de la nature des stimuli [118, 121].

Réverbération

De nombreuses études ont montré que les sons réverbérés sont perçus plus externalisés que les mêmes sons non réverbérés [118, 121, 138]. L'ajout de réflexions précoces sur un son anéchoïque permet d'augmenter efficacement l'externalisation. L'ajout additionnel de réflexions tardives permet d'augmenter encore davantage l'externalisation, mais de façon moins prononcée [118]. En comparant des stimuli avec des réverbérations plus ou moins longues, Catic *et al.* ont mis en évidence une augmentation de l'externalisation proportionnelle à la longueur de la

réverbération jusqu'à 80 ms [144]. Pour des réverbérations plus longues que 80 ms, l'externalisation stagne.

L'influence du contenu spectral de la réverbération, notamment de la modification due aux HRTFs, n'a pas été établie de façon claire. Les résultats de certaines études montrent que le filtrage de la réverbération par des HRTFs n'améliore pas l'externalisation [142, 145] alors que les résultats d'autres études suggèrent que la présence d'informations binaurales dans le signal réverbéré est importante pour obtenir une externalisation convaincante [121, 144].

L'influence de la réverbération sur l'externalisation est à mettre en relation avec le réalisme. Une réverbération incohérente avec la scène sonore et les attentes de l'auditeur crée une divergence qui nuit au réalisme de la scène et, par conséquent, à l'externalisation [120, 146].

Modalité visuelle

Les informations disponibles visuellement ont une influence sur l'externalisation. Celle-ci est meilleure quand l'auditeur peut voir un objet cohérent positionné à l'emplacement virtuel de la source simulée que si l'auditeur n'a pas d'information visuelle [120]. Typiquement, si une enceinte silencieuse se trouve dans la salle, l'auditeur peut associer le son à l'enceinte, alors même que le son provient du casque, améliorant ainsi l'externalisation. Même sans objet particulier, la salle que visualise l'auditeur lors de l'écoute peut influencer l'externalisation. Si l'auditeur visualise une salle dont les caractéristiques sont cohérentes avec la réverbération simulée du signal sonore diffusé au casque, l'externalisation est meilleure que s'il visualise une salle complètement incohérente avec la réverbération du son [147, 148]. Cette visualisation peut être liée à la salle dans laquelle se trouve l'auditeur lors de l'écoute, ou à une salle visualisée au travers d'un casque de réalité virtuelle.

L'importance d'une modalité visuelle corrélée avec la stimulation sonore peut expliquer le fait que l'externalisation est moins bonne pour une source située à 0° devant l'auditeur que pour une source située à 180° derrière l'auditeur. En effet, lors d'une écoute naturelle, l'auditeur est généralement capable de voir l'objet qui diffuse un son devant lui alors que l'absence de cette modalité lors d'une écoute de la même source au casque crée une disparité qui peut expliquer une augmentation de l'internalisation du son. En revanche, cette disparité n'existe pas pour une source située derrière l'auditeur puisque l'objet diffusant le son n'est pas visible même en écoute naturelle.

L'influence de la modalité visuelle est compatible avec l'idée que, de manière générale, les conditions d'écoute réalistes favorisent l'externalisation.

Suivi de mouvement

L'utilisation d'un système de head-tracking pour adapter les signaux audio aux mouvements de tête de l'auditeur telle que décrit dans la section 1.3.2 permet d'obtenir une écoute plus réaliste, notamment en modifiant la position des sources afin qu'elles restent fixes du point de vue de l'environnement extérieur, comme lors d'une écoute naturelle. Certaines études suggèrent que l'utilisation d'un tel système ne permet pas d'améliorer l'externalisation [118] ou seulement de manière faible [149]. Dans ces études, les sujets devaient évaluer la distance de stimuli écoutés soit pendant un mouvement de tête, soit sans mouvement de tête. Les résultats en distance étaient ensuite convertis en externalisation (les problématiques liées à une telle conversion sont discutées dans la section 1.4.2). Cependant, ces deux études ont utilisé des stimuli très courts (3 s) qui ne permettaient pas forcément aux auditeurs de bénéficier de l'ensemble des indices dérivés des mouvements de tête. De plus, les résultats d'externalisation de ces deux études sont moyennés pour des positions azimutales tout autour de l'auditeur. Comme l'externalisation pour les sources latérales est déjà relativement élevée (*cf.* section 1.4.1), il est possible que le fait de moyenner les résultats pour toutes les positions avant l'analyse statistique ait masqué une potentielle différence significative d'externalisation entre les stimuli avec vs. sans mouvement de tête.

Des études plus récentes ont examiné l'influence des mouvements de tête avec head-tracking pour des stimuli plus longs (8 s) et en considérant les différentes positions en azimut de façon individuelle [137, 150, 151]. Les résultats de ces études montrent que le fait d'effectuer des mouvements de tête avec head-tracking permet d'améliorer l'externalisation de façon importante pour les sources frontale et arrière, et de manière modérée voire nulle pour les sources latérales. Lorsqu'un auditeur effectue un mouvement de tête avec head-tracking, il explore un panel de HRTFs bien plus important que lors d'une écoute statique et on pourrait s'attendre à ce que le choix du set de HRTFs modifie l'influence du head-tracking. Les résultats d'une de ces études précédentes suggèrent que l'amélioration d'externalisation due au head-tracking n'est significative qu'en restitution binaurale non-individualisée [137]. Le head-tracking permet alors d'élever l'externalisation jusqu'au niveau d'externalisation obtenu pour un contenu individualisé. Ce résultat est important puisqu'un des principaux facteurs d'internalisation des sons lors de restitutions binaurales est lié à l'utilisation de HRTFs générique (*cf.* section 1.4.1). Si, dès lors que la reproduction binaurale est effectuée de façon dynamique, les mouvements de tête permettent d'obtenir une externalisation satisfaisante, quel que soit le set de HRTFs utilisé, il n'est plus indispensable d'individualiser la restitution binaurale pour que chaque auditeur

puisse avoir une externalisation satisfaisante.

Une étude de Brimijoin *et al.* [137] a montré que des mouvements de tête avec head-tracking permettent d'améliorer l'externalisation *pendant* le mouvement de tête, cependant il est crucial de déterminer si des améliorations substantielles peuvent être observées même lorsque l'auditeur n'effectue pas de mouvement de tête volontaire. En effet, l'intérêt pratique du head-tracking serait fortement réduit s'il permettait d'améliorer l'externalisation uniquement lorsque l'auditeur bouge la tête. Cela signifie que les auditeurs devraient bouger la tête en permanence pour écouter un contenu binaural avec une externalisation satisfaisante, ce qui n'est pas envisageable.

König et Sussmann [152] ont montré que, même en écoute passive et statique, les auditeurs effectuent très régulièrement des mouvements de tête involontaires avec des amplitudes allant jusqu'à 5° d'azimut. De tels mouvements (avec head-tracking) pourraient permettre d'obtenir une amélioration d'externalisation suffisante pour une reproduction binaurale satisfaisante et réalisable puisque les auditeurs n'auraient pas de mouvement particulier à effectuer. Dans l'étude de Wersenyi [153], il a été démontré que l'émulation de petits mouvements de tête de 2° augmentait les taux d'externalisation pour 20 % des sujets.

Les études ayant comparé l'externalisation avec et sans head-tracking de sources écoutées par des auditeurs "fixes" (pas de mouvements volontaires, mais micro mouvements possibles) suggèrent que l'externalisation est légèrement meilleure avec un système de head-tracking [137, 150], mais reste néanmoins inférieure à l'externalisation obtenue avec de grands mouvements de tête. Il est intéressant de noter que ces études ont également évalué l'externalisation sans head-tracking des sources écoutées par des auditeurs en mouvement et, dans ce cas-là, l'externalisation décroît drastiquement. Il semble que même si l'utilisation d'un système de head-tracking en écoute statique permet d'améliorer l'externalisation, cette amélioration soit trop faible pour obtenir une externalisation satisfaisante.

Une autre solution pour que le head-tracking soit efficace pour améliorer l'externalisation en évitant que l'auditeur ne soit contraint de bouger la tête en permanence serait que cette amélioration perdure même après l'arrêt du mouvement de tête (persistance de l'effet des mouvements de tête). C'est ce qu'ont évalué Hendrickx *et al.* [150, 151] en demandant aux sujets de noter l'externalisation de sources sonores juste après le mouvement de tête. Les stimuli duraient 8 s. Pendant les 5.5 premières secondes, les sujets devaient soit effectuer un mouvement de tête, soit rester statique et pendant les 3.5 s restantes, ils devaient toujours rester statiques et évaluer

l'externalisation. Les résultats ont montré que l'amélioration d'externalisation perdurait après le mouvement. Cependant, les sujets devaient évaluer l'externalisation *immédiatement* après l'arrêt du mouvement de tête et pour une seule source (voix d'homme). Dans des applications concrètes utilisant la restitution binaurale, les contenus audio peuvent durer bien plus longtemps que quelques secondes et plusieurs sources peuvent être présentes. Aucune étude n'a exploré quelle serait l'amélioration d'externalisation due à des mouvements de tête avec head-tracking plusieurs dizaines de secondes après l'arrêt des mouvements et pour des sources sonores variées.

Si, dès lors que la restitution binaurale est effectuée de façon dynamique, les mouvements de tête permettent d'obtenir une externalisation satisfaisante comme le suggèrent les études précédemment, alors il n'est plus nécessaire de développer des systèmes d'individualisation des HRTFs pour les applications grand public, et il suffit de mettre en place des systèmes de synthèse binaurale dynamique dans les appareils de lecture audio grand public. Bien que la mise en place de tels systèmes demande une évolution des appareils de lecture actuels, elle reste néanmoins plus légère que la mesure de HRTFs individuelles pour chaque auditeur.

1.4.2 Mesure et problématique

La revue de Best *et al.* [132] détaille les différentes méthodes qui ont été utilisées afin de mesurer l'externalisation.

Dans une première méthode, par exemple utilisée par Brimijoin *et al.* [137], les sujets doivent simplement indiquer s'ils perçoivent la source sonore à l'intérieur ou à l'extérieur de leur tête (exemple : figure 1.25 A). Dans une autre méthode, les sujets doivent faire la distinction entre des stimuli comportant de potentielles différences d'externalisation. Par exemple, dans l'étude de Kulkarni & Colburn [154], les sujets devaient distinguer des stimuli naturels, produits par une source réellement présente dans l'environnement, de stimuli virtuels, diffusés au casque, dont les paramètres variaient. Quand les sujets ne pouvaient pas faire la différence, l'externalisation était considérée comme maximale. Plus les sujets arrivaient à distinguer les stimuli virtuels des stimuli naturels, plus l'externalisation était considérée comme étant faible. Les tests perceptifs utilisant ces méthodes binaires ont pour avantage d'être très simples pour les sujets : ils ne requièrent que peu d'explications et permettent ainsi de s'affranchir partiellement de potentielles différences de compréhension d'un sujet à l'autre. Cependant, ces méthodes ne permettent aucune évaluation d'externalisation intermédiaire et sont peu utilisées.

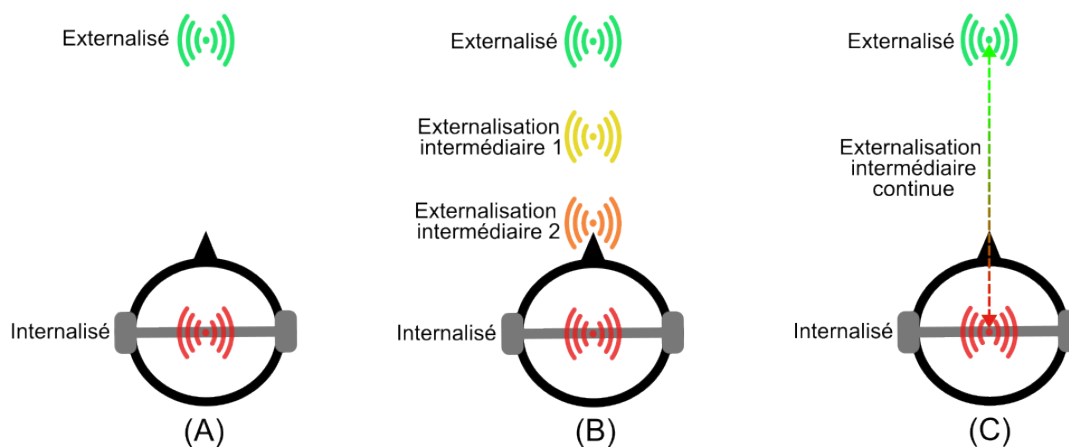


FIGURE 1.25 – Différentes méthodes d'évaluation de l'externalisation : (A) choix binaire, (B) échelle discrète (le nombre de points intermédiaires peut varier), (C) échelle continue. Image adaptée de [132].

Dans la méthode la plus couramment utilisée, les sujets doivent évaluer l'externalisation de chaque stimulus à l'aide d'une échelle continue (*e.g.* [139, 142], figure 1.25 C) ou discrète (*e.g.* [147, 150, 151], figure 1.25 B), avec pour externalisation minimale le centre de la tête. Différents labels ont été proposés pour donner des points de référence le long des échelles. Par exemple, les échelles utilisées par Hendrickx *et al.* [150, 151] contenaient des notions d'externalisation et de distance ("la source est au centre de ma tête", "la source est externalisée et lointaine"). L'échelle utilisée par Hartmann et Wittenberg [139] comprenait, en plus de la notion d'externalisation, une notion de compacité ("La source n'est pas bien externalisée. Elle est à mon oreille, ou sur mon crâne, ou très diffuse") et la confusion de localisation ("La source est externalisée, mais elle est diffuse ou au mauvais endroit."). D'autres études ont, pour leur part, utilisé une échelle continue sans label intermédiaire [121, 140].

De manière générale, l'utilisation d'une échelle d'externalisation implique que la notion d'externalisation soit expliquée aux sujets. La compréhension de cette notion peut ne pas être aisée pour des sujets n'étant pas habitués à l'écoute critique et peut varier d'un individu à l'autre. Cette problématique peut entraîner des biais importants et complexes à quantifier lors de tests perceptifs. Pour s'affranchir de ce problème, certaines études utilisent une échelle de distance [118, 136, 144, 155]. Bien que plus facile à présenter aux sujets, ce type d'échelle n'est pas idéal puisque la distance et l'externalisation ne sont pas forcément liées. En effet, un son perçu très proche de l'auditeur peut être perçu parfaitement externalisé et sera pourtant considéré comme étant peu externalisé dans un test utilisant une échelle de distance.

D'autres méthodes alternatives à l'évaluation perceptive des sujets sur une échelle continue ou discrète ont été explorées. Dans une étude de Yuan *et al.* [156], les sujets devaient, à chaque essai, classer comparativement quatre stimuli du plus externalisé au moins externalisé. Une autre méthode consiste à demander aux sujets de pointer les positions de sources virtuelles sur une interface visuelle représentant l'espace autour du sujet (*e.g.* [157-160], figure 1.26). L'externalisation des sources est définie à partir de la position que leur a attribuée le sujet.

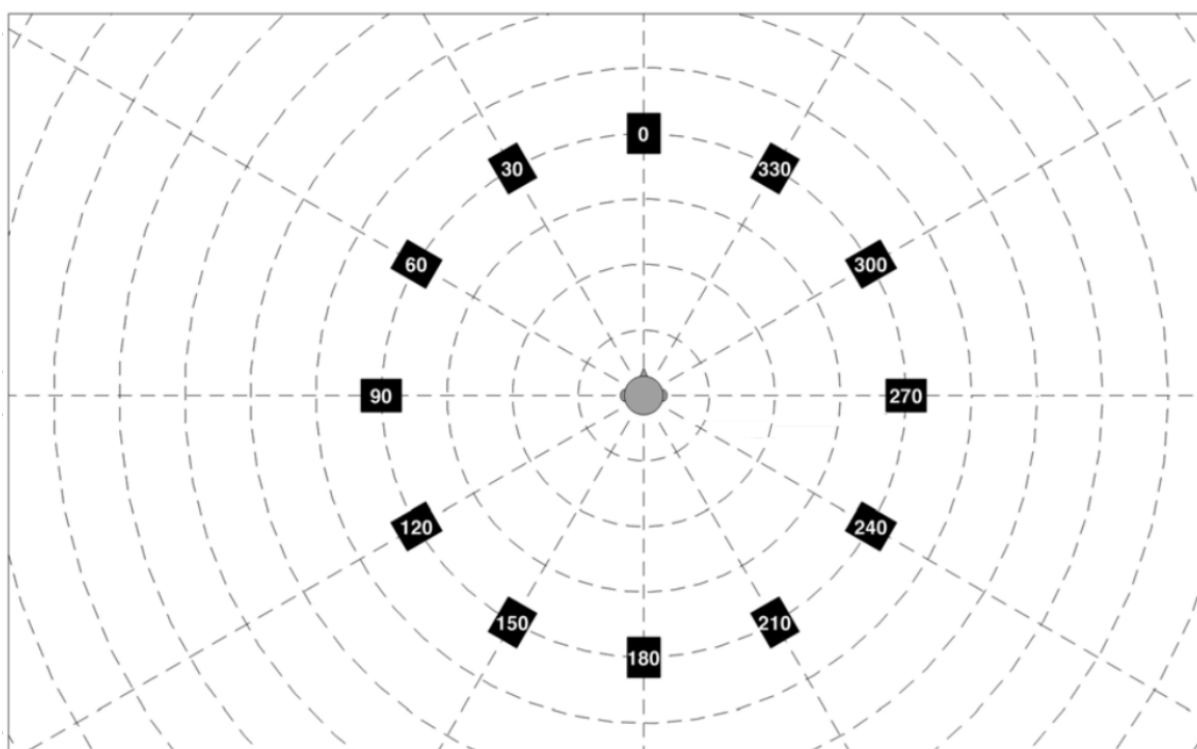


FIGURE 1.26 – Interface utilisée par les sujets pour pointer la position perçue des sources virtuelles dans l'étude [158], les rectangles noirs représentent des enceintes placées dans la salle de test et servent de repère.

Best *et al.* [132] ont mis en évidence plusieurs problèmes liés à ces méthodes de mesure. Pour faciliter l'évaluation de l'externalisation par les sujets, certaines études incluent des références visuelles telles que des haut-parleurs silencieux positionnés aux emplacements des sources virtuelles. Pour donner leurs réponses, les sujets peuvent simplement indiquer si le son semble provenir d'un haut-parleur plus ou moins proche, il n'est alors pas nécessaire de leur expliquer la notion d'externalisation. Cependant, l'ajout d'une modalité visuelle constitue un biais important puisque celle-ci modifie la perception d'externalisation (*cf.* section 1.4.1).

Pour des études testant différents azimuts de sources, l'utilisation d'une échelle unique peut poser problème. Best *et al.* proposent l'exemple d'un test perceptif comparant des stimuli frontaux à des stimuli latéraux. Quand l'externalisation est minimale, les sons frontaux seront probablement perçus au centre de la tête, tandis que les sons latéraux seront perçus au niveau de l'oreille ipsilatérale. En utilisant une même échelle d'externalisation pour ces deux situations, avec le centre de la tête comme référence pour une externalisation minimale, un auditeur pourrait attribuer des notes légèrement plus élevées aux stimuli latéraux qu'aux stimuli frontaux, même lorsque leur externalisation est minimale, puisque ceux-ci ne sont pas perçus au centre de la tête, mais du côté d'où provient le son.

Dans les méthodes où le sujet doit pointer la position perçue des sources, les confusions de localisation (qui sont courantes en restitution binaurale [134]), si elles ne sont pas prises en compte, peuvent donner lieu à des résultats d'externalisation erronés lors du calcul de la moyenne des répétitions d'un même stimulus. En effet, il est possible qu'un sujet pointe des positions externalisées, mais parfois à l'avant et parfois à l'arrière. Alors même que le stimulus était correctement externalisé, le calcul de la moyenne donnera l'impression que ce stimulus était peu externalisé, avec une image sonore étalée.

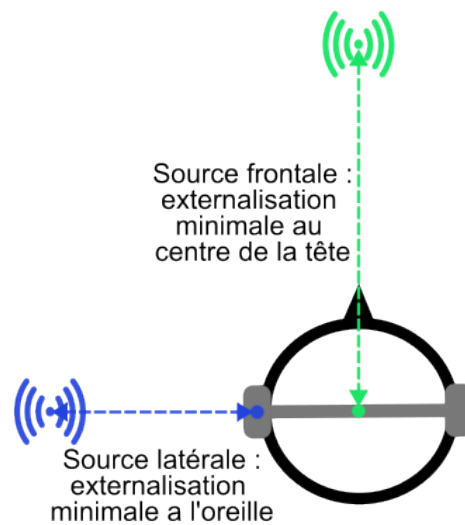


FIGURE 1.27 – Problème lié à l'utilisation d'une échelle unique pour des sources frontales et latérales. Image adaptée de [132].

Ainsi, la multiplicité des méthodes d'évaluation perceptive d'externalisation et les problèmes associés ne permettent pas de définir une méthode parfaitement robuste et fiable pour évaluer l'externalisation. Quelle que soit la méthode utilisée, l'approche perceptive constitue

une barrière qui, dans le cas d'études explorant des différences subtiles, peut entraver la collecte de données. Pour s'affranchir de cette barrière, il faudrait pouvoir étudier l'externalisation d'un point de vue objectif, par exemple en définissant une activité électrique cérébrale caractéristique de l'externalisation.

1.4.3 Externalisation et activité cérébrale

Externalisation et IRMf

Un nombre réduit d'études ont comparé l'activité cérébrale générée par des stimuli externalisés vs. internalisés. Une étude utilisant la IRMf, réalisée par Hunter *et al.* [161], a comparé les variations hémodynamiques du cerveau induites par des phrases courtes restituées en stéréo classique vs. binaural non-individualisé. Les résultats ont mis en évidence une activation plus importante du planum temporale pour les phrases restituées en binaural (voir figure 1.28). Le planum temporale est un gyrus du lobe temporal. Il s'agit d'une région de forme triangulaire, située juste derrière le cortex auditif. Cette zone est impliquée dans le traitement du langage et dans la perception de la distance [162]. Une autre étude IRMf réalisée par Callan *et al.* [163] a comparé des bruits blancs monophoniques avec des bruits blancs stéréophoniques et binauraux individualisés. Comme pour l'étude de Hunter *et al.*, les résultats ont montré une activation accrue du planum temporale pour les bruits blancs binauraux.

Ces résultats pourraient indiquer que l'externalisation est traitée dans la même zone du cerveau que la perception de la distance et serait donc liée à la distance. Cependant, un problème commun aux deux études complique cette interprétation. En effet, dans ces deux études, en plus de l'enregistrement IRMf, les sujets ont passé un test perceptif pour vérifier que les différents types de stimuli étaient perçus différemment. Dans l'étude de Hunter *et al.*, ils devaient indiquer quel stimulus était binaural dans une série de comparaisons par paire opposant à chaque fois un stimulus binaural à un stimulus stéréo. Les sujets ont été capables de reconnaître les stimuli binauraux avec une précision de 100%. Dans l'étude de Callan *et al.*, les sujets devaient simplement indiquer, à chaque essai, s'il s'agissait d'un stimulus mono, stéréo ou binaural. Les sujets étaient capables de distinguer les trois conditions avec un taux de réussite d'environ 60 %, ce qui est significativement supérieur au hasard (33 %).

Ces tests perceptifs permettent de vérifier que les stimuli mono, stéréo et binauraux étaient différenciables, mais ne permettent pas d'affirmer que les stimuli binauraux étaient effectivement perçus plus externalisés que les autres. En effet, les sujets auraient très bien pu distinguer les différents types de stimuli en utilisant les différences spectrales sans pour autant percevoir de

différences d'externalisation. Ainsi, l'activation accrue du planum tempore observée dans ces études pourrait être due à d'autres facteurs que les différences d'externalisation inter-stimulus. Cette observation nécessite d'être complétée par d'autres résultats d'études dans lesquelles les variations d'externalisation inter-stimulus sont vérifiées.

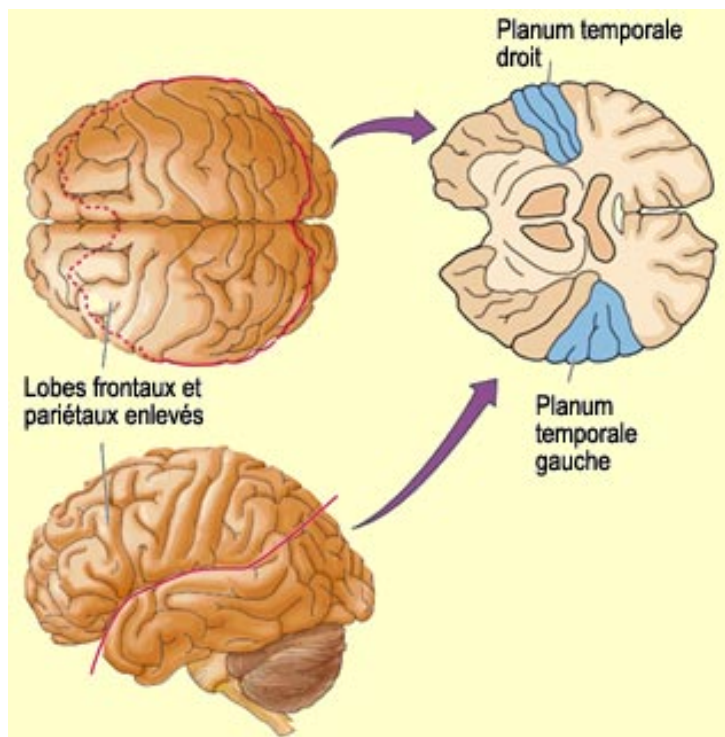


FIGURE 1.28 – Représentation du planum tempore. Image extraite de [164]

Externalisation et ERPs

Les quelques études qui ont comparé des stimuli non externalisés vs. externalisés n'avaient pas pour objectif explicite d'étudier l'activité cérébrale liée à l'externalisation. Cependant, trois études ont mis en évidence des différences qui pourraient être attribuées à des différences d'externalisation.

Nicol *et al.* [165] ont mis en place un protocole *oddball* (standard - déviant) visuel lors duquel les sujets écoutaient une scène sonore réaliste par rapport à l'environnement dans lequel ils se trouvaient, diffusée soit en stéréo classique soit en binaural.

Un protocole *oddball* visuel se présente comme une succession d'images dont la majorité est "standard" (*e.g.* un damier parfaitement distribué : figure 1.30 A) et dont une petite propor-

tion est "déviante" (*e.g.* un damier avec une inversion de case, figure 1.30). Les sujets doivent indiquer quand ils détectent un stimulus déviant et une composante ERP P300 apparaît à la présentation de chaque stimulus déviant. La fréquence et l'amplitude de la P300 peuvent varier en fonction des paramètres du stimulus déviant, mais également de l'état psychologique du sujet telle que son attention. De plus, la précision de réponse (pourcentage de réponses correctes) et le temps de réaction peuvent être enregistrés.

Dans l'étude de Nicol *et al.*, bien qu'il s'agissait d'un protocole oddball visuel, la P300 était modulée par le type de contenu audio qu'écoutait le sujet (stéréophonique ou binaural). Les résultats ont montré que, lorsque les sujets écoutaient une scène sonore binaurale, la première partie de la P300, définie comme la P3a (*cf.* section 1.1.3), avait une amplitude plus importante que quand ils écoutaient une scène sonore stéréophonique. Parallèlement, les résultats comportementaux ont montré que le temps de réaction des sujets était plus long lors de la restitution binaurale.

La P3a est décrite comme reflétant la surprise dans des changements d'attention, même de façon multimodale [166, 167]. Bien que n'étant pas directement reliés à l'externalisation, les résultats de l'étude de Nicol *et al.* pourraient indiquer que les sujets étaient davantage surpris à la présentation du stimulus déviant lorsqu'ils écoutaient la scène sonore binaurale et étaient donc davantage immergés dans la scène sonore. Comme l'externalisation est souvent associée à l'immersion [132], cet accroissement d'amplitude de la P300 pourrait être dû à des différences d'externalisation.

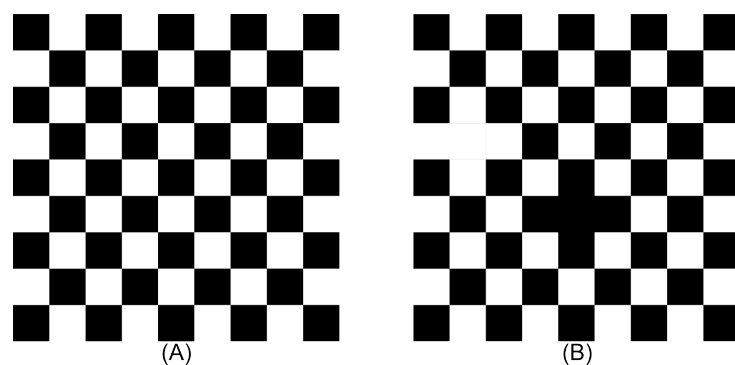


FIGURE 1.29 – Exemple de stimuli standard (A) et déviant (B) pour un protocole oddball.

Dans une étude utilisant la MEG, Palomäki *et al.* [168] ont comparé les ERPs générés par des salves de bruit blanc (*burst noise*) diffusées à huit positions différentes et spatialisées de

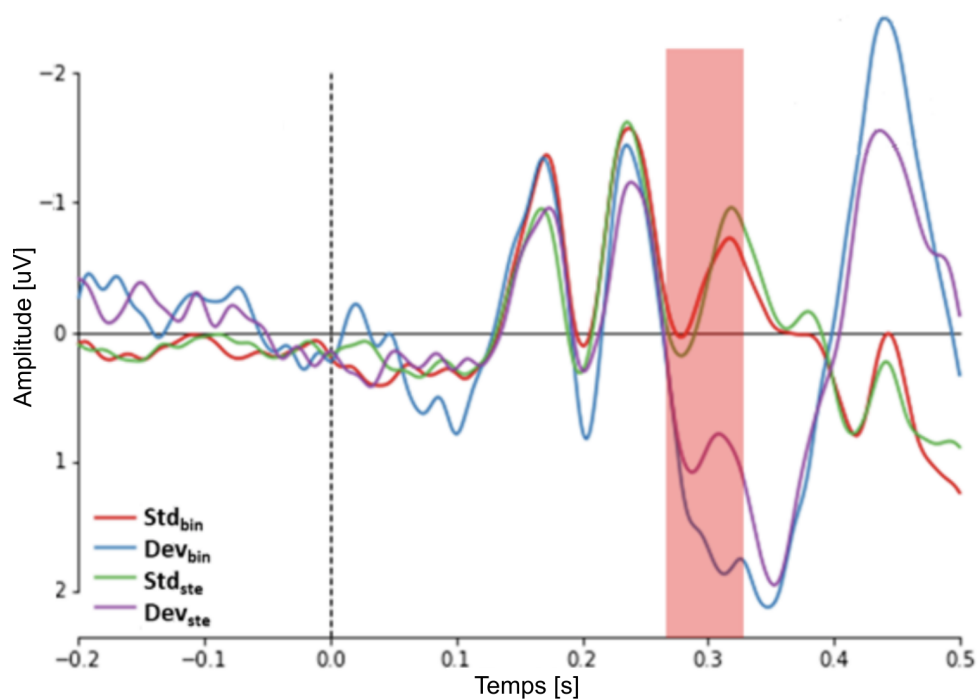


FIGURE 1.30 – P300 modulé par les différents modes d'écoute dans l'étude de Nicol *et al.* [165]. Les courbes bleue et violette correspondent respectivement au déviant binaural et au déviant stéréo, les courbes rouge et verte correspondent respectivement au standard binaural et au standard stéréo. Les courbes bleue et violette diffèrent de manière significative dans la fenêtre temporelle surlignée en rose (de 280 ms à 319 ms).

cinq façons différentes. Les méthodes de spatialisation utilisées étaient 1) des ITDs, 2) des ILDs, 3) une combinaison d'ITDs et d'ILDs, 4) du binaural non-individualisé ou 5) du binaural individualisé. Les trois premières méthodes sont des variations de spatialisation stéréophonique et produisent généralement des sons pas ou peu externalisés. La quatrième méthode peut produire des sons externalisés, mais avec une grande variabilité inter-sujet. La cinquième méthode peut produire des sons correctement externalisés. En complément de l'enregistrement MEG, les sujets avaient pour tâche perceptive de localiser la provenance des sons. Les résultats ont montré que, indépendamment de l'angle d'incidence, les stimuli diffusés en binaural individualisé produisaient une plus grande amplitude de la composante N1m (équivalent MEG de la N1) que les stimuli en binaural non individualisé et que les stimuli stéréophoniques. De plus, l'amplitude de la N1m était plus grande pour les stimuli binauraux non-individualisés que pour les stimuli stéréophoniques. En considérant que les stimuli diffusés en binaural individualisé étaient plus externalisés que ceux diffusés en binaural non-individualisé, eux-mêmes plus externalisés que les stimuli stéréophoniques, ces résultats peuvent suggérer que l'externalisation se traduit par une amplitude accrue de la composante N1m.

Une étude de Getzman et Lewald [17] a montré que l'équivalent EEG de la N1m, la N1, pouvait également être modulée par le mode de spatialisation. Dans cette étude, les sujets écoutaient des bruits blancs continus comportant deux phases. Pendant la première phase, les stimuli étaient localisés à 0° d'azimut et étaient immobiles, et lors de la deuxième phase les stimuli se déplaçaient rapidement de 90° vers la gauche ou vers la droite. Comme ces stimuli comportaient un changement interne, les ERPs résultants présentaient un complexe N1 - P2 synchronisé avec le début du stimulus, suivi d'un ACC dont les composantes cN1 et cP2 étaient synchronisées avec le début du mouvement (*cf.* section 1.1.3). Il y avait quatre modes de spatialisation : restitution sur enceintes (pour laquelle les stimuli devaient être parfaitement externalisés), restitution binaurale non-individualisée (pour laquelle les stimuli devaient être externalisés, mais moins qu'en écoute sur enceintes), restitution stéréophonique avec des ITDs et restitution stéréophonique avec des ILDs (pour lesquelles l'externalisation devait être médiocre). En plus de l'enregistrement EEG, les sujets devaient indiquer la direction de déplacement (droite ou gauche) et leurs temps de réaction étaient enregistrés.

Les résultats ont montré que la latence des composantes cN1 et cP2 était plus courte lors des restitutions sur enceintes et binaurale que lors des restitutions stéréophoniques. De plus, l'amplitude de la cN1 était plus grande lors de la restitution sur enceintes que lors des restitutions binaurale ou stéréophoniques (voir figure 1.31). Les résultats ERP étaient en accord avec les

résultats perceptifs indiquant que les temps de réaction étaient plus courts lors de la restitution sur enceintes que lors de la restitution binaurale, eux même plus courts que lors des restitutions stéréophoniques. Ces résultats pourraient suggérer que plus un son est externalisé, plus il produit un complexe N1 - P2 précoce et d'amplitude élevée (au moins dans le cas d'un complexe cN1 - cP2).

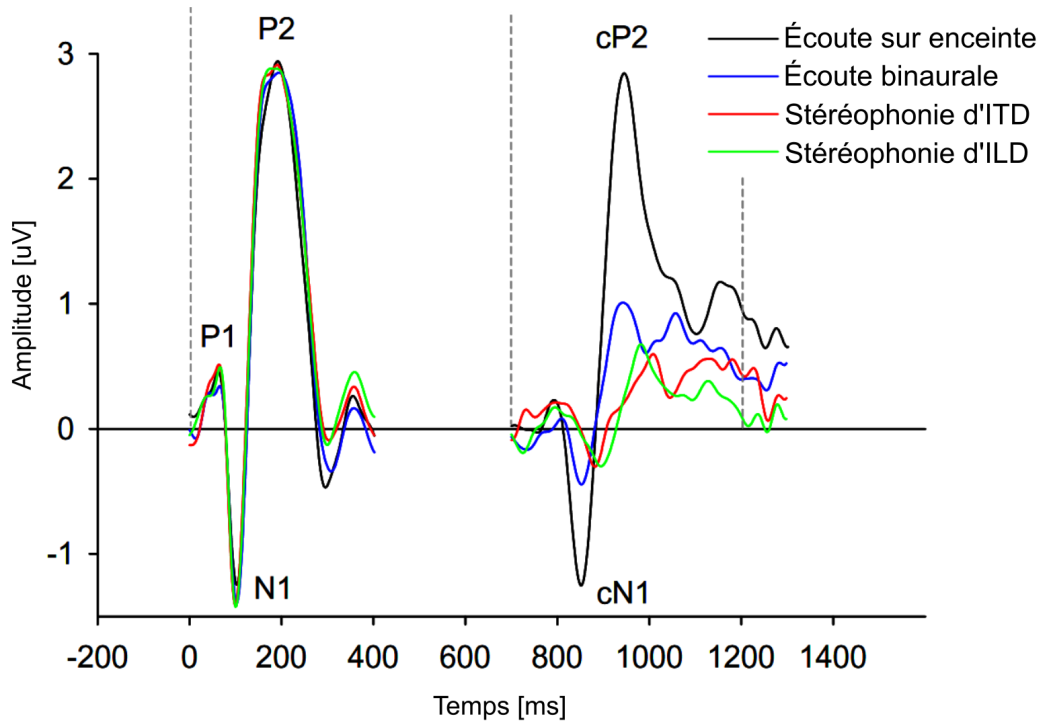


FIGURE 1.31 – ERP modulé par les différents modes d'écoute dans l'étude de Getzmann et Lewald [17].

Ces trois études utilisant des ERPs comportent cependant le même problème que les études utilisant l'IRMf discutées précédemment : il n'est pas possible d'associer avec certitude les différences observées sur les ERPs à des différences d'externalisation puisque l'externalisation des sources n'a pas été évaluée de façon perceptive. Il est tout à fait possible que les différences observées soient dues à des différences de contenu spectral ou de modalité tactile. En effet, les différents modes de restitution utilisés dans ces études (restitution sur enceintes, binaurale individualisée, binaurale non-individualisée et stéréophonique) induisent typiquement des différences d'externalisation, mais également des différences de contenu spectral (HRTFs individualisées vs. HRTFs non-individualisées vs. pas de filtrage HRTFs) et de perception tac-

tile de par la présence ou non du casque sur la tête de l'auditeur (écoute sur enceintes vs. écoute au casque).

1.4.4 Objectifs de la thèse

Il a été présenté précédemment dans cette section l'importance de l'externalisation pour une restitution binaurale convaincante et les paramètres qui influent sur la perception de l'externalisation. Dans la littérature, l'externalisation a été étudiée principalement d'un point de vue perceptif. Il n'existe que peu d'études ayant utilisé des techniques objectives pour explorer cette dernière et aucune d'entre elles n'a utilisé de protocole permettant d'associer catégoriquement un marqueur objectif d'activité cérébrale à l'externalisation. Il a également été présenté l'apport des systèmes de head-tracking pour améliorer l'externalisation. Cependant, l'efficacité de cette amélioration et sa robustesse dans le temps n'ont pas encore été établies de façon claire.

Le but de ce travail de thèse a été d'explorer l'externalisation en restitution binaurale avec head-tracking. Dans un premier temps, une étude composée d'une expérience comportementale et d'une expérience EEG a été réalisée, avec pour objectif de comparer des résultats perceptifs et électrophysiologiques de stimuli plus ou moins externalisés afin d'établir un marqueur objectif de l'externalisation. Un système de head-tracking a été utilisé pour faire varier l'externalisation d'un stimulus à l'autre. Les résultats de cette première étude ont soulevé plusieurs problématiques concernant l'influence de la nature des sources sonores et du temps d'exposition à ces sources sur la persistance de l'amélioration d'externalisation obtenue après des mouvements de tête avec head-tracking. Une deuxième étude, composée de trois expériences perceptives, a donc été réalisée afin d'explorer les relations entre la nature des sources sonores, le temps d'exposition à ces sources, et l'amélioration d'externalisation induite par le head-tracking.

ÉTUDE COMPARATIVE COMPORTEMENTALE VS. EEG DE L'EXTERNALISATION EN RESTITUTION BINAURALE DYNAMIQUE

Ce chapitre présente la première étude menée dans le cadre de la thèse visant à trouver un corrélat neuronal de l'externalisation en mettant en parallèle des résultats comportementaux et EEG. Cette étude a été publiée dans la revue *Hearing Research* [169].

2.1 Introduction

La section 1.4.2 a présenté les limitations liées à l'étude de l'externalisation à l'aide de tests comportementaux. Bien que certaines études utilisant des méthodes objectives aient considéré l'externalisation comme un facteur modulant potentiellement l'activité cérébrale, aucune n'a mis en évidence un corrélat clair de l'externalisation. En effet, l'étude de Palomäki *et al.* [168] ainsi que l'étude de Getzman et Lewald [17] ont comparé l'activité cérébrale induite par des stimuli plus ou moins externalisés, mais la présence de différences inter-stimulus autres que l'externalisation, notamment de différences spectrales, empêche d'attribuer de façon catégorique les différences d'activité cérébrale observées à des différences d'externalisation (Voir section 1.4.3 pour une revue détaillée de ces études). La mise en évidence d'un tel corrélat pourrait permettre de faciliter les futures études de l'externalisation.

La première étude de cette thèse avait pour objectif de trouver un marqueur objectif de l'externalisation à l'aide de l'EEG. Deux expériences - une expérience EEG et une expérience comportementale - ont été menées par les mêmes sujets dans les mêmes conditions expérimentales. Pour s'assurer que le seul facteur de variation était l'externalisation, les stimuli étaient iden-

tiques d'un essai à l'autre (emplacement identique, contenu spectral identique et délivrés par des dispositifs de restitution identiques) mais ils étaient précédés soit d'un mouvement de tête avec head-tracking, soit d'un mouvement de tête sans head-tracking, soit d'aucun mouvement de tête. En effet, comme des études antérieures ont montré que les améliorations d'externalisation persistent une fois que le sujet a cessé de bouger la tête (*cf.* section 1.4.1), les stimuli précédés de mouvements de tête avec head-tracking devraient être davantage externalisés que les stimuli précédés de mouvements de tête sans head-tracking et que les stimuli sans mouvement de tête. En considérant les résultats des études précédentes présentés dans la section 1.4.3, l'hypothèse de la présente étude était que les ERPs générés par des stimuli plus externalisés ont une plus grande amplitude et une latence plus courte que les ERPs générés par des stimuli moins externalisés. En particulier, des différences sur les composantes N1 et P2 étaient attendues, car ce sont ces dernières qui ont été modifiées par des différences d'externalisation dans les études précédentes [17, 168].

2.2 Matériel et méthode

2.2.1 Participants

Vingt sujets (8 hommes, 12 femmes ; âge moyen 25 ; écart-type 6,7 ; intervalle 18-39 ans) ont participé à l'étude. Pour tenir compte de potentielles différences de main dominante, tous les sujets ont répondu au questionnaire *Finders Handedness* [170]. Seize sujets étaient droitiers et quatre gauchers. Tous ont déclaré avoir une audition normale. L'étude était conforme au Code d'éthique de l'Association médicale mondiale (Déclaration d'Helsinki), publié dans le *British Medical Journal* (18 juillet 1964) et a été approuvée par le Comité Consultatif Institutionnel pour la Protection des Personnes dans la Recherche (CICPPR) d'IMT Atlantique (enregistré sous le numéro IRB00013722 auprès du bureau américain de protection des humains dans la recherche [171]). En outre, tous les sujets ont donné leur consentement éclairé écrit pour participer à l'étude.

2.2.2 Stimulus

Chaque essai consistait en un extrait de 5,5 s du poème "l'Albatros" (v.13-14) de Charles Baudelaire suivi d'un train de dix répétitions d'un pseudo-mot bi-syllabique non sémantique d'une durée d'une seconde : "boru". En prenant en compte la réverbération, le mot "boru" durait 600 ms. Un silence de 600 ms était inséré entre la phrase et le premier "boru", ainsi qu'entre

chaque "boru". L'intervalle inter-stimulus (IIS) était donc de 600 ms et l'intervalle entre les débuts de deux stimuli consécutifs était de 1200 ms. La durée totale de l'essai était de 17,5 s. La figure 2.1 montre une représentation de la composition d'un essai. Il a déjà été démontré que les pseudo-mots bisyllabiques tels que "boru" pouvaient générer de manière fiable des ERPs (e.g. [172-174]). En outre, un pré-test informel a été réalisé sur six sujets afin de tester différents mots possibles (sémantiques, non sémantiques, monosyllabiques, bi-syllabiques et plus longs). Le mot "boru" générait l'ERP le plus net en moyenne sur l'ensemble des sujets.

La situation idéale aurait été, après chaque extrait de poème, d'avoir le mot "boru" une seule fois. Ainsi, chaque essai aurait généré un seul ERP *single trial* (cf. section 1.1.3), indépendant de la réponse des autres essais. Cependant, afin d'obtenir des ERPs exploitables pour étudier les composantes N1 et P2, il est nécessaire de moyennner un grand nombre d'ERPs *single trial*. Si chaque essai n'avait contenu qu'un "boru", il aurait fallu effectuer un grand nombre de répétitions d'essais, ce qui aurait rendu le test trop long (cf. section 1.1.3). Afin d'obtenir un nombre d'ERPs *single trial* suffisant dans une durée de test réalisable, le mot "boru" était répété dix fois dans chaque essai. Ainsi, un essai permettait de générer dix ERPs *single trial* sans avoir besoin de répéter la première partie de l'essai (l'extrait de poème).

Ce choix n'était pas idéal puisque, au sein d'un essai, les ERPs *single trial* n'étaient pas indépendants. Leur position au sein du train de répétitions pouvait avoir une influence sur l'activité électrique résultante. On pouvait, par exemple, s'attendre à ce que le premier "boru" juste après l'extrait de poème ne génère pas la même activité que le dernier "boru". Ce choix de regrouper les "boru" par train de dix a néanmoins été adopté, car il semblait être le meilleur compromis pour avoir un nombre de répétitions suffisamment élevé dans une durée de test réalisable.

2.2.3 Enregistrement et traitement du stimulus sonore

Le son a été enregistré par une locutrice dans une petite chambre anéchoïque avec un microphone Neumann U87 et une interface audio RME babyface, connectée à un ordinateur Windows utilisant Protools comme logiciel d'enregistrement. Le mot "boru" était composé de 400 ms de son direct suivi de 200 ms de réverbération.

Une réverbération réaliste et pouvant s'adapter aux mouvements induits par le head-tracking a été générée à l'aide de la réponse impulsionnelle ambisonique de 4e ordre d'une salle existante et a été convoluée avec l'enregistrement anéchoïque de la voix. Seule la partie réverbérée résultant de cette convolution a été conservée et mixée au son direct anéchoïque original. Cette réponse impulsionnelle avait une résolution de 24 bits et un taux d'échantillonnage de 48 kHz [175]. L'ajout de réverbération était une précaution pour éviter que les stimuli soient

tous perçus complètement internalisés (*cf.* section 1.4.1), ce qui aurait empêché d'obtenir des différences d'externalisation inter-stimuli.

L'enregistrement de la réponse impulsionnelle a été effectué dans une petite salle de concert ("l'Ubu" de Rennes; temps de réverbération moyen $\approx 0,46$ s; rapport champ direct / champ réverbéré $\approx 6,08$ dB), à l'aide de la sphère de microphones EigenMike EM32, à une distance de 150 cm d'un haut-parleur Genelec 8040. Un signal exponentiel sinusoïdal de 10 s a été utilisé pour les mesures. Les réponses impulsionnelles mesurées ont été débruitées à l'aide de la procédure décrite dans [176]. Ensuite, un filtre FIR a été utilisé pour compenser la réponse en fréquence du haut-parleur dans les mesures de 60 Hz à 16 kHz. Le filtre d'égalisation a été dérivé d'une mesure de la réponse impulsionnelle du haut-parleur effectuée dans une chambre anéchoïque à l'aide d'un microphone omnidirectionnel dans l'axe du haut-parleur. Pendant l'expérience, la réverbération ambisonique et le son direct étaient binauralisés indépendamment, puis mixés en temps réel à l'aide du logiciel "SPAT" de l'IRCAM [177]. La mesure de HRTFs individualisées étant un processus long et compliqué (*cf.* section 1.3.1), la grande majorité des contenus binauraux sont produits à l'aide de HRTFs non-individualisées. Par conséquent, les stimuli utilisés dans la présente étude étaient binauralisés à l'aide de HRTFs non-individualisées.

La section 1.3.3 a présenté des études antérieures montrant que le système auditif est capable de s'adapter aux indices binauraux et monauraux créés par le filtrage HRTF non individualisé. Après une phase d'apprentissage, la localisation et l'externalisation d'une source virtuelle sont meilleures pour un auditeur. D'autres études ont même mis en évidence le fait que cet apprentissage peut se faire de façon non intentionnelle, simplement à force d'exposition à un set de HRTFs particulier [178, 179]. Pour minimiser un potentiel effet d'apprentissage qui aurait compliqué l'étude des mouvements de tête, trois sets de HRTFs différents étaient utilisés alternativement : les sets n°1004, n°1068, et n°1100 de la bibliothèque de l'IRCAM "BILI" [180].

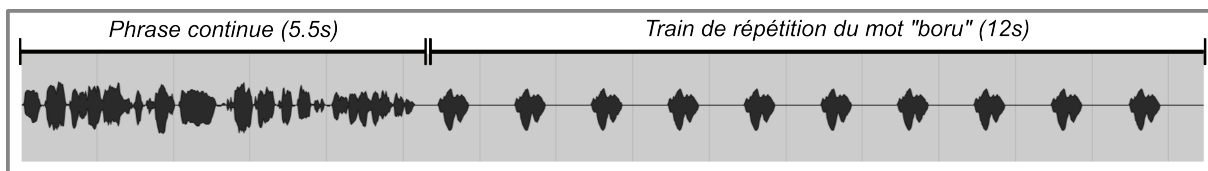


FIGURE 2.1 – Représentation de la composition d'un stimulus.

2.2.4 Système de reproduction audio

Les expériences ont été menées dans une salle insonorisée à double paroi et faiblement éclairée. Les stimuli étaient présentés au moyen d'un casque (Sennheiser HD650) que les sujets n'étaient autorisés à retirer à aucun moment de l'expérience afin d'éviter des modifications de HPTFs au cours de l'expérience (*cf.* section 1.3.1). Le niveau de pression sonore a été réglé à 65 dBA aux oreilles des sujets en plaçant le casque sur une tête artificielle (Neumann KU100 [104]). La tête artificielle a été calibrée à l'aide d'un pisonphone (Brüel & Kjør Type 4231 [181]) à 1 kHz. La lecture, l'affichage de l'interface et l'acquisition des données étaient contrôlés à l'aide d'un logiciel implémenté en Max/MSP sur un ordinateur Windows connecté à une carte son RME Babyface. Le module "SPAT" [177] utilisé pour la binauralisation était également implémenté dans Max/MSP. Un écran déporté, connecté à l'ordinateur, était utilisé pour afficher l'interface de test pour les sujets.

Les mouvements de tête étaient pris en compte par une rotation de la source sonore (pour le son direct) et une rotation de la matrice ambisonique (pour la réverbération) entraînant une rotation de la position du son direct et de la réverbération directionnelle dans la direction opposée à celle de la tête, ce qui permettait de maintenir l'ensemble de la scène sonore fixe par rapport au monde extérieur, tel que décrit dans la section 1.3.2. Un head-tracker "T3" de *Feichter Audio* [182] était placé au-dessus du casque et envoyait les informations de position de la tête (*yaw*, *pitch* et *roll*) au logiciel de binauralisation. Ces informations étaient enregistrées pour pouvoir vérifier que les sujets effectuaient correctement les mouvements de tête. La position des stimuli était actualisée en réponse aux mouvements de tête à une fréquence de 100 Hz. La latence totale du système était en moyenne de 60 ms (± 5 ms), ce qui correspond au seuil de détectabilité observé pour les meilleurs auditeurs (*cf.* section 1.3.2). En outre, Wenzel *et al.* [112] ont constaté que des latences pouvant aller jusqu'à 500 ms ne modifiaient pas l'externalisation perçue.

2.2.5 Enregistrement EEG

Pour enregistrer le signal EEG, une installation légère avec peu d'électrodes a été conçue afin que les sujets puissent effectuer facilement des mouvements de tête tout en minimisant les mouvements des fils électriques. Une carte Bluetooth "Cyton" [183] de *Open BCI* sur laquelle étaient connectées six électrodes cupules "goldcup" [184] a été utilisée. Chaque électrode était remplie de gel conducteur et fixée sur la tête du sujet à l'aide de bande adhésive. De plus, un bonnet en tissu était placé sur la tête pour maintenir toutes les électrodes en place. Trois élec-

trodes étaient positionnées aux emplacements C3, Cz et C4 du système international 10-20 (cf. section 1.1.2), où les composantes N1 et P2 ont généralement la plus grande amplitude [185]. L'enregistrement de ces composantes a été privilégié afin de pouvoir comparer les résultats de la présente étude avec ceux des études précédentes (cf. section 2.1). Une électrode était placée à Fp2 pour contrôler l'activité oculaire et deux électrodes étaient placées aux emplacements A1 et A2 (sur les mastoïdes), pour servir de référence. La figure 2.2 montre le placement des électrodes. La synchronisation entre les stimuli audio et l'EEG était réalisée à l'aide d'une courte impulsion générée par la carte son et connectée à la carte "Cyton" à l'aide d'un montage bipolaire. La carte "Cyton" était placée sur le sujet afin de minimiser les mouvements des câbles. Tous les signaux étaient envoyés par Bluetooth à un ordinateur Linux et enregistrés avec le logiciel *OpenBCI* [186]. Tous les canaux étaient enregistrés avec une fréquence d'échantillonnage de 250 Hz.

La figure 2.3 présente un synoptique des installations pour le test comportemental et pour le test EEG.



FIGURE 2.2 – Placement des électrodes utilisées pour l'expérience EEG.

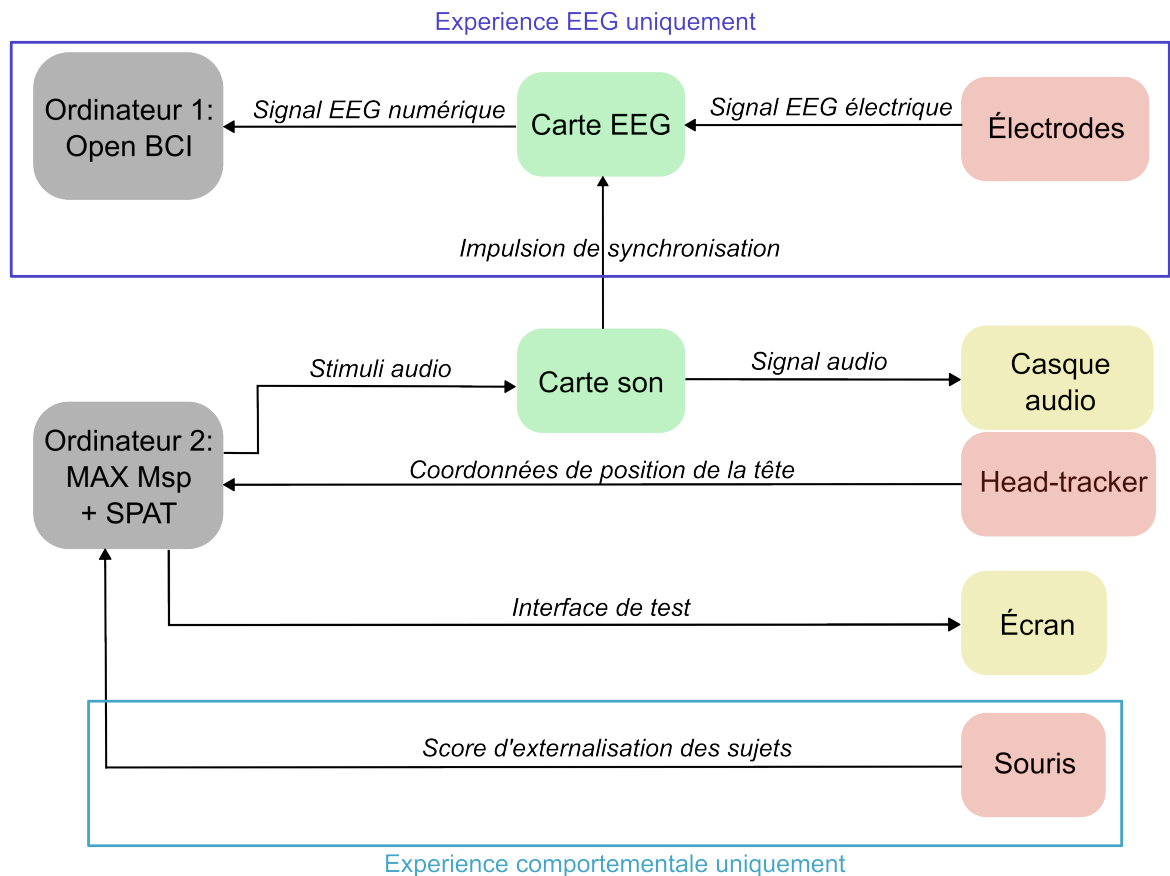


FIGURE 2.3 – Synoptique de l’installation pour les tests. La partie encadrée en bleu foncé était propre au test EEG, la partie encadrée en bleu clair était propre au test comportemental. Les autres parties étaient communes aux deux tests.

2.2.6 Procédure

La procédure utilisée était une adaptation du protocole utilisé dans l'étude de Hendrickx *et al.* [150]. la présentation des stimuli était divisée en deux phases :

1. Les sujets écoutaient un stimulus binaural lors duquel ils devaient :
 - soit garder la tête immobile,
 - soit effectuer un mouvement de tête sans head-tracking,
 - soit effectuer un mouvement de tête avec head-tracking.
2. Ils écoutaient un stimulus binaural en gardant la tête immobile. Les mesures comportementales et EEG étaient effectuées au cours de cette seconde phase.

Ce protocole a permis de s'assurer que les stimuli d'un essai à l'autre était "acoustiquement identiques" (localisation, contenu spectral et dispositifs de restitution identiques), mais avec un degré d'externalisation variable selon que la mesure était précédée ou non de mouvements de tête lors de la première phase. En effet, des études antérieures ayant montré que l'externalisation persiste une fois que le sujet a cessé de bouger la tête [150, 151], les stimuli précédés de mouvements avec head-tracking devaient être plus externalisés que les stimuli non précédés de mouvements avec head-tracking.

Pendant la phase 1 de chaque stimulus, les sujets devaient soit effectuer un mouvement horizontal de tête, soit rester immobiles. Pendant les mouvements de tête, le dispositif de head-tracking était soit actif, soit inactif. Par conséquent, la procédure expérimentale comprenait trois *conditions de mouvement* différentes (voir figure 2.4) :

- *SF* : "Statique" (head-tracking inactif), "Fixe" (pas de mouvement)
- *SM* : "Statique" (head tracking inactif), "Mouvement" (mouvement de tête)
- *DM* : "Dynamique" (head-tracking actif), "Mouvement" (mouvement de tête)

Il était demandé aux sujets de garder leur tête droite dans une position naturelle lors de l'écoute des stimuli. Pour les conditions *SM* et *DM*, ils devaient effectuer le mouvement de tête en suivant ce déroulement :

1. Pendant la phrase de 5.5 s (phase 1), les sujets devaient tourner la tête à 50° d'un côté, puis à 50° de l'autre côté avant de revenir en position de repos, à 0°, tête droite. Il était demandé à tous les sujets d'effectuer le même mouvement et d'être revenu à la position de repos juste avant la fin de la phrase.

2. Pendant le train de répétitions de "boru" (phase 2), les sujets devaient rester immobiles, en position de repos.

Pour la condition *SF*, les sujets devaient rester immobiles, en position de repos, pendant la phase 1 et la phase 2 des stimuli. Il était donc attendu que les stimuli dans la condition *DM* soient plus externalisés que les stimuli dans les conditions *SF* et *SM*. En outre, les résultats d'une étude précédente [150] suggèrent que l'externalisation est plus faible lorsque l'auditeur effectue un mouvement de tête sans head-tracking que quand l'auditeur reste immobile. On pourrait donc s'attendre à ce que la condition *SM* soit moins bien externalisée que la condition *SF*, ce qui permettrait d'obtenir 3 niveaux d'externalisation distincts.

Pour vérifier si le set de HRTFs pouvait avoir une influence sur l'externalisation, la procédure expérimentale comprenait également une variable *HRTFs* comprenant 3 sets de HRTFs différents. L'utilisation de 3 sets différents permettait en outre de minimiser un potentiel effet d'apprentissage d'un set de HRTFs unique (*cf.* section 2.2.4).

Les sources sonores situées sur les côtés de l'auditeur sont généralement correctement externalisées avec et sans mouvement de tête (*cf.* section 1.4.1). Dans la présente étude, afin qu'il existe une différence d'externalisation entre *DM* et les autres conditions, il était nécessaire de placer les stimuli en position azimutale 0°, devant les sujets.

Comme l'objectif de l'étude était de découvrir des marqueurs les plus objectifs possibles de l'externalisation, l'expérience EEG a été menée avant l'expérience comportementale, sans attirer l'attention des sujets sur l'externalisation. Au cours de l'expérience comportementale, comme les sujets devaient évaluer l'externalisation, leur attention était évidemment focalisée sur l'externalisation.

Avant chaque expérience (EEG et comportementale), les sujets devaient effectuer une phase d'entraînement pour apprendre à exécuter correctement les mouvements de tête. Avant chaque essai, des indications visuelles étaient affichées à l'écran pour indiquer aux sujets d'exécuter le mouvement de la tête (pour les conditions *SM* ou *DM*) ou de rester statiques (*SF*).

Spécificités de l'expérience EEG

Pendant l'expérience EEG, lorsqu'un essai était terminé, le suivant commençait automatiquement après une pause de 1200 ms. Chacune des 10 répétitions du mot "boru" générant un ERP et était considérée comme un stimulus. Chaque essai pour une combinaison de [condition de mouvement × HRTF] était répétée douze fois, soit un nombre total de 120 "boru" pour chaque combinaison de [condition de mouvement × HRTF]. Le nombre total d'essais était de 3

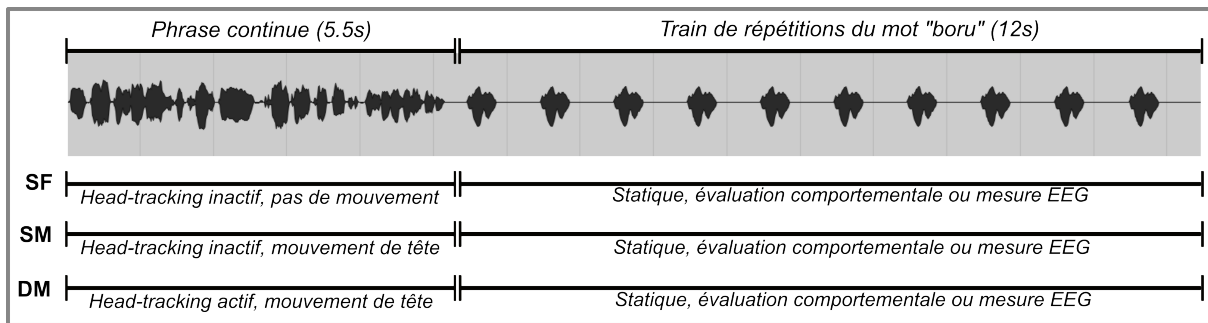


FIGURE 2.4 – Représentation des trois conditions de mouvement.

[conditions de mouvement] \times 3 [HRTFs] \times 12 [répétition d'essai] = 108 essais. Les trois conditions de mouvement étaient testées dans un ordre aléatoire, différent pour chaque sujet. Pour minimiser un éventuel effet d'apprentissage d'un set de HRTFs, le set de HRTFs changeait toujours entre deux essais successifs (en alternant entre les trois sets de HRTFs). Les 108 essais étaient divisés en trois blocs de 36 essais, avec de courtes pauses entre chaque bloc.

Spécificités de l'expérience comportementale

Au cours de l'expérience comportementale, après chaque essai, les sujets devaient déplacer un curseur sur l'interface de l'ordinateur pour évaluer l'externalisation ressentie pendant la deuxième partie de l'essai (*i.e.* le train de "boru"). L'échelle du curseur allait de 0 ("source complètement internalisée") à 100 ("source parfaitement externalisée"). Comme discuté dans la section 1.4.2, aucune échelle n'est optimale. Cette échelle pseudocontinue a été choisie pour permettre aux sujets d'affiner au maximum leurs réponses et ainsi d'obtenir les scores d'externalisation les plus précis possibles (figure 2.5). Comme l'externalisation n'est pas forcément une notion instinctive pour des auditeurs non experts, la description suivante de l'externalisation était donnée aux sujets avant de commencer le test. Cette description était simple et brève afin que l'évaluation reste naturelle et aussi spontanée que possible :

L'externalisation est le fait de percevoir le son comme s'il provenait de l'extérieur de notre tête, comme si la source sonore était présente dans la pièce. L'interface de réponse va de 0 à 100. 0 correspond à un son totalement internalisé, provenant de l'intérieur de votre tête, au centre, comme une voix intérieure. 100 correspond à une source parfaitement externalisée, comme lorsque quelqu'un vous parle dans la réalité.

Contrairement à l'expérience EEG lors de laquelle chaque "boru" était considéré comme un stimulus, chaque essai (*i.e.* chaque train de "boru") de l'expérience comportementale ne donnait lieu qu'à une réponse et était donc considéré comme un seul stimulus. Chaque combinaison de [condition de mouvement × HRTF] était répétée dix fois, soit un total de 3 [conditions de mouvement] × 3 [HRTF] × 10 [répétition d'essai] = 90 essais. Comme dans l'expérience EEG, les trois conditions de mouvement étaient testées dans un ordre aléatoire, différent pour chaque sujet. Pour minimiser un éventuel effet d'apprentissage d'un set de HRTFs, le set de HRTFs changeait toujours entre deux essais successifs (en alternant entre les trois sets de HRTF).

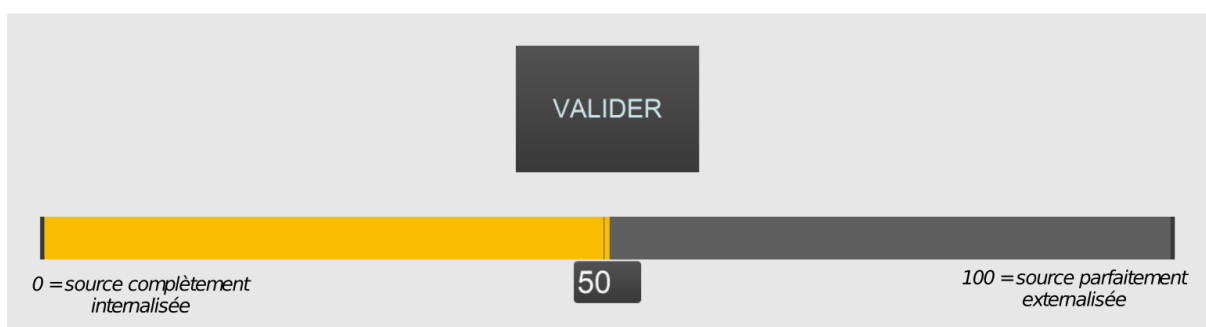


FIGURE 2.5 – Interface de réponse utilisée par les sujets dans l'expérience comportementale.

2.2.7 Pré-traitement de l'EEG et création des epochs

Le prétraitement et l'analyse de l'EEG ont été effectués à l'aide de la bibliothèque Python *MNE*, version 1.4.1 [187]. Pour toutes les étapes mentionnées ci-dessous, les paramètres recommandés par la littérature EEG et repris par les fonctions de *MNE*, ont été utilisés.

Le signal brut a été re-référencé sur la moyenne des signaux des électrodes A1 et A2. Le re-référencement des signaux issus des électrodes "utiles" par rapport au signal issu d'une électrode la plus inactive possible est nécessaire pour obtenir des électroencéphalogrammes exploitables, et permet de mieux distinguer les différences évoquées comme des écarts par rapport à la ligne de base. La moyenne des électrodes des mastoïdes a été choisie comme référence, car les mastoïdes sont placées sur une partie osseuse conduisant peu de courant, et elles sont éloignées des électrodes "utiles" localisées sur le sommet du crâne. (*cf.* section 1.1.2).

Le bruit du secteur a été supprimé à l'aide d'un filtre coupe-bande aux fréquences 50 Hz, 100 Hz et 150 Hz. Pour obtenir des ERPs de qualité, les dérives lentes ont été supprimées en appliquant un filtre passe-haut de 0,1 Hz comme c'est couramment le cas (*cf.* section 1.1.3).

Comme décrit dans la section 1.1.2, les clignements d'oeil représentent un artefact très courant lors des enregistrements EEG qu'il convient de supprimer. Les segments contenant des clignements d'oeil ont été identifiés sur le signal brut de l'électrode Fp2 à l'aide du module *find-eog-event* [188] de MNE et supprimés lors de la création des epochs (pour une description du principe d'epochs, voir section 1.1.3).

Comme décrit dans la section 1.1.3, le choix de la fenêtre de prélèvement des epochs dépend des composantes que l'on souhaite étudier et est limité par la durée de l'intervalle inter-stimuli (IIS). Cette fenêtre a été définie de -200 à 500 ms autour du début du stimulus (chaque répétition de "boru"). Une durée de 500 ms après le début du stimulus était suffisante pour pouvoir étudier les composantes N1, P2 et potentiellement des composantes cN1 et cP2 plus tardives (*cf.* section 1.1.3). D'autre part, une ligne de base de 200 ms préstimulus était suffisante pour obtenir des ERP clairs en réduisant l'influence de la dérive lente (*cf.* section 1.1.3). L'IIS étant de 600 ms, il n'y avait aucun risque de recouvrement entre epochs successifs.

Même en supprimant le bruit du secteur, les segments contenant des clignements d'oeil, un certain nombre d'epochs étaient inutilisables, car ils étaient contaminés par des artefacts de mouvements des sujets ou de mouvements des électrodes. Le module *auto-reject* [189] de MNE a été utilisé pour supprimer ces epochs contaminés. Ce module calcule automatiquement des seuils spécifiques pour chaque électrode de chaque sujet. Les segments dont l'amplitude dépasse ce seuil sont rejetés. Une inspection manuelle des epochs ainsi rejetés a été effectuée pour les données de chaque sujet pour s'assurer que le module *auto-reject* fonctionnait correctement pour tous les sujets et ne supprimait pas d'epochs utilisables. Les données de cinq sujets ont été écartés des analyses, car leur taux de rejets était trop élevé ($\geq 20\%$). Pour les autres sujets, le taux de rejets moyen était de 11% des epochs. Les epochs rejetés étaient en moyenne uniformément répartis à la fois à l'intérieur de chaque stimulus et globalement tout au long de l'expérience.

Un filtre passe-bas avec une fréquence de coupure de 40 Hz a été appliqué pour maximiser le rapport signal / bruit pour l'analyse ERP [190]. D'autres valeurs de fréquence de coupure ont été explorées (25 Hz, 50 Hz et 100 Hz) mais n'ont pas été retenues. Le filtre avec une fréquence de coupure à 25 Hz était trop destructif et risquait de supprimer des différences d'amplitude entre les conditions expérimentales. Les filtres à 50 Hz et 100 Hz n'étaient pas assez efficaces pour supprimer le bruit de mesure. Une fréquence de coupure de 40 Hz permettait une bonne suppression du bruit sans altérer les composantes ERPs. De plus, 40 Hz est une fréquence de coupure couramment utilisée dans des études ERPs similaires à la présente étude [165, 191].

Enfin, une normalisation de la ligne de base a été effectuée en calculant le *z-score* des epochs

prétraitées en fonction de la ligne de base, à l'aide du module *baseline.rescale* [192] de MNE. La méthode *z-score* a été préférée à une simple soustraction de la ligne de base pour mieux prendre en compte les différences individuelles dans les amplitudes ERP.

2.2.8 Analyses statistiques des ERPs

Les premières analyses ont suggéré que la position du "boru" au sein du train de répétitions de "boru" avait un impact sur l'ERP. Il est apparu que les "boru" situés vers la fin du train de répétitions généraient une activité d'amplitude absolue plus faible que les "boru" situés vers le début du train. Cet effet semblait être un effet de suppression due à la répétition et sera discuté dans la section 2.4.2. Néanmoins, l'observation de cet effet, qui n'avait pas complètement été anticipé lors de la mise en place du protocole, a conduit à considérer qu'il était nécessaire d'étudier l'influence de la position du "boru" au sein du train de répétitions. Cependant, étudier l'ERP pour chacune des dix positions possibles du "boru" au sein du train de répétitions aurait été impossible, il n'y aurait pas eu suffisamment d'époches par positions pour obtenir des ERPs exploitables. De plus, bien que les "boru" éloignés (par exemple le premier et le dernier de chaque train) étaient significativement différents, les "boru" proches (par exemple le premier et le deuxième de chaque train) étaient similaires. Ainsi, les trois premiers "boru" de chaque train ont été moyennés ensemble en tant que *groupe de répétitions 1*, les trois suivants en tant que *groupe de répétitions 2* et les trois derniers en tant que *groupe de répétitions 3*. Afin de conserver un nombre égal de répétitions de "boru" par groupe de répétitions, le 10ème "boru" de chaque train n'a pas été inclus. De cette manière, la variable *groupe de répétitions* a été définie.

Pour pouvoir comparer statistiquement les valeurs de latence et d'amplitude des composantes ERPs entre conditions expérimentales, il était nécessaire de créer, pour chaque condition, l'ERP de la condition en moyennant les époques correspondant à cette condition. Idéalement, il aurait fallu moyenné ensemble les époques correspondant à chaque combinaison de conditions expérimentales : [condition de mouvement × HRTF × groupe de répétitions]. Néanmoins, il n'y aurait pas eu suffisamment d'époches par combinaison de conditions pour obtenir des ERPs de qualité suffisamment bonne pour étudier les composantes N1 et P2. En effet, ces composantes sont généralement de taille modérée et nécessitent un grand nombre de répétitions (*cf.* section 1.1.3). Il a donc fallu regrouper dans un même ERP moyen les époques correspondants à différentes combinaisons de conditions.

Différentes manières de regrouper les epochs ont été explorées. Les regroupements qui semblaient les plus pertinents et qui ont été utilisés pour effectuer l'analyse statistique ont été définis de la manière suivante :

- Afin d'explorer les différences entre les conditions de mouvement tout en contrôlant l'effet des groupes de répétitions, les epochs ont été moyennés sur les sets de HRTFs pour chaque condition de mouvement et chaque groupe de répétitions. Le nombre d'epochs par combinaison de [condition de mouvement × groupe de répétitions] était le suivant : $3 \text{ (HRTF)} \times 12 \text{ (répétition de l'essai)} \times 3 \text{ (répétition du "boru" au sein d'un groupe)} = 108$ epochs (moins le taux de rejet de $\approx 11\%$).
- Afin d'étudier l'effet des sets de HRTFs, les epochs ont été moyennés sur les dix répétitions de "boru" pour chaque condition de mouvement et chaque set de HRTFs. Le nombre d'epochs par combinaison de [condition de mouvement × HRTF] était le suivant : $12 \text{ (répétition de l'essai)} \times 10 \text{ (répétition du "boru" au sein d'un essai)} = 120$ epochs (moins le taux de rejet de $\approx 11\%$).

Analyse fenêtrée

Pour tester les différences entre les conditions, deux approches statistiques ont été utilisées. Tout d'abord, une analyse fenêtrée a été réalisée, dans laquelle des fenêtres temporelles de composantes ERP d'intérêt ont été définies, puis les valeurs de latence et d'amplitude des pics de ces composantes ont été extraites sur chaque électrode. Les fenêtres temporelles ont été définies de manière suffisamment large pour inclure les composantes individuelles de chaque sujet, en procédant à l'inspection visuelle des ERP moyens par sujet. Les fenêtres temporelles suivantes ont été définies : 70-100 ms (correspondant à la N1), 90-160 ms (correspondant à la P2) et 170-230 ms. Cette dernière fenêtre temporelle correspond au pic cP2 faisant partie du complexe de changement auditif (ACC). L'ACC apparaît généralement après le complexe N1-P2 pour les stimuli complexes tels que les syllabes de la parole (*cf.* section 1.1.3). Il est donc logique que le mot "boru" ait généré un ACC. Pour chaque ensemble de valeurs (latence ou amplitude) de chaque composante (N1, P2, cP2), deux ANOVAs à mesures répétées ont été réalisées :

1. Conditions de mouvement [*SF, SM, DM*] × groupes de répétitions [*groupe de répétitions 1, groupe de répétitions 2, groupe de répétitions 3*].
2. Conditions de mouvement [*SF, SM, DM*] × HRTF [*HRTF 1, HRTF 2, HRTF 3*].

Des tests *post-hoc* de Holm Bonferroni ont ensuite été effectués sur les résultats significatifs.

Analyse "data-driven"

Deuxièmement, une analyse "data-driven" a été réalisée. La méthode non paramétrique proposée par Maris en 2007 [193] a été utilisée. Elle est largement adoptée par la littérature EEG et son principe est résumé ici. Cette méthode permet d'étudier les différences entre chaque paire de conditions de mouvement ou de groupes de répétitions à l'aide de tests non-paramétriques effectués sur les valeurs d'amplitude de chaque point temporel. Comme un grand nombre de comparaisons a été effectué, il était nécessaire de corriger les valeurs de p . Une correction de Bonferroni (ou même de Holm Bonferroni) aurait été beaucoup trop conservatrice sur un aussi grand nombre de comparaisons et aurait abouti à un nombre très réduit de valeurs de p significatives, une correction par cluster a donc été effectuée. Pour cela, des ensembles de points significatifs ont été définis, les *clusters*. Ces clusters ont été formés en regroupant dans un même cluster tous les points significatifs temporellement voisins. En effet, si plusieurs points temporels successifs sont significatifs, il est probable que la significativité observée soit due à une différence entre conditions plutôt qu'à un faux positif. Ainsi, un nombre réduit de clusters a été défini et les seuils de significativité ont été corrigés pour chaque cluster plutôt que pour chaque point temporel. Cette méthode statistique "data-driven" a pour avantage de détecter les effets statistiques à une résolution temporelle élevée sur la totalité des ERPs, et pas uniquement sur des fenêtres temporelles prédéfinies. Dans la présente étude, cette analyse statistique a été effectuée avec le module statistique du logiciel MNE.

2.3 Résultats

2.3.1 Résultats comportementaux

L'analyse des données de mouvements de tête a montré que les sujets ont, en moyenne, très bien respecté les instructions de mouvements de tête. Les graphiques de mouvements, par sujet, sont présentés dans la figure A.1 de l'annexe A.

Les données récoltées en utilisant une échelle continue sans label intermédiaires, comme c'était le cas de l'échelle utilisée dans la présente étude, sont susceptibles de contenir des différences de moyenne globale et de dynamique entre sujets. En effet, certains sujets n'utilisent qu'une petite partie de l'échelle, donnant lieu à des scores très condensés, alors que d'autres utilisent la totalité de la dynamique disponible. Si les tests statistiques prévus sont des tests paramétriques, il est nécessaire de compenser les différences de dynamique entre les sujets en effectuant une normalisation par *z-transform*. Cette normalisation permet de faire en sorte que

les données de chaque sujet aient la même moyenne globale et le même écart type, tout en conservant les différences entre conditions. Ainsi, les conditions peuvent être moyennées entre tous les sujets.

Les données étaient normalement distribuées et un test de Mauchly a révélé que la sphéricité était respectée. Les données, normalisées par *z-transform*, ont été soumises à une ANOVA à mesures répétées : condition de mouvement [*SF*, *SM*, *DM*] × HRTF × répétition d'essai. Un effet simple de la condition de mouvement ($F[2, 38] = 17,8; p < 0,001$) a été constaté. Un test *post-hoc* de Holm Bonferroni a montré que l'externalisation était plus importante dans la condition *DM* que dans les conditions *SF* ($p < 0,001$) et *SM* ($p < 0,001$). Aucune différence significative entre les conditions *SF* et *SM* n'est apparue (figure 2.6). Il n'y a pas eu d'effet du set de HRTFs ou de la répétition de l'essai et aucune interaction n'était significative.

Afin d'explorer les différences individuelles, les effets des conditions de mouvement sur les scores d'externalisation bruts (*i.e.* sans normalisation par *z-transform*) ont été étudiés sujet par sujet (figure 2.7). Les mouvements avec head-tracking ont amélioré l'externalisation modérément (>15 points) pour six sujets (1, 4, 7, 8, 9 et 13) et légèrement (>5 points) mais toujours de manière significative pour six autres sujets (2, 3, 10, 11, 14 et 15). Pour trois sujets, les mouvements avec head-tracking n'ont pas amélioré l'externalisation (sujets 5, 6 et 12). On remarque qu'il existe des différences d'externalisation importantes entre les sujets. Par exemple, le sujet 10 a donné des scores d'externalisation globalement très faibles, et avec des écarts-types très faibles, ce qui traduit une faible sensibilité à l'externalisation, mais une grande régularité dans sa perception. Ainsi, même si la différence d'externalisation entre la condition *DM* et les deux autres conditions est faible, elle est significative. En revanche, le sujet 5 a donné des scores d'externalisation globalement élevés, mais avec des écarts-types importants et aucune différence significative entre les conditions.

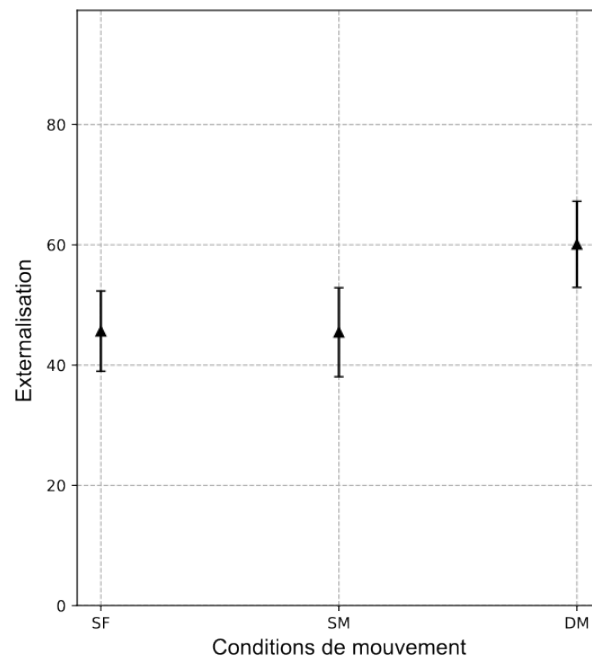


FIGURE 2.6 – Scores d'externalisation moyennés et normalisés (z -transform) sur tous les sujets pour chaque condition de mouvement (SF : pas de mouvement de tête, head-tracking inactif, SM : mouvement de tête, head-tracking inactif, DM : mouvement de tête, head-tracking actif). Les barres d'erreur indiquent l'intervalle de confiance à 95%.

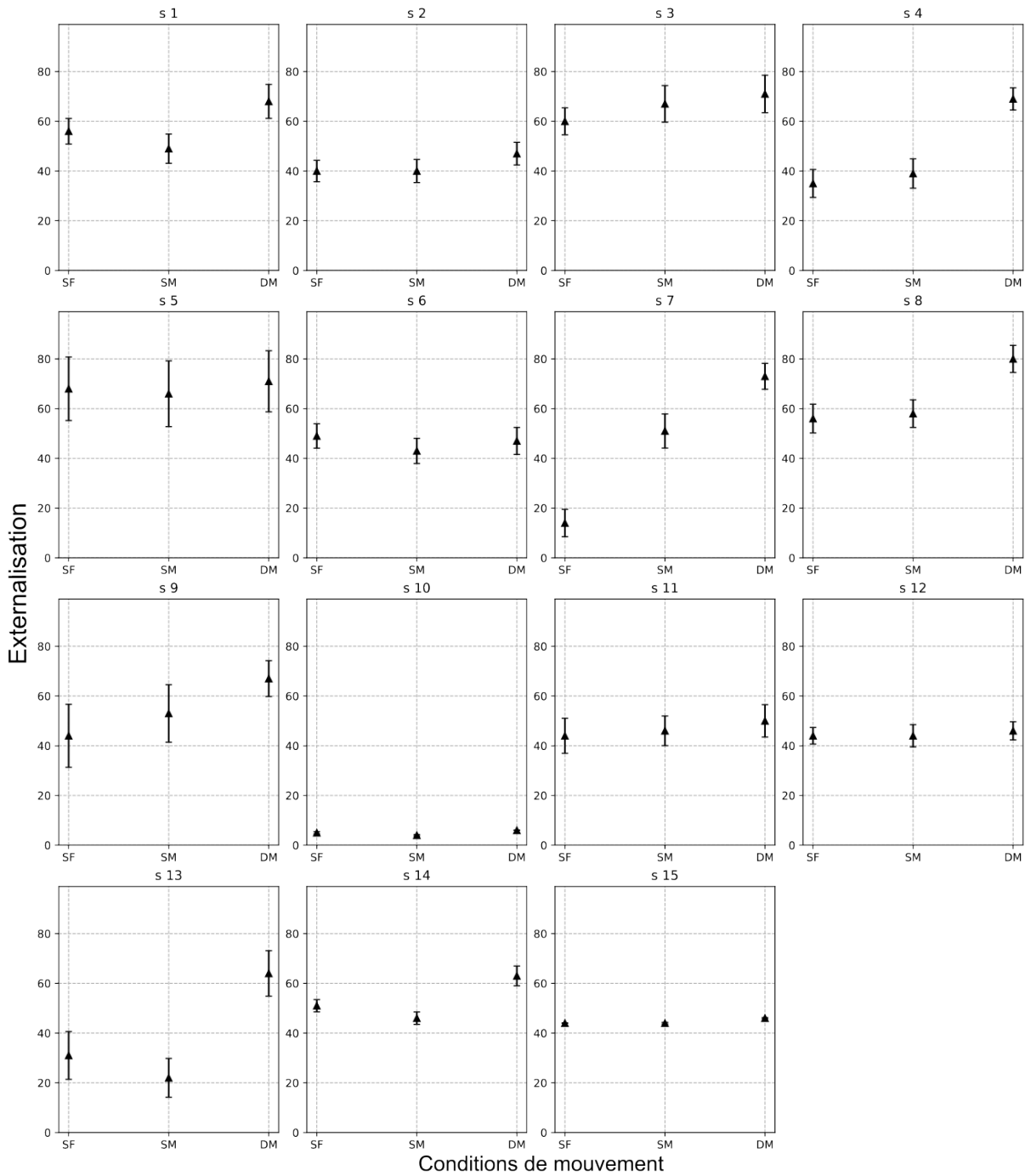


FIGURE 2.7 – Scores d'externalisation moyennés (non normalisés) par sujet pour chaque condition de mouvement (*SF* : pas de mouvement de tête, head-tracking inactif, *SM* : mouvement de tête, head-tracking inactif, *DM* : mouvement de tête, head-tracking actif). Les barres d'erreur indiquent l'intervalle de confiance à 95%.

2.3.2 Résultats EEG

Condition de mouvement

Dans l'analyse fenêtrée, un effet simple de la condition de mouvement a été trouvé pour l'amplitude de la N1 ($F[2, 28] = 5.01$; $p = 0.014$) : un test *post-hoc* a indiqué que l'amplitude de *SM* était plus négative que les amplitudes de *SF* et de *DM* (figure 2.8). L'interaction [groupe de répétitions \times condition de mouvement] était significative pour l'amplitude de la cP2 ($F[2, 28] = 16.73$; $p < 0.001$) : un test *post-hoc* a indiqué que, pour le *groupe de répétitions 1*, l'amplitude de *SF* était moins négative que celle de *SM* et de *DM* (figure 2.9).

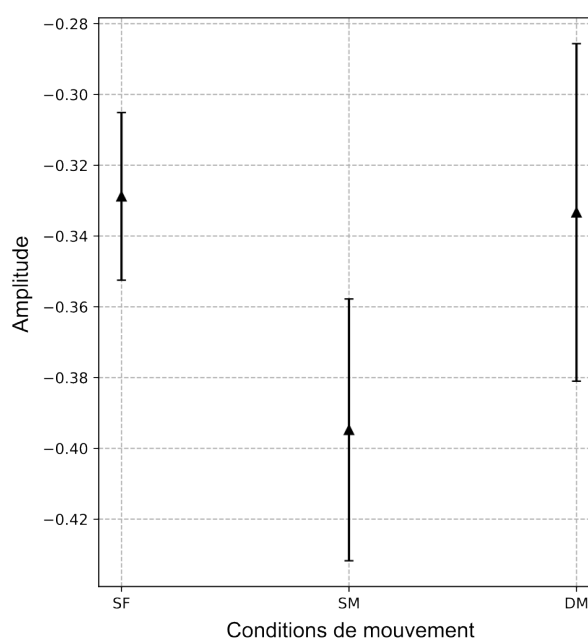


FIGURE 2.8 – Moyenne de l'amplitude des pics, normalisée par *z-transform*, pour la composante N1, sur Cz, par condition de mouvement, avec les groupes de répétitions regroupés. Les barres d'erreur indiquent l'intervalle de confiance à 95%. Unités d'amplitude arbitraires.

Une tendance similaire a été observée dans l'analyse "data-driven" (figure 2.10 et figure 2.11). Un cluster significatif a été trouvé dans la fenêtre temporelle correspondant à la N1 (80-116 ms), pour le *groupe de répétitions 1* : l'amplitude de *SM* était plus négative que l'amplitude de *SF* (pic $t = -3.1$, $p = 0.043$). Malgré l'utilisation de différentes approches statistiques et l'examen de fenêtres temporelles autres que celles correspondant aux composantes d'intérêt (N1 et P2), aucune différence statistique n'a pu être trouvée entre la condition de mouvement *DM* et les deux autres conditions *SF* et *SM*.

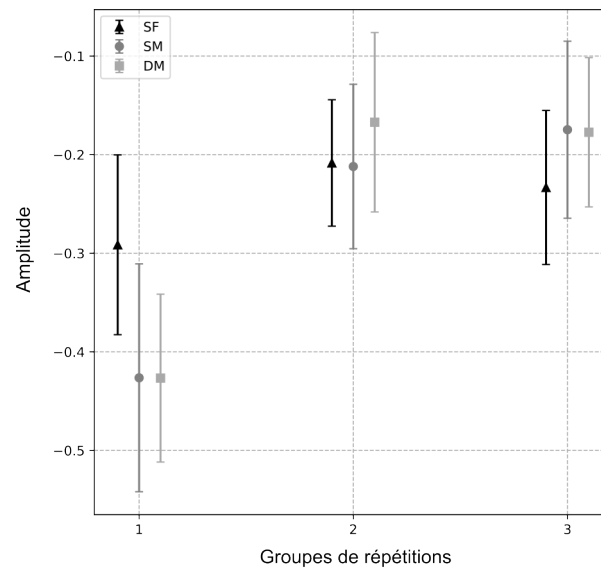


FIGURE 2.9 – Moyenne sur tous les sujets de l'amplitude des pics, normalisée par z -transform, pour la composante cP2, sur Cz, par condition de mouvement et groupe de répétitions. Les barres d'erreur indiquent l'intervalle de confiance à 95%. Unités d'amplitude arbitraires.

Pour vérifier l'absence d'effet, des ANOVAs bayésiennes à mesures répétées [194] ont été réalisées sur les latences et les amplitudes. Les méthodes de statistiques bayésiennes étant relativement nouvelles, une courte présentation de leur principe est donnée dans l'annexe B. Les résultats ont fourni des preuves modérées à fortes de l'absence d'effet de la condition de mouvement pour : la latence de la N1 ($DF_{01} = 8.704$), la latence de la P2 ($DF_{01} = 12,572$), l'amplitude de la P2 ($DF_{01} = 13.085$) et l'amplitude de la cP2 ($DF_{01} = 10.120$). Pour l'amplitude de la N1, un *post-hoc* a fourni une preuve modérée de l'absence de différence entre SF et DM ($DF_{01} = 5,846$). Par conséquent, l'ANOVA bayésienne a fourni des preuves modérées à fortes de l'absence de différences entre DM et les deux autres conditions de mouvement.

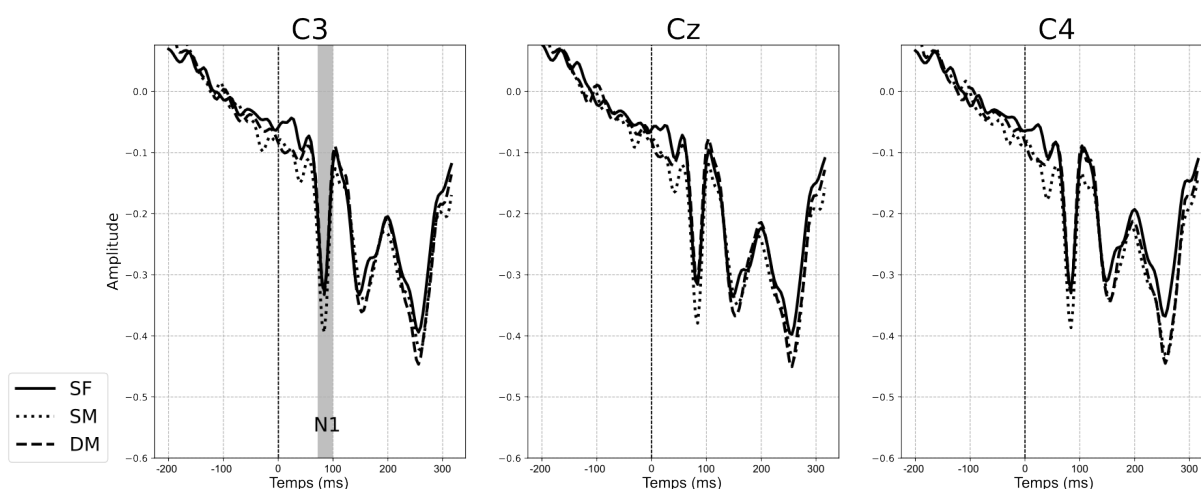


FIGURE 2.10 – ERP (moyennés sur tous les sujets) comparant les 3 conditions de mouvement (*SF* : pas de mouvement de tête, head-tracking inactif, *SM* : mouvement de tête, head-tracking inactif, *DM* : mouvement de tête, head-tracking actif), avec les groupes de répétitions regroupés, sur les 3 électrodes centrales. Unités d’amplitude arbitraires. La zone grise montre la fenêtre temporelle à partir de laquelle ont été extraites les valeurs de latence et d’amplitude de la composante N1.

Groupe de répétitions

Dans l’analyse fenêtrée, les effets constatés pour les groupes de répétitions sont les suivants :

- Pour la latence de la P2 ($F[2, 28] = 8.30$; $p = 0.001$) : un test *post-hoc* a indiqué que le *groupe de répétitions 1* avait une latence plus courte que les groupes de répétitions 2 et 3.
- Pour l’amplitude de la P2 ($F[2, 28] = 8.67$; $p = 0.001$) : un test *post-hoc* a indiqué que l’amplitude du *groupe de répétitions 1* était plus négative celles des groupes de répétitions 2 et 3 (figure 2.12)
- L’interaction [groupe de répétitions \times condition de mouvement] était significative pour l’amplitude de la cP2 ($F[2, 28] = 16.73$; $p < 0.001$) : un test *post-hoc* a indiqué que, pour *SM* et *DM*, l’amplitude du *groupe de répétitions 1* était plus négative que celles des groupes de répétitions 2 et 3. (figure 2.9).

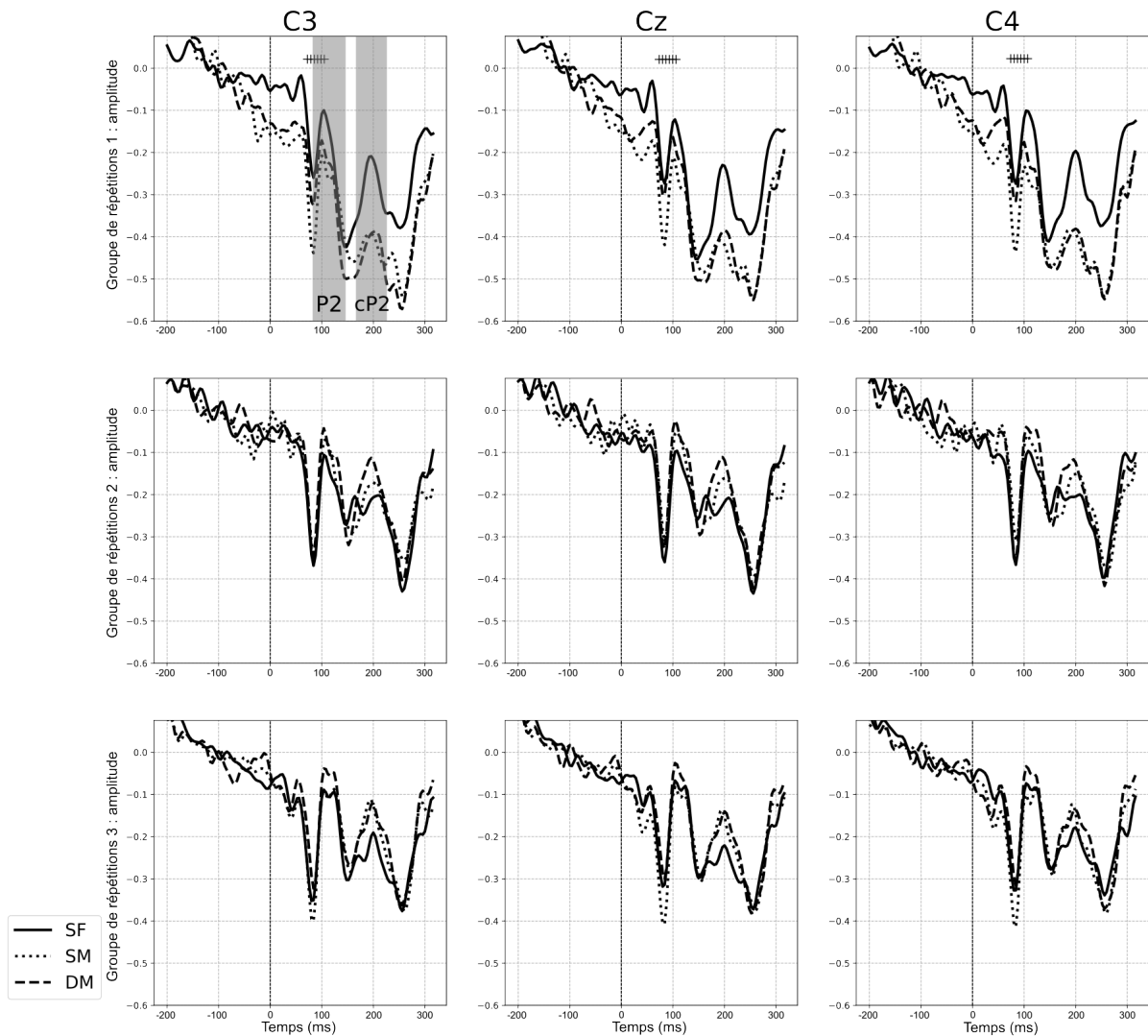


FIGURE 2.11 – ERPs (moyennés sur tous les sujets) comparant les 3 conditions de mouvement pour chaque groupe de répétitions séparément, sur les 3 électrodes centrales. Unités d'amplitude arbitraires. Les zones grises montrent les fenêtres temporelles à partir desquelles ont été extraites les valeurs de latence et d'amplitude des composantes P2 et cP2. Les "+" indiquent les points temporels où des différences significatives entre les conditions de mouvement *SF* et *SM* ont été trouvées par l'analyse de cluster.

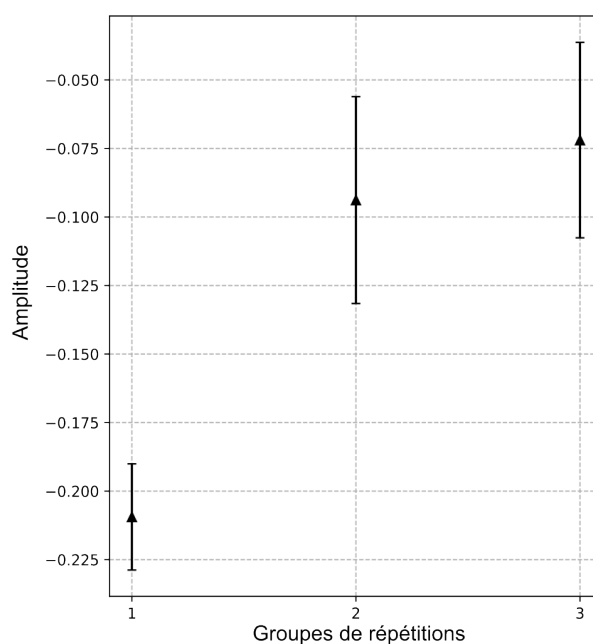


FIGURE 2.12 – Moyenne sur tous les sujets de l’amplitude des pics, normalisée par *z-transform*, pour la composante P2, sur Cz, par groupe de répétitions, conditions de mouvement regroupés. Les barres d’erreur indiquent l’intervalle de confiance à 95%. Unités d’amplitude arbitraires.

Une tendance similaire a été observée dans l’analyse "data-driven", comme le montre la figure 2.13. Les clusters significatifs étaient les suivants :

- Entre 132 et 176 ms, pour SF, l’amplitude du groupe de répétitions 1 était plus négative que celle du groupe de répétition 2 (peak $t = -0.29$, $p = 0.005$).
- Entre 0 et 32ms, pour SM, l’amplitude du *groupe de répétitions 1* était plus négative que celle du *groupe de répétition 2* (pic $t = -2.9$, $p < 0.001$).
- Entre 104 et 316 ms, pour SM, l’amplitude du *groupe de répétitions 1* était plus négative que celle du *groupe de répétition 3* (pic $t = -2.1$, $p < 0.001$).
- Entre 104 et 308 ms, pour DM, l’amplitude du *groupe de répétitions 1* était plus négative que celles des *groupes de répétitions 2* (pic $t = -1.4$, $p < 0.001$) et 3 (pic $t = -1.7$, $p < 0.001$).

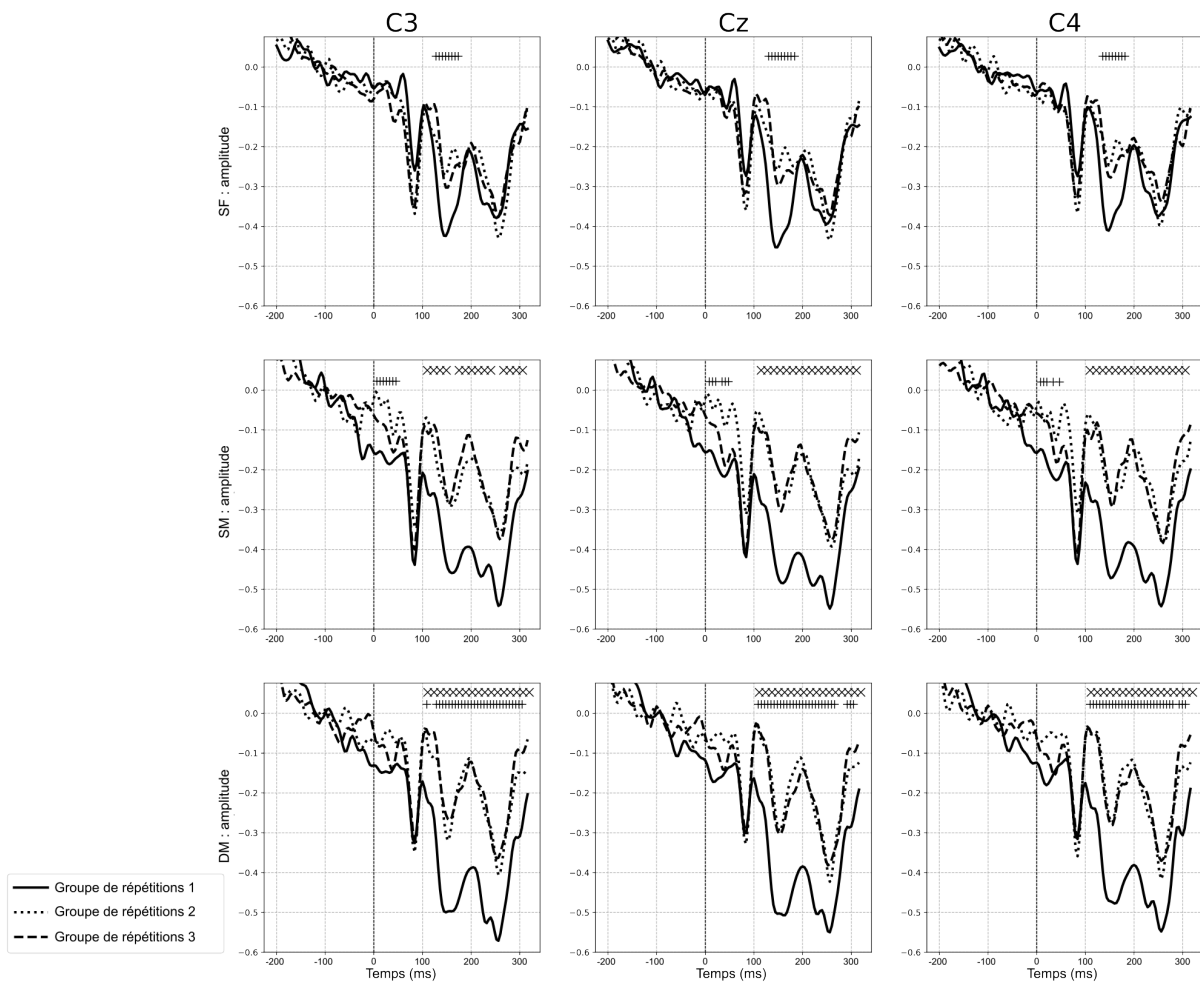


FIGURE 2.13 – ERP comparant les 3 groupes de répétitions pour chaque conditions de mouvement séparément, sur les 3 électrodes centrales. Unités d'amplitude arbitraires. Les "+" indiquent les points temporels où des différences significatives entre les *groupes de répétitions 2 et 1* ont été trouvées par l'analyse de cluster. Les "X" indiquent les points temporels où des différences significatives entre les *groupes de répétitions 3 et 1* ont été trouvées par l'analyse de cluster.

2.4 Discussion

Dans cette étude, une expérience comportementale et une expérience EEG ont été menées afin de trouver des corrélats électrophysiologiques de l'externalisation perçue à l'écoute de stimuli vocaux. Il était attendu que les stimuli dans la condition de mouvement *DM* (head-tracking actif, avec mouvement de tête) soient plus externalisés que les stimuli dans les autres conditions de mouvement [150, 151]. La littérature suggérait également que les composantes N1 et P2 de l'ERP générées par les stimuli dans la condition *DM* auraient une plus grande amplitude et une latence plus courte que celles générées dans les autres conditions de mouvement [17, 168].

2.4.1 Constats comportementaux

Comme dans les études précédentes [150, 151], des améliorations significatives dues aux mouvements de tête ont été observées pour la plupart des sujets, avec des amplitudes très variables d'un sujet à l'autre. La faible proportion de sujets ne rapportant aucune amélioration d'externalisation après les mouvements de tête avec head-tracking (3 sur 15 dans la présente étude) est également comparable aux études précédentes.

Les analyses statistiques ont montré que la condition avec mouvement de tête et head-tracking actif (*DM*) induisait une meilleure externalisation que les conditions sans head-tracking (*SF* et *SM*). Ce résultat est similaire à celui d'études antérieures qui ont examiné la persistance de l'effet des mouvements de tête [150, 151]. Cependant, l'amélioration d'externalisation induite par le mouvement de tête visible dans la présente étude est moins importante ce qui aurait pu être attendu au regard des études précédentes. Dans les études de Hendrickx *et al.* les sujets devaient évaluer l'externalisation de stimuli immédiatement après un mouvement de tête avec head-tracking, ou sans mouvement de tête pour plusieurs angles d'azimut. Pour un angle de 0°, l'amélioration d'externalisation induite par les mouvements de tête était de +38% [150] et pouvait monter jusqu'à +73% [151]. Dans la présente étude, l'amélioration d'externalisation constatée est de l'ordre de +15%. Cette comparaison est à prendre avec prudence, car, dans les études de Hendrickx *et al.* l'échelle utilisée était une échelle discrète à 6 points et les pourcentages d'amélioration discutée ici ont été calculé à partir des résultats obtenus sur l'échelle discrète. Cette moins bonne externalisation constatée dans la présente étude pourrait s'expliquer par la nature du stimulus utilisé. Le stimulus pendant le mouvement de tête était une phrase continue et naturelle, tandis que le stimulus pendant l'évaluation de l'externalisation était constitué de la répétition du pseudo-mot ("boru") conçu pour générer des ERPs clairs et exploi-

tables. Bien que récité par la même voix que la phrase, ce train de pseudo-mots était peu naturel et constituait une rupture de continuité avec la phrase précédente, créant ainsi une rupture de cohérence au sein de l'extrait. En effet, plusieurs sujets ont signalé que, même s'ils avaient une bonne externalisation pendant le mouvement de tête (donc pendant la phrase naturelle), l'externalisation chutait de manière significative lorsque le train de mots répétés commençait.

Comme décrit dans la section 1.4.1, les conditions d'écoute réalistes et cohérentes par rapport aux attentes de l'auditeur favorisent systématiquement une bonne externalisation. Il est donc possible que l'incohérence entre le train de mots répétés et la phrase naturelle ait favorisé une baisse d'externalisation expliquant les résultats moins élevés que dans les études précédentes. Cependant, peu d'études ont étudié l'influence de la nature du son et de sa cohérence par rapport aux attentes de l'auditeur sur l'externalisation [121, 137], et d'autres études seraient donc nécessaires pour vérifier cette hypothèse.

2.4.2 Constats EEG

Les résultats de l'expérience EEG peuvent être résumés comme suit :

- La condition avec mouvement de tête et head-tracking inactif (*SM*) a induit une amplitude plus négative de la composante N1 que la condition avec mouvement de tête et head-tracking actif (*DM*) et que la condition sans mouvement de la tête (*SF*).
- Les conditions avec mouvement de tête (*SM* et *DM*) ont induit une amplitude plus négative sur les composantes après 100 ms que la condition sans mouvement la tête (*SF*).
- Les stimuli présentés au milieu et à la fin du train de répétitions ont eu une amplitude réduite de toutes les composantes par rapport aux stimuli présentés au début du train.

Suppression due à la répétition

Le fait que le *groupe de répétitions 1* ait, après 100 ms, une amplitude plus négative que les *groupes 2* et *3* provient probablement d'un effet de suppression dû à la répétition du mot "boru" comme décrit dans la littérature (*e.g.* [195-197]). Lorsqu'une stimulation est répétée, on observe une diminution de l'activité cérébrale liée à cette stimulation entraînant une diminution de l'amplitude des ERPs associés. Cet effet de suppression peut être observé dès la première répétition d'une stimulation et s'accroît avec le nombre de répétitions. Lorsqu'une stimulation différente intervient, cet effet est "réinitialisé" et la stimulation répétée génère de nouveau des ERPs avec la même amplitude que lors de la toute première stimulation. On peut observer cet

effet sur les ERPs de la présente étude : le premier "boru" de chaque train a généré l'activité électrique cérébrale la plus importante, et chaque "boru" suivant dans le même train a généré une activité de plus en plus faible.

Considérations sur les artefacts de mouvement potentiels

Les lignes de base des ERP dans les figures 2.10, 2.11 et 2.13 montrent une dérive négative pour toutes les conditions. Cette dérive négative peut également être observée sur les données EEG brutes, en particulier après les mouvements de tête. Une inspection visuelle des epochs et des données EEG brutes a révélé que cette dérive lente semble être composée de deux artefacts de dérive négative distincts. Tout d'abord, une dérive lente négative globale qui apparaît sur tous les epochs, indépendamment des mouvements de tête. Deuxièmement, une dérive négative plus forte, probablement due à un artefact électrophysiologique lié au mouvement de tête lui-même (muscles faciaux ou mouvement du câble). Ces deux dérives négatives distinctes peuvent être facilement observées sur les graphiques de la figure 2.11. En examinant les lignes de base des *groupes de répétitions 2* et *3*, on peut observer la dérive lente négative globale (toutes les conditions de mouvement diminuent d'environ 0,1 unité). En examinant les lignes de base *du groupe de répétitions 1*, la même dérive lente globale peut encore être observée sur la condition de mouvement *SF*. Cependant, on peut voir que les conditions de mouvement *SM* et *DM* (avec mouvements de tête) montrent une diminution négative plus forte (d'environ 0,25 unité), probablement due à l'effet des mouvements de tête. Le changement d'amplitude provoqué par la dérive lente globale semble plutôt modéré et avoir affecté toutes les conditions de la même manière. Il est donc très probable que cette dérive lente globale n'ait pas eu d'influence sur les différences statistiques entre les conditions. En revanche, la dérive négative plus forte, probablement due aux mouvements de tête, a affecté différemment les conditions et pourrait avoir eu une influence sur les résultats statistiques.

Dans la condition *SM* en particulier, l'analyse par clusters a révélé une différence précoce entre les *groupes de répétitions 1* et *2*, qui pourrait être due à cette ligne de base bruitée, alors qu'aucun effet précoce de ce type n'a été trouvé dans la condition *DM*. Néanmoins, il est possible que la ligne de base de la première répétitions du "boru" ait été décalée en raison de la dérive lente due aux mouvements de tête.

Afin de vérifier si cette dérive lente a eu ou non un effet sur les différences entre conditions de mouvement, plusieurs analyses ont été effectuées :

- Des tests statistiques supplémentaires ont été réalisés sur les données filtrées avec des filtres passe-haut avec des fréquences de coupure plus élevées que la fréquence de coupure

standard de 0,1 Hz. Des fréquences de coupure de 1 Hz, 2 Hz et 3 Hz ont été testées. Avec des fréquences de coupure aussi élevées, la dérive négative était très diminuée (1 Hz et 2 Hz) ou même complètement supprimée (3 Hz) et des résultats similaires à ceux obtenus pour les données filtrées à 0,1 Hz ont été obtenus (voir l'annexe C pour le détail de ces analyses).

- Ensuite, les signaux issus des électrodes des mastoïdes ont été inclus dans l'analyse par cluster. Aucun cluster significatif n'a été trouvé. Ce résultat indique que la dérive négative du mouvement de tête n'a pas induit de différence significative entre conditions pour ces électrodes. Or, les électrodes mastoïdiennes étaient affectées par la dérive lente des mouvements de tête de la même façon que les autres électrodes, mais ne présentaient que très peu d'activité par ailleurs. Ainsi, comme la dérive lente des mouvements de tête n'a pas affecté les différences entre conditions au niveau des électrodes mastoïdiennes, il n'y a pas de raison qu'elle ait affecté les différences entre conditions au niveau des autres électrodes.
- Pour finir, une analyse sans la première répétition du "boru" de chaque train a été effectuée et un schéma de résultats similaires aux résultats de l'analyse incluant la première répétition du "boru" a été trouvé, avec des différences entre les conditions moins marquées.

Dans l'ensemble, bien que l'influence des artefacts dus aux mouvements de tête ne puisse pas être totalement exclue, les analyses complémentaires suggèrent que les effets constatés dans les ERP sont imputables aux groupes de répétitions et aux conditions expérimentales *per se*.

Effets du mouvement de tête

Pour les composantes après la N1, le fait que l'amplitude du *groupe de répétitions 1* pour les conditions de mouvement *SM* et *DM* était plus négative que pour la condition *SF* pourrait être un effet de traitement neuronal lié à la proximité temporelle du mouvement de la tête. Plus précisément, le fait que les sujets venaient d'arrêter de bouger la tête aurait pu modifier le traitement neuronal du stimulus sonore qui suivait le mouvement. Si l'on considère cette possibilité, cette modification du traitement ne peut pas être attribuée à un changement d'externalisation puisque, dans l'expérience comportementale, la condition de mouvement *DM* a induit une différence d'externalisation par rapport à *SF*, alors que *SM* n'a pas induit de telles différences. Il est possible que cet effet soit lié à l'activité du système vestibulaire, qui peut être induite par le mouvement de tête et qui se poursuit même après l'arrêt du mouvement [198, 199].

Le fait que, pour la N1, la condition *SM* ait une amplitude plus négative que les conditions *SF* et *DM* peut provenir d'une différence de référentiel spatial. Dans les deux conditions *SF* et *DM*, le son était cohérent par rapport à l'environnement extérieur au sujet tout au long de la lecture, comme il le serait en condition d'écoute naturelle. En effet, dans la condition *SF*, il n'y avait pas de mouvement de tête et le son était constamment devant les sujets. Dans *DM*, les sujets effectuaient un mouvement de tête, mais le head-tracking permettait au son de rester fixe par rapport au monde extérieur, *i.e.* le son était cohérent dans un référentiel "terrestre" comme le serait une source sonore réelle en condition d'écoute naturelle. Cependant, pour *SM*, lorsque les sujets effectuaient le mouvement de tête, le son se déplaçait avec le casque et devenait incohérent avec l'environnement extérieur, tout en étant cohérent dans le référentiel de la "tête de l'auditeur" (le son n'était plus "fixe" comme l'aurait été une source sonore réelle). Cette différence de référentiel entre *SM* et les deux autres conditions pourrait avoir induit des différences de traitement neuronal et de perception des stimuli, comme cela a été rapporté dans des études antérieures [185, 200]. Ces études ont obtenu des ERPs différents en comparant des stimuli sonores qui étaient cohérents soit dans un référentiel "terrestre", soit dans un référentiel "tête de l'auditeur".

Comparaison avec les résultats comportementaux

Les résultats comportementaux ont montré que la condition de mouvement *DM* induisait une meilleure externalisation que les conditions de mouvement *SF* et *SM*, mais aucune différence entre la condition de mouvement *DM* et les conditions de mouvement *SF* et *SM* n'a pu être trouvée dans les résultats EEG. Cette différence peut s'expliquer par le fait que l'attention des sujets était dirigée vers l'externalisation pendant l'expérience comportementale, mais pas pendant l'expérience EEG. En effet, dans l'expérience EEG, les sujets avaient simplement pour instruction d'écouter le son (écoute passive) alors que dans l'expérience comportementale, les sujets avaient pour instruction d'évaluer l'externalisation (écoute active). Étant donné que l'engagement que demande la tâche (écoute passive ou active) peut avoir une influence sur les ERPs [201], la différence d'instruction entre les expériences comportementales et EEG pourrait expliquer l'absence d'un corrélat neuronal de l'externalisation dans les ERPs résultants. Une autre explication possible est que les différences d'externalisation entre les conditions de mouvement auraient pu être trop faibles pour être observées dans les ERPs. En effet, les différences de score d'externalisation entre les conditions de mouvement dans l'expérience comportementale étaient modérées (≈ 45 points pour *SF* et *SM* contre ≈ 60 points pour *DM* soit ≈ 15 points de différence). Il est donc possible que la différence d'activité neuronale induite lors de

la condition de mouvement *DM* ait été trop subtile pour pouvoir être observée sur les ERPs

Comparaison avec les études précédentes

Deux études comparant les ERPs générés par des stimuli plus ou moins externalisés ont été présentées dans la section 1.4.3. Les résultats de ces études ont servi de point de départ pour la mise en place de la présente étude.

Dans l'étude de Getzmann et Lewald [17], des différences entre des stimuli externalisés (écoute en champ libre et restitution binaurale non individualisée) et des stimuli non externalisés (restitution stéréophonique au casque) ont été constatées sur la latence de la N1 et de la P2 dans une tâche où les sujets devaient identifier les mouvements d'une source. Dans la présente étude, des stimuli externalisés (*DM*) ont été comparés à des stimuli moins externalisés (*SF* et *SM*) dans le cadre d'une écoute passive. Aucune différence entre *DM* et *SF-SM* n'a été trouvée dans les ERPs.

Cette différence avec les résultats précédents pourrait provenir d'une différence d'attention du sujet (identification de mouvement vs. écoute passive), mais pourrait également provenir de la nature des stimuli. En effet, dans l'étude de Getzmann et Lewald, outre les différences d'externalisation, il existait une différence de dispositifs de restitution (champ libre vs restitution au casque) entre les types de stimuli qui n'existait pas dans la présente étude. Cette différence de dispositifs de restitution implique des différences d'informations visuelles et tactiles (casque sur la tête) qui auraient pu avoir une influence sur les ERPs. De plus, dans l'étude de Getzmann et Lewald, la restitution binaurale a été comparée à la restitution stéréophonique au casque, ce qui implique des différences importantes de contenu fréquentiel qui n'existaient pas dans la présente étude et qui auraient pu également influencer les ERPs.

Dans l'étude de Palomaki *et. al.* [168] utilisant la MEG, des différences entre des stimuli plus ou moins externalisés (restitution binaurale individualisée vs. restitution binaurale non individualisée vs. restitution stéréophonique au casque) ont été constatées sur l'amplitude de la N1m dans une tâche de localisation, alors qu'aucun effet de ce type n'a été constaté dans la présente étude. Cette différence pourrait également être due à une différence d'attention du sujet (tâche de localisation vs. écoute passive) ou à des différences de contenu fréquentiel entre la restitution binaurale et la restitution stéréophonique plutôt qu'à des différences d'externalisation.

Ainsi, dans les études précédentes, les différences d'ERPs entre les stimuli plus ou moins externalisés pouvaient être causées par d'autres facteurs que l'externalisation. L'absence de telles différences entre les stimuli dans la présente étude pourrait expliquer pourquoi aucune différence d'ERPs propre à l'externalisation n'a été trouvée.

2.5 Conclusion et perspectives

Dans cette étude, une expérience comportementale et une expérience EEG ont été menées dans le but de mettre en lumière des corrélats neuronaux de l'externalisation à l'aide d'ERPs. Étudier l'externalisation d'un point de vue objectif, avec des ERPs, est complexe, car il est nécessaire de comparer des stimuli comportant uniquement des différences d'externalisation, sinon les potentielles différences observées sur les ERPs ne peuvent pas être catégoriquement attribuées à des différences d'externalisation. Or, une manipulation de l'externalisation passe généralement par une manipulation d'autres paramètres tels que le contenu spectral ou le système de restitution. Pour contourner ce problème dans la présente étude, différents degrés d'externalisation ont été obtenus en faisant précéder les mesures EEG et comportementales 1) de mouvements de tête avec head-tracking, 2) de mouvements de tête sans head-tracking et 3) d'aucun mouvement de tête. Ainsi, lors de la mesure EEG, les stimuli pouvaient être plus ou moins externalisés tout en étant "acoustiquement identiques". La tâche lors de l'expérience comportementale consistait en une évaluation subjective de l'externalisation, tandis que l'expérience EEG consistait en une simple écoute passive. Hormis cette différence de tâche, les deux expériences étaient similaires afin de pouvoir mettre en relation leurs résultats.

L'exécution d'un mouvement de tête, que le head-tracking soit actif ou non, a augmenté l'amplitude des composantes ERPs après 100 ms, ce qui pourrait être un effet de traitement neuronal lié à la proximité temporelle du mouvement de la tête, pouvant potentiellement être attribué au système vestibulaire. En outre, le fait de bouger la tête alors que le head-tracking était inactif a augmenté l'amplitude de la composante N1, qui pourrait être un marqueur d'une rupture de cohérence dans le son.

D'autre part, contrairement à ce qui était attendu des études précédentes, aucun corrélat neuronal spécifique de l'externalisation n'a été trouvé dans les ERPs, ce qui pourrait être dû à plusieurs choses. Tout d'abord, il est possible qu'un potentiel marqueur de l'externalisation ne soit pas observable sur les potentiels étudiés dans la présente étude. Il pourrait être intéressant d'appliquer la méthode développée afin de comparer des stimuli plus ou moins externalisés et "acoustiquement identiques" dans un protocole permettant d'étudier des composantes plus tardives. Ceci nécessiterait d'utiliser des ISI plus longs et d'adapter le test pour avoir une longueur acceptable. Une autre possibilité serait d'étudier les composantes générées par des protocoles "standard-déviant" telles que la MMN ou la P300. La MMN peut être générée sans que la déviance soit consciemment détectée par le sujet, alors que la P300 apparaît dans des tâches de détection de la déviance (*cf.* section 1.1.3). L'étude de ces composantes pourrait permettre de

définir si l'externalisation est liée à l'attention. Autrement dit, est-ce qu'il est nécessaire d'attirer l'attention des sujets sur l'externalisation pour observer des différences d'externalisation ? Par ailleurs, il est possible que les différences observées sur les composantes ERPs entre les stimuli plus ou moins externalisés dans les études précédentes n'aient pas été causées par l'externalisation, mais plutôt par des différences de contenu spectral ou de système de restitution, ou alors que les différences d'externalisation entre les stimuli de la présente étude aient été trop faibles pour être observées sur les ERPs.

En effet, l'expérience comportementale a permis de vérifier que les mouvements de tête avec head-tracking augmentaient l'externalisation, mais cette augmentation était modérée comparée à ce qui aurait pu être attendu au regard des études précédentes utilisant un protocole similaire. Deux explications possibles de cette différence peuvent être émises :

1. La constitution peu naturelle des stimuli, conçus pour générer efficacement des ERPs, aurait pu entraîner une rupture de cohérence et de réalisme et ainsi favoriser une diminution de l'externalisation.
2. Comme les stimuli post-mouvement étaient plus longs que dans les études précédentes, l'amélioration d'externalisation obtenue immédiatement après les mouvements de tête aurait pu diminuer au cours de l'exposition statique (*i.e.* sujet immobile) à ces stimuli.

Cependant, peu d'études ont exploré l'influence de la nature du son et de sa cohérence par rapport aux attentes de l'auditeur sur l'externalisation et il n'existe pas de résultats dans la littérature permettant de valider ces hypothèses. Ce constat a mené à la réalisation d'une deuxième étude visant à comparer l'externalisation de stimuli plus ou moins longs, naturels et cohérents.

ÉTUDE DE LA PERSISTANCE DE L'EFFET DES MOUVEMENTS DE TÊTE SUR L'EXTERNALISATION EN RESTITUTION BINAURALE DYNAMIQUE

Ce chapitre présente la deuxième étude menée dans le cadre de la thèse visant à caractériser l'évolution de l'externalisation après l'arrêt de mouvements de tête avec head-tracking.

3.1 Introduction

Les résultats de l'étude présentée dans le chapitre 2 ont montré que des mouvements de tête avec head-tracking permettaient d'améliorer l'externalisation après l'arrêt des mouvements. Les résultats de cette étude ont également soulevé la question de l'influence de la durée et de la nature des sources sonores sur cette amélioration d'externalisation. En effet, contrairement à ce qui aurait pu être attendu au regard des résultats des études précédentes [150, 151], l'amélioration d'externalisation présentée dans le chapitre 2 n'était pas substantielle, mais relativement modérée. Ceci pourrait être expliqué par la nature des stimuli. Dans les études précédentes, les stimuli diffusés après le mouvement de tête, pendant lesquels les sujets devaient évaluer l'externalisation, étaient relativement courts (3 s). De plus, les stimuli pendant et après les mouvements de tête étaient parfaitement cohérents (extraits de poème). En revanche, dans l'étude présentée dans la section 2, le stimulus diffusé après les mouvements de tête était plus long (train de répétition du mot "boru" d'une durée de 12 s) et n'était pas complètement cohérent avec le stimulus diffusé pendant le mouvement de tête (extrait de poème). Il est donc possible que :

- L'amélioration d'externalisation obtenue avec les mouvements de tête ait décliné au cours

des 12 s après l'arrêt du mouvement

- L'amélioration d'externalisation obtenue avec les mouvements de tête ait chuté, car le stimulus post-mouvement était incohérent avec le stimulus pendant le mouvement et donc ne correspondait pas à ce à quoi s'attendaient les sujets (*cf.* section 1.4.1).

La section 1.4.1 a décrit de quelle façon l'utilisation de l'effet des mouvements de tête pour améliorer l'externalisation pourrait permettre de s'affranchir du problème d'internalisation liée à la restitution binaurale non-individualisée. Pouvoir obtenir une externalisation satisfaisante en restitution binaurale non-individualisée permettrait de faciliter grandement la diffusion de contenus binauraux. Pour cela, il est nécessaire que l'effet des mouvements de tête soit robuste dans le temps et permette d'améliorer l'externalisation de toutes les sources d'une scène sonore, pas uniquement des sources entendues pendant les mouvements de tête.

Pour répondre à ces problématiques, trois expériences ont été réalisées et sont présentées dans ce chapitre.

La première expérience a permis d'explorer l'externalisation d'une source sonore diffusée immédiatement après des mouvements de tête ou après des mouvements de tête suivis d'une période de 31 s pendant laquelle les sujets devaient garder la tête immobile (appelée partie *distraction*). La source diffusée pendant les mouvements de tête et pendant l'évaluation de l'externalisation (appelée source *initiale + évaluation*) était toujours la même. En revanche, la source lors de la partie *distraction* (appelée *distracteur*) était soit de même nature que la source *initiale + évaluation* soit de nature différente. Le but était de voir si l'externalisation était aussi élevée 31 s après l'arrêt des mouvements de tête qu'immédiatement après les mouvements et si la nature du *distracteur* diffusée pendant ces 31 s avait une influence sur celle-ci.

La deuxième expérience a exploré l'externalisation d'une source diffusée après des mouvements de tête suivis d'une partie *distraction* de durée et de nature variable. La source *initiale + évaluation* était toujours la même. En revanche, la partie *distraction* contenait deux sources diffusées l'une après l'autre. Le but était d'étudier en détail de quelle façon l'amélioration d'externalisation due aux mouvements de tête évolue au cours du temps après l'arrêt des mouvements quand le *distracteur* contient plusieurs sources.

Pour finir, la troisième expérience a exploré l'externalisation d'une source diffusée immédiatement après des mouvements de tête, mais de nature différente de la source diffusée pendant les mouvements de tête. Le but était de vérifier si l'amélioration d'externalisation obtenue lors de mouvements de tête effectués sur une source donnée permettait d'externaliser une source différente après l'arrêt des mouvements.

3.2 Expérience 1

Le but de cette première expérience était d'explorer si l'amélioration d'externalisation due aux mouvements de tête avec head-tracking à l'écoute d'une voix de femme pouvait persister pendant un temps relativement long après que les sujets ont arrêté de bouger la tête.

Une partie *distraction* était insérée immédiatement après les mouvements de tête. Le *distrupteur* pouvait être soit la même voix de femme que celle diffusée lors du mouvement de tête, soit une voix d'homme, soit du silence, soit du bruit rose de façon à explorer une potentielle influence de la nature du *distrupteur* sur la persistance de l'externalisation. Après la partie *distraction*, la voix de femme initiale était diffusée de nouveau et les sujets devaient évaluer l'externalisation lors de ce dernier extrait de voix de femme. Cette procédure a été établie pour vérifier si le mouvement de tête effectué en début de stimulus permettait toujours d'améliorer l'externalisation de la voix de femme, malgré la présence de la partie *distraction*. En d'autres termes, si le sujet bouge la tête en écoutant une source *A*, puis écoute (avec la tête immobile) une source *B*, puis écoute de nouveau la source *A* (avec la tête immobile), est-ce qu'il bénéficie toujours de l'amélioration d'externalisation due aux mouvements de tête effectués lors de la première écoute de la source *A* ?

Étant donné que l'étude présentée dans le chapitre 2 a obtenu des scores d'externalisation plus faibles que ce qui aurait pu être attendu au regard des études précédentes (qui ont utilisé des stimuli post-mouvement plus courts) [150, 151], on peut s'attendre à ce que l'externalisation des stimuli diffusés après des mouvements de tête suivis d'une partie *distraction* soit plus faible que celle des stimuli diffusés immédiatement après des mouvements de tête (hypothèse **H1**). En outre, on peut s'attendre à ce que l'externalisation soit plus importante lorsque la source *distrupteur* et la source *initiale + évaluation* sont similaires que lorsqu'elles sont différentes (hypothèse **H2**). En effet, lorsque les sources sont différentes, la cohérence globale de l'extrait sonore est rompue et l'extrait est moins susceptible de correspondre aux attentes de l'auditeur.

3.2.1 Matériel et méthode

Participants

Vingt sujets (16 hommes, 4 femmes ; âge moyen 26 ; écart-type 3,1 ; intervalle 21-33 ans) ont participé à l'étude. Aucun d'entre eux n'avait participé à l'étude présentée dans le chapitre 1. Tous ont déclaré avoir une audition normale. L'étude était conforme au Code d'éthique de l'Association médicale mondiale (Déclaration d'Helsinki), publié dans le *British Medical Jour-*

nal (18 juillet 1964) et a été approuvée par le Comité d'éthique institutionnel d'IMT Atlantique (enregistré sous le numéro IRB00013722 auprès du bureau américain de protection des humains dans la recherche [171]). En outre, tous les sujets ont donné leur consentement éclairé écrit pour participer à l'étude.

Stimuli

Il y avait six stimuli au total. Quatre d'entre eux étaient divisés en trois parties :

1. Un extrait de 5,2 s du poème "l'Albatros" (v.1-2) de Charles Baudelaire lu par une femme (appelé partie *initiale* du stimulus);
2. un *distracteur* de 31 s;
3. un extrait de 5,5 s du poème "l'Albatros" (v.15-14) de Charles Baudelaire lu par la même femme (appelé partie *évaluation* du stimulus).

Deux stimuli supplémentaires, composés uniquement des parties 1) et 3), servaient de référence et d'ancre basse (appelés respectivement *ref* et *ancre*). Le contenu audio de ces deux stimuli était identique. Ils étaient néanmoins considérés comme des stimuli différents, car, dans *ref*, les sujets devaient effectuer un mouvement de tête pendant la partie *initiale* alors que, dans *ancre*, ils devaient rester statiques (*i.e.* ne pas bouger la tête). La comparaison du score d'externalisation obtenu pour ces deux stimuli a permis de quantifier l'amélioration de l'externalisation induite par les mouvements de tête.

Il y avait quatre *distracteurs* différents :

- un autre extrait du même poème lu par la même femme (le stimulus contenant ce *distracteur* est nommé *fem*)
- le même extrait de poème que pour *fem* mais lu par un homme (le stimulus contenant ce *distracteur* est nommé *hom*)
- un silence (le stimulus contenant ce *distracteur* est nommé *sil*)
- un bruit rose (le stimulus contenant ce *distracteur* est nommé *rose*)

La figure 3.1 montre une représentation visuelle des stimuli.

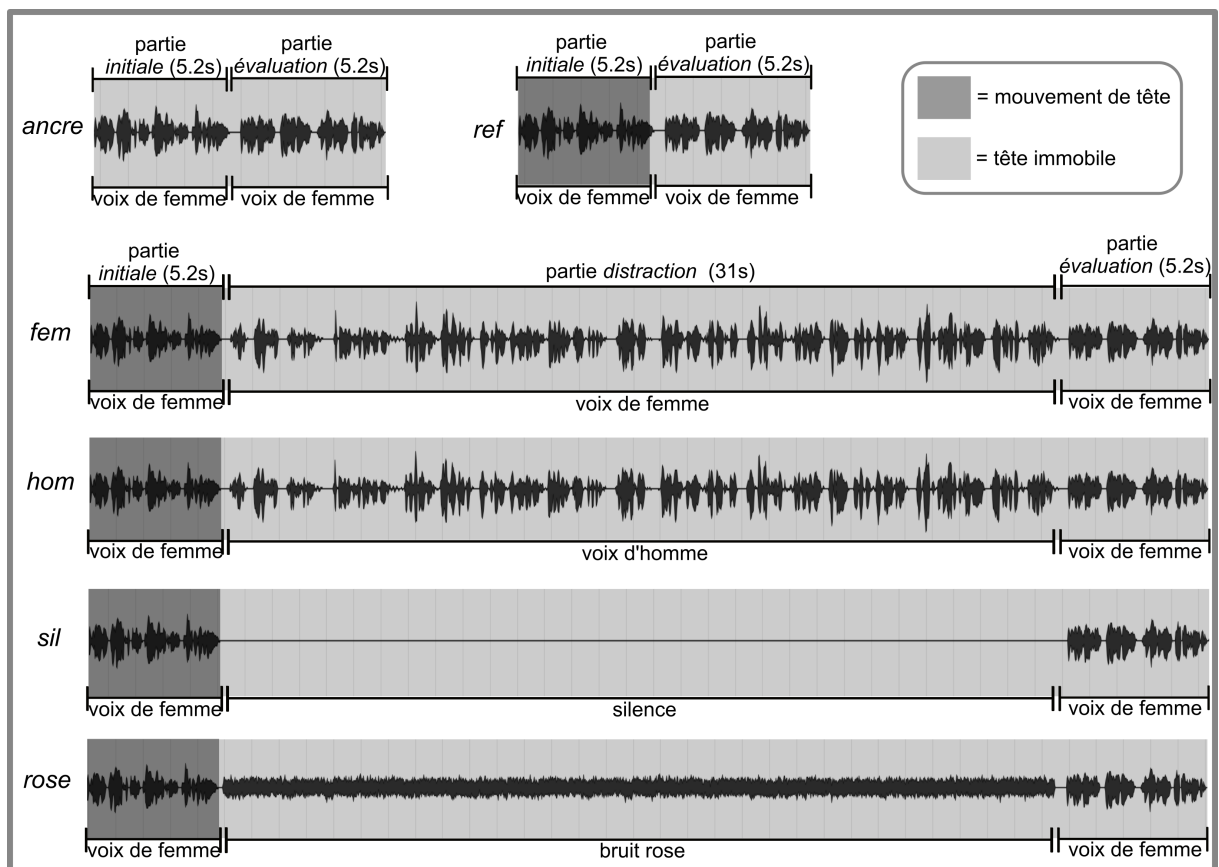


FIGURE 3.1 – Représentation des stimuli utilisés dans l'expérience 1.

Enregistrement et traitement des stimuli sonores

Le son a été enregistré par une locutrice et un locuteur dans une petite chambre anéchoïque avec un microphone Neumann U87 et une interface audio RME babyface, connectée à un ordinateur Windows utilisant Protools comme logiciel d'enregistrement. La voix de la locutrice a été enregistrée en première. Le locuteur a pu ainsi écouter les enregistrements de voix de femme pour imiter, dans une certaine mesure, la prosodie de la locutrice, tout en gardant une élocution naturelle. Le but était que les prosodies des stimuli de voix de femme et de voix d'homme soient les plus similaires possibles pour limiter un potentiel effet de la prosodie sur l'externalisation. Le bruit rose a été généré dans Protools avec une résolution de 24 bits et une fréquence de 48 kHz.

Une réverbération réaliste et pouvant s'adapter aux mouvements induits par le head-tracking a été générée à l'aide de la même réponse impulsionnelle ambisonique que celle utilisée dans le chapitre 2. Seule la partie réverbérée résultant de cette convolution a été conservée et mixée au son direct anéchoïque original. L'ajout de réverbération était une précaution pour éviter que les stimuli soient tous perçus complètement internalisés (*cf.* section 1.4.1), ce qui aurait empêché d'obtenir des différences d'externalisation inter-stimuli. Pendant l'expérience, la réverbération ambisonique et le son direct étaient binauralisés indépendamment, puis mixés en temps réel à l'aide du logiciel "SPAT" de l'IRCAM [177].

Les stimuli étaient binauralisés à l'aide de HRTFs non-individualisées. Pour minimiser un potentiel effet d'apprentissage d'un set de HRTFs, trois sets différents étaient utilisés alternativement : les sets n°1004, n°1068, et n°1100 de la bibliothèque de l'IRCAM "BILI" [180].

Appareils et système de reproduction audio

Les expériences ont été menées dans une salle insonorisée à double paroi et faiblement éclairée. Le système de reproduction audio ainsi que le système de head-tracking étaient identiques à l'étude présentée dans le chapitre 1. La lecture du son, l'affichage de l'interface et l'acquisition des données étaient contrôlés à l'aide d'un logiciel implémenté en Max/MSP sur un ordinateur Windows connecté à une carte son RME Babyface. Le module "SPAT" utilisé pour la binauralisation était également implémenté dans Max/MSP. Les informations visuelles nécessaires au sujet étaient affichées sur un écran.

Procédure

Pendant la partie *initiale*, les sujets devaient effectuer un mouvement de tête de droite à gauche entre $\pm 60^\circ$ sur le plan horizontal, sauf pour le stimulus *ancree* où ils devaient garder la tête immobile dans une position naturelle. Pendant la partie *distraction* et pendant la partie *évaluation*, les sujets devaient toujours garder la tête immobile.

Après la fin de chaque stimulus, les sujets devaient évaluer l'externalisation perçue pendant les dernières 5,5 s (*i.e.* pendant la partie *évaluation*), en déplaçant un curseur sur une échelle pseudocontinue à 100 points affichée sur un écran d'ordinateur, dont les extrémités étaient intitulées "source complètement internalisée" (0) et "source parfaitement externalisée" (100). La même description de l'externalisation que celle utilisée dans l'étude du chapitre 2 était donnée aux sujets avant de commencer le test. Cette description était simple et brève afin que l'évaluation reste naturelle et aussi spontanée que possible. Le stimulus suivant commençait dès que les sujets validaient leur réponse.

Cette procédure a permis :

- De vérifier si les sources sonores diffusées immédiatement après des mouvements de tête (*ref*) étaient plus externalisées que les sources diffusées sans mouvement de tête (*ancree*).
- De comparer la persistance de l'effet des mouvements de tête sur l'externalisation des sources sonores diffusées immédiatement après le mouvement de tête (*ref*) vs. 36,5 s (31 s de partie *distraction* + 5,5 s de partie *évaluation*) après la fin du mouvement de tête (*fem, hom, sil, rose*).

Les différents types de *distracteurs* ont permis d'étudier si l'amélioration d'externalisation induite par les mouvements de tête sur la voix de femme persistait dans le temps :

- Lorsque la cohérence de l'extrait sonore était respectée (*fem*).
- Lorsque la cohérence de l'extrait sonore était légèrement rompue (*hom*).
- Lorsque la cohérence de l'extrait sonore était complètement rompue (*rose*).
- Lorsqu'il n'y avait aucune continuité sonore (*sil*).

L'objectif de cette expérience n'était pas d'examiner si l'amélioration d'externalisation induite par des mouvements de tête sur une source *A* peut aider à externaliser une source *B*, mais plutôt de vérifier dans quelle mesure l'amélioration d'externalisation induite par les mouvements de tête pour une source *A* persiste dans le temps une fois que le sujet a cessé de bouger la tête malgré des changements temporaires de source sonore dans le stimulus.

Avant l'expérience, tous les sujets devaient effectuer une phase d'entraînement pour s'habituer à la tâche et apprendre à effectuer correctement le mouvement de tête. Les mouvements de tête étaient enregistrés pendant l'expérience. Les six stimuli (*ref*, *ancr*, *fem*, *hom*, *sil*, *rose*) étaient testés dans un ordre aléatoire, différent pour chaque sujet. Pour minimiser un éventuel effet d'apprentissage d'un set de HRTFs, le set de HRTFs changeait entre les essais successifs (en alternant entre les 3 sets de HRTFs). Chaque combinaison de [stimulus × HRTF] a été répétée 3 fois pour un total de [6 stimulus × 3 HRTF × 3 répétitions] = 54 essais. La durée totale de l'expérience variait de 45 à 60 minutes.

3.2.2 Résultats

L'analyse des données de mouvements de tête a montré que les sujets se conformaient en moyenne très bien aux instructions données et qu'ils effectuaient correctement les mouvements. Les graphiques de mouvement, par sujet, sont présentés dans la figure A.2 de l'annexe A.

Une analyse de la distribution des scores d'externalisation a révélé que les données n'étaient pas normalement distribuées. Des méthodes d'analyse statistique non paramétriques ont donc été choisies. La figure 3.2 montre le score moyen d'externalisation des sujets pour chaque stimulus. Bien que des méthodes non paramétriques aient été utilisées, il a été choisi de montrer, sur toutes les figures de l'étude, les moyennes avec des intervalles de confiance, car cette représentation est plus lisible et est généralement préférée à une représentation plus adaptée aux méthodes non paramétriques, telle que les boîtes à moustaches.

Un test de Friedman effectué sur toutes les données a révélé un effet du stimulus ($\chi^2(5) = 32,2$; $p < 0,001$) et aucun effet des sets de HRTFs ou des répétitions ($p \gg 0,05$). Des tests de Wilcoxon ont été effectués pour comparer chaque paire de stimulus avec une correction FDR. Le FDR (*False Discovery Rate*) permet de quantifier le nombre de faux positifs parmi les résultats significatifs pour pouvoir les supprimer. Quand il y a un nombre important de comparaisons, cette méthode peut être plus adaptée que les méthodes Bonferroni ou Holm Bonferroni réputées trop strictes [202].

Les résultats ont montré que les scores d'externalisation étaient significativement plus élevés pour *ref* ($p < 0.001$), *hom* ($p < 0.001$), *sil* ($p < 0.001$) et *rose* ($p < 0.001$) que pour *ancr*. En outre, les scores d'externalisation étaient significativement plus élevés pour *ref* ($p < 0.001$), *hom* ($p = 0.013$), *sil* ($p = 0.044$) et *rose* ($p = 0.047$) que pour *fem*.

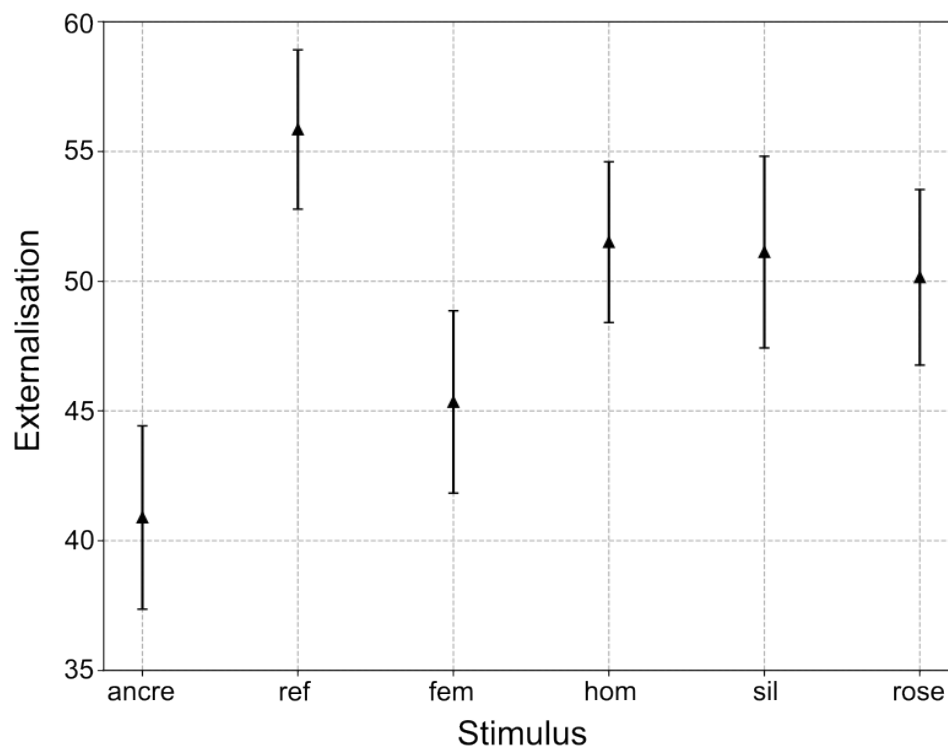


FIGURE 3.2 – Scores d'externalisation moyennés sur tous les sujets pour chaque stimulus. Les barres d'erreur indiquent l'intervalle de confiance à 95%.

Pour vérifier l'absence d'effet entre les comparaisons non significatives, des tests de Wilcoxon bayésiens [203] ont été effectués. Les résultats ont fourni une preuve modérée de l'absence d'effet [204] entre *ancree* et *fem* ($BF_{01} = 6.4$). Les résultats ont fourni des preuves fortes de l'absence d'effet entre *hom*, *sil* et *rose* ($10 < BF_{01} < 12$). Toutefois, les résultats n'ont pas permis de prouver l'absence d'effet entre *ref* et *hom* ou *sil* ou *rose* ($0,3 < BF_{01} < 3$).

En résumé, les résultats montrent deux groupes au sein desquels les stimuli partagent des scores d'externalisation similaires : *ancree* et *fem* ont les scores les plus bas, *ref*, *hom*, *sil*, et *rose* ont les scores les plus élevés.

3.2.3 Discussion

Comme prévu, les stimuli diffusés immédiatement après les mouvements de tête (*ref*) étaient plus externalisés que ceux diffusés sans mouvement de tête (*ancree*). On remarque néanmoins que les moyennes d'externalisation peuvent paraître modérées (entre 40 et 55 environ) de même que les différences significatives entre stimuli (environ 15 points de différence au maximum). Il pourrait sembler que la différence entre *ancree* et *ref* n'est pas particulièrement substantielle et donc que les mouvements de tête ne sont pas très efficaces pour améliorer l'externalisation immédiatement après les mouvements. Néanmoins, l'échelle continue utilisée ne comportait pas de labels intermédiaires et les sujets n'ont probablement pas tous utilisé l'échelle de la même façon. Il est compliqué de savoir quel niveau d'externalisation ces moyennes entre 40 et 55 points traduisent et si cet écart de 15 points entre conditions représentent une différence d'externalisation importante ou non. Il est cependant possible de comparer ces moyennes d'externalisation aux moyennes obtenues dans l'étude de Leclère *et al.* [121] utilisant la même échelle d'externalisation. Dans cette étude, une source vocale, située à 0° , avec une réverbération de salle de classe (comparable à la réverbération utilisée dans la présente étude) a obtenu une externalisation moyenne d'environ 38 en HRTFs non-individualisées et d'environ 50 en HRTFs individualisées. On remarque que cette différence d'externalisation est du même ordre de grandeur que la différence entre *ancree* et *ref* de la présente étude. L'amélioration d'externalisation due aux mouvements de tête est donc comparable au gain apporté par l'individualisation des HRTFs dans l'étude de Leclère *et al.* Les scores d'externalisation absolu obtenus dans cette étude (entre 35 et 50 points) sont également comparables aux scores obtenus dans la présente étude.

Le fait que le stimulus *fem* ait reçu des scores d'externalisation similaires à ceux du stimulus *ancree* indique que l'amélioration d'externalisation obtenue immédiatement après un mouvement de tête avec head-tracking s'effondre en moins de 36,5 s (31 s de partie *distraction* + 5,5 s de partie *évaluation*) après la fin du mouvement. Il convient de noter que cette tendance a été observée uniquement pour le stimulus *fem*, c'est-à-dire pour le seul *distracteur* qui était similaire à la source *initiale* + *évaluation*. Par conséquent, **H1** a été validé pour *fem* uniquement. Le protocole utilisé dans cette première expérience n'a pas permis de savoir précisément combien de temps il fallait pour que le niveau d'externalisation perçu immédiatement après un mouvement de tête retombe au niveau d'externalisation perçu sans mouvement de tête. Ce point de jonction se produisait quelque part entre 5,5 s et 31 s après la fin du mouvement.

Les stimuli *hom*, *sil* et *rose* ont reçu des scores d'externalisation similaires, tous plus élevés que *ancree* et *fem*. Ces résultats indiquent que, lorsque le *distracteur* différait de la source *initiale* + *évaluation*, l'amélioration d'externalisation obtenue après un mouvement de tête persistait mieux dans le temps que lorsque le *distracteur* était similaire à la source *initiale* + *évaluation*, contrairement à ce qui était attendu de l'hypothèse **H2**. Le fait que le *distracteur* ne différait que légèrement ou de manière plus importante de la source *initiale* + *évaluation* ne semble pas avoir influencé l'externalisation. En effet, *hom* a été évalué de manière similaire à *rose* et *sil*, alors qu'on aurait pu s'attendre à des tendances similaires entre *fem* et *hom*, étant donné qu'il s'agissait de stimuli vocaux utilisant les mêmes mots et une prosodie similaire.

Étant donné que **H1** a été validé pour le cas où le *distracteur* était similaire à la source *initiale* + *évaluation*, une deuxième expérience a été menée pour confirmer cet effet et pour étudier plus en détail l'influence de la longueur du *distracteur* sur l'externalisation.

3.3 Expérience 2

L'hypothèse **H1** soutenait l'idée que l'externalisation des sources sonores diffusées immédiatement après un mouvement de tête était plus élevée que celle des sources diffusées après un mouvement de tête et une période de *distraction*. Les résultats de l'expérience 1 suggèrent que cette hypothèse est vraie lorsque le *distracteur* est similaire à la source *initiale* + *évaluation* et fausse lorsque le *distracteur* est différent de la source *initiale* + *évaluation*. L'objectif de l'expérience 2 était donc d'étudier comment l'externalisation évolue dans le temps, avec des *distracteurs* similaires ou différents de la source *initiale* + *évaluation*, à l'aide de deux hypothèses supplémentaires :

- **H3** : quand le *distracteur* est similaire à la source *initiale* + *évaluation*, un *distracteur* plus long se traduit par des scores d'externalisation plus faible.
- **H4** : quand le *distracteur* est différent de la source *initiale* + *évaluation*, les scores d'externalisation sont aussi élevés que lorsqu'il n'y a pas de *distracteur*, quelle que soit la longueur du *distracteur*.

3.3.1 Matériel et méthode

Participants

Vingt sujets (17 hommes, 3 femmes ; âge moyen 27 ; écart-type 3,5 ; intervalle 20-35 ans) ont participé à l'étude. Un d'entre eux avait participé à l'expérience 1 présentée précédemment dans ce chapitre. Les autres n'avaient participé à aucune autre expérience effectuée dans le cadre de cette thèse. Tous ont déclaré avoir une audition normale. L'étude était conforme au Code d'éthique de l'Association médicale mondiale (Déclaration d'Helsinki), publié dans le *British Medical Journal* (18 juillet 1964) et a été approuvée par le Comité d'éthique institutionnel d'IMT Atlantique (enregistré sous le numéro IRB00013722 auprès du bureau américain de protection des humains dans la recherche [171]). En outre, tous les sujets ont donné leur consentement éclairé écrit pour participer à l'étude.

Stimuli

L'expérience comportait huit stimuli construits selon le schéma suivant :

1. Un extrait de 5,2 s du poème "l'Albatros" (v.1-2) de Charles Baudelaire lu par une femme (la même partie *initiale* que dans l'expérience 1).
2. Un extrait de longueur variable lu par un homme.
3. Un extrait de longueur variable lu par la même femme. Les 5,5 dernières secondes de cet extrait étant toujours la même phrase au cours de laquelle les sujets devaient évaluer l'externalisation (la même partie *évaluation* que dans l'expérience 1 : "l'Albatros" (v.15-16)).

La durée des parties 2) et 3) était de 5,5 s, 15 s ou 25 s (respectivement désignées par *C*, *M* et *L*, pour Court, Moyen et Long) et toutes les combinaisons ont été testées, sauf le cas où les parties 2) et 3) dureraient toutes deux 25 s (ce qui aurait rendu le test perceptif trop long). Le tableau 3.1 montre le détail de chaque stimulus. La partie 3) des stimuli a été appelée

distracteur + évaluation parce qu'elle contenait les dernières 5,5 s de voix de femme utilisées pour l'évaluation, parfois précédées d'une portion de voix de femme utilisée comme *distracteur* (stimuli *CM*, *MM*, *LM*, *CL* et *ML*). Pour plus de clarté, même dans le cas où cette partie 3) ne durait que 5,5 s (stimuli *CC*, *MC* et *LC*) et ne contenait donc que la partie *évaluation* des stimuli (sans aucun *distracteur* supplémentaire), elle est tout de même appelée partie *distracteur + évaluation*.

Label	Partie <i>initiale</i>	<i>Distracteur</i>	<i>Distracteur</i> <i>+ évaluation</i>	Durée totale
	Femme	Homme	Femme	
<i>CC</i>	5.2	5.5	5.5	16.7
<i>MC</i>	5.2	15	5.5	25.7
<i>LC</i>	5.2	25	5.5	35.7
<i>CM</i>	5.2	5.5	15	25.7
<i>MM</i>	5.2	15	15	35.2
<i>LM</i>	5.2	25	15	45.2
<i>CL</i>	5.2	5.5	25	35.7
<i>ML</i>	5.2	15	25	45.2

TABLE 3.1 – Tableau présentant la composition des huit stimuli. Les temps sont indiqués en secondes. Les sujets devaient évaluer l'externalisation ressentie au cours des 5,5 dernières secondes de la partie *distracteur + évaluation*.

L'enregistrement et le traitement du son, le système de reproduction sonore et de head-tracking ainsi que la lecture, l'affichage de l'interface et l'acquisition des données étaient identiques à l'expérience 1.

Procédure

La présentation des stimuli était similaire à celle de l'expérience 1. Les sujets devaient effectuer le mouvement de tête pendant la première phrase du stimulus (la partie *initiale*). Contrairement à l'expérience 1, il n'y avait pas de condition statique servant d'ancre basse. Pendant le reste du stimulus, les participants devaient rester statiques. Après chaque présentation, ils devaient évaluer l'externalisation perçue pendant la dernière phrase (la partie *évaluation*) de

la même manière que dans l'expérience 1. Avant l'expérience, tous les sujets devaient effectuer la même phase d'entraînement que dans l'expérience 1. Les mouvements de tête étaient enregistrés pendant l'expérience.

Les huit stimuli (*CC*, *CM*, *MC*, *CL*, *MM*, *LC*, *ML* et *LM*) étaient testés dans un ordre aléatoire, différent pour chaque sujet. Pour minimiser un potentiel effet d'apprentissage d'un set de HRTFs, le set de HRTFs changeait entre les essais successifs. Chaque combinaison de [stimulus \times HRTF] a été répétée 3 fois pour un total de 8 [stimulus] \times 3 [HRTF] \times 3 [répétition] = 72 essais. La durée totale de l'expérience variait de 45 à 60 minutes.

Conformément à **H4**, il était attendu que tous les stimuli avec une durée de voix de femme courte au sein de la partie *distracteur + évaluation* (*CC*, *MC*, *LC*) seraient aussi bien externalisés que le stimulus *ref* de l'expérience 1, quelle que soit la durée de voix d'homme. De plus, conformément à **H3**, il était attendu que plus la durée de la voix de femme serait longue au sein de la partie *distracteur + évaluation*, plus les scores d'externalisation seraient faibles, quelle que soit la durée de la voix d'homme

3.3.2 Résultats

L'analyse des données de mouvements de tête a montré que les sujets se conformaient en moyenne très bien aux instructions données et qu'ils effectuaient correctement les mouvements de tête. Les graphiques de mouvement, par sujet, sont présentés dans la figure A.3 de l'annexe A.

La figure 3.3 montre les scores d'externalisation moyennés sur tous les sujets par stimulus. Une analyse de la distribution des données a révélé que les données n'étaient pas normalement distribuées. Des méthodes d'analyse statistique non paramétriques ont donc été choisies. Un test de Friedman effectué sur toutes les données a révélé un effet du stimulus ($\chi^2(7) = 164$; $p < 0.001$) et aucun effet des sets de HRTFs ou des répétitions. Des tests de Wilcoxon ont été effectués pour comparer chaque paire de stimulus avec une correction FDR. Pour vérifier l'absence d'effet entre les comparaisons non significatives, des tests de Wilcoxon bayésiens [203] ont été effectués.

Trois groupes de stimuli sont apparus, appelés A, B et C dans la figure 3.3 :

- A) Tous les stimuli avec une durée courte de voix de femme (*CC*, *MC*, *LC*) ont reçu des scores d'externalisation similaires les uns aux autres (tous $BF_{01} > 5$) [204] et ils ont tous reçu des scores d'externalisation plus élevés que les autres stimuli ($p < 0,05$).

- B) Tous les stimuli avec une durée moyenne de voix de femme (*CM*, *MM*, *LM*) ont reçu des scores d’externalisation similaires les uns aux autres ($BF_{01} > 5$) et ils ont tous reçu des scores d’externalisation plus élevés que les stimuli avec une durée longue de voix de femme ($p < 0.05$).
- C) Les deux stimuli avec une durée longue de voix de femme (*CL*, *ML*) ont reçu des scores d’externalisation similaires ($BF_{01} > 5$).

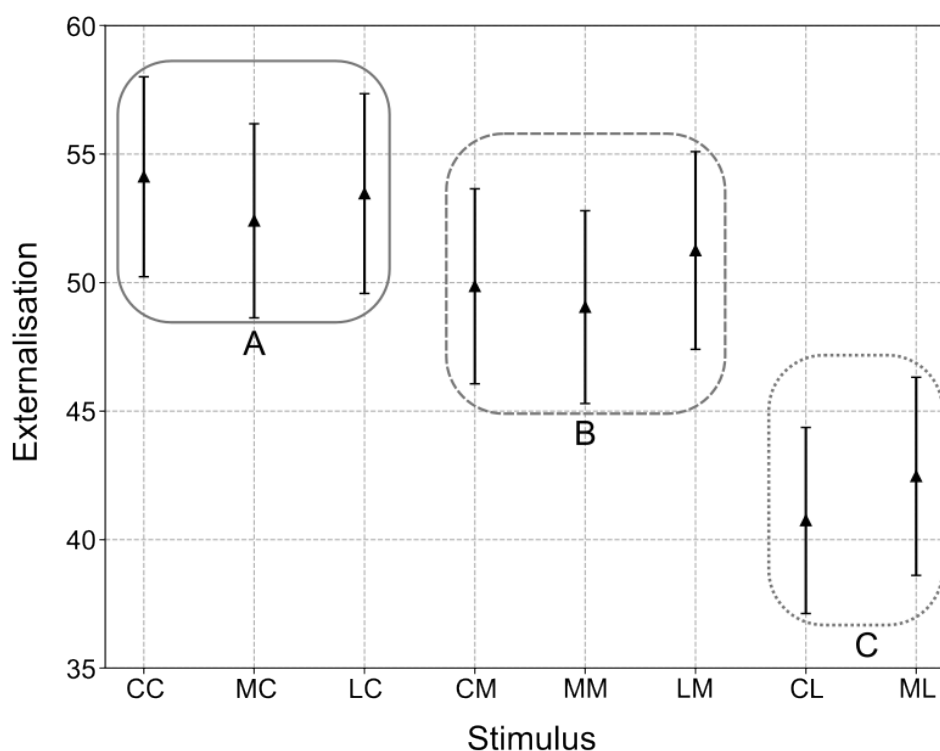


FIGURE 3.3 – Scores d’externalisation moyennés sur tous les sujets pour chaque stimulus. Les barres d’erreur indiquent l’intervalle de confiance à 95%. Groupe A : tous les stimuli avec une durée courte de voix de femme. B : tous les stimuli avec une durée moyenne de voix de femme. C : tous les stimuli avec une durée longue de voix de femme.

Notons que les résultats obtenus pour le groupe A sont très similaires aux résultats obtenus pour le stimulus *ref* dans l’expérience 1. De même, les résultats obtenus pour le groupe C sont très similaires à ceux obtenus pour le stimulus *ancree* dans l’expérience 1. Des tests de Mann-Whitney bayésiens [205] ont confirmé l’absence d’effets (preuves modérées) entre les résultats du groupe A et *ref*, ainsi qu’entre les résultats du groupe C et *ancree*. Ces comparaisons

confirment que les stimuli avec une durée courte de voix de femme étaient aussi bien externalisés que lorsqu'il n'y avait pas de partie *distraction*. Elles confirment également que les stimuli avec une durée longue de voix de femme étaient aussi peu externalisés que lorsqu'il n'y avait pas de mouvement de tête.

3.3.3 Discussion

Conformément à **H3** et **H4**, les résultats suggèrent que, après avoir effectué un mouvement de tête sur une source *A* :

- plus on écoute cette même source *A* en condition statique, plus l'externalisation diminue
- l'ajout d'une source *B* différente, quelle que soit sa durée, n'a aucun impact sur l'externalisation lors de la reprise de l'écoute de la source *A* en condition statique.

Pour les expériences 1 et 2, dans les cas où le *distracteur* était différent de la source *initial + évaluation*, il est possible que l'externalisation ait été surestimée pendant la partie *évaluation* en comparaison à une faible externalisation potentiellement associée au *distracteur*. Cette explication suggérerait qu'effectuer un mouvement de tête en écoutant une source *A* ne permettrait pas d'améliorer l'externalisation d'une source *B* diffusée immédiatement après le mouvement de tête. Cependant, ni les protocoles utilisés dans cette étude ni la littérature ne permettent de savoir quelle était l'externalisation des différentes sources diffusées immédiatement après le mouvement de tête (la voix d'homme et le bruit rose).

En effet, la plupart des études antérieures explorant l'effet du head-tracking sur l'externalisation n'ont utilisé qu'une seule source tout au long de leurs expériences (soit de la parole [118, 150, 151], soit du bruit [112]). Quelques études ont exploré l'externalisation de scènes sonores complexes [206, 207]. Toutefois, dans ces expériences, l'externalisation a été évaluée pour l'ensemble de la scène sonore et non pour des sources individuelles. En particulier, aucune étude n'a examiné la persistance de l'effet des mouvements de tête pour une source différente de celle utilisée pendant les mouvements de tête.

C'est pourquoi une troisième expérience a été réalisée afin d'étudier l'influence d'un changement de source sonore sur la persistance de l'effet des mouvements de tête avec head-tracking.

3.4 Expérience 3

Les résultats des expériences 1 et 2 ont conduit à l'hypothèse selon laquelle l'exécution d'un mouvement de tête sur une source ne permettrait pas de mieux externaliser des sources différentes qui viendraient immédiatement après le mouvement de la tête. L'objectif de l'expérience 3 était d'étudier cette hypothèse, appelée **H5**.

En effet, on pourrait s'attendre à ce que les indices spatiaux obtenus en effectuant des mouvements de tête sur une source *A* ne soient pas pertinents lors de l'écoute d'une source *B* dont le contenu spectral est différent et que l'externalisation qui en résulte soit plus faible.

3.4.1 Matériel et méthode

Participants

Vingt sujets (16 hommes, 4 femmes ; âge moyen 26 ; écart-type 3,8 ; intervalle 20-35 ans) ont participé à l'étude. Sept d'entre eux avaient participé à l'expérience 2 présentée précédemment dans ce chapitre. Les autres n'avaient participé à aucune autre expérience effectuée dans le cadre de cette thèse. Tous ont déclaré avoir une audition normale. L'étude était conforme au Code d'éthique de l'Association médicale mondiale (Déclaration d'Helsinki), publié dans le *British Medical Journal* (18 juillet 1964) et a été approuvée par le Comité d'éthique institutionnel d'IMT Atlantique (enregistré sous le numéro IRB00013722 auprès du bureau américain de protection des humains dans la recherche [171]). En outre, tous les sujets ont donné leur consentement éclairé écrit pour participer à l'étude.

Stimuli

L'expérience comportait neuf stimuli construits selon le schéma suivant :

1. Une partie *initiale* de 5,2 s qui était :
 - soit un extrait du poème "l'Albatros" (v.1-2) lu par une femme,
 - soit le même extrait de poème lu par un homme,
 - soit un bruit rose.
2. Une partie *évaluation* de 5,5 s qui était :
 - soit un extrait du poème "l'Albatros" (v.15-16) lu par une femme,
 - soit le même extrait de poème lu par un homme,
 - soit un bruit rose.

Le tableau 3.2 montre le détail de chaque stimulus.

Label	Partie initiale (5,2 s)	Partie évaluation (5.5 s)
<i>FF</i>	Voix de femme	Voix de femme
<i>FH</i>	Voix de femme	Voix d'homme
<i>FR</i>	Voix de femme	Bruit rose
<i>HF</i>	Voix d'homme	Voix de femme
<i>HH</i>	Voix d'homme	Voix d'homme
<i>HR</i>	Voix d'homme	Bruit rose
<i>RF</i>	Bruit rose	Voix de femme
<i>RH</i>	Bruit rose	Voix d'homme
<i>RR</i>	Bruit rose	Bruit rose

TABLE 3.2 – Tableau présentant la composition des neuf stimuli. F : voix de femme. H : voix d'homme. R : bruit rose.

L'enregistrement et le traitement du son, le système de reproduction sonore et de head-tracking ainsi que la lecture, l'affichage de l'interface et l'acquisition des données étaient similaires aux expériences 1 et 2.

Procédure

Tous les stimuli étaient testés en condition statique (*i.e.* sans mouvement de tête pendant la partie *initiale*) ainsi qu'en condition dynamique (*i.e.* avec un mouvement de tête pendant la partie *initiale*). Pendant la partie *évaluation*, les sujets devaient toujours rester statiques. Après chaque présentation, ils devaient évaluer l'externalisation perçue pendant la partie *évaluation* de la même manière que dans les expériences 1 et 2.

Avant l'expérience, tous les sujets devaient suivre une phase d'entraînement similaire à celle des expériences 1 et 2. Les mouvements de tête étaient enregistrés pendant l'expérience. Les neuf stimuli (*FF*, *FH*, *FR*, *HF*, *HH*, *HR*, *RF*, *RH* et *RR*) et les deux conditions de mouvement (statique vs. dynamique) ont été testés dans un ordre aléatoire, différent pour chaque sujet. Pour minimiser un potentiel effet d'apprentissage d'un set de HRTFs, le set de HRTFs changeait entre les essais successifs. Chaque combinaison de [stimulus × mouvement × HRTF] a été répétée 3 fois pour un total de 9 [stimulus] × 2 [mouvement] × 3 [HRTF] × 3 [répétition] = 162 essais. La durée totale de l'expérience variait de 40 à 55 minutes.

Dans la condition dynamique, il était attendu que les trois stimuli cohérents (*i.e.* les stimuli qui avaient des sources similaires pendant leur partie *initiale* et leur partie *évaluation* : *FF*, *HH* et *RR*) soient plus externalisés que tous les stimuli incohérents (*i.e.* les stimuli qui avaient des sources différentes pendant leur partie *initiale* et leur partie *évaluation* : *FH*, *FR*, *HF*, *HR*, *RF*, *RH*).

En condition statique, il était attendu que tous les stimuli aient une externalisation similaire entre eux et faible.

3.4.2 Résultats

L'analyse des données de mouvements de tête a montré que les sujets se conformaient en moyenne très bien aux instructions données et qu'ils effectuaient correctement les mouvements de tête. Les graphiques de mouvement, par sujet, sont présentés dans la figure A.4 de l'annexe A.

La figure 3.4 montre les scores d'externalisation moyens pour chaque stimulus. Un test de Friedman effectué sur toutes les données a révélé que les scores d'externalisation étaient significativement plus élevés en condition dynamique ($\chi^2(8) = 32.2$; $p < 0.001$). Il n'y a eu aucun effet des sets de HRTFs ou des répétitions ($p \gg 0.05$). Les stimuli ont eu des scores d'externalisation moyens de 44,9 en condition statique et de 52,1 en condition dynamique.

Pour vérifier l'absence d'effet entre les stimuli au sein d'une même condition de mouvement, des tests de Wilcoxon bayésiens ont été effectués [203] entre chaque paire de stimuli. Les résultats ont fourni des preuves fortes à très fortes de l'absence d'effet entre les stimuli ($BF_{01} > 10$) [204].

3.4.3 Discussion

Après un mouvement de tête, les stimuli incohérents étaient aussi bien externalisés que les stimuli cohérents. Ces résultats suggèrent que l'exécution d'un mouvement de tête est aussi efficace pour améliorer l'externalisation d'une source différente de celle diffusée pendant le mouvement de tête que pour améliorer l'externalisation d'une source similaire (**H5** réfuté).

Au regard de ces résultats, on pourrait anticiper qu'un mouvement de tête puisse améliorer l'externalisation de toute source diffusée immédiatement après. Néanmoins, cette expérience n'a testé que trois sources. Il serait donc pertinent d'élargir l'éventail des sources examinées afin de confirmer la généralité de cet effet.

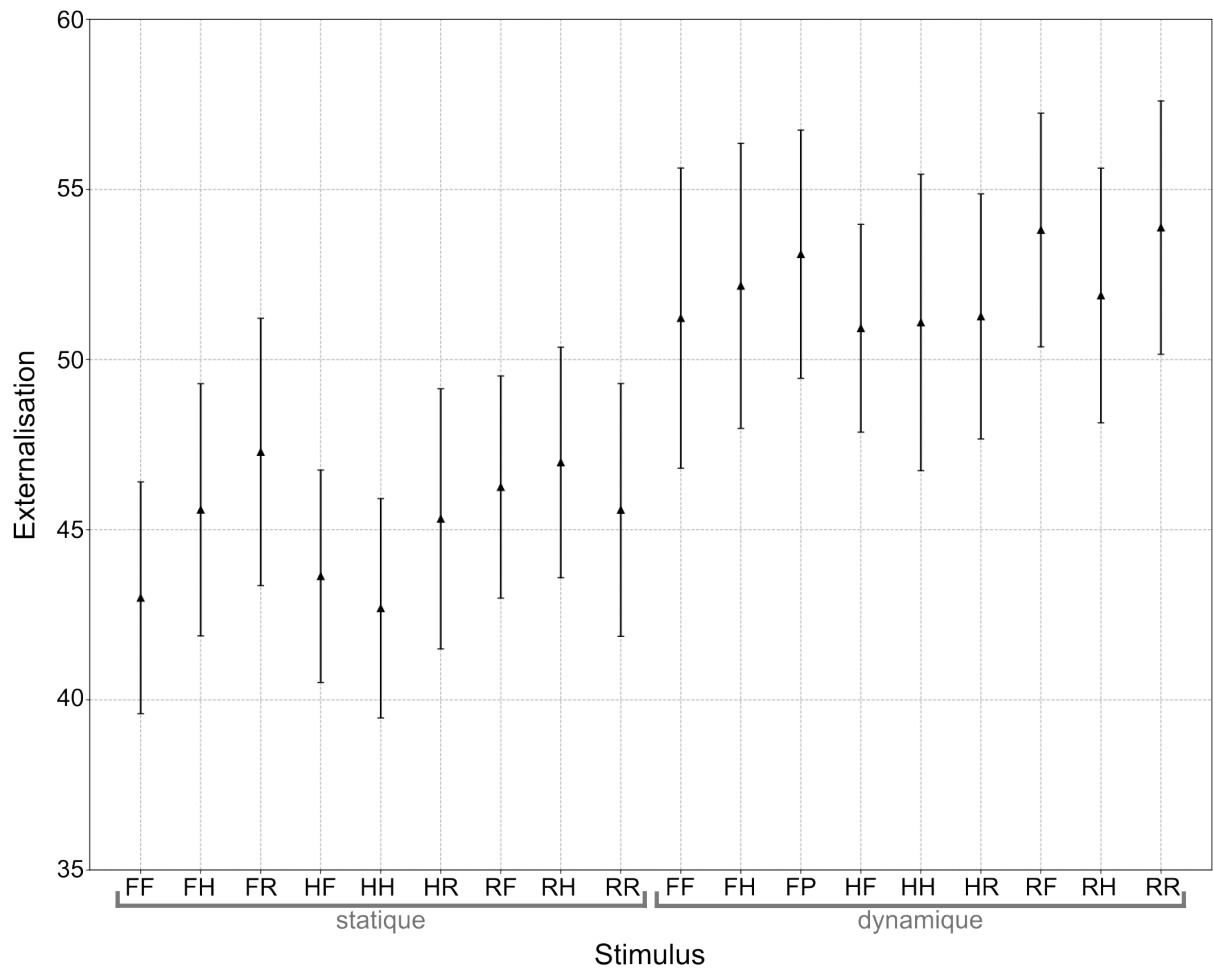


FIGURE 3.4 – Scores d'externalisation moyennés sur tous les sujets pour chaque stimulus. Les barres d'erreur indiquent l'intervalle de confiance à 95%.

En prenant en compte les résultats de cette expérience, il semble que l'externalisation ne dépende pas uniquement d'une source particulière, mais implique un processus plus complexe entre plusieurs sources présentes dans la scène sonore. Dans ce cas, il serait plus pertinent d'évaluer l'externalisation de l'ensemble de la scène sonore plutôt que l'externalisation d'une source particulière lorsqu'on étudie l'externalisation d'une scène contenant plus d'une source. Cela pourrait expliquer pourquoi, dans les quelques études portant sur l'externalisation de scènes sonores complexes, les sujets devaient évaluer l'externalisation de l'ensemble de la scène et non de sources individuelles [206, 207]. Toutefois, dans ces études, l'externalisation n'était qu'un des nombreux paramètres testés et n'a pas été examinée en détail. Par conséquent, des études supplémentaires portant sur des scènes sonores complexes seraient nécessaires pour vérifier cette considération.

3.5 Discussion générale

Puisque **H5** a été réfuté, les résultats des expériences 1 et 2 ne peuvent pas être expliqués uniquement par le fait que les sources *distracteur* différaient de la source *initiale + évaluation*.

Les résultats des trois expériences suggèrent les effets suivants :

1. Un mouvement de tête effectué en écoutant une source *A* permet de mieux externaliser n'importe quelle source diffusée immédiatement après le mouvement.
2. Après un mouvement de tête effectué en écoutant une source *A*, l'exposition statique à la même source *A* conduit à une diminution progressive de l'amélioration d'externalisation due au mouvement.
3. Après un mouvement de tête effectué lors de l'écoute d'une source *A* suivi d'une exposition statique à une source *B*, L'amélioration d'externalisation due au mouvement reste effective lors la réécoute statique de la source *A*. L'amélioration d'externalisation obtenue pour la source *A* est "conservé" au cours de l'écoute statique de la source *B*.

L'explication de ces effets peut s'avérer complexe car l'effet 1. suggère que l'externalisation est indépendante de la nature des sources, alors que les effets 2. et 3. suggèrent que l'externalisation d'une source dépend de l'ensemble des sources dans la scène sonore.

Une explication possible pourrait venir de l'hypothèse selon laquelle la perception de l'externalisation d'une source sonore est basée sur l'intégration des indices perceptifs les plus pertinents stockés dans la mémoire à court terme.

Lorsqu'une source sonore similaire à celle que les sujets devaient évaluer était diffusée auparavant au sein du même stimulus (expériences 1 et 2), le système auditif aurait pu s'appuyer sur les indices propres à cette source pour créer une perception d'externalisation. L'externalisation aurait été élevée si, lors de l'écoute la plus récente de la source, celle-ci était bien externalisée (mouvement de tête) et faible si, lors de l'écoute la plus récente de la source, celle-ci n'était pas bien externalisée (exposition statique). Lorsqu'il n'y avait pas de source sonore similaire à celle que les sujets devaient évaluer au sein du même stimulus (expérience 3), le système auditif aurait utilisé la seule information d'externalisation disponible, qui était soit la source correctement externalisée entendue lors du mouvement de tête précédent (condition dynamique), soit la source peu externalisée entendue lors de l'écoute statique précédente (condition statique), même si cette source était différente de celle qu'ils devaient évaluer.

En effet, il a été démontré que la perception d'autres caractéristiques sonores repose sur une intégration du son entendu lors des dernières quelques secondes [208], en particulier dans les tâches faisant appel à l'attention [209]. En outre, les résultats de deux études antérieures réalisées par Begault et Wenzel [118, 149] et portant sur l'externalisation de sons très courts (3 s) diffusés lors d'un mouvement de tête avec head-tracking pourraient s'expliquer par cette perception de l'externalisation basée sur une intégration. En effet, ces deux études suggèrent que les mouvements de tête ne renforcent l'externalisation que très légèrement [149] ou même pas du tout [118]. Ces résultats sont surprenants, car plusieurs études récentes [137, 150, 151] utilisant des stimuli plus longs ont démontré que l'exécution de mouvements de tête est efficace pour améliorer l'externalisation. Il est possible que les stimuli utilisés dans ces deux études contradictoires de Begault et Wenzel aient été si courts que le système auditif des auditeurs n'avait pas le temps d'effectuer l'intégration nécessaire pour percevoir correctement l'externalisation, ce qui aurait pu causer l'obtention de scores d'externalisation faibles même dans les conditions avec mouvements de tête. L'exploration d'une perception de l'externalisation basée sur l'intégration pourrait faire l'objet de travaux futurs.

Il est intéressant de noter que l'effet 2. permet d'expliquer les scores d'externalisation modérément élevés obtenus dans l'étude du chapitre 2. En effet, dans cette étude, les stimuli diffusés après les mouvements de tête, pendant lesquels les sujets devaient évaluer l'externalisation, étaient relativement longs (12 s). Il est possible que l'externalisation ait été plus élevée immédiatement après le mouvement de tête, mais ait décliné au cours des 12 secondes d'écoute statique post-mouvement de tête. Il est donc possible que les scores d'externalisation résultants étaient moins élevés qu'ils n'auraient été avec des stimuli post-mouvement plus courts, comme c'était

le cas dans les études précédentes [150, 151].

3.6 Conclusion

Dans l'étude présentée dans ce chapitre, trois expériences ont été menées pour déterminer si l'amélioration d'externalisation induite par des mouvements de tête était robuste dans le temps en fonction de la nature et de la durée des sources sonores.

Dans la première expérience, les sujets devaient évaluer l'externalisation de sources présentes immédiatement après des mouvements de tête avec head-tracking ou après des mouvements de tête suivis d'une partie *distraction* (sans mouvement de tête) de durée fixe (31 s). La même source était diffusée pendant les mouvements de tête et pendant l'évaluation (appelée source *initiale + évaluation*) et l'évaluation était toujours effectuée pendant que les sujets étaient immobiles. Le *distracteur* (*i.e.* la source pendant la partie *distraction*) était soit similaire, soit légèrement différent, soit très différent de la source *initiale + évaluation*. Une condition statique supplémentaire, sans aucun mouvement de tête, a été ajoutée en tant qu'ancre basse, pour laquelle l'externalisation était médiocre. Les résultats ont montré que l'externalisation était meilleure immédiatement après des mouvements de tête que sans mouvement de tête. En outre, après la partie *distraction*, l'externalisation de la source *initiale + évaluation* restait aussi élevée qu'immédiatement après le mouvement de tête si le *distracteur* était différent de la source *initiale + évaluation*. En revanche, lorsque *distracteur* était similaire à la source *initiale + évaluation*, l'externalisation retombait au niveau d'externalisation obtenu sans mouvement de tête.

Dans la seconde expérience, les sujets devaient évaluer l'externalisation de sources sonores diffusées après des mouvements de tête suivis d'une partie *distraction* de durée variable. Le *distracteur* était constitué de deux sources, diffusées l'une après l'autre : un extrait de voix d'homme (*i.e.* différent de la voix de femme *initiale*) suivi d'un extrait de voix de femme (*i.e.* similaire à la voix de femme *initiale*). Les résultats ont montré que lorsque le *distracteur* contenait un extrait court de voix de femme, l'externalisation était robuste dans le temps, quelle que soit la durée de voix d'homme. Cependant, lorsque le *distracteur* contenait un extrait plus long de voix de femme, l'externalisation diminuait progressivement pour retomber au niveau d'externalisation obtenu sans mouvement de tête, quelle que soit la durée de voix d'homme.

Dans les deux premières expériences, la source pendant laquelle les sujets devaient évaluer l'externalisation était similaire à la source diffusée pendant les mouvements de tête (toujours

une voix de femme). La troisième expérience a testé l'effet des mouvements de tête sur l'externalisation pour une source *évaluation* différente de la source *initiale* diffusée pendant les mouvements de tête afin d'étudier si l'amélioration d'externalisation due aux mouvements de tête lors de l'écoute d'une source pouvait permettre de mieux externaliser une source différente. Les résultats ont montré que l'externalisation était aussi élevée lorsque la source *évaluation* était différente de la source *initiale* que lorsque la source *évaluation* était similaire à la source *initiale*.

Les résultats des trois expériences révèlent trois effets distincts. 1) Un mouvement de tête pendant l'écoute d'une source améliore l'externalisation de toute source diffusée immédiatement après. 2) Après un mouvement de tête pendant l'écoute d'une source *A*, une écoute statique de cette même source *A* réduit progressivement l'amélioration d'externalisation induite par le mouvement. 3) Si, après un mouvement de tête en écoutant une source *A*, on écoute statiquement une source *B*, l'amélioration d'externalisation pour la source *A* est préservée lors de sa réécoute statique. L'amélioration d'externalisation obtenue pour la source *A* a été "conservée" au cours de l'écoute statique de la source *B*.

CONCLUSION GÉNÉRALE

Résultats

Les travaux présentés dans cette thèse ont permis d'explorer l'externalisation en écoute binaurale avec head-tracking d'un point de vue subjectif et d'un point de vue objectif afin de mettre en évidence plusieurs effets du head-tracking sur l'externalisation.

Une externalisation convaincante est nécessaire pour une restitution binaurale de qualité et l'utilisation d'un dispositif de head-tracking permet d'améliorer celle-ci pendant et après le mouvement de tête. Afin d'essayer de trouver un marqueur objectif de l'externalisation et d'étudier l'influence de la nature des stimuli sur l'externalisation après des mouvements de tête avec head-tracking, deux études ont été réalisées.

Dans la première étude, une expérience comportementale et une expérience EEG ont été menées pour identifier un corrélat neuronal de l'externalisation en utilisant les potentiels évoqués (ERPs). Une méthode innovante a été conçue afin de s'assurer que la seule différence entre les stimuli était l'externalisation, de façon à pouvoir attribuer sans ambiguïté de potentielles différences d'ERPs à des différences d'externalisation. Afin d'obtenir différents degrés d'externalisation, les mesures EEG et comportementales étaient précédées : 1) de mouvements de tête avec head-tracking, 2) de mouvements de tête sans head-tracking, ou 3) d'aucun mouvement de tête. Ainsi les stimuli étaient "acoustiquement identiques", mais variaient en termes d'externalisation. Lors de l'expérience comportementale, les participants ont évalué subjectivement l'externalisation, tandis que l'expérience EEG consistait en une écoute passive. À l'exception de cette différence, les deux expériences étaient similaires pour permettre la comparaison de leurs résultats.

Aucun corrélat neuronal spécifique à l'externalisation n'a été trouvé dans les ERPs, probablement en raison de la différence de tâche entre les deux expériences (évaluation de l'externalisation vs. écoute passive) ou de différences d'externalisation trop faibles pour être détectées.

En effet, l'expérience comportementale a révélé que l'amélioration d'externalisation due à la persistance de l'effet des mouvements de tête était moins marquée que ce que les études précédentes, utilisant un protocole similaire, laissaient espérer. Il est possible que la longueur et

l'incohérence des stimuli post-mouvement, conçus pour générer efficacement des ERPs, aient entraîné une diminution de l'externalisation. Aucune étude à notre connaissance n'a exploré l'impact de la nature et de la cohérence du son sur la persistance de l'externalisation due aux mouvements de tête. Ce constat a conduit à la réalisation d'une deuxième étude visant à comparer l'externalisation de stimuli plus ou moins longs et cohérents.

Dans la deuxième étude, trois expériences ont été menées pour déterminer la robustesse de l'amélioration d'externalisation induite par les mouvements de tête, en fonction de la nature et de la durée des sources sonores.

Dans la première expérience, les sujets devaient évaluer l'externalisation de sources présentées après une écoute statique (sans mouvement de tête), immédiatement après des mouvements de tête, ou après des mouvements de tête suivis d'une période de *distraction* de 31 secondes. La source initiale, diffusée pendant les mouvements de tête, était toujours similaire à la source finale, que les sujets devaient évaluer. Cependant, la nature du *distracteur* variait : il était soit similaire, soit légèrement différent, soit très différent de la source initiale. Les résultats ont permis de vérifier que l'externalisation était meilleure immédiatement après un mouvement de tête qu'après une écoute statique. De plus, après une période de *distraction*, l'externalisation restait élevée si le *distracteur* était différent de la source initiale, mais retombait au niveau d'externalisation obtenu sans mouvement de tête si le *distracteur* était similaire.

Dans la deuxième expérience, les sujets évaluaient l'externalisation des sources sonores après des mouvements de tête suivis d'une période de *distraction* de durée variable. Le *distracteur* était constitué de deux sources, diffusées l'une après l'autre : une voix d'homme (donc différente de la voix de femme initiale) suivie d'une voix de femme (similaire à la source initiale). Les résultats ont montré que lorsque le *distracteur* contenait un extrait court de voix de femme, l'externalisation était robuste dans le temps, quelle que soit la durée de voix d'homme. En revanche, lorsque le *distracteur* contenait un extrait plus long de voix de femme, l'externalisation diminuait progressivement pour retomber au niveau d'externalisation obtenu sans mouvement de tête, quelle que soit la durée de voix d'homme.

Dans les deux premières expériences, la source dont l'externalisation été évaluée était toujours une voix de femme, similaire à celle entendue pendant les mouvements de tête. La troisième expérience a testé l'externalisation après des mouvements de tête pour une source différente de celle entendue initialement, afin de voir si l'amélioration due aux mouvements de tête s'appliquait à une source différente. Les résultats ont montré que l'externalisation restait élevée même si la source après le mouvement était différente de la source initiale.

Les résultats des trois expériences suggèrent que : 1) un mouvement de tête en écoutant une source améliore l'externalisation de toute source diffusée immédiatement après, 2) après des mouvements de tête, l'exposition statique à la même source que celle diffusée pendant les mouvements entraîne une diminution progressive de l'externalisation, et 3) après un mouvement de tête avec une source A suivi d'une exposition statique à une source B, l'amélioration d'externalisation pour la source A est maintenue.

Ces effets permettent une compréhension plus fine du fonctionnement de la persistance de l'effet des mouvements de tête et permettent, dans une certaine mesure, d'expliquer les résultats d'externalisation modérément élevés obtenus dans l'étude du chapitre 2. Il semblerait effectivement que la nature incohérente des stimuli utilisés dans cette étude n'ait pas affecté négativement les scores d'externalisation, mais que leur longueur ait entraîné une diminution d'externalisation.

Perspectives

Le concept développé dans le protocole de l'étude du chapitre 2, visant à utiliser la persistance de l'externalisation due aux mouvements de tête pour comparer des stimuli plus ou moins externalisés mais "acoustiquement identiques" par ailleurs, semblait pertinent. Malheureusement, le protocole n'a pas permis de mettre en lumière des corrélats neuronaux de l'externalisation. Il pourrait donc être intéressant d'établir un protocole similaire, mais en prenant en compte les résultats de l'étude du chapitre 3 afin de créer des stimuli qui bénéficieraient mieux de la persistance de l'effet des mouvements de tête. Il pourrait alors être possible d'obtenir des différences d'externalisation entre conditions plus marquées et ainsi de, peut-être, observer un corrélat neuronal de l'externalisation. Les résultats de l'étude 2 suggèrent que l'utilisation de stimuli plus courts pourrait permettre d'obtenir des différences d'externalisation plus marquées. Des stimuli plus courts pourrait en outre permettre de limiter l'effet de suppression due à la répétition relevé dans l'étude 1. La contrepartie serait de devoir augmenter le nombre d'essais totaux afin d'obtenir suffisamment de répétition d'ERP.

En dehors de la nature des stimuli, une différence importante existait entre l'expérience comportementale et l'expérience EEG de la première étude. En effet, lors de l'expérience comportementale, l'attention des sujets était dirigée vers l'externalisation, puisque ceux-ci devaient l'évaluer, alors que ce n'était pas le cas dans l'expérience EEG. Il est possible que le fait de demander aux sujets d'évaluer l'externalisation soit à l'origine des différences d'externalisation observées dans l'expérience comportementale. Cette hypothèse est cohérente avec les résul-

tats des études précédentes utilisant un protocole similaire dans lesquelles les sujets devaient également évaluer l'externalisation. Il pourrait alors être intéressant de comparer des résultats comportementaux et EEG de stimuli plus ou moins externalisés au sein de protocoles "standard-déviant" afin d'étudier les potentiels MMN et P300. Un marqueur de l'externalisation sur ces potentiels pourrait ouvrir des pistes de recherche sur la relation entre l'externalisation et l'attention.

La deuxième étude, en plus d'avoir donné des pistes d'explication pour les résultats de la première étude, a permis de mettre en lumière plusieurs effets de la nature des sources sonores et de la durée d'exposition à ces sources sur l'évolution de l'externalisation après un mouvement de tête avec head-tracking. Cependant, un nombre relativement limité de sources a été utilisé et il pourrait être intéressant d'étendre ces effets en testant d'autres sources. Pour aller plus loin, la plupart des études précédentes ont évalué l'externalisation de stimuli ne contenant qu'une seule source sonore. Il pourrait être intéressant de travailler avec des stimuli composés de scène sonore complexe, comportant plusieurs sources. On pourrait ainsi demander aux sujets d'évaluer l'externalisation de chaque source individuellement d'une part, et d'évaluer l'externalisation de l'ensemble de la scène sonore d'autre part. Cette démarche pourrait permettre de voir si l'externalisation est un percept qui est propre à chaque source, ou si l'externalisation d'une source est dépendante de l'externalisation des autres sources dans la scène sonore comme les résultats de l'étude 2 semblent suggérés.

Pour finir, afin que ces effets puissent être pris en compte lors de la création de contenus binauraux concrets, il est nécessaire de vérifier leur validité dans des cadres plus écologiques, avec des mouvements de tête libres plutôt qu'imposés et avec l'ajout d'autres modalités telles que la vision. En effet, l'ajout d'une modalité visuelle a une influence importante sur l'externalisation. Il est donc nécessaire de vérifier si l'apport du head-tracking mis en évidence dans les études précédentes, et développées dans cette thèse, est encore pertinent face à l'apport de la modalité visuelle.

ANNEXE A : DONNÉES DES MOUVEMENTS DE TÊTE

Cette annexe présente les données de mouvements de tête enregistrées pendant chaque expérience. Pour chaque sujet, les mouvements de tous les essais en condition dynamiques d'une part (*i.e.* les conditions dans lesquelles les sujets devaient effectuer le mouvement de tête), et en condition statique d'autre part (*i.e.* les conditions dans lesquelles les sujets ne devaient pas bouger la tête) ont été moyennés pour chaque pas temporel. La direction du mouvement n'était pas imposée, les sujets étaient libres de commencer leurs mouvements à droite ou à gauche. Certains sujets commençaient toujours leurs mouvements du même côté, tandis que d'autres alternaient à droite ou à gauche. Cette liberté avait été laissée aux sujets pour garder des mouvements naturel, sans imposer de contraintes inutiles. En effet, que les sujets tournent la tête à gauche puis à droite ou l'inverse, ils balayaient les mêmes HRTFs et bénéficiaient des mêmes indices spatiaux et spectraux. Ainsi, pour une meilleure lisibilité des graphiques, les essais pour lesquels les sujets commençaient leurs mouvements à gauche ont été inversés, donnant l'impression que tous les sujets commençaient toujours leurs mouvements à droite. Les graphiques de chacune des expériences sont présentés par ordre d'apparition dans le manuscrit.

L'observation de ces graphiques indique que, dans toutes les expériences, les sujets ont très bien respecté les consignes. Tous les sujets étaient immobiles lors des essais en condition statique. De plus, tous les sujets effectuaient un mouvement correct lors des essais en condition dynamique, avec parfois des différences modérées d'amplitude angulaire entre les sujets.

Graphiques de mouvements du chapitre 2

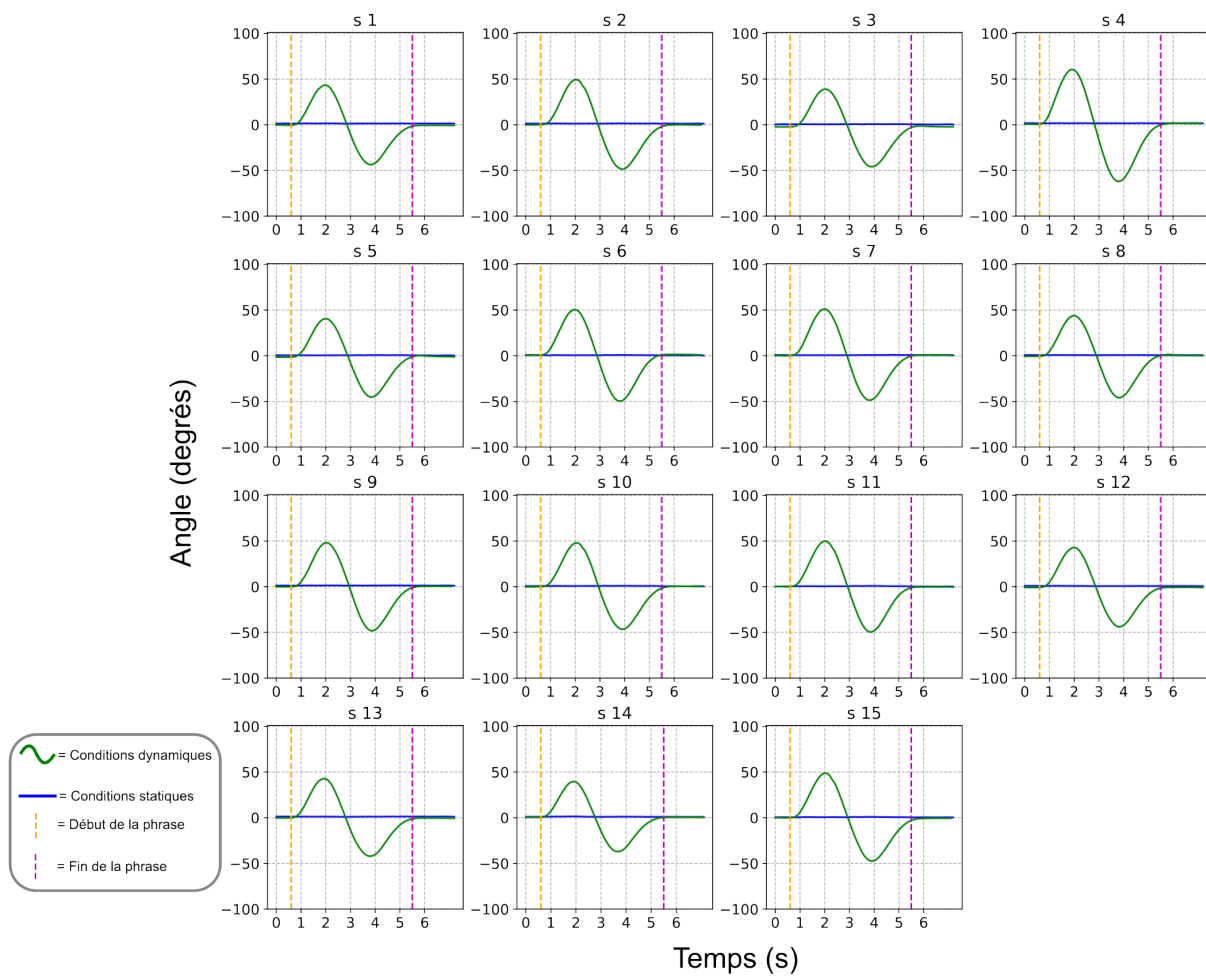


FIGURE A.1 – Chapitre 2 : mouvements de tête par condition statique ou dynamique, pour chaque sujet.

Graphiques de mouvements du chapitre 3, expérience 1

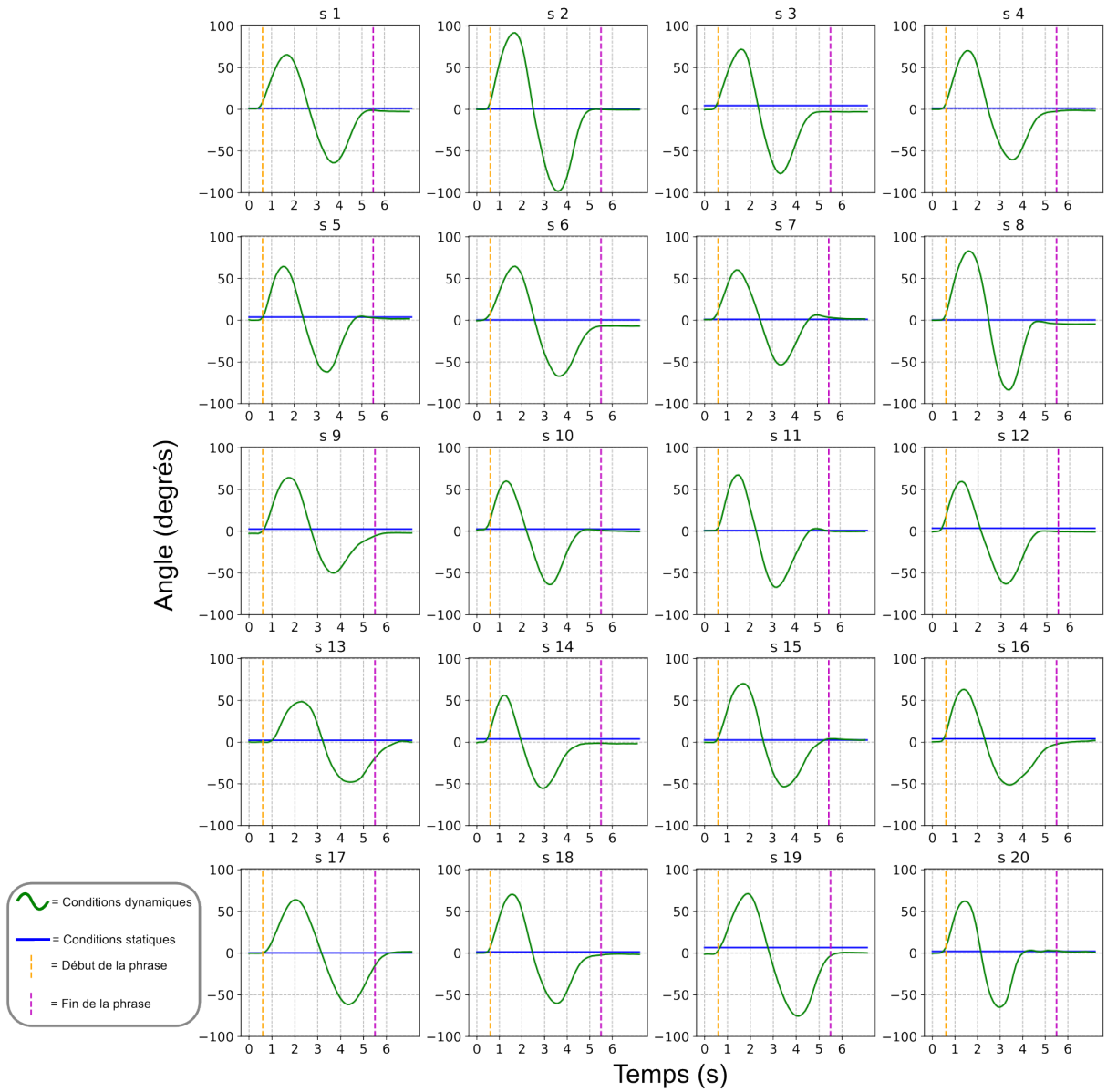


FIGURE A.2 – Chapitre 3, expérience 1 : mouvements de tête par condition statique ou dynamique, pour chaque sujet.

Graphiques de mouvements du chapitre 3, expérience 2

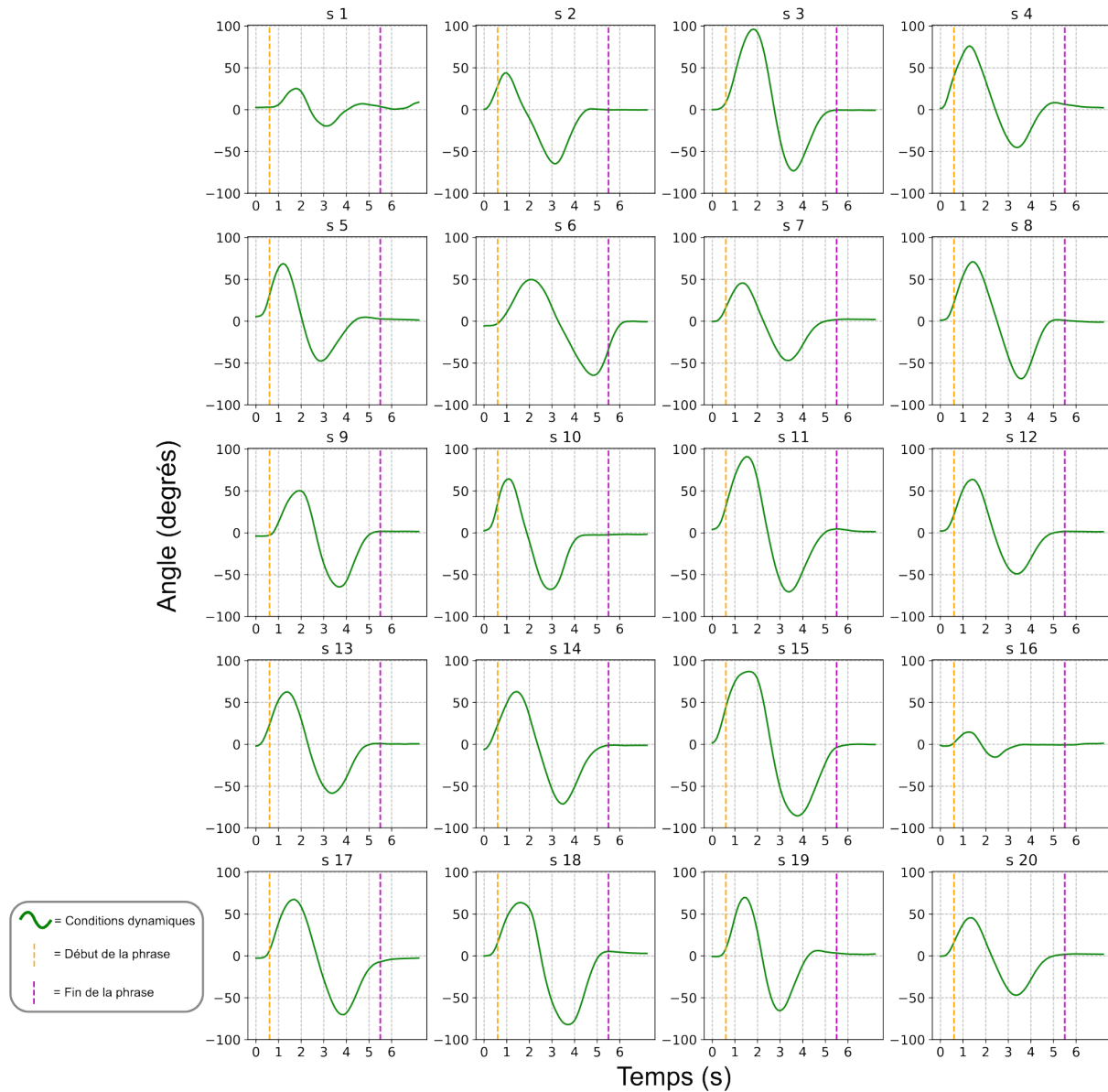


FIGURE A.3 – Chapitre 3, expérience 2 : mouvements de tête pour chaque sujet. Il n’y avait pas de condition statique dans cette expérience. L’amplitude des mouvements des sujets 1 et 16 était relativement faible. Les données d’externalisation de ces sujets ont tout de même été conservées dans l’analyse statistique, car elles étaient cohérentes avec celles des autres sujets. Il semble que la faible amplitude des mouvements n’ait pas affecté l’externalisation de façon différente que si les mouvements avaient été plus amples.

Graphiques de mouvements du chapitre 3, expérience 3

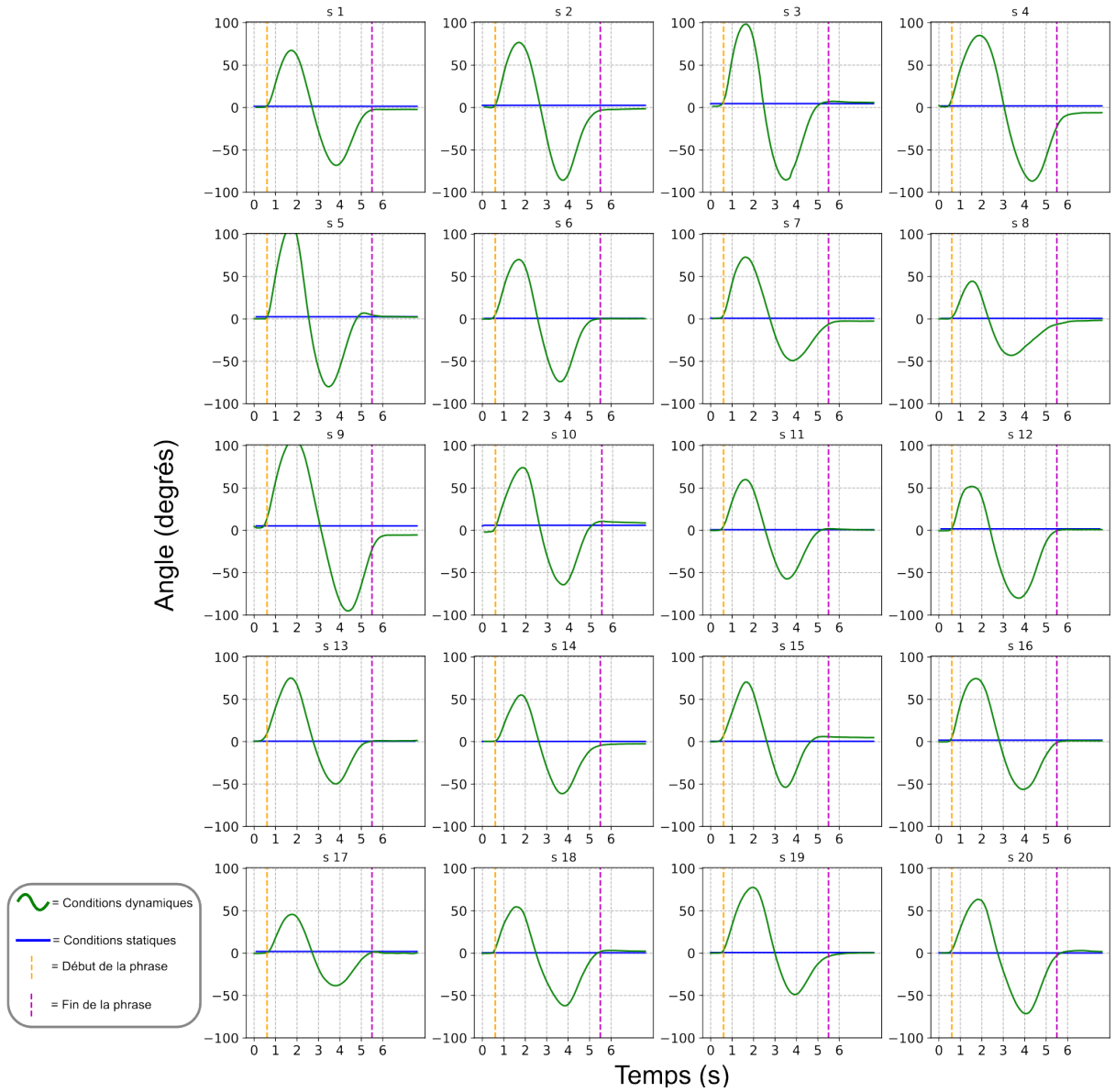


FIGURE A.4 – Chapitre 3, expérience 3 : mouvements de tête par condition statique ou dynamique, pour chaque sujet.

ANNEXE B : STATISTIQUES BAYÉSIENNES

Cette annexe présente brièvement le principe des statistiques bayésiennes et leur apport par rapport aux statistiques fréquentistes.

Les tests bayésiens comparent la probabilité qu'une hypothèse H_1 explique mieux la distribution des données que l'hypothèse nulle H_0 . Prenons l'exemple d'un test t bayésien cherchant à voir si une condition expérimentale a induit des différences entre deux sets de données. Les hypothèses sont :

- H_0 : la condition expérimentale n'a pas affecté les données (hypothèse nulle)
- H_1 : la condition expérimentale a affecté les données. Autrement dit, les différences entre les données des deux sets sont dues à la condition expérimentale.

Au regard de la distribution des données, la probabilité p que chaque hypothèse explique les données est calculée. À partir de ces probabilités, le facteur de Bayes (BF pour "Bayes Factor") est calculé. Ce facteur compare la probabilité de deux hypothèses. Par exemple, BF_{10} ¹ compare la probabilité p_1 que l'hypothèse H_1 explique les données à la probabilité p_0 de l'hypothèse nulle H_0 . Un $BF_{10} > 1$ est alors la preuve que l'hypothèse H_1 explique mieux les données que l'hypothèse H_0 . Autrement dit, un $BF_{10} > 1$ est la preuve que la condition expérimentale a affecté les données. À l'inverse, un $BF_{10} < 1$ est la preuve que la condition expérimentale n'a pas affecté les données. Autrement dit, un $BF_{10} < 1$ est la preuve de l'absence d'effet de la condition expérimentale.²

Plus la valeur de BF est élevée, plus la preuve est forte. Jeffreys [204] a établi une échelle permettant de quantifier la force de la preuve obtenue en fonction de la valeur de BF (cf. tableau B.1).

1. À l'inverse, BF_{01} compare H_0 par rapport à H_1 . $BF_{01} = 1/BF_{10}$

2. Il est important de noter que cet exemple de comparaison par paire n'inclut que deux hypothèses, dans le cas d'une ANOVA par exemple, il y aura autant d'hypothèses qu'il n'y a de variable dépendante est donc autant de BF que de comparaisons possibles entre variables.

Les tests fréquentistes permettent uniquement de valider ou de rejeter une hypothèse³. Les tests bayésiens ont pour avantage de permettre une appréciation plus nuancée de la validité d'une hypothèse.⁴

BF_{10}	Preuve
> 100	Preuve extrême pour H_1
30 - 100	Preuve très forte pour H_1
10 - 30	Preuve forte pour H_1
3 - 10	Preuve modérée pour H_1
1 - 3	Preuve anecdotique pour H_1
1	Aucune preuve
1 - 0,33	Preuve anecdotique pour H_0
0,33 - 0,1	Preuve modérée pour H_0
0,1 - 0,033	Preuve forte pour H_0
0,033 - 0,001	Preuve très forte pour H_0
< 0,001	Preuve extrême pour H_0

TABLE B.1 – Echelle de BF .

3. valeur de $p < 0,05$ = hypothèse validée ; valeur de $p > 0,05$ = hypothèse rejetée

4. $BF_{10} > 3$ = preuve modérée : l'hypothèse est probablement valide, mais à prendre avec précaution ; $BF_{10} > 100$ = preuve extrêmement forte : il n'y a aucun doute que l'hypothèse est valide.

ANNEXE C : ANALYSES SUPPLÉMENTAIRES DES DONNÉES EEG

C.1 Méthode

Cette annexe présente les analyses supplémentaires effectuées pour tenter de réduire l'effet du décalage de la baseline dans l'étude présentée dans le chapitre 2. Les données ont été filtrées avec des filtres passe-haut avec des fréquences de coupure à 1 Hz, 2 Hz et 3 Hz. La figure C.1 permet de comparer les ERPs des 3 conditions de mouvement (SF, SM et DM) pour chaque filtre.

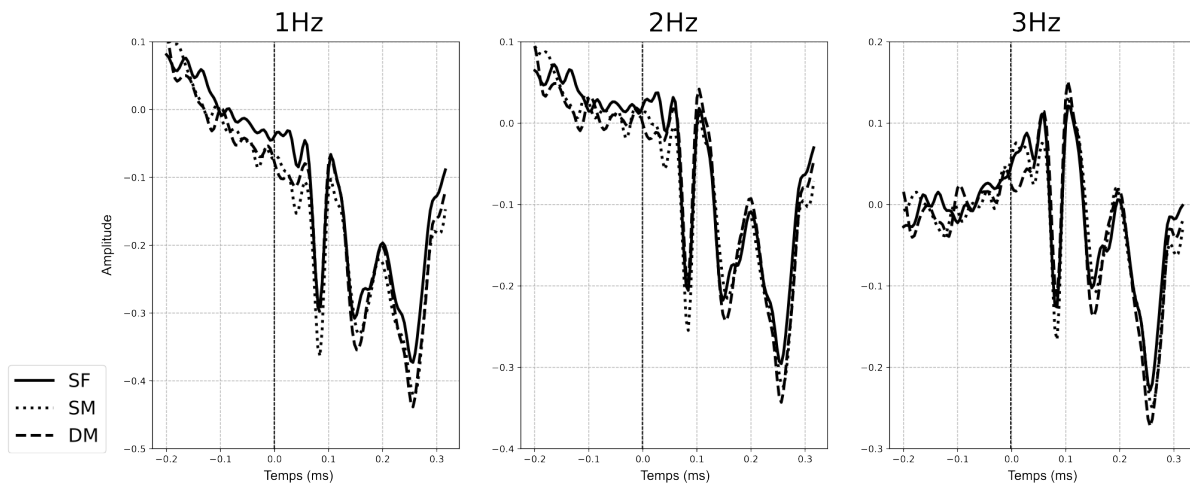


FIGURE C.1 – ERPs des trois conditions de mouvement (SF, SM et DM), enregistrés sur l'électrode Cz, avec des filtres passe-haut à 1 Hz, 2 Hz et 3 Hz. Unités d'amplitude arbitraires

Pour les données issues de chaque filtre, la même analyse statistique que pour les données filtrées à 0,1 Hz (correspondant aux données présentées dans le chapitre 2) a été réalisée. Pour les fenêtres temporelles des composantes N1, P2 et cP2, les valeurs maximales d'amplitude et

de latence ont été extraites. Pour chaque ensemble de valeurs, une ANOVA à mesures répétées a été réalisée (conditions de mouvement [*SF*, *SM*, *DM*] × groupes de répétitions [*groupe de répétitions 1*, *groupe de répétitions 2*, *groupe de répétitions 3*]). Des tests *post-hoc* ont été effectués sur les comparaisons significatives. La plupart des comparaisons qui étaient significatives pour les données originales filtrées à 0,1 Hz l'étaient également pour les données filtrées à 1 Hz, 2 Hz et 3 Hz.

C.2 Résultats

Le tableau C.1 compare les valeurs de F et les valeurs de p des ANOVA pour les données originales (0,1 Hz) avec les valeurs de F et les valeurs de p des ANOVA pour les données filtrées avec des fréquences de coupure plus élevées. Chaque fois qu'une comparaison était significative pour les données filtrées à 1 Hz, 2 Hz et 3 Hz, un test *post-hoc* a été effectué pour voir si l'effet était similaire à celui trouvé pour les données filtrées à 0,1 Hz, ce qui a toujours été le cas. Pour la N1, l'amplitude de *SM* était plus négative que celles de *SF* et de *DM* pour toutes les valeurs de filtrage. Pour la latence de la P2, la latence du *groupe de répétitions 1* était plus courte que celles des *groupes de répétition 2* et *3* pour toutes les valeurs de filtre. Pour l'amplitude de la P2, l'amplitude du *groupe de répétitions 1* était plus négative que celles des *groupes de répétition 2* et *3* pour les données filtrées à 1 Hz. Pour les données filtrées à 2 Hz et 3 Hz, cet effet n'apparaissait pas du tout. En ce qui concerne l'amplitude de la cP2, pour *SM* et *DM*, l'amplitude du *groupe de répétitions 1* était plus négative que celles des *groupes de répétition 2* et *3* pour les données filtrées à 1 Hz et 2 Hz. Pour les données filtrées à 3 Hz, cet effet existait également, mais toutes les comparaisons n'étaient pas significatives dans le test *post-hoc*. Aucune autre comparaison n'était significative pour les données filtrées à 1 Hz, 2 Hz et 3 Hz.

Fréquences	Amplitude N1		Latence P2		Amplitude P2		Amplitude cP2	
	F	p	F	p	F	p	F	p
0.1 Hz	5.010	0.014	8.300	0.001	8.670	0.001	16.730	0.001
1 Hz	3.429	0.047	7.615	0.002	6.766	0.004	3.707	0.010
2 Hz	5.200	0.012	9.396	0.001	0.603	0.554	8.457	0.001
3 Hz	4.900	0.015	7.039	0.003	0.069	0.934	4.232	0.005

TABLE C.1 – Résultats des ANOVAs pour différents filtres passe-haut effectuées sur la latence et l'amplitude des composantes d'intérêt de l'étude du chapitre 2.

L'analyse effectuée sur les données filtrées avec des fréquences de coupure plus élevées montre un schéma de résultats similaire à celui trouvé avec les données originales (filtrées à 0,1 Hz). En filtrant les données avec des filtres passe-haut destructifs (2 Hz et 3 Hz), certaines différences d'amplitude ont été perdues, mais les tendances principales ont été conservées. Ces résultats suggèrent que la dérive négative qui apparaît sur les données originales n'a pas affecté de manière significative les résultats originaux, à l'exception, peut-être, de l'amplitude de la P2.

BIBLIOGRAPHIE

1. STINSON, M. R. & LAWTON, B. W., Specification of the geometry of the human ear canal for the prediction of sound-pressure level distribution, en, *The Journal of the Acoustical Society of America* **85**, 2492-2503, ISSN : 0001-4966, 1520-8524, <https://pubs.aip.org/jasa/article/85/6/2492/798846/Specification-of-the-geometry-of-the-human-ear> (2024) (juin 1989).
2. PUJOL, R., *Audition - Oreille - Cochlée* fr, <http://www.cochlea.eu> (2024).
3. *Aperçu de la structure et de la fonction du neurone (leçon)* fr, <https://fr.khanacademy.org/science/biologie-a-l-ecole/x5047ff3843d876a6:bio-5e-annee-sciences-de-base/x5047ff3843d876a6:bio-5-1h-le-neurone-et-le-systeme-nerveux-description/a/overview-of-neuron-structure-and-function> (2024).
4. SIMON, PERROT, X. & MERTENS, P., Anatomie fonctionnelle du nerf cochléaire et du système auditif central, fr, *Neurochirurgie* **55**, 120-126, ISSN : 00283770, <https://linkinghub.elsevier.com/retrieve/pii/S002837700900037X> (2024) (avr. 2009).
5. KAWASE, T. & TAKASAKA, T., The effects of contralateral noise on masked compound action potential in humans, en, *Hearing Research* (1995).
6. MICHEYL, C., PERROT, X. & COLLET, L., Relationship between auditory intensity discrimination in noise and olivocochlear efferent system activity in humans. eng, *Behavioral neuroscience* **111**, Place : United States, 801-807, ISSN : 0735-7044 (août 1997).
7. MAY, B. J., BUDELIS, J. & NIPARKO, J. K., Behavioral Studies of the Olivocochlear Efferent System : Learning to Listen in Noise, en, *Archives of Otolaryngology-Head & Neck Surgery* **130**, 660, ISSN : 0886-4470, <http://archotol.jamanetwork.com/article.aspx?doi=10.1001/archotol.130.5.660> (2024) (mai 2004).
8. KANDEL, E., SCHWARTZ, J., JESSELL, T., SIEGELBAUM, S. & HUDSPETH, A., *Principles of neural science* (McGraw-hill, 2012).

-
9. NUNEZ, P. L. & SRINIVASAN, R., *Electric Fields of the Brain : The neurophysics of EEG* ISBN : 978-0-19-505038-7, <https://doi.org/10.1093/acprof:oso/9780195050387.001.0001> (Oxford University Press, jan. 2006).
 10. RÖTTGER, S., SCHRÖGER, E., GRUBE, M., GRIMM, S. & RÜBSAMEN, R., Mismatch negativity on the cone of confusion, en, *Neuroscience Letters* **414**, 178-182, ISSN : 03043940, <https://linkinghub.elsevier.com/retrieve/pii/S0304394006013371> (2021) (mars 2007).
 11. HAENSCHEL, C., Event-Related Brain Potential Correlates of Human Auditory Sensory Memory-Trace Formation, en, *Journal of Neuroscience* **25**, 10494-10501, ISSN : 0270-6474, 1529-2401, <https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.1227-05.2005> (2022) (nov. 2005).
 12. GETZMANN, S. & LEWALD, J., Shared Cortical Systems for Processing of Horizontal and Vertical Sound Motion, en, *Journal of Neurophysiology* **103**, 1896-1904, ISSN : 0022-3077, 1522-1598, <https://www.physiology.org/doi/10.1152/jn.00333.2009> (2021) (avr. 2010).
 13. DUCOMMUN, C. Y. *et al.*, Segregated Processing of Auditory Motion and Auditory Location : An ERP Mapping Study, en, *NeuroImage* **16**, 76-88, ISSN : 10538119, <https://linkinghub.elsevier.com/retrieve/pii/S1053811902910621> (2021) (mai 2002).
 14. GAMBLE, M. L. & LUCK, S. J., N2ac : An ERP component associated with the focusing of attention within an auditory scene : The N2ac component, en, *Psychophysiology* **48**, 1057-1068, ISSN : 00485772, <https://onlinelibrary.wiley.com/doi/10.1111/j.1469-8986.2010.01172.x> (2021) (août 2011).
 15. NÄÄTÄNEN, R. *et al.*, Frequency and location specificity of the human vertex NI wave, en, *Electroencephalography and clinical Neurophysiology*, 523-531 (1987).
 16. ALTMANN, C. F., BLEADOWSKI, C., WIBRAL, M. & KAISER, J., Processing of location and pattern changes of natural sounds in the human auditory cortex, en, *NeuroImage* **35**, 1192-1200, ISSN : 10538119, <https://linkinghub.elsevier.com/retrieve/pii/S1053811907000365> (2021) (avr. 2007).
 17. GETZMANN, S. & LEWALD, J., Effects of natural versus artificial spatial cues on electrophysiological correlates of auditory motion, en, *Hearing Research* **259**, 44-54, ISSN :

-
- 03785955, <https://linkinghub.elsevier.com/retrieve/pii/S0378595509002457> (2021) (jan. 2010).
18. BERGER, H., Über das Elektrenkephalogramm des Menschen, *Archiv für Psychiatrie und Nervenkrankheiten* **87**, 527-570, ISSN : 1433-8491, <https://doi.org/10.1007/BF01797193> (déc. 1929).
 19. *MNE — MNE 1.7.0 documentation* <https://mne.tools/stable/index.html> (2024).
 20. *Neurolite en*, <https://shop.neurolite.ch/en/> (2024).
 21. *Ambu, électrodes cupules fr*, <https://www.ambu.fr/neurologie/contamination-croisee/electrodes-cupules-y-a-t-il-un-risque-pour-vos-patients> (2024).
 22. *Subdermal Needle Electrode en-US*, <https://www.spesmedica.com/en/eeg/subdermal-needle-electrode/> (2024).
 23. JASPER, H. H., The Ten-Twenty Electrode System of the International Federation, *Electroencephalography and Clinical Neurophysiology*, 371-375 (1958).
 24. KLEM, G. H., LÜDERS, H. O., JASPER, H. H. & ELGER, C, in *Recommendations for the Practice of Clinical Neurophysiology : Guidelines of the International Federation of Clinical Physiology* (1999).
 25. *Electrode arrangement according to the international 10/20 system en*, <https://www.ternimed.de/EEG-Electrode-arrangement-according-to-the-international-10/20-system> (2024).
 26. KANOGA, S. & MITSUKURA, Y., en, in *Electroencephalography* (éd. SITTIPRAPAPORN, W.) (InTech, nov. 2017), ISBN : 978-953-51-3637-8 978-953-51-3638-5, <http://www.intechopen.com/books/electroencephalography/review-of-artifact-rejection-methods-for-electroencephalographic-systems> (2024).
 27. COOPER, R., OSSELTON, J. W. & SHAW, J. C., *Elektroenzephalographie : Technik und Methoden* Stuttgart : Gustav Fischer Verlag (1984).
 28. HASTING, A., Syntax in a blink : Early and automatic processing of syntactic rules as revealed by event-related brain potentials, *MPI Series in Human Cognitive and Brain Sciences*, v.97 (2008) (2008).

-
29. LUCK, S. J., *An Introduction to the Event-Related Potential Technique* 2nd edition, ISBN : 978-0-262-52585-5 (MIT Press, 2014).
 30. WILLIAMS, H. L., TEPAS, D. I. & MORLOCK, H. C., Evoked Responses to Clicks and Electroencephalographic Stages of Sleep in Man, en, *Science* **138**, 685-686, ISSN : 0036-8075, 1095-9203, <https://www.science.org/doi/10.1126/science.138.3541.685> (2024) (nov. 1962).
 31. HILLYARD, S. A., HINK, R. F., SCHWENT, V. L. & PICTON, T. W., Electrical signs of selective attention in the human brain. eng, *Science (New York, N.Y.)* **182**, Place : United States, 177-180, ISSN : 0036-8075 (oct. 1973).
 32. NÄÄTÄNEN, R. & PICTON, T., The N1 Wave of the Human Electric and Magnetic Response to Sound : A Review and an Analysis of the Component Structure, en, *Psychophysiology* **24**, 375-425, ISSN : 0048-5772, 1469-8986, <https://onlinelibrary.wiley.com/doi/10.1111/j.1469-8986.1987.tb00311.x> (2021) (juill. 1987).
 33. HACKLEY, S. A., WOLDORFF, M. & HILLYARD, S. A., Cross-modal selective attention effects on retinal, myogenic, brainstem, and cerebral evoked potentials. eng, *Psychophysiology* **27**, Place : United States, 195-208, ISSN : 0048-5772 (mars 1990).
 34. SCHWARTZE, M., FARRUGIA, N. & KOTZ, S. A., Dissociation of formal and temporal predictability in early auditory evoked potentials, en, *Neuropsychologia* **51**, 320-325, ISSN : 00283932, <https://linkinghub.elsevier.com/retrieve/pii/S0028393212004095> (2024) (jan. 2013).
 35. in, *Electrophysiology of Mind : Event-related Brain Potentials and Cognition* 1-26 (Oxford University Press, sept. 1996), ISBN : 978-0-19-852416-8, <https://doi.org/10.1093/acprof:oso/9780198524168.001.0001>.
 36. NÄÄTÄNEN, R., SIMPSON, M. & LOVELESS, N., Stimulus deviance and evoked potentials, en, *Biological Psychology* **14**, 53-98, ISSN : 03010511, <https://linkinghub.elsevier.com/retrieve/pii/0301051182900175> (2024) (fév. 1982).
 37. NÄÄTÄNEN, R., GAILLARD, A. W. K. & MÄNTYSALO, S., Early selective-attention effect on evoked potential reinterpreted, *Acta Psychologica* **42**, 313-329, ISSN : 0001-6918, <https://www.sciencedirect.com/science/article/pii/0001691878900069> (1978).
 38. NÄÄTÄNEN, R., PAAVILAINEN, P., TIITINEN, H., JIANG, D. & ALHO, K., Attention and mismatch negativity. eng, *Psychophysiology* **30**, Place : United States, 436-450, ISSN : 0048-5772 (sept. 1993).

-
39. SUTTON, S., BRAREN, M., ZUBIN, J. & JOHN, E. R., Evoked-potential correlates of stimulus uncertainty. eng, *Science (New York, N.Y.)* **150**, Place : United States, 1187-1188, ISSN : 0036-8075 (nov. 1965).
 40. DONCHIN, E., Surprise ! ... Surprise ?, *Psychophysiology* (1981).
 41. HRUBY, T. & MARSALEK, P., Event-Related Potentials - the P3 Wave, en, *Acta Neurobiologiae Experimentalis* **63**, 55-63, ISSN : 1689-0035, 0065-1400, <https://ane.pl/index.php/ane/article/view/1455> (2024) (mars 2003).
 42. SCHRÖGER, E., A Neural Mechanism for Involuntary Attention Shifts to Changes in Auditory Stimulation, en, *Journal of Cognitive Neuroscience* **8**, 527-539, ISSN : 0898-929X, 1530-8898, <https://direct.mit.edu/jocn/article/8/6/527/3233/A-Neural-Mechanism-for-Involuntary-Attention> (2024) (nov. 1996).
 43. SCHRÖGER, E. & WOLFF, C., Attentional orienting and reorienting is indicated by human event-related brain potentials : en, *NeuroReport* **9**, 3355-3358, ISSN : 0959-4965, <http://journals.lww.com/00001756-199810260-00003> (2024) (oct. 1998).
 44. SCHRÖGER, E. & WOLFF, C., Behavioral and electrophysiological effects of task-irrelevant sound change : a new distraction paradigm, en, *Cognitive Brain Research* **7**, 71-87, ISSN : 09266410, <https://linkinghub.elsevier.com/retrieve/pii/S0926641098000135> (2024) (juill. 1998).
 45. POLICH, J., Updating P300 : An integrative theory of P3a and P3b, en, *Clinical Neurophysiology* **118**, 2128-2148, ISSN : 13882457, <https://linkinghub.elsevier.com/retrieve/pii/S1388245707001897> (2024) (oct. 2007).
 46. MARTIN, B. A. & BOOTHROYD, A., Cortical, auditory, event-related potentials in response to periodic and aperiodic stimuli with the same spectral envelope, *Ear and Hearing* (1999).
 47. JONES, S. & PEREZ, N., The auditory 'C-process' : analyzing the spectral envelope of complex sounds, en, *Clinical Neurophysiology* **112**, 965-975, ISSN : 13882457, <https://linkinghub.elsevier.com/retrieve/pii/S1388245701005156> (2022) (juin 2001).
 48. DIMITRIJEVIC, A. *et al.*, Intensity changes in a continuous tone : Auditory cortical potentials comparison with frequency changes, en, *Clinical Neurophysiology* **120**, 374-383, ISSN : 13882457, <https://linkinghub.elsevier.com/retrieve/pii/S1388245708012510> (2024) (fév. 2009).

-
49. DIMITRIJEVIC, A. *et al.*, Auditory cortical N100 in pre- and post-synaptic auditory neuropathy to frequency or intensity changes of continuous tones, en, *Clinical Neurophysiology* **122**, 594-604, ISSN : 13882457, <https://linkinghub.elsevier.com/retrieve/pii/S1388245710006279> (2024) (mars 2011).
 50. HARRIS, K. C., MILLS, J. H. & DUBNO, J. R., Electrophysiologic correlates of intensity discrimination in cortical evoked potentials of younger and older adults. eng, *Hearing research* **228**, Place : Netherlands, 58-68, ISSN : 0378-5955 (juin 2007).
 51. MARTIN, B. A. & BOOTHROYD, A., Cortical, auditory, evoked potentials in response to changes of spectrum and amplitude, en, *The Journal of the Acoustical Society of America* **107**, 2155-2161, ISSN : 0001-4966, <http://asa.scitation.org/doi/10.1121/1.428556> (2022) (avr. 2000).
 52. FAN, Z.-T. *et al.*, Acoustic Change Complex Evoked by Horizontal Sound Location Change in Young Adults With Normal Hearing, en, *Frontiers in Neuroscience* **16**, 908989, ISSN : 1662-453X, <https://www.frontiersin.org/articles/10.3389/fnins.2022.908989/full> (2023) (juin 2022).
 53. ROHRBAUGH, J. W., DONCHIN, E. & ERIKSEN, C. W., Decision making and the P300 component of the cortical evoked response, en, *Perception & Psychophysics* **15**, 368-374, ISSN : 0031-5117, 1532-5962, <http://link.springer.com/10.3758/BF03213960> (2024) (mars 1974).
 54. CAMPBELL, T., WINKLER, I. & KUJALA, T., N1 and the mismatch negativity are spatiotemporally distinct ERP components : Disruption of immediate memory by auditory distraction can be related to N1, en, *Psychophysiology* **44**, 530-540, ISSN : 0048-5772, 1469-8986, <https://onlinelibrary.wiley.com/doi/10.1111/j.1469-8986.2007.00529.x> (2022) (juill. 2007).
 55. BUCHWALD, J. S. *et al.*, Midlatency auditory evoked responses : PI abnormalities in adult autistic subjects, en, *Electroencephalography and clinical Neurophysiology* (1992).
 56. HARI, R., SAMS, M. & JÄRVILEHTO, T., Auditory evoked transient and sustained potentials in the human EEG : I. Effects of expectation of stimuli, en, *Psychiatry Research* **1**, 297-306, ISSN : 01651781, <https://linkinghub.elsevier.com/retrieve/pii/0165178179900118> (2024) (déc. 1979).

-
57. McCALLUM, W., FARMER, S. F. & POCOCK, P. V., The effects of physical and semantic incongruities on auditory event-related potentials, *Electroencephalography and clinical Neurophysiology* (1984).
 58. ALTMANN, C. F. *et al.*, Categorical speech perception during active discrimination of consonants and vowels, en, *Neuropsychologia* **64**, 13-23, ISSN : 00283932, <https://linkinghub.elsevier.com/retrieve/pii/S002839321400311X> (2024) (nov. 2014).
 59. BRATTICO, E., JACOBSEN, T., DE BAENE, W., GLERAN, E. & TERVANIEMI, M., Cognitive vs. affective listening modes and judgments of music – An ERP study, en, *Biological Psychology* **85**, 393-409, ISSN : 03010511, <https://linkinghub.elsevier.com/retrieve/pii/S0301051110002322> (2024) (déc. 2010).
 60. POIKONEN, H. *et al.*, Event-related brain responses while listening to entire pieces of music, en, *Neuroscience* **312**, 58-73, ISSN : 03064522, <https://linkinghub.elsevier.com/retrieve/pii/S0306452215009896> (2024) (jan. 2016).
 61. GUILLON, P., *Individualisation des indices spectraux pour la synthèse binaurale : recherche et exploitation des similarités inter-individuelles pour l'adaptation ou la reconstruction de HRTF* thèse de doct. (Université du Maine, 2009).
 62. RAYLEIGH, L., XII. *On our perception of sound direction*, en, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **13**, 214-232, ISSN : 1941-5982, 1941-5990, <https://www.tandfonline.com/doi/full/10.1080/14786440709463595> (2024) (fév. 1907).
 63. WOODWORTH, R. S. & SCHLOSBERG, H., En, in *Experimental psychology* Holt, Rinehart and Winston, 349 -361 (New York, 1954).
 64. BERTHOMIEU, G., *Influence de la position d'une source sur le niveau sonore perçu* fr, thèse de doct. (Université de Bretagne Occidentale, 2019).
 65. MOORE, B. C. J., *An introduction to the psychology of hearing* BRILL, ISBN : ISBN 978-1-78052-038-4. (Leiden, Netherlands, 2012).
 66. MILLS, A. W., in *Foundations of modern auditory theory* Academic Press, 303 -348 (New York, 1972), ISBN : 978-0-12-691901-1.
 67. STERN, R. M., BROWN, J. G. & WANG, D., in *Computational auditory scene analysis* IEEE Press Editorial Board, 147 -185 (A JOHN WILEY & SONS, INC., 2006).

-
68. HARTMANN, W. M., How We Localize Sound, en, *Physics Today* **52**, 24-29, ISSN : 0031-9228, 1945-0699, <https://pubs.aip.org/physicstoday/article/52/11/24/410870/How-We-Localize-SoundRelying-on-a-variety-of-cues> (2024) (nov. 1999).
69. MIDDLEBROOKS, J. C. & GREEN, D. M., Sound Localization by Human Listeners, en, *Annual Review of Psychology* (1991).
70. WENZEL, E. M., BERGAULT, D. R. & GODFROY-COOPER, M., en, in *Immersive sound : the art and science of binaural and multi-channel audio* (éd. ROGINSKA, A. & GELUSO, P.) First published, 5 -40 (Routledge, Taylor & Francis Group, New York London, 2018), ISBN : 978-1-138-90001-1 978-1-138-90000-4.
71. WALLACH, H., On Sound Localization, en, *Journal of the Acoustical Society of America* (1939).
72. VAN SOEST, P. J., Rictingshooren bij sinusoornige geluidstrillingen, *Physica* (1929).
73. BURGER, J. F., Front-back discrimination of the hearing system, *Acustica* (1958).
74. WIGHTMAN, F. L. & KISTLER, D. J., Resolution of front-back ambiguity in spatial hearing by listener and source movement, en, *The Journal of the Acoustical Society of America* **105**, 2841-2853, ISSN : 0001-4966, 1520-8524, <https://pubs.aip.org/jasa/article/105/5/2841/554518/Resolution-of-front-back-ambiguity-in-spatial> (2024) (mai 1999).
75. SHINN-CUNNINGHAM, B. G., SANTARELLI, S. & KOPCO, N., Tori of confusion : Binaural localization cues for sources within reach of a listener, en, *The Journal of the Acoustical Society of America* **107**, 1627-1636, ISSN : 0001-4966, 1520-8524, <https://pubs.aip.org/jasa/article/107/3/1627/554959/Tori-of-confusion-Binaural-localization-cues-for> (2024) (mars 2000).
76. PLACK, C. J., *The Sense of Hearing* 4 th, en, ISBN : 978-1-03-229951-8 (Routledge, New York, NY, 2024).
77. SHAW, E. A. G. & TERANISHI, R., Sound Pressure Generated in an External-Ear Replica and Real Human Ears by a Nearby Point Source, en, *The Journal of the Acoustical Society of America* **44**, 240-249, ISSN : 0001-4966, 1520-8524, <https://pubs.aip.org/jasa/article/44/1/240/760466/Sound-Pressure-Generated-in-an-External-Ear> (2024) (juill. 1968).

-
78. SHAW, E. A. G., En, in *Binaural and Spatial Hearing in Real and Virtual Environments* (éd. GILKEY, R. H. & ANDERSON, T. R.) 1st, 25 -47 (Psychology Press, New York, NY, 1997), ISBN : 978-1-315-80634-1.
 79. ALGAZI, V. R., DUDA, R. O., DURAISWAMI, R., GUMEROV, N. A. & TANG, Z., Approximating the head-related transfer function using simple geometric models of the head and torso, en, *The Journal of the Acoustical Society of America* **112**, 2053-2064, ISSN : 0001-4966, 1520-8524, <https://pubs.aip.org/jasa/article/112/5/2053/545972/Approximating-the-head-related-transfer-function> (2024) (nov. 2002).
 80. BUTLER, R. A. & BELENDIUK, K., Spectral cues utilized in the localization of sound in the median sagittal plane, en, *The Journal of the Acoustical Society of America* **61**, 1264-1269, ISSN : 0001-4966, 1520-8524, <https://pubs.aip.org/jasa/article/61/5/1264/626721/Spectral-cues-utilized-in-the-localization-of> (2024) (mai 1977).
 81. BATTEAU, D. W., The role of the pinna in human localization, en, *Proceedings of the Royal Society B* (1967).
 82. STEVENS, S. S. & GUIRAO, M., Loudness, Reciprocity, and Partition Scales, en, *The Journal of the Acoustical Society of America* **34**, 1466-1471, ISSN : 0001-4966, 1520-8524, <https://pubs.aip.org/jasa/article/34/9B/1466/640174/Loudness-Reciprocity-and-Partition-Scales> (2024) (sept. 1962).
 83. ZAHORIK, P., Assessing auditory distance perception using virtual acoustics, en, *The Journal of the Acoustical Society of America* **111**, 1832-1846, ISSN : 0001-4966, 1520-8524, <https://pubs.aip.org/jasa/article/111/4/1832/547181/Assessing-auditory-distance-perception-using> (2024) (avr. 2002).
 84. VON BEKESY, G., *Experiments in hearing* en (Mcgraw Hill Book Company, 1960).
 85. MERSHON, D. H. & KING, L. E., Intensity and reverberation as factors in the auditory perception of egocentric distance, en, *Perception & Psychophysics* (1975).
 86. LARSEN, E., IYER, N., LANSING, C. R. & FENG, A. S., On the minimum audible difference in direct-to-reverberant energy ratioa), en, *J. Acoust. Soc. Am.* **124** (2008).
 87. BRUNGART, D. S., DURLACH, N. I. & RABINOWITZ, W. M., Auditory localization of nearby sources. II. Localization of a broadband source, en, *The Journal of the Acoustical Society of America* **106**, 1956-1968, ISSN : 0001-4966, 1520-8524, <https://pubs.aip.org/>

-
- [jasa/article/106/4/1956/556634/Auditory-localization-of-nearby-sources-II](https://pubs.aip.org/jasa/article/106/4/1956/556634/Auditory-localization-of-nearby-sources-II) (2024) (oct. 1999).
88. BRUNGART, D. S. & RABINOWITZ, W. M., Auditory localization of nearby sources. Head-related transfer functions, en, *The Journal of the Acoustical Society of America* **106**, 1465-1479, ISSN : 0001-4966, 1520-8524, <https://pubs.aip.org/jasa/article/106/3/1465/553213/Auditory-localization-of-nearby-sources-Head> (2024) (sept. 1999).
89. BRUNGART, D., *Auditory parallax effects in the HRTF for nearby sources* en, in *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. WASPAA'99 (Cat. No.99TH8452)* (IEEE, New Paltz, NY, USA, 1999), 171-174, ISBN : 978-0-7803-5612-2, <http://ieeexplore.ieee.org/document/810877/> (2024).
90. INGÅRD, U., A Review of the Influence of Meteorological Conditions on Sound Propagation, en, *The Journal of the Acoustical Society of America* **25**, 405-411, ISSN : 0001-4966, 1520-8524, <https://pubs.aip.org/jasa/article/25/3/405/623238/A-Review-of-the-Influence-of-Meteorological> (2024) (mai 1953).
91. BLAUERT, J., in *Spatial Hearing : The Psychophysics of Human Sound Localization* MIT press (1996), ISBN : 978-0-262-26868-4.
92. LITTLE, A. D., MERSHON, D. H. & COX, P. H., Spectral Content as a Cue to Perceived Auditory Distance, en, *Perception* **21**, 405-416, ISSN : 0301-0066, 1468-4233, <http://journals.sagepub.com/doi/10.1068/p210405> (2024) (juin 1992).
93. BLAUERT, J., Ein versuch zum richtungshören bei gleichzeitiger optischer stimulation (an experiment in directional hearing with simultaneous optical stimulation), *Acustica* (1970).
94. CARLILE, S., LEONG, P. & HYAMS, S., The nature and distribution of errors in sound localization by human listeners, en, *Hearing Research* **114**, 179-196, ISSN : 03785955, <https://linkinghub.elsevier.com/retrieve/pii/S0378595597001615> (2024) (déc. 1997).
95. *LISTEN HRTF DATABASE* <http://recherche.ircam.fr/equipes/salles/listen/index.html> (2024).

-
96. PERNAUX, J. M., *Spatialisation du son par les techniques binaurales : application aux services de télécommunications*. thèse de doct. (Institut National Polytechnique, Grenoble, 2003).
 97. KATZ, B., *Measurement and calculation of individual head-related transfer function using a boundary element model including the measurement and effect of skin and hair impedance* thèse de doct. (Pennsylvania State University, 1998).
 98. KAHANA, Y., *Numerical modelling of the head-related transfer function* thèse de doct. (University of Southampton, 2000).
 99. MCANALLY, K. & MARTIN, R., Variability in the headphone-to-ear-canal transfer function, *Journal of the Audio Engineering Society* **50**, 263-266 (2002).
 100. KULKARNI, A. & COLBURN, H. S., Variability in the characterization of the headphone transfer-function, en, *The Journal of the Acoustical Society of America* **107**, 1071-1074, ISSN : 0001-4966, 1520-8524, <https://pubs.aip.org/jasa/article/107/2/1071/555412/Variability-in-the-characterization-of-the> (2024) (fév. 2000).
 101. PAQUIER, M. & KOEHL, V., Discriminability of the placement of supra-aural and circumaural headphones, en, *Applied Acoustics* **93**, 130-139, ISSN : 0003682X, <https://linkinghub.elsevier.com/retrieve/pii/S0003682X15000341> (2024) (juin 2015).
 102. *Binaural Microphone - In-Ear Microphone | Type 4101 | Brüel & Kjør* <https://www.bksv.com/en/transducers/acoustic/binaural/binaural-microphone> (2024).
 103. *3Dio : Professional Binaural Microphones* <https://3diosound.com/> (2024).
 104. *KU 100 - Dummy Head* <https://www.neumann.com/en-de/products/microphones/ku-100/> (2024).
 105. *GRASSound&Vibration* <http://kemar.us/> (2024).
 106. CHRISTENSEN, F., MØLLER, H., MINNAAR, P., POLGSTIES, J. & OLESEN, S., *Interpolating between head-related transfer functions measured with low directional resolution* en, in *Proceedings of the 107th Audio Engineering Society Convention* (New York, NY, 1999).
 107. HARTUNG, K., BRAASCH, J. & STERBING, S. J., *Comparison of Different Methods for the Interpolation of Head-Related Transfer Functions in Audio Engineering Society Conference : 16th International Conference : Spatial Sound Reproduction* (mars 1999), <https://www.aes.org/e-lib/browse.cfm?elib=8026>.

-
108. NISHINO, T., KAJITA, S., TAKEDA, K. & ITAKURA, F., *Interpolating head related transfer functions in the median plane* en, in *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. WASPAA'99 (Cat. No.99TH8452)* (IEEE, New Paltz, NY, USA, 1999), 167-170, ISBN : 978-0-7803-5612-2, <http://ieeexplore.ieee.org/document/810876/> (2024).
109. LARCHER, V., *Techniques de spatialisation du son pour la réalité virtuelle* thèse de doct. (Université de Paris VI, 2001).
110. SANDVAD, J., *Dynamic Aspects of Auditory Virtual Environments* in *Audio Engineering Society Convention 100* (Copenhagen, Denmark, mai 1996), <https://www.aes.org/e-lib/browse.cfm?elib=7547>.
111. BRUNGART, D. S., KORDIK, A. J. & SIMPSON, B. D., Effects of Headtracker Latency in Virtual Audio Displays, en, *J. Audio Eng. Soc.* **54** (2006).
112. WENZEL, E. M., Effect of increasing system latency on localization of virtual sounds with short and long duration, en (2001).
113. *Immersive sound : the art and science of binaural and multi-channel audio* First published, en (éd. ROGINSKA, A. & GELUSO, P.) ISBN : 978-1-138-90001-1 978-1-138-90000-4 (Routledge, Taylor & Francis Group, New York London, 2018).
114. ARMSTRONG, C., THRESH, L., MURPHY, D. & KEARNEY, G., A Perceptual Evaluation of Individual and Non-Individual HRTFs : A Case Study of the SADIE II Database, en, *Applied Sciences* **8**, 2029, ISSN : 2076-3417, <http://www.mdpi.com/2076-3417/8/11/2029> (2024) (oct. 2018).
115. HUR, Y., LEE, S.-P., PARK, Y.-c. & YOUN, D.-h., *Efficient Individualization of HRTF Using Critical-Band Based Spectral Cues Control* in *Audio Engineering Society Convention 124* (mai 2008), <https://www.aes.org/e-lib/browse.cfm?elib=14577>.
116. MØLLER, H., SØRENSEN, M. F., JENSEN, C. B. & HAMMERSHØI, D., Binaural Technique : Do We Need Individual Recordings ?, *J. Audio Eng. Soc.* **44**, 451-469, <https://www.aes.org/e-lib/browse.cfm?elib=7897> (1996).
117. WENZEL, E. M., ARRUDA, M., KISTLER, D. J. & WIGHTMAN, F. L., Localization using non-individualized head-related transfer functions, en, *The Journal of the Acoustical Society of America* **94**, 111-123, ISSN : 0001-4966, <http://asa.scitation.org/doi/10.1121/1.407089> (2023) (juill. 1993).

-
118. BEGAULT, D. R. & WENZEL, E. M., Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source, en, *J. Audio Eng. Soc.* **49**, 14 (2001).
 119. KIM, S.-M. & CHOI, W., On the externalization of virtual sound images in headphone reproduction : A Wiener filter approach), en, *J. Acoust. Soc. Am.* **117**, 10 (2005).
 120. WERNER, S., KLEIN, F., MAYENFELS, T. & BRANDENBURG, K., *A summary on acoustic room divergence and its effect on externalization of auditory events* en, in *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)* (IEEE, Lisbon, Portugal, juin 2016), 1-6, ISBN : 978-1-5090-0354-9, <http://ieeexplore.ieee.org/document/7498973/> (2023).
 121. LECLÈRE, T., LAVANDIER, M. & PERRIN, F., On the externalization of sound sources with headphones without reference to a real source, en, *The Journal of the Acoustical Society of America* **146**, 2309-2320, ISSN : 0001-4966, <http://asa.scitation.org/doi/10.1121/1.5128325> (2021) (oct. 2019).
 122. HOFMAN, P. M., Relearning sound localization with new ears, en, *nature neuroscience* **1** (1998).
 123. BLUM, A, KATZ, B & WARUSFEL, O, *Eliciting adaptation to non-individual HRTF spectral cues with multi-modal training* en, in (Strasbourg, France, 2004).
 124. PARSEIHIAN, G. & KATZ, B. F. G., Rapid head-related transfer function adaptation using a virtual auditory environment, en, *The Journal of the Acoustical Society of America* **131**, 2948-2957, ISSN : 0001-4966, <http://asa.scitation.org/doi/10.1121/1.3687448> (2022) (avr. 2012).
 125. ZAHORIK, P, TAM, C, WANG, K, BANGAYAN, P & SUNDARESWARAN, V, Localization Accuracy In 3-D Sound Displays : The Role Of Visual-Feedback Training, en, *Proc. Advanced Displays and Interactive Displays ARL Fed. Lab. 5th Ann. Symp*, 17-22 (2002).
 126. HONDA, A. *et al.*, Transfer effects on sound localization performances from playing a virtual three-dimensional auditory game, en, *Applied Acoustics* **68**, 885-896, ISSN : 0003682X, <https://linkinghub.elsevier.com/retrieve/pii/S0003682X06001824> (2024) (août 2007).
 127. JEFFRESS, L. A. & TAYLOR, R. W., Lateralization vs Localization, *The Journal of the Acoustical Society of America* (1960).
 128. THOMPSON, S. P., On binaural audition, Part I, *Philosophical Magazine*, 274-277 (1977).

-
129. THOMPSON, S. P., On binaural audition, Part II, *Philosophical Magazine*, 383-391 (1978).
130. THOMPSON, S. P., On binaural audition, Part III, *Philosophical Magazine*, 351-355 (1981).
131. WEBER, E. H., Ueber die Umst€and durch welche man geleitet wird manche Empfindungen auf au ssere Objecte zu beziehen, *Berichte Ueber die Verhandlungen der Koeniglichen Saechsischen Gesellschaft der Wissenschaften (Mathematisch-Physische Klasse)* (1949).
132. BEST, V., BAUMGARTNER, R., LAVANDIER, M., MAJDAK, P. & KOPČO, N., Sound Externalization : A Review of Recent Research, en, *Trends in Hearing* **24**, 233121652094839, ISSN : 2331-2165, 2331-2165, <http://journals.sagepub.com/doi/10.1177/2331216520948390> (2021) (jan. 2020).
133. PLENGE, G., On the differences between localization and lateralization, en, *The Journal of the Acoustical Society of America* **56**, 944-951, ISSN : 0001-4966, 1520-8524, <https://pubs.aip.org/jasa/article/56/3/944/761749/On-the-differences-between-localization-and> (2024) (sept. 1974).
134. DURLACH, N. I. *et al.*, On the Externalization of Auditory Images, en, *Presence : Teleoperators and Virtual Environments* **1**, 251-257, ISSN : 1054-7460, <https://direct.mit.edu/pvar/article/1/2/251-257/58771> (2022) (jan. 1992).
135. SIMON, L. S. R., ZACHAROV, N. & KATZ, B. F. G., Perceptual attributes for the comparison of head-related transfer functions, en, *The Journal of the Acoustical Society of America* **140**, 3623-3632, ISSN : 0001-4966, 1520-8524, <https://pubs.aip.org/jasa/article/140/5/3623/679161/Perceptual-attributes-for-the-comparison-of-head> (2024) (nov. 2016).
136. BEGAULT, D. R. & WENZEL, E. M., Headphone Localization of Speech, en, *Human Factors : The Journal of the Human Factors and Ergonomics Society* **35**, 361-376, ISSN : 0018-7208, 1547-8181, <http://journals.sagepub.com/doi/10.1177/001872089303500210> (2024) (juin 1993).
137. BRIMIJOIN, W. O., HARTMA, W. O., BOYD, A. W. & AKEROYD, M. A., The Contribution of Head Movement to the Externalization and Internalization of Sounds, en, *PLoS ONE* **8** (éd. ALAIN, C.) e83068, ISSN : 1932-6203, <https://dx.plos.org/10.1371/journal.pone.0083068> (2022) (déc. 2013).

-
138. KATES, J. M., AREHART, K. H., MURALIMANOVAR, R. K. & SOMMERFELDT, K., Externalization of remote microphone signals using a structural binaural model of the head and pinna, en, *The Journal of the Acoustical Society of America* **143**, 2666-2677, ISSN : 0001-4966, 1520-8524, <https://pubs.aip.org/jasa/article/143/5/2666/671848/Externalization-of-remote-microphone-signals-using> (2024) (mai 2018).
139. HARTMANN, W. M. & WITTENBERG, A., On the externalization of sound images, en, *The Journal of the Acoustical Society of America* **99**, 3678-3688, ISSN : 0001-4966, <http://asa.scitation.org/doi/10.1121/1.414965> (2022) (juin 1996).
140. BOYD, A. W., WHITMER, W. M., SORAGHAN, J. J. & AKEROYD, M. A., Auditory externalization in hearing-impaired listeners : The effect of pinna cues and number of talkers, en, *The Journal of the Acoustical Society of America* **131**, EL268-EL274, ISSN : 0001-4966, 1520-8524, <https://pubs.aip.org/jasa/article/131/3/EL268/994322/Auditory-externalization-in-hearing-impaired> (2024) (mars 2012).
141. BAUMGARTNER, R. *et al.*, Asymmetries in behavioral and neural responses to spectral cues demonstrate the generality of auditory looming bias, en, *Proceedings of the National Academy of Sciences* **114**, 9743-9748, ISSN : 0027-8424, 1091-6490, <https://pnas.org/doi/full/10.1073/pnas.1703247114> (2024) (sept. 2017).
142. HASSAGER, H. G., GRAN, F. & DAU, T., The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment, en, *The Journal of the Acoustical Society of America* **139**, 2992-3000, ISSN : 0001-4966, <http://asa.scitation.org/doi/10.1121/1.4950847> (2023) (mai 2016).
143. LI, S., SCHLIEPER, R. & PEISSIG, J., The Role of Reverberation and Magnitude Spectra of Direct Parts in Contralateral and Ipsilateral Ear Signals on Perceived Externalization, en, *Applied Sciences* **9**, 460, ISSN : 2076-3417, <http://www.mdpi.com/2076-3417/9/3/460> (2024) (jan. 2019).
144. CATIC, J., SANTURETTE, S. & DAU, T., The role of reverberation-related binaural cues in the externalization of speech, en, *The Journal of the Acoustical Society of America* **138**, 1154-1167, ISSN : 0001-4966, <http://asa.scitation.org/doi/10.1121/1.4928132> (2022) (août 2015).
145. JIANG, Z., SANG, J., ZHENG, C. & LI, X., The effect of pinna filtering in binaural transfer functions on externalization in a reverberant environment, en, *Applied Acoustics* **164**, 107257, ISSN : 0003682X, <https://linkinghub.elsevier.com/retrieve/pii/S0003682X19309223> (2024) (juill. 2020).

-
146. KLEIN, F., WERNER, S. & MAYENFELS, T., Influences of Training on Externalization of Binaural Synthesis in Situations of Room Divergence, en, *Journal of the Audio Engineering Society* **65**, 178-187, ISSN : 15494950, <http://www.aes.org/e-lib/browse.cfm?elib=18553> (2022) (mars 2017).
 147. GIL-CARVAJAL, J. C., CUBICK, J., SANTURETTE, S. & DAU, T., Spatial Hearing with Incongruent Visual or Auditory Room Cues, en, *Scientific Reports* **6**, 37342, ISSN : 2045-2322, <https://www.nature.com/articles/srep37342> (2024) (nov. 2016).
 148. UDESEN, J., PIECHOWIAK, T. & GRAN, F., The Effect of Vision on Psychoacoustic Testing with Headphone-Based Virtual Sound, en, *Journal of the Audio Engineering Society* **63**, 552-561, ISSN : 15494950, <http://www.aes.org/e-lib/browse.cfm?elib=17841> (2024) (août 2015).
 149. WENZEL, E., *The relative contribution of interaural time and magnitude cues to dynamic sound localization* en, in *Proceedings of 1995 Workshop on Applications of Signal Processing to Audio and Acoustics* (IEEE, New Paltz, NY, USA, 1995), 80-83, ISBN : 978-0-7803-3064-1, <http://ieeexplore.ieee.org/document/482963/> (2024).
 150. HENDRICKX, E. *et al.*, Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis, en, *The Journal of the Acoustical Society of America* **141**, 2011-2023, ISSN : 0001-4966, <http://aip.scitation.org/doi/10.1121/1.4978612> (2021) (mars 2017).
 151. HENDRICKX, E. *et al.*, Improvement of Externalization by Listener and Source Movement Using a "Binauralized" Microphone Array, en, *Journal of the Audio Engineering Society* **65**, 589-599, ISSN : 15494950, <http://www.aes.org/e-lib/browse.cfm?elib=19174> (2021) (août 2017).
 152. KÖNIG, G. & SUSSMANN, W., Zum Richtungshören in der Median-Sagittal-Ebene (On directional hearing in the medial-sagittal planes), *Archiv für Ohren-, Nasen- und Kehlkopfheilkunde* **167**, 303-307, ISSN : 1434-4726, <https://doi.org/10.1007/BF02107754> (déc. 1955).
 153. WERSENYI, G., Effect of Emulated Head-Tracking for Reducing Localization Errors in Virtual Audio Simulation, en, *IEEE Transactions on Audio, Speech, and Language Processing* **17**, 247-252, ISSN : 1558-7916, <http://ieeexplore.ieee.org/document/4749455/> (2024) (fév. 2009).

-
154. KULKARNI, A. & COLBURN, H. S., Role of spectral detail in sound-source localization, en, *Nature* **396**, 747-749, ISSN : 0028-0836, 1476-4687, <https://www.nature.com/articles/25526> (2024) (déc. 1998).
 155. CATIC, J., SANTURETTE, S., BUCHHOLZ, J. M., GRAN, F. & DAU, T., The effect of interaural-level-difference fluctuations on the externalization of sound, en, *The Journal of the Acoustical Society of America* **134**, 1232-1241, ISSN : 0001-4966, <http://asa.scitation.org/doi/10.1121/1.4812264> (2023) (août 2013).
 156. YUAN, Y., XIE, L., FU, Z.-H., XU, M. & CONG, Q., Sound image externalization for headphone based real-time 3D audio, en, *Frontiers of Computer Science* **11**, 419-428, ISSN : 2095-2228, 2095-2236, <http://link.springer.com/10.1007/s11704-016-6182-2> (2024) (juin 2017).
 157. CUBICK, J., BUCHHOLZ, J. M., BEST, V., LAVANDIER, M. & DAU, T., Listening through hearing aids affects spatial perception and speech intelligibility in normal-hearing listeners, en, *The Journal of the Acoustical Society of America* **144**, 2896-2905, ISSN : 0001-4966, <https://pubs.aip.org/asa/jasa/article/144/5/2896-2905/646365> (2023) (nov. 2018).
 158. HASSAGER, H. G., WIINBERG, A. & DAU, T., Effects of hearing-aid dynamic range compression on spatial perception in a reverberant environment, en, *The Journal of the Acoustical Society of America* **141**, 2556-2568, ISSN : 0001-4966, 1520-8524, <https://pubs.aip.org/jasa/article/141/4/2556/1059155/Effects-of-hearing-aid-dynamic-range-compression> (2024) (avr. 2017).
 159. HASSAGER, H. G., MAY, T., WIINBERG, A. & DAU, T., Preserving spatial perception in rooms using direct-sound driven dynamic range compression, en, *The Journal of the Acoustical Society of America* **141**, 4556-4566, ISSN : 0001-4966, 1520-8524, <https://pubs.aip.org/jasa/article/141/6/4556/916350/Preserving-spatial-perception-in-rooms-using> (2024) (juin 2017).
 160. ROBINSON, P. & XIANG, N., Design, construction, and evaluation of a 1 :8 scale model binaural manikin, en, *The Journal of the Acoustical Society of America* **133**, EL162-EL167, ISSN : 0001-4966, 1520-8524, <https://pubs.aip.org/jasa/article/133/3/EL162/913530/Design-construction-and-evaluation-of-a-1-8-scale> (2024) (mars 2013).

-
161. HUNTER, M. D. *et al.*, A neural basis for the perception of voices in external auditory space, en, *Brain* **126**, 161-169, ISSN : 0006-8950, 1460-2156, <https://academic.oup.com/brain/article-lookup/doi/10.1093/brain/awg015> (2024) (jan. 2003).
162. KOPČO, N. *et al.*, Neuronal representations of distance in human auditory cortex, en, *Proceedings of the National Academy of Sciences* **109**, 11019-11024, ISSN : 0027-8424, 1091-6490, <https://pnas.org/doi/full/10.1073/pnas.1119496109> (2023) (juill. 2012).
163. CALLAN, A., CALLAN, D. E. & ANDO, H., Neural correlates of sound externalization, en, *NeuroImage* **66**, 22-27, ISSN : 10538119, <https://linkinghub.elsevier.com/retrieve/pii/S1053811912010609> (2021) (fév. 2013).
164. *Le cerveau à tous les niveaux!* https://lecerveau.mcgill.ca/flash/i/i_10/i_10_cr/i_10_cr_lan/i_10_cr_lan.html (2024).
165. NICOL, R., DUFOR, O., GROS, L., RUEFF, P. & FARRUGIA, N., *EEG measurement of binaural sound immersion* en, in, Artwork Size : 6 pages Publisher : EAA Spatial Audio Signal Processing Symposium (Paris, 2019), 6 pages, <https://hal.archives-ouvertes.fr/SASP/hal-02275183> (2023).
166. ALHO, K., ESCERA, C., DÍAZ, R., YAGO, E. & SERRA, J. M., Effects of involuntary auditory attention on visual task performance and brain activity : en, *NeuroReport* **8**, 3233-3237, ISSN : 0959-4965, <http://journals.lww.com/00001756-199710200-00010> (2024) (oct. 1997).
167. KATAYAMA, J. & POLICH, J., Auditory and visual P300 topography from a 3 stimulus paradigm, en, *Clinical Neurophysiology* **110**, 463-468, ISSN : 13882457, <https://linkinghub.elsevier.com/retrieve/pii/S1388245798000352> (2024) (mars 1999).
168. PALOMÄKI, K. J., TIITINEN, H., MÄKINEN, V., MAY, P. J. & ALKU, P., Spatial processing in human auditory cortex : The effects of 3D, ITD, and ILD stimulation techniques, en, *Cognitive Brain Research* **24**, 364-379, ISSN : 09266410, <https://linkinghub.elsevier.com/retrieve/pii/S0926641005000480> (2021) (août 2005).
169. COLAS, T., FARRUGIA, N., HENDRICKX, E. & PAQUIER, M., Sound externalization in dynamic binaural listening : A comparative behavioral and EEG study, en, *Hearing Research* **440**,

-
- 108912, ISSN : 03785955, <https://linkinghub.elsevier.com/retrieve/pii/S0378595523002241> (2023) (déc. 2023).
170. NICHOLLS, M. E., THOMAS, N. A., LOETSCHER, T. & GRIMSHAW, G. M., The Flinders Handedness survey (FLANDERS) : A brief measure of skilled hand preference, en, *Cortex* **49**, 2914-2926, ISSN : 00109452, <https://linkinghub.elsevier.com/retrieve/pii/S0010945213000385> (2022) (nov. 2013).
171. *Office for Human Research Protections* en, Text, Last Modified : 2024-05-03T20 :54 :08-0400, mars 2022, <https://www.hhs.gov/ohrp/index.html> (2024).
172. SCHAPKIN, S. A., GUSEV, A. N. & KUHL, J., Categorization of unilaterally presented emotional words : an ERP analysis, *Acta Neurobiol. Exp.*, 17-28 (2000).
173. NAUMANN, E., MAIER, S., DIEDRICH, O., BECKER, G. & BARTUSSEK, D., Structural, semantic, and emotion-focused processing of neutral and negative nouns : event-related potential correlates, *Journal of Psychophysiology*, 158-172 (1997).
174. PERRIN, F. *et al.*, Neural mechanisms involved in the detection of our first name : a combined ERPs and PET study, en, *Neuropsychologia* **43**, 12-19, ISSN : 00283932, <https://linkinghub.elsevier.com/retrieve/pii/S0028393204001629> (2022) (jan. 2005).
175. SALMON, F., HENDRICKX, ÉPAIN, N. & PAQUIER, M., The Influence of Vision on the Perceived Differences Between Sound Spaces, en, *Journal of the Audio Engineering Society* **68**, 522-531, ISSN : 15494950, <https://www.aes.org/e-lib/browse.cfm?elib=20888> (2023) (sept. 2020).
176. CABRERA, D., LEE, D., YADAV, M. & MARTENS, W. L., Decay Envelope Manipulation of Room Impulse Responses : Techniques for Auralization and Sonification, en (2011).
177. CARPENTIER, T., NOISTERNIG, M. & WARUSFEL, O., Twenty Years of Ircam Spat : Looking Back, Looking Forward, en, 8 (2015).
178. MENDONÇA, C., CAMPOS, G., DIAS, P. & SANTOS, J. A., Learning Auditory Space : Generalization and Long-Term Effects, en, *PLoS ONE* **8** (éd. MALMIERCA, M.) e77900, ISSN : 1932-6203, <https://dx.plos.org/10.1371/journal.pone.0077900> (2022) (oct. 2013).
179. MENDONÇA, C. *et al.*, On the improvement of localization accuracy with non-individualized HRTF-based sounds, en, *Journal of the Audio Engineering Society*, 821-830 (oct. 2013).

-
180. RUGELES OSPINA, F., EMERIT, M. & KATZ, B. F., *The three-dimensional morphological database for spatial hearing research of the BiLi project* en, in (Pittsburgh, Pennsylvania, 2015), 050001, <http://asa.scitation.org/doi/abs/10.1121/2.0000050> (2023).
181. *Calibrateur Acoustique | TYPE 4231| Brüel & Kjør* <https://www.bksv.com/fr/transducers/acoustic/calibrators/sound-calibrator-4231> (2024).
182. *T3 | Produits audio 3D pour les professionnels fr-FR*, <http://feichter-audio.com/produits/diffusion/t3/> (2023).
183. *Cyton Biosensing Board (8-channels)* en, <https://shop.openbci.com/products/cyton-biosensing-board-8-channel> (2024).
184. *Gold Cup Electrodes* en, <https://shop.openbci.com/products/openbci-gold-cup-electrodes> (2024).
185. ALTMANN, C. F., WILCZEK, E. & KAISER, J., Processing of Auditory Location Changes after Horizontal Head Rotation, en, *Journal of Neuroscience* **29**, 13074-13078, ISSN : 0270-6474, 1529-2401, <https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.1708-09.2009> (2021) (oct. 2009).
186. *The OpenBCI GUI | OpenBCI Documentation* en, juin 2024, <https://openbci.github.io/Software/OpenBCISoftware/GUIDocs/> (2024).
187. GRAMFORT, A. *et al.*, MNE software for processing MEG and EEG data, en, *NeuroImage* **86**, 446-460, ISSN : 10538119, <https://linkinghub.elsevier.com/retrieve/pii/S1053811913010501> (2022) (fév. 2014).
188. *mne.preprocessing.find_eog_events — MNE 1.4.1 documentation* https://mne.tools/stable/generated/mne.preprocessing.find_eog_events.html (2023).
189. JAS, M., ENGEMANN, D., RAIMONDO, F., BEKHTI, Y. & GRAMFORT, A., Automated rejection and repair of bad trials in MEG/EEG, en, 5 (2016).
190. WIDMANN, A., SCHRÖGER, E. & MAESS, B., Digital filter design for electrophysiological data – a practical approach, en, *Journal of Neuroscience Methods* **250**, 34-46, ISSN : 01650270, <https://linkinghub.elsevier.com/retrieve/pii/S0165027014002866> (2023) (juill. 2015).

-
191. HEINE, L. *et al.*, Virtually spatialized sounds enhance auditory processing in healthy participants and patients with a disorder of consciousness, en, *Scientific Reports* **11**, 13702, ISSN : 2045-2322, <http://www.nature.com/articles/s41598-021-93151-6> (2021) (déc. 2021).
192. *mne.baseline.rescale* — *MNE 1.4.1 documentation* <https://mne.tools/stable/generated/mne.baseline.rescale.html> (2023).
193. MARIS, E. & OOSTENVELD, R., Nonparametric statistical testing of EEG- and MEG-data, en, *Journal of Neuroscience Methods* **164**, 177-190, ISSN : 01650270, <https://linkinghub.elsevier.com/retrieve/pii/S0165027007001707> (2023) (août 2007).
194. KEYSERS, C., GAZZOLA, V. & WAGENMAKERS, E.-J., Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence, en, *Nature Neuroscience* **23**, 788-799, ISSN : 1097-6256, 1546-1726, <http://www.nature.com/articles/s41593-020-0660-4> (2022) (juill. 2020).
195. PICTON, T., WOODS, D. & PROULX, G., Human auditory sustained potentials. II. Stimulus relationships, en, *Electroencephalography and Clinical Neurophysiology* **45**, 198-210, ISSN : 00134694, <https://linkinghub.elsevier.com/retrieve/pii/S0013469478900044> (2022) (août 1978).
196. BUDD, T., BARRY, R. J., GORDON, E., RENNIE, C. & MICHIE, P., Decrement of the N1 auditory event-related potential with stimulus repetition : habituation vs. refractoriness, en, *International Journal of Psychophysiology* **31**, 51-68, ISSN : 01678760, <https://linkinghub.elsevier.com/retrieve/pii/S0167876098000403> (2022) (déc. 1998).
197. GARRIDO, M. I. *et al.*, Repetition suppression and plasticity in the human brain, en, *NeuroImage* **48**, 269-279, ISSN : 10538119, <https://linkinghub.elsevier.com/retrieve/pii/S1053811909006661> (2023) (oct. 2009).
198. NOLAN, H. *et al.*, Neural correlates of oddball detection in self-motion heading : A high-density event-related potential study of vestibular integration, en, *Experimental Brain Research* **219**, 1-11, ISSN : 0014-4819, 1432-1106, <http://link.springer.com/10.1007/s00221-012-3059-y> (2023) (mai 2012).
199. NAKUL, E., BARTOLOMEI, F. & LOPEZ, C., Vestibular-Evoked Cerebral Potentials, en, *Frontiers in Neurology* **12**, 674100, ISSN : 1664-2295, <https://www.frontiersin.org/articles/10.3389/fneur.2021.674100/full> (2023) (sept. 2021).

-
200. ALTMANN, C. F., GETZMANN, S. & LEWALD, J., Allocentric or Craniocentric Representation of Acoustic Space : An Electrotomography Study Using Mismatch Negativity, en, *PLoS ONE* **7** (éd. MALMIERCA, M. S.) e41872, ISSN : 1932-6203, <https://dx.plos.org/10.1371/journal.pone.0041872> (2011) (juill. 2012).
201. SCHWARTZE, M., ROTHERMICH, K., SCHMIDT-KASSOW, M. & KOTZ, S. A., Temporal regularity effects on pre-attentive and attentive processing of deviance, en, *Biological Psychology* **87**, 146-151, ISSN : 03010511, <https://linkinghub.elsevier.com/retrieve/pii/S030105111100069X> (2011) (avr. 2011).
202. BENJAMINI, Y. & HOCHBERG, Y., Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing, en, *Journal of the Royal Statistical Society : Series B (Methodological)* **57**, 289-300, ISSN : 00359246, <https://onlinelibrary.wiley.com/doi/10.1111/j.2517-6161.1995.tb02031.x> (2011) (jan. 1995).
203. VAN DOORN, J., LY, A., MARSMAN, M. & WAGENMAKERS, E.-J., Bayesian rank-based hypothesis testing for the rank sum test, the signed rank test, and Spearman's, en, *Journal of Applied Statistics* **47**, 2984-3006, ISSN : 0266-4763, 1360-0532, <https://www.tandfonline.com/doi/full/10.1080/02664763.2019.1709053> (2020) (déc. 2020).
204. JEFFREYS, H., *Theory of probability* (Oxford University Press, 1961).
205. CHECHILE, R. A., A Bayesian analysis for the Mann-Whitney statistic, en, *Communications in Statistics - Theory and Methods* **49**, 670-696, ISSN : 0361-0926, 1532-415X, <https://www.tandfonline.com/doi/full/10.1080/03610926.2018.1549247> (2020) (fév. 2020).
206. PIKE, C. W., *Evaluating the Perceived Quality of Binaural Technology* en, thèse de doct. (University of York, 2019), <https://etheses.whiterose.ac.uk/24022/>.
207. LORENZ, A., *Impact of Head-Tracking on the listening experience of binaural music* en, thèse de doct. (Fraunhofer Institute for Integrated Circuits, 2020), file:///C:/Users/pesbr/Downloads/200513_main.pdf.
208. COWAN, N., On short and long auditory stores. *Psychological bulletin* (1984).
209. ZIMMERMANN, J. F., MOSCOVITCH, M. & ALAIN, C., Attending to auditory memory, en, *Brain Research* **1640**, 208-221, ISSN : 00068993, <https://linkinghub.elsevier.com/retrieve/pii/S0006899315008963> (2016) (juin 2016).

Titre : Externalisation en restitution binaurale non-individualisée avec head-tracking : aspects objectifs et perceptifs

Mot clés : Restitution binaurale, Externalisation, Head-tracking, EEG, aERPs

Résumé : La restitution binaurale est une méthode de restitution du son au casque visant à simuler une écoute naturelle. Souvent, les sons restitués au casque semblent provenir de l'intérieur de la tête. Pour une restitution binaurale convaincante, les sons doivent paraître venir de l'extérieur (externalisés), comme dans la réalité. L'utilisation d'un système de suivi des mouvements de tête (head-tracking) peut permettre d'améliorer l'externalisation pendant et après les mouvements. Cette thèse examine l'effet de persistance de l'amélioration d'externalisation après des mouvements de tête. Une première étude a combiné une expérience comportementale et une expérience EEG pour identifier un corrélat neuronal de l'externalisation.

Une méthode innovante, exploitant cet effet de persistance de l'externalisation, a été développée afin de comparer des stimuli "acoustiquement identiques", mais avec différents niveaux d'externalisation. Les résultats n'ont pas révélé de corrélat neuronal, mais ont soulevé plusieurs questions sur l'influence de la nature et de la durée des sources sonores sur l'effet de persistance de l'externalisation. Une deuxième étude, comportant trois expériences comportementales, a été menée afin d'aborder ces questions. Les résultats ont montré que l'amélioration d'externalisation après des mouvements de tête fonctionne pour différentes sources, mais diminue après une dizaine de seconde.

Title: Externalization in non-individualized binaural listening with head-tracking: a objective and perceptive aspects

Keywords: Binaural restitution, Externalization, Head-tracking, EEG, aERPs

Abstract: Binaural reproduction is a method of sound playback through headphones aimed at simulating natural listening. Often, sounds played through headphones seem to originate from inside the head. For convincing binaural reproduction, sounds must appear to come from outside (externalized), as they do in reality. Using a head-tracking system can enhance externalization during and after head movements. This thesis examines the persistence of the head movements after-effect on externalization. The first study combined a behavioral experiment and an EEG experiment to identify a neural correlate of external-

ization. An innovative method leveraging this after-effect was developed to compare "acoustically identical" stimuli with different levels of externalization. The results did not reveal a neural correlate but raised several questions about the influence of the nature and duration of sound sources on the head movements after-effect on externalization. A second study, involving three behavioral experiments, was conducted to address these questions. The results showed that the improvement in externalization after head movements works for various sound sources but decreases after about ten seconds .