UNIVERSITÉ DE LORRAINE | BIBLIOTHÈQUES UNIVERSITAIRES

AVERTISSEMENT
DROIT D'AUTEUR

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document. Toute contrefaçon, plagiat ou reproduction illicite encourt une poursuite pénale.

Contact bibliothèque : ddoc-theses-contact@univ-lorraine.fr
(Cette adresse ne permet pas de contacter les auteurs)

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

UNIVERSITÉ DE LORRAINE
34 COURS LÉOPOLD — CS 25233
54052 NANCY CEDEX
TÉL : 03 54 50 54 00
CONTACT@UNIV-LORRAINE.FR
WWW.UNIV-LORRAINE.FR

UNIVERSITÉ DE LORRAINE

École doctorale IAEM Lorraine

Stimulations sonores pour les interfaces cerveau-ordinateur : confort utilisateur et détectabilité

THÈSE

présentée et soutenue publiquement le 16 juillet 2026

pour l'obtention du

Doctorat de l'Université de Lorraine

(mention informatique)

par

Lenaïg Guého

Composition du jury

Présidente :	Armelle Brun, Professeure des universités, Université de Lorraine
Rapporteurs :	Nicolas Farrugia, Professeur des universités, Imt Atlantique Raphaëlle Roy, Professeure des universités, Isae-supaero
Examinatrices :	Léa Pillette, Chargée de recherche, Irisa - cnrs/Université de Rennes Armelle Brun, Professeure des universités, Université de Lorraine
Directeur :	Patrick Henaff, Professeur des universités, Bretagne inp (invité)
Co-directrice :	Rozenn Nicol, Ingénieure de recherche, Orange Innovation, France
Encadrants :	Laurent Bougrain, Maître de conférences, Université de Lorraine Cyril Plapous, Ingénieur de recherche, Orange Innovation, France (invité)

Laboratoire Lorrain de Recherche en Informatique et ses Applications — UMR 7503

Mis en page avec la classe thesul.

Remerciements

À l'issue de mon stage, j'avais décidé de remercier tous les Jean-Michel d'Orange. Toutefois, il est de rigueur de changer de prénom pour cette thèse et de remercier plutôt les Patrick, d'Orange ou d'ailleurs.

Ainsi, je tiens tout d'abord à remercier Patrick H. et Laurent d'avoir accepté le package "encadrement de thèse + Lenaïg". Merci Patrick pour ton accompagnement tout au long de cette thèse et pour tes petites anecdotes, toujours rigolotes. Merci Laurent pour tous ces échanges scientifiques, mais aussi pour les parties de Ricochet Robots (je garde espoir de te battre un jour). Je souhaite ensuite remercier Rozenn et Cyril, qui ont accepté d'encadrer mon stage il y a quatre ans et ont été assez fous (ou assez sympas) pour poursuivre l'aventure en supervisant ma thèse. Merci Rozenn pour ton humanité, ta gentillesse, ta bienveillance et ton oreille d'or ; ce fut un honneur de travailler avec toi. Merci Cyril pour ton soutien, aussi bien dans le travail que dans les galères migraineuses ou le choix des LEGO à acheter (je ne te remercie pas pour le budget qui va avec par contre, même si la LEGO-thérapie semble fonctionner). Merci à tous les deux, j'espère que vous avez autant apprécié que moi cette garde alternée !

Je souhaite également remercier Raphaëlle ROY et Nicolas FARRUGIA qui ont accepté de relire cette thèse, ainsi qu'Armelle BRUN et Léa PILLETTE, d'avoir consenti à faire partie de mon jury. Merci à tous pour les échanges très enrichissants lors de la soutenance et, plus particulièrement, à Léa, qui m'a suivie durant ces trois années. Merci pour ton intérêt envers mon travail, tes encouragements et ta bienveillance.

Cette thèse aurait été bien plus compliquée sans les collègues d'Orange, qui m'ont soutenue, pendant les pauses surtout ! On revient donc aux Patrick... Merci Patrick D.M. pour tous ces partages, ces cafés et ces rires. Merci Julian d'avoir accepté de dire au revoir à ta tranquillité pour partager ton bureau, de m'avoir écoutée, orientée et de t'être assuré du bon déroulement de ma thèse [Moreira, 2023] [1] (en revanche, les plantes ne te remercient pas). Merci également à Loïc A. pour l'aide technique (les réparations de clés *Bluetooth* et de bonnets EEG notamment) et les parties de baby-foot. Merci aussi à Cédric d'avoir mis en place des pistes cyclables me permettant d'arriver entière (ou presque) au travail. Merci aux jeunes d'Orange, même si ils sont peu nombreux, au moins ils sont sympas ! Merci à Joël GPT pour toutes les informations partagées, sur le fonctionnement d'Orange et de la vie. Merci à Totocho pour les réunions RH, indispensables au bon déroulement de nos thèses. Merci à Cicine pour tous ces rires et ces découvertes (je pense que mon corps est encore traumatisé par le pilates hamac). Merci également à Aminata, Hichem, Loïc F. et Rayane. Que seraient les jeunes d'Orange sans les stagiaires et les alternants ? Merci Alan, Katell Q., Paul... Et surtout, un immense merci au singe savant, le *big brain* ultime : Henrique ! Merci pour tous ces échanges, ce travail et cette présence. À toi de jouer maintenant, j'ai déjà hâte d'assister à ta soutenance ! Plus largement, merci à toute l'équipe NUX, aux personnes du couloir, ou même des autres bâtiments, pour tous les échanges et les nombreuses parties de baby-foot.

Les débuts de semaine n'auraient pas été aussi faciles sans le yoga du lundi midi, puis le CrossFit du mardi midi, qui permet de s'assurer de ne plus vouloir quitter sa chaise de bureau de la semaine... Merci à mes acolytes : Carine, Nadia et les Patrick, A. et B. Merci également à Fred P., qui m'a permis d'occuper les quelques midis restants (ceux non consacrés au baby-foot) en bricolant au Ti-lieu, merci pour ta patience et ton partage de connaissances.

Pour en finir avec Orange, merci à tous ceux qui ont ~~été forcés~~ accepté de participer à mes expériences, parfois même plusieurs fois. Je sais désormais qu'il est possible de travailler chez

1. Moreira, J. (2023). Eh t'as fait le Pédantix aujourd'hui ? Doctoral dissertation, Lannion, Orange Labs.

Orange en ayant un encéphalogramme plat, mais je ne dévoilerai pas le nom de famille des Patrick qui m'ont permis de le démontrer !

Après le CrossFit du mardi midi, direction Lannion Athlétisme le soir. Merci à tous pour votre soutien dans l'effort, qu'il soit physique ou intellectuel, et pour tous ces bons moments qui font oublier la thèse (c'est forcément plus difficile d'y penser quand on ne peut plus respirer). Je tiens à remercier chaleureusement Pierre pour son soutien et sa bienveillance, sans oublier sa moustache et son style vestimentaire d'un autre siècle. Merci également à son fidèle compagnon Jean, pour ses blagues pas toujours drôles ! Et merci à tous ceux avec qui j'ai apprécié faire des kilomètres (et surtout discuter) : Alain, Blandine, Cécile, Christine, Fred, Manu, Vincent... Merci aussi aux jeunes coureurs (plus ou moins réguliers) de Pierre (là aussi, ils se font rares), devenus amis au fil des bornes : Juliette et Hugo, Florie, Madame de Bir et Louise. Merci pour ces footings, ces blessures, ces pizza-plages et ces coups à boire. Enfin, même si on a davantage échangé de *kudos* que de kilomètres ces trois dernières années, merci aux Copains Coureurs, qui se sont toujours souciés de l'avancée de mon travail. Et parce que je ne faisais manifestement pas assez d'activités pour oublier la thèse, j'ai décidé de faire de la musique le mercredi. C'est donc l'heure (ou le jour) de remercier mes compagnons de musique pour tous ces moments de partage et de rigolade, et notamment les personnes jouant du plus bel instrument du monde : les saxophonistes ! Merci Arnaud, Edith, Jérémy et Stéphanie.

Il faut bien avouer que j'ai quand même raté un certain nombre de répés à cause de mes séjours à Nancy (bien que ceux-ci se soient progressivement raréfiés). Merci aux filles du Loria, Faustine, Laëtitia, Sarah et Valérie, pour leur accueil, leur soutien et leur gentillesse, qui ont rendu ces déplacements bien plus agréables.

Après les collègues, les sportifs et les artistes, place à ceux qui ne rentrent pas dans des cases, les "copains d'Antoine" : Taz et Sylvain (bon, allez, je vous mets dans la deuxième catégorie, je sais que ça vous fera plaisir), Mamie Vivi et Papy Viny, Ewen et Thalie. Merci pour les parties de Mario Kart, les soirées puzzles infinies et les dégustations de chips... Je voudrais également remercier tous les copains que les études supérieures m'ont permis de rencontrer. Merci d'abord à Paul, présent depuis le début. À ton tour d'aller au bout de l'aventure maintenant, mais reviens un jour quand même hein ! Merci à FitGab, Léo et Flokie ; cette thèse, c'est aussi grâce à vous et à toutes ces heures passées à la BU (et au bar). Quelles auraient été ces études sans un Erasmus ? Merci les Franchouillards ! Merci Coco, toujours là au quotidien (malheureusement), malgré la distance (j'avais pourtant essayé d'aller le plus loin possible). Merci aussi les gars : Cyril, Loann et Victor ; il est temps d'organiser notre réunion annuelle ! Merci également à Len, comme quoi faire un deuxième master n'était finalement pas complètement inutile.

Il est maintenant temps de remercier ceux qui ont largement contribué à ma survie pendant ces trois années... Merci à la famille Panpan pour la fourniture du carburant : Papy et Mamie Panpan pour les crêpes, Papa et Maman Panpan pour le jus de pomme, le cidre la gnôle et, plus généralement, tout ce qui pousse dans leur jardin. Et bien-sûr, un grand merci aussi à ma famille, Guého et Le Guennan, pour son soutien tout au long de ces années. Je tiens en premier lieu à remercier mes grands-parents. Ensuite, même si je n'ai aucun doute sur le fait que j'aurais été beaucoup plus tranquille sans eux, je me serais certainement bien plus ennuyée... Je suis donc dans l'obligation de remercier mes frères, qui m'ont forgé un caractère capable d'endurer trois années de thèse ! Enfin, merci à mon père (on peut dire que j'ai trouvé maintenant) et à ma mère, pour son soutien tout au long de ma scolarité. Merci d'avoir cru en nous, de nous avoir poussé et d'avoir respecté nos choix (je t'avais bien dit que la fac était une meilleure idée que la prépa). Pour terminer, je souhaite remercier Antoine, sans qui rien n'aurait été possible. Merci pour ton soutien inconditionnel, merci de m'avoir nourrie, autant l'estomac que le cœur.

Table des matières

Table des figures **vii**

Liste des tableaux **xi**

Nomenclature **xiii**

Acronymes **xvii**

Introduction générale **xix**

1 Objectif xx

2 Contributions xx

3 Structure du manuscrit xx

Partie I Interactions cerveau-machine : contexte et problématique **1**

Chapitre 1
Aires cérébrales impliquées dans le traitement auditif

1.1 Introduction 3

1.2 Organisation générale du cerveau humain 4

1.3 Mécanismes cérébraux du traitement de la parole 5

1.4 Mécanismes cérébraux du traitement de la musique 9

1.5 Mécanismes cérébraux de la discrimination des catégories sonores 11

1.6 Mécanismes cérébraux de la localisation sonore 12

1.7 Variabilité interindividuelle du traitement cérébral des sons 13

1.7.1 Influence de l'expertise musicale 13
1.7.2 Influence de la latéralité 13
1.7.3 Influence du sexe biologique 15
1.8 Influence des substances psychoactives sur le traitement cérébral 15
1.8.1 Effets de la caféine et de la théine 15
1.8.2 Effets de la nicotine 16
1.8.3 Effets de l'alcool 16
1.8.4 Effets des drogues et des médicaments 16
1.9 Conclusion 16

Chapitre 2
Les interfaces cerveau-ordinateur (*Brain-Computer Interfaces* ou BCI)

2.1 Introduction 19
2.2 Classification des BCI 20
2.3 Les méthodes d'imagerie cérébrale 21
2.3.1 Méthodes invasives 21
2.3.2 Méthodes non invasives 22
2.4 Marqueurs neurophysiologiques utilisés dans les BCI actives et réactives . . . 24
2.4.1 Marqueurs neurophysiologiques utilisés dans les BCI actives 25
2.4.2 Marqueurs neurophysiologiques utilisés dans les BCI réactives 26
2.5 Conclusion 34

Chapitre 3
Analyse des signaux ElectroEncéphaloGraphiques (EEG)

3.1 Introduction 37
3.2 Signaux EEG dans un contexte BCI 38
3.2.1 Acquisition des signaux EEG 38
3.2.2 Organisation des données EEG 39
3.2.3 Propriétés des signaux EEG 40
3.2.4 Classification des signaux EEG 42
3.2.5 Stratégie d'analyse des signaux EEG 42
3.3 Prétraitements des signaux EEG 43
3.3.1 Filtrage passe-bande 44
3.3.2 Extraction temporelle 45
3.4 Réduction de dimension 45
3.4.1 Notations des *epochs* 45

3.4.2 Analyse en composantes principales (*Principal Component Analysis* ou PCA) . . . 46
3.4.3 *Common Spatial Pattern* (CSP) . . . 47
3.4.4 Analyse discriminante linéaire (*Linear Discriminant Analysis* ou LDA) 48
3.5 Détection de la fréquence de modulation . . . 49
3.5.1 Rapport Signal sur Bruit (RSB) . . . 49
3.5.2 Méthodes de classification . . . 50
3.6 Conclusion . . . 66

Partie II Contributions **67**

Chapitre 4
A la recherche d'alternatives aux sons purs : mise en place d'une expérience

4.1 Introduction . . . 69
4.1.1 Deux préoccupations originales : les facteurs acoustiques et les facteurs humains . . . 70
4.2 Travaux antérieurs . . . 70
4.2.1 Paradigme expérimental . . . 73
4.2.2 Stimuli auditifs . . . 73
4.2.3 Dispositifs matériels . . . 74
4.2.4 Participants . . . 75
4.2.5 Classification . . . 75
4.2.6 Performances . . . 75
4.3 Protocole expérimental . . . 76
4.3.1 Objectifs et hypothèses . . . 76
4.3.2 Paradigme expérimental . . . 76
4.3.3 Stimuli auditifs . . . 77
4.3.4 Collecte de métadonnées . . . 81
4.3.5 Dispositifs matériels . . . 82
4.3.6 Participants . . . 83
4.4 Facteurs acoustiques . . . 83
4.4.1 Casque d'écoute . . . 84
4.4.2 Audiomètre . . . 88
4.5 Facteurs individuels . . . 93
4.5.1 Questionnaires . . . 93
4.6 Conclusion . . . 99

Chapitre 5
Evaluation de la détectabilité de la fréquence de modulation selon le signal porteur

5.1 Introduction et présentation du contexte 101
5.2 RSB 103
5.2.1 Prétraitements des signaux EEG préalables au calcul du RSB 104
5.2.2 Détectabilité de la fréquence de modulation 107
5.3 Classification 111
5.3.1 Prétraitements des signaux EEG préalables à l'application de classifieurs 111
5.3.2 Résultats de détection de la fréquence de modulation 114
5.4 Cartographie cérébrale 122
5.5 Analyse critique des propriétés des stimulations 125
5.6 Conclusion 126

Chapitre 6
Evaluation perceptive et caractérisation des participants

6.1 Introduction 129
6.2 Evaluation perceptive des stimuli auditifs et de l'expérimentation 130
6.2.1 Classement des stimuli par ordre de préférence 130
6.2.2 Acceptabilité des stimuli 131
6.2.3 Perception de la sonie des stimuli 132
6.2.4 Acceptabilité et impact émotionnel de l'expérimentation 133
6.3 Profils des participants 133
6.4 Conclusion 134

Chapitre 7
Conclusion et perspectives

7.1 Synthèse générale 137
7.2 Perspectives 139
7.2.1 Conception de nouveaux signaux porteurs 139
7.2.2 Influence de la fréquence de modulation, du niveau sonore et de la durée de stimulation sur les performances 139
7.2.3 Analyse de la variabilité interindividuelle 140

Contributions

Bibliographie

Table des figures

1.1 Lobes externes du cerveau humain (figure tirée de [Wikipédia, 2006a]). 4

1.2 Visualisation des différentes aires composant le cortex auditif. 5

1.3 Visualisation du faisceau arqué reliant l'aire de Broca et l'aire de Wernicke (figure tirée de [Le cerveau à tous les niveaux, nd]). 6

1.4 Schéma des différentes régions cérébrales pouvant être impliquées dans le traitement des sons. 7

1.5 Termes directionnels anatomiques utilisés pour décrire la localisation des différentes régions cérébrale (figure tirée de [neuromedia, nd]). 8

1.6 Schéma des différentes régions cérébrales impliquées dans le traitement des sons musicaux. 9

1.7 Schéma des différentes régions cérébrales impliquées dans le traitement de stimuli acoustiques de différentes natures, présentant des variations de leurs caractéristiques temporelles et spectrales. 10

1.8 Schéma des différentes régions cérébrales impliquées dans la discrimination de catégories de sons. 11

1.9 Schéma des différentes régions cérébrales impliquées dans le traitement auditif spatial. 12

1.10 Régions cérébrales présentant une plus grande centralité lors de l'écoute de sons musicaux chez des participants ayant une plus grande expertise musicale. 13

1.11 Asymétries anatomiques de certaines régions cérébrales associées à la latéralité manuelle. 14

2.1 Schéma de fonctionnement d'une BCI réactive. 20

2.2 Illustration des ordres de grandeur des résolutions temporelle et spatiale de différentes méthodes d'imagerie cérébrale non invasives. 22

2.3 Illustration d'un paradigme SSVEP. Quatre stimuli visuels (ronds) clignotent à des fréquences distinctes (10, 12, 15 et 17 Hz) [Si-Mohammed et al., 2022]. L'activité cérébrale est enregistrée à l'aide d'un système EEG à 24 canaux, équipé d'électrodes passives positionnées selon le système international 10–20. La figure présente la réponse cérébrale d'un sujet induite par ces stimulations visuelles (courbe bleue), ainsi que l'activité enregistrée en l'absence de stimulation (courbe pointillée verte). Les flèches oranges indiquent les fréquences des stimulations visuelles. 28

2.4 Un son modulé en amplitude à 40 Hz est présenté à un sujet. Son activité cérébrale est enregistrée à l'aide d'un système EEG à 24 canaux, équipé d'électrodes passives positionnées selon le système international 10–20. La figure illustre la réponse cérébrale induite par cette stimulation auditive (courbe bleue), comparée à l'activité enregistrée en l'absence de stimulation (courbe verte). Les flèches oranges indiquent la fréquence de modulation du stimulus, ainsi que sa première harmonique. 29
2.5 Création d'une modulation en amplitude (en orange) à partir de deux signaux sinusoïdaux : le signal porteur (en vert) et le signal modulant (en bleu), avec f_c la fréquence du signal porteur et A_y, l'amplitude du signal modulé. 30
2.6 Système international 10–20 utilisé pour le positionnement des électrodes EEG (figure tirée de [Wikimedia commons, 2017]). 33

3.1 Exemple d'organisation hiérarchique d'une acquisition de l'activité EEG dans un contexte BCI : 2 sessions, chacune constituée de 2 *runs*, chaque *run* regroupant N essais correspondant à des répétitions d'une ou de plusieurs tâches (ou stimulations). 39
3.2 Capture d'écran de la bibliothèque EEGLAB montrant les signaux EEG obtenus après échantillonnage et quantification sur 24 canaux (*channels*), pour une durée de 5 secondes. 40
3.5 Principe de la SVM : l'hyperplan séparateur maximise la marge, définie comme la distance entre l'hyperplan et les échantillons les plus proches appartenant à chaque classe, appelés **vecteurs de support**. 52
3.6 Représentation du plan tangent à une sphère, en un point. Ce plan constitue une approximation euclidienne locale de la surface, illustrant la différence entre géométrie euclidienne (locale) et géométrie riemannienne (globale) (figure tirée de [Wikipédia, 2007]). 54
3.7 Espace tangent $\mathcal{T}_p$ au point $\mathbf{P}$. $\mathbf{S}_i$ correspond à la projection de $\mathbf{P}_i$ sur l'espace tangent via l'application logarithmique riemannienne ($\text{Log}_{\mathbf{P}}$) (figure tirée de [Barachant et al., 2010b]). 55
3.8 La matrice de covariance $\mathbf{P}_x$ de la nouvelle *epoch* $\mathbf{X}_x$ est comparée aux moyennes riemanniennes $\mathbf{P}_k$ de chaque classe $k = 1, \ldots, 4$ à l'aide de la distance riemannienne (figure tirée de [Bleuzé et al., 2021]). 57
3.9 Phase de test de la méthode FgMDM. 58
3.10 Exemple d'un réseau de neurones artificiel à trois couches : une couche d'entrée (verte), une couche cachée (bleue) et une couche de sortie (jaune). Les cercles représentent les neurones, et les connexions entre eux correspondent aux poids du modèle (figure tirée de [Wikipédia, 2006b]). 59
3.11 Illustration des principaux types de couches d'un réseau de neurones convolutif : convolution, *pooling* (sous-échantillonnage) et *fully connected* (entièrement connectée) (figure tirée de [Shervine Amidi, nd]). 60
3.12 Architecture du modèle EEGNet (figure tirée de [Lawhern et al., 2018]). 61
3.13 Architecture du modèle SCN (figure tirée de [Schirrmeister et al., 2017]). 63
3.14 Architecture du modèle GREEN (figure tirée de [Paillard et al., 2024]). 63
3.15 Evaluation WS par validation croisée à $\beta = 5$ blocs, pour chaque sujet (n) et chaque session (i). À chaque itération, 4 blocs sont utilisés pour l'apprentissage, tandis que le bloc restant est utilisé pour le test. La performance finale du modèle correspond à la moyenne des performances obtenues sur l'ensemble des itérations. 65

3.16 Evaluation CS. Les données de tous les sujets, à l'exception d'un seul (n), sont utilisées pour l'apprentissage, tandis que les données du sujet exclu servent exclusivement au test. 65

4.1 Participant équipé d'un casque EEG et d'écouteurs intra-auriculaires lors de la phase de test. 77
4.2 Représentations temporelles et spectrogrammes (échelle en dB) des signaux porteurs (à gauche) et des stimuli auditifs correspondants (à droite), tels qu'ils sont perçus par les participants (i.e. sans application du filtre de correction du casque). Les gains utilisés pour égaliser l'intensité perçue des stimuli sont également appliqués aux signaux porteurs, permettant une comparaison cohérente entre les différentes conditions. Pour les représentations temporelles, une échelle d'amplitude différente est utilisée pour le ronronnement de chat, dont l'amplitude plus élevée réduisait la visibilité des autres stimuli. Pour le son pur, l'échelle fréquentielle est limitée à 2000 Hz, l'énergie du signal étant concentrée autour de 1000 Hz. Dans les autres cas, elle est restreinte à 10000 Hz, y compris pour les signaux porteurs présentant des composantes au-delà de cette fréquence, afin de faciliter la comparaison avec les stimuli correspondants. 79
4.3 Courbes isosoniques définies par la norme ISO 226 :2023 (*International Organization for Standardization*) [ISO, 2023]. 80
4.4 Illustration des deux conditions expérimentales : à gauche, chaque oreille reçoit un stimulus de 50 phones, conduisant à une sonie globale perçue de 60 phones grâce au gain binaural ; à droite, chaque oreille reçoit un stimulus de 56 phones, conduisant à une sonie globale perçue de 66 phones. 81
4.5 Vue d'ensemble du déroulement de l'expérimentation, comprenant une phase pré-test, une phase de test et une phase post-test. Les phases pré- et post-test sont dédiées à la collecte de métadonnées sur les participants, ainsi qu'à la collecte de leurs retours subjectifs relatifs aux stimulations et à l'expérimentation. La phase de test consiste à mesurer l'activité cérébrale des participants en réponse aux cinq stimuli. 81
4.8 Source étalon [Brüel & Kjaer, 2022] utilisée pour la calibration du HATS, équipée d'un adaptateur de calibration permettant le couplage avec l'entrée d'une oreille du HATS. 85
4.11 Réponse en fréquence cible correspondant à une écoute en champ libre. 87
4.13 Interface graphique de l'audiomètre développé sous MATLAB. 89
4.14 Seuils auditifs de référence définis par la norme ISO 389, également appelés **zéros audiométriques** [Sterkers-Artières et Vincent, 2014; ISO, 2023]. 90
4.15 Evolution de la perte auditive (en dB HL) en fonction de l'âge, illustrant l'effet de la presbyacousie, particulièrement marquée dans les hautes fréquences (figure tirée de [Fillon, 2004]). 91
4.16 Interface graphique du questionnaire de latéralité développé sous MATLAB. 94
4.17 Interface graphique du questionnaire de musicalité développé sous MATLAB. . . . 96
4.18 Interface graphique du questionnaire général développé sous MATLAB. 97
4.19 Interface graphique du questionnaire final développé sous MATLAB. 98

5.1 Pour chaque participant (n), les réponses cérébrales associées à 10 répétitions de chaque stimulus ont été enregistrées pour chaque *run*, soit un total de 20 répétitions pour les 2 *runs* de chaque session (cf. section 4.3.2). 102

5.2 Répartition des 48 participants entre les deux conditions expérimentales : 24 participants ont été exposés à des stimuli égalisés à 60 phones et 24 autres, à des stimuli égalisés à 66 phones. 102
5.3 Etapes du calcul des deux types de RSB : RSB_{BRUT} et RSB_{CSP}. 104
5.4 Expérimentation menée à 60 phones : évolution du RSB_{CSP} (dB) en fonction de la largeur de bande des filtres passe-bande, centrés sur 40 et 80 Hz, appliqués aux signaux EEG lors du prétraitement. * indique $p < 0.05$, ** indique $p < 0.01$ et *** indique $p < 0.001$ (test des rangs signés de Wilcoxon avec correction de Bonferroni). 106
5.5 Valeurs des RSB_{BRUT} et RSB_{CSP} pour chaque stimulus, moyennées sur l'ensemble des sujets. *** indique $p < 0{,}001$ (test des rangs signés de Wilcoxon avec correction de Bonferroni). 108
5.6 Périodogramme du premier filtre spatial CSP appliqué aux signaux EEG prétraités, sans filtrage passe-bande. Les filtres CSP sont calculés, pour un sujet et un *run* donnés, à partir des *epochs* associées à chaque stimulation, ainsi qu'à la condition de silence (courbes vertes). 110
5.7 Expérimentation menée à 60 phones : influence du type de filtre et de la largeur de bande sur les performances de classification, en utilisant des fenêtres de 2 secondes sans recouvrement. Les résultats sont moyennés sur l'ensemble des sujets. 112
5.8 Expérimentation menée à 60 phones : influence de la taille des fenêtres pour différentes configurations de recouvrement avec le filtre de Bessel optimal (ordre 4, $B = 0{,}1$ Hz), ainsi que sans recouvrement pour différentes largeurs de bande avec un filtre de Bessel d'ordre 4. Les résultats sont moyennés sur l'ensemble des sujets. 113
5.9 Vue d'ensemble des *pipelines* utilisés dans cette étude. KPCA : *Kernel Principal Component Analysis*, LR : *Logistic Regression*, SVM : *Support Vector Machine*, LDA : *Linear Discriminant Analysis*, MDM : *Minimum-Distance-to-Mean*, FgMDM : *Fisher Geodesic Minimum-Distance-to-Mean*, RiSVM : *Riemannian Support Vector Machine*, CSP : *Common Spatial Pattern*, SCN : *ShallowConvNet*, GREEN : *Gabor Riemannian* EEGNet, DL : *Deep Learning*. 115
5.12 Moyenne des motifs spatiaux (valeurs absolues des poids associées à chaque électrode) sur l'ensemble des sujets, pour tous les stimuli confondus en opposition à la condition de silence. 123
5.13 A gauche, expérimentation menée à 60 phones ; à droite, expérimentation menée à 66 phones. Moyenne des motifs spatiaux (valeurs absolues des poids associées à chaque électrode) sur l'ensemble des sujets, pour chaque stimulus en opposition à la condition de silence. 124

6.1 Retours subjectifs des participants relatifs à la perception des stimuli auditifs, incluant leur ordre de préférence et leur acceptabilité. 131
6.2 Proportion de participants ayant indiqué, pour chaque stimulus, l'avoir perçu comme plus fort que les autres. * indique $p < 0.05$, ** indique $p < 0.01$ et *** indique $p < 0.001$ (test des rangs signés de Wilcoxon). 132
6.3 Emotion dominante rapportée par les participants avant et après la phase de test, pour chaque niveau de sonie (i.e. 60 et 66 phones). 133
6.4 Description quantitative et qualitative des participants à l'expérimentation. . . . 134

Liste des tableaux

4.1 Synthèse des études sur les BCI basées sur les SSAEP. Le terme **dichotique** désigne un mode de présentation des stimuli auditifs dans lequel les deux oreilles (gauche : G/droite : D) reçoivent des stimulations différentes, par opposition au mode **diotique**, où le même stimulus est présenté aux deux oreilles. Les nombres de sessions, de *runs* et d'essais correspondent aux totaux rapportés dans chaque étude, toutes conditions confondues. Suite dans le tableau 4.2. 71
4.2 Suite du tableau 4.1. Synthèse des études sur les BCI basées sur les SSAEP. Dans la mesure EEG, l'activité cérébrale est captée sous forme de différences de potentiel entre chaque électrode et une électrode de référence. Une électrode de masse est également utilisée afin de stabiliser les mesures et de réduire les interférences électriques, notamment le bruit du réseau. 72
4.3 Seuils auditifs de référence exprimés en dB SPL et correspondance, pour chaque fréquence, entre les niveaux sonores de calibration en dB HL et en dB SPL. 91

5.1 Hyperparamètres utilisés pour la recherche par grille lors de l'évaluation WS. . . . 116
5.2 Hyperparamètres des modèles de DL. 117
5.3 Critères utilisés pour comparer les propriétés spectrales et temporelles des signaux porteurs. 126

6.1 Probabilités, pour chaque stimulus, d'être classé comme stimulus préféré, à chaque niveau de sonie, ainsi que pour les deux niveaux de sonie confondus, selon le modèle de Plackett–Luce. 131
6.2 Test des rangs signés de Wilcoxon : *p-values* comparant, pour chaque niveau sonore, la proportion de participants jugeant la sinusoïde désagréable à celle observée pour les autres stimuli. 132

Nomenclature

$\boldsymbol{\alpha}_i$	multiplicateur de Lagrange
$\boldsymbol{\omega}$	vecteur des coefficients de régression
$\boldsymbol{\psi}$	vecteur de poids normal à l'hyperplan
Δ	réponse droite
δ	distance riemannienne
ℓ	log-vraisemblance
Γ	réponse gauche
γ	géodésique reliant deux points de la variété riemannienne
κ	fonction noyau
λ	valeur propre
$\mathbb{P}$	probabilité conditionnelle
$\mathbf{\Lambda}$	matrice diagonale des valeurs propres
$\mathbf{A}$	matrice des motifs spatiaux
$\mathbf{F}$	matrice des composantes principales
$\mathbf{P}$	point de référence sur la variété
$\mathbf{P}_i$	matrice de covariance associée à $\mathbf{X}_i$
$\mathbf{P}_k$	moyenne riemannienne des matrices de covariance de la classe k
$\mathbf{P}_r$	matrice de covariance de référence
$\mathbf{q}_r$	variable canonique associée aux signaux de référence
$\mathbf{q}_x$	variable canonique associée aux données EEG
$\mathbf{R}_s$	matrice des signaux de référence
$\mathbf{S}$	matrice diagonale des valeurs singulières
$\mathbf{S}_B$	matrice de dispersion inter-classe
$\mathbf{S}_i$	vecteur tangent associé à $\mathbf{P}_i$
$\mathbf{S}_W$	matrice de dispersion intra-classe
$\mathbf{U}$	matrice des vecteurs singuliers gauches

$\mathbf{V}$	matrice des vecteurs singuliers droits
$\mathbf{v}$	vecteur de projection
$\mathbf{w}$	filtres spatiaux
$\mathbf{W}_{\text{CSP}}$	matrice de projection du CSP contenant les filtres spatiaux
$\mathbf{W}_{\text{LDA}}$	matrice de projection de la LDA
$\mathbf{X}$	matrice de données représentant les enregistrements cérébraux
$\mathbf{x}_i$	vecteur des observations correspondant aux mesures simultanées sur l'ensemble des électrodes à l'instant i
$\mathbf{X}_i^{(k)}$	matrice de données représentant la i-ème *epoch* de la classe k
$\mathbf{X}_{\text{CSP}}$	matrice des données projetées par le CSP
$\mathbf{X}_{\text{LDA}}$	matrice des données projetées par la LDA
$\mathbf{Z}$	matrice de nouvelles observations
$\mathbf{z}$	nouvelle observation
$\mathcal{H}$	espace de Hilbert
$\mathcal{L}$	fonction de vraisemblance
$\mathcal{M}$	variété riemannienne des matrices SPD
$\mathcal{P}(C)$	espace des matrices SPD
$\mathcal{T}_p$	espace tangent à la variété en $\mathbf{P}$
μ	moyenne
Ω	étalement spectral
Φ	centroïde spectral
ϕ	transformation non linéaire vers un espace de Hilbert
ψ_0	terme de biais
ρ	coefficient de corrélation
σ	écart-type
τ	décalage temporel
Θ	score de latéralité
$\tilde{\mathbf{r}}$	combinaison linéaire des signaux de référence
$\tilde{\mathbf{x}}$	combinaison linéaire des données EEG
Ξ	signal sinusoïdal
ζ	dimension de l'espace tangent
A	amplitude
A_c	amplitude du signal porteur

A_m amplitude du signal modulant

A_y amplitude du signal modulé

B bande passante

b indice de *bin*

b_1 indice inférieur de la bande fréquentielle

b_2 indice supérieur de la bande fréquentielle

C nombre total d'électrodes

c signal porteur en fonction du temps

d durée

f fréquence

f_b fréquence associée au *bin* b

f_c fréquence du signal porteur

f_e fréquence d'échantillonnage

f_m fréquence de modulation

f_s fréquence de stimulation

G gain en décibels

g gain linéaire

H nombre d'harmoniques

h hyperplan séparateur

I_k nombre total d'*epochs* de la classe k

J fonction objectif

K nombre total de classes

k indice de classe

l nombre de filtres spatiaux retenus

m signal modulant en fonction du temps

N nombre total d'échantillons temporels

N_i nombre d'échantillons temporels dans la i-ème *epoch*

$p_{\boldsymbol{\omega}}$ probabilité prédite d'appartenance à la classe positive

P_{bruit} puissance spectrale des enregistrements EEG à la fréquence d'intérêt en l'absence de stimulus

$P_{\text{entrée}}$ puissance du signal d'entrée

P_{signal} puissance spectrale des enregistrements EEG à la fréquence d'intérêt en présence du stimulus

P_{sortie} puissance du signal de sortie

RF	réponse en fréquence
S	nombre total de fréquences de stimulation
s_b	valeur du spectre au *bin* b
t	temps
th	seuil de décision
v	nombre de composantes principales retenues
x_n	valeur du signal à l'indice n
y	signal modulé en fonction du temps
y_i	étiquette associée à la i-ème observation

Acronymes

BCI	*Brain-Computer Interface*
BP	*BereitschaftsPotential*
CCA	*Canonical Correlation Analysis*
CEI	Commission Electrotechnique Internationale
CNN	*Convolutional Neural Network*
CS	*Cross-Subject*
CSP	*Common Spatial Pattern*
DL	*Deep Learning*
ECoG	ElectroCorticoGraphie
EDF	*European Data Format*
EEG	ElectroEncéphaloGraphie
EHI	*Edinburgh Handedness Inventory*
ELU	*Exponential Linear Unit*
ERD	*Event-Related Desynchronization*
ERP	*Event-Related Potentials*
ERS	*Event-Related Synchronization*
FBCSP	*Filter-Bank Common Spatial Pattern*
FC	*Fully Connected*
FIR	*Finite Impulse Response*
FgMDM	*Fisher Geodesic Minimum-Distance-to-Mean*
GREEN	*Gabor Riemannian* EEGNet
HATS	*Head and Torso Simulator*
HL	*Hearing Level*
IIR	*Infinite Impulse Response*
IRM	Imagerie par Résonance Magnétique
ISO	*International Organization for Standardization*
ITU	*International Telecommunication Union*
KPCA	*Kernel Principal Component Analysis*
LDA	*Linear Discriminant Analysis*
LOO	*Leave-One-Out*
LOSO	*Leave-One-Subject-Out* ou *Leave-One-Session-Out*
LR	*Logistic Regression*
MDM	*Minimum-Distance-to-Mean*
MEA	*MicroElectrode Arrays*
MEG	MagnétoEncéphaloGraphie
MOABB	*Mother Of All Bci Benchmarks*
MSI	*Musical Sophistication Index*

NIRS	*Near-InfraRed Spectroscopy*
ORL	Oto-Rhino-Laryngologiste
PCA	*Principal Component Analysis*
ReLU	*Rectified Linear Unit*
RKHS	*Reproducing Kernel Hilbert Space*
RP	*Readidness Potential*
RSB	Rapport Signal sur Bruit
RiSVM	*Riemannian Support Vector Machine*
SCM	*Sample Covariance Matrix*
SCN	ShallowConvNet
SCP	*Slow Cortical Potentials*
SEP	*Sensory Evoked Potentials*
SOM	*Self-Organizing Map*
SPD	*Symmetric and Positive-Definite*
SPL	*Sound Pressure Level*
SSAEP	*Steady-State Auditory Evoked Potentials*
SSMEP	*Steady-State Somatosensory Evoked Potentials*
SSVEP	*Steady-State Visually Evoked Potentials*
SVD	*Singular Value Decomposition*
SVM	*Support Vector Machine*
TEP	Tomographie par Emission de Positons
TS	*Tangent Space*
TTD	*Thought Translation Device*
VEP	*Visual Evoked Potential*
WS	*Within-Session*

Introduction générale

L'analyse de l'activité cérébrale d'un individu permet d'identifier les processus cognitifs en cours et d'exploiter ces informations pour contrôler des dispositifs externes. C'est le principe des interfaces cerveau-ordinateur (*Brain-Computer Interface* ou BCI). Parmi les différentes catégories de BCI, les approches réactives reposent sur l'analyse des réponses cérébrales induites par des stimuli externes [Wolpaw, 2013]. Une approche largement étudiée consiste à exploiter les potentiels évoqués stationnaires, c'est-à-dire les réponses cérébrales occasionnées par une stimulation sensorielle périodique [Higashi et al., 2011]. Ces réponses peuvent être mesurées de manière non invasive, à l'aide de l'ElectroEncéphloGraphie (EEG), qui enregistre l'activité électrique à la surface du cuir chevelu [Wolpaw, 2013].

Les BCI réactives sont majoritairement basées sur des stimulations visuelles. En pratique, lorsqu'un utilisateur fixe un stimulus clignotant à une fréquence donnée, cette fréquence, ainsi que ses harmoniques, peuvent être détectées dans ses signaux cérébraux [Vialatte et al., 2010]. Leur identification permet alors de déterminer le stimulus ciblé et de déclencher l'action correspondante [Kim et al., 2007; Heo et al., 2017]. Cependant, ces systèmes, appelés **BCI basées sur les potentiels évoqués stationnaires visuels** (*Steady-State Visually Evoked Potentials* ou SSVEP), présentent plusieurs limites. Ils ne sont pas utilisables par tous [Kim et al., 2011a] et peuvent induire une fatigue visuelle. Par ailleurs, ces approches soulèvent des enjeux de confidentialité, dans la mesure où les stimuli sont généralement présentés sur des écrans visibles par des tiers. L'utilisation d'un casque de réalité étendue peut atténuer cette contrainte, mais constitue une solution plus contraignante sur le plan matériel.

Dans ce contexte, les BCI auditives apparaissent comme un complément, voire une alternative prometteuse. En particulier, les traitements cérébraux auditifs étant généralement plus rapides que les traitements visuels [Pascual-Leone et al., 1992], l'utilisation de stimulations sonores pourrait potentiellement améliorer la réactivité du système. Ces interfaces reposent notamment sur les potentiels évoqués stationnaires auditifs (*Steady-State Auditory Evoked Potentials* ou SSAEP), générés en réponse à des sons modulés en amplitude, analogues aux stimulations visuelles clignotantes. Cette modulation résulte de la combinaison d'un signal porteur, correspondant au son perçu par l'utilisateur, et d'un signal modulant, sous la forme d'une onde sinusoïdale dont la fréquence, appelée **fréquence de modulation**, est détectable dans l'activité cérébrale. Toutefois, la majorité des travaux existants s'appuie sur l'utilisation de sons purs comme signaux porteurs [Lopez et al., 2009; Xie et al., 2021; Kim et al., 2011a; Kaongoen et Jo, 2017; Higashi et al., 2011]. Or, leur utilisation prolongée soulève des questions importantes en termes de confort auditif et d'acceptabilité dans des contextes d'usage réel, alors que ces aspects ont été peu étudiés jusqu'à présent. Dès lors, afin de pérenniser l'usage des BCI auditives, il est nécessaire, d'une part, de concevoir des stimulations sonores à la fois confortables pour l'utilisateur et détectables dans l'activité cérébrale. D'autre part, le développement de méthodes de détection performantes permettant d'identifier les fréquences de modulation dans les signaux EEG apparaît indispensable.

1 Objectif

L'objectif de cette thèse est d'évaluer des stimulations auditives pour les BCI, tant d'un point de vue perceptif que fonctionnel. Plus précisément, elle vise à analyser l'influence de la nature des stimuli sur la perception des utilisateurs et sur la détectabilité des fréquences de modulation dans leurs signaux EEG, c'est-à-dire sur leur capacité à émerger de l'activité cérébrale enregistrée.

2 Contributions

Cette thèse propose un cadre expérimental pour l'évaluation de stimulations sonores dans un contexte de BCI, en intégrant conjointement des aspects acoustiques, physiologiques et perceptifs.

L'étude repose sur l'enregistrement de signaux EEG chez 48 participants exposés à différents sons modulés en amplitude à 40 Hz [Galambos et al., 1981; Kaongoen et Jo, 2017]. Les stimuli considérés comprennent un son pur (utilisé comme référence), des sons naturels (chants des cigales [Heo et al., 2017], ronronnement de chat), ainsi qu'un bruit Brownien (caractérisé par un contenu spectral proche de certains sons naturels, tels que les bruits d'eau). Ces stimulations ont été sélectionnées car elles sont susceptibles d'offrir un meilleur confort auditif que des signaux sinusoïdaux. Elles ont été présentées à deux niveaux de sonie (60 et 66 phones), afin d'évaluer l'influence du niveau sonore sur la détectabilité de la fréquence de modulation. Les signaux EEG ont été acquis à l'aide d'un casque comportant 24 électrodes, offrant une résolution spatiale suffisante pour analyser la distribution des réponses cérébrales. L'un des objectifs étant notamment d'identifier les électrodes contribuant le plus à la détection de la fréquence de modulation. La détectabilité de la fréquence de modulation dans les données EEG a été évaluée en termes de Rapport Signal sur Bruit (RSB) et de performances de classification. Plusieurs méthodes d'apprentissage automatique récentes, incluant des approches riemanniennes et des réseaux d'apprentissage profond, ont été comparées, afin d'établir un référentiel des stratégies de détection des SSAEP.

Un soin particulier a été apporté à la caractérisation de la chaîne de reproduction sonore, de manière à garantir une restitution fidèle des stimuli auditifs, notamment par la mesure de la réponse en fréquence de l'ensemble des dispositifs utilisés, à l'aide d'un simulateur de tête et de torse (*Head and Torso Simulator* ou HATS), permettant de contrôler la pression acoustique au niveau du conduit auditif. Par ailleurs, les capacités auditives des participants ont été évaluées en amont de l'expérience afin de détecter d'éventuelles déficiences. À cette fin, un test auditif spécifique, adaptable et réutilisable dans de futures études, a été développé en remplacement des dispositifs habituellement employés. Les participants ont également complété des questionnaires permettant d'établir leur profil et d'identifier les facteurs susceptibles d'influencer leurs réponses cérébrales. Enfin, leurs ressentis vis-à-vis des stimulations, tels que la préférence, l'inconfort ou la sonie perçue, ont été recueillis et analysés.

3 Structure du manuscrit

Le manuscrit est structuré en deux parties : un état de l'art suivi d'une étude expérimentale.

Dans la première partie, le chapitre 1 donne une vue d'ensemble du cerveau humain, avant de mettre l'accent sur les régions impliquées dans le traitement auditif, ainsi que sur les facteurs individuels influençant ce traitement. Le chapitre 2 est consacré aux interfaces cerveau-ordinateur. Il détaille les différents types de BCI, les techniques d'enregistrement de l'activité cérébrale, ainsi que les principaux marqueurs neurophysiologiques exploités, notamment les potentiels évoqués

stationnaires et plus particulièrement, les potentiels évoqués stationnaires auditifs. Le chapitre 3 traite de l'analyse des signaux EEG, en abordant leurs propriétés, les étapes de prétraitement, les méthodes de réduction de dimension et les techniques de détection de la fréquence de modulation.

Dans la seconde partie, le chapitre 4 décrit le protocole expérimental et les choix méthodologiques associés. Il prend en compte à la fois les facteurs acoustiques et humains, en caractérisant l'ensemble de la chaîne de reproduction des stimuli et en évaluant le profil, ainsi que les perceptions de chaque participant. Le chapitre 5 présente les résultats expérimentaux. Il étudie l'influence des étapes de prétraitement et évalue la détection de la fréquence de modulation en fonction du type de stimulus et du niveau sonore de présentation, tout en explorant la distribution spatiale des réponses cérébrales associées aux stimuli sonores. Le chapitre 6 s'intéresse au profil des participants et analyse leurs perceptions de l'expérimentation, notamment en ce qui concerne les stimuli. Enfin, le dernier chapitre (chapitre 7) synthétise les principales contributions de cette thèse et ouvre plusieurs perspectives, notamment en formulant des recommandations pour la conception de BCI auditives basées sur les SSAEP.

Première partie

Interactions cerveau-machine : contexte et problématique

1

Aires cérébrales impliquées dans le traitement auditif

Sommaire

1.1 Introduction 3
1.2 Organisation générale du cerveau humain 4
1.3 Mécanismes cérébraux du traitement de la parole 5
1.4 Mécanismes cérébraux du traitement de la musique 9
1.5 Mécanismes cérébraux de la discrimination des catégories sonores 11
1.6 Mécanismes cérébraux de la localisation sonore 12
1.7 Variabilité interindividuelle du traitement cérébral des sons . . 13
1.7.1 Influence de l'expertise musicale 13
1.7.2 Influence de la latéralité 13
1.7.3 Influence du sexe biologique 15
1.8 Influence des substances psychoactives sur le traitement cérébral 15
1.8.1 Effets de la caféine et de la théine 15
1.8.2 Effets de la nicotine 16
1.8.3 Effets de l'alcool 16
1.8.4 Effets des drogues et des médicaments 16
1.9 Conclusion 16

1.1 Introduction

L'analyse des réponses cérébrales aux stimuli auditifs constitue un élément important pour le développement d'interfaces cerveau-ordinateur auditives. L'interprétation de ces réponses repose sur la compréhension des mécanismes neurobiologiques impliqués dans le traitement de sons, qu'il s'agisse de signaux simples (e.g. sinusoïdaux) ou complexes (e.g. parole, musique). Elle dépend également de la répartition spatiale des sources sonores dans l'environnement acoustique. En outre, les caractéristiques anatomiques et fonctionnelles propres à chaque individu sont susceptibles d'influencer le traitement auditif et la perception sonore. À cela s'ajoutent des facteurs liés à l'état physiologique et aux consommations, tels que la prise de substances psychoactives (e.g. caféine, nicotine, alcool, médicaments, etc.), qui peuvent moduler l'activité cérébrale. La prise en compte de l'ensemble de ces facteurs apparaît ainsi essentielle pour garantir une interprétation

fiable de l'activité cérébrale enregistrée, identifier avec précision les régions cérébrales impliquées et optimiser le positionnement des capteurs d'enregistrement.

Dans ce but, ce chapitre vise à caractériser les principales structures cérébrales engagées dans le traitement auditif. Après une présentation générale de l'organisation anatomique du cerveau et du système auditif (section 1.2), les régions et les processus impliqués dans le traitement de la parole (section 1.3) et de la musique (section 1.4), ainsi que la discrimination (section 1.5) et la localisation des sons (section 1.6) sont détaillés. Les facteurs individuels pouvant influencer l'anatomie et/ou l'activation des aires cérébrales, ainsi que la perception des sons, sont ensuite présentés en section 1.7. Enfin, la section 1.8 explore l'influence potentielle de plusieurs substances sur les réponses cérébrales enregistrées.

1.2 Organisation générale du cerveau humain

Le cerveau humain est composé de deux hémisphères (gauche et droit) séparés par un profond sillon (la fissure longitudinale), mais connectés par des voies nerveuses (la voie la plus large est le corps calleux) [Carlach et al., 2017]. À la surface de ces hémisphères se trouve le cortex, une couche de substance grise impliquée dans le traitement de l'information. Cette substance est principalement composée des corps cellulaires des neurones (i.e. leur partie centrale, contenant notamment le noyau) et de leurs dendrites, des prolongements impliqués dans la réception et la transmission de l'information nerveuse.

Chaque hémisphère est divisé en quatre lobes principaux : le frontal, le pariétal, le temporal et l'occipital (figure 1.1). Chaque lobe est associé à une ou plusieurs fonctions, bien qu'il y ait un certain chevauchement fonctionnel entre eux.

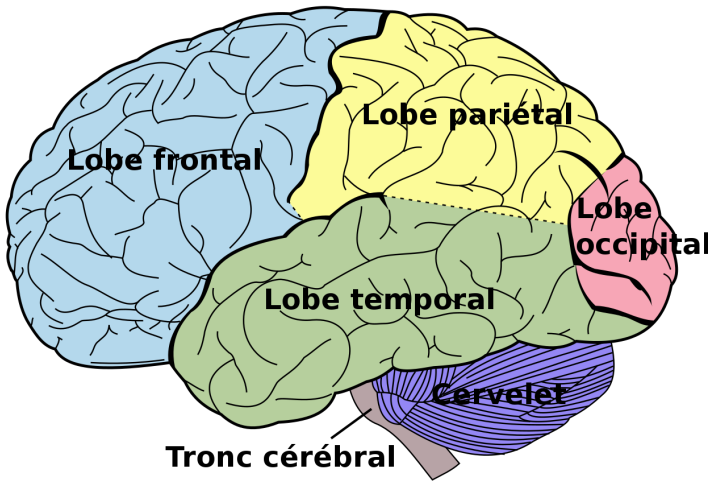


FIGURE 1.1 – Lobes externes du cerveau humain (figure tirée de [Wikipédia, 2006a]).

L'aire de traitement visuel du cerveau se situe dans le lobe occipital. Le cortex auditif, quant à lui, est divisé en une douzaine d'aires entourant la surface supérieure du lobe temporal.

Chez l'humain, les principales composantes corticales du système auditif se situent dans un plan frontal qui s'étend du conduit auditif au lobe temporal supérieur [Patterson et Johnsrude, 2008] (figure 1.2). Comme d'autres aires sensorielles, le cortex auditif est organisé hiérarchiquement en aires primaires et secondaires [Pinel, 2007]. La surface supérieure du lobe temporal (correspondant aux aires 41 et 42 définies par Korbinian Brodmann en 1909 [Strotzer, 2009]) est

le siège des aires auditives primaires. Le cortex auditif primaire est entouré d'au moins six autres zones, anatomiquement distinctes, impliquées dans le traitement et l'encodage des informations auditives, formant ainsi un sous-réseau de régions cérébrales [Viceic et al., 2006]. L'examen des régions composant ce sous-réseau indique toutefois que la plupart d'entre elles ne sont pas directement impliquées dans le traitement auditif lui-même, mais participent plutôt à des fonctions cognitives associées, telles que la mémoire ou le langage [Bonetti et al., 2021]. Les régions qui entourent le cortex auditif primaire constituent le cortex auditif secondaire. Celui-ci comprend notamment le planum temporale, situé sur la face supérieure du gyrus temporal supérieur. Un **gyrus** désigne un ensemble de replis sinueux du cortex cérébral, délimité par des sillons plus profonds qui marquent la surface du cerveau. Le planum temporale est associé à l'aire 22 de Brodmann [Golestani, 2014], dont une partie recouvre l'aire de Wernicke (figure 1.3).

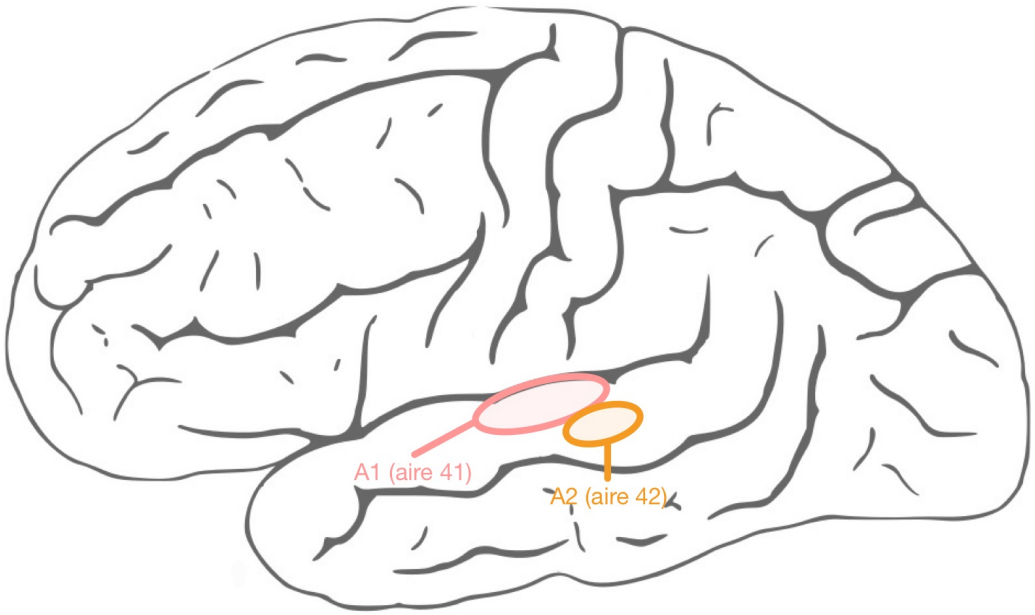


FIGURE 1.2 – Visualisation des différentes aires composant le cortex auditif.

1.3 Mécanismes cérébraux du traitement de la parole

Le son d'un mot entendu est d'abord traité dans le cortex auditif primaire. Au niveau de l'hémisphère gauche, autour du sillon latéral (qui délimite le lobe temporal du lobe frontal), s'organise un réseau neuronal impliqué dans la compréhension et la production orale du langage. À l'extrémité frontale de cette boucle se situe l'aire de Broca (figure 1.3), habituellement associée à la production du langage. À l'autre extrémité, plus précisément dans la partie supérieure et postérieure du lobe temporal, se situe l'aire de Wernicke, associée au traitement des paroles entendues, autrement dit à la compréhension du langage [Patterson et Johnsrude, 2008]. L'aire de Broca et de Wernicke sont connectées par un ensemble de fibres nerveuses, appelé **faisceau arqué**. Bien que les termes d'aires de Broca et de Wernicke soient communément utilisés, les limites de ces régions ne sont pas clairement définies et elles peuvent varier d'une personne à l'autre, elles ne peuvent pas être décrites de façon précise en se référant uniquement aux aires de Brodmann. En effet, si pour plusieurs auteurs, l'aire de Broca correspond aux régions 44 et 45 de Brodmann, pour d'autres, il s'agit quelques fois uniquement de l'aire 44 ou 45, ou

même parfois de l'aire 44, 45 et 47. Par ailleurs, ces aires peuvent aussi être impliquées dans d'autres fonctions que le langage. Le concept initial de l'aire de Broca a ainsi évolué, passant d'un "centre du langage" à celui d'une région intégrée au sein d'un réseau cérébral complexe. Cette région participe non seulement aux traitements sémantiques, syntaxiques et phonologiques, mais également à certaines tâches non-reliées au langage, telles que le contrôle du mouvement. Ces observations suggèrent l'existence possible de subdivisions fonctionnelles au sein de cette aire.

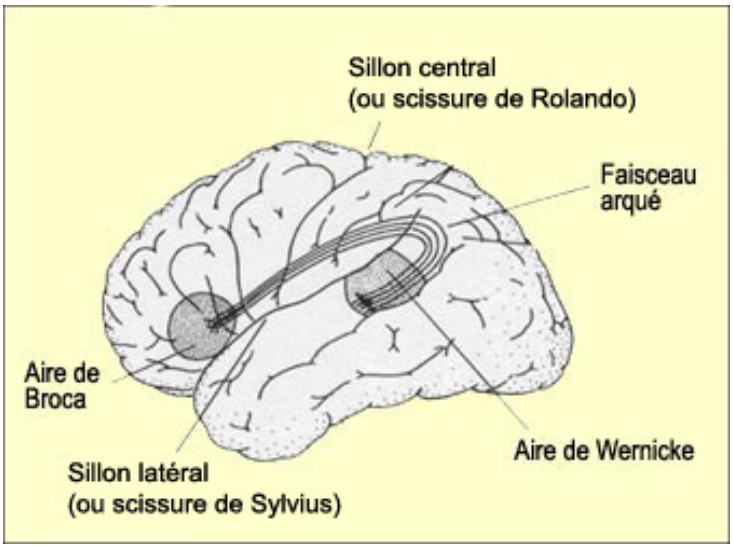


FIGURE 1.3 – Visualisation du faisceau arqué reliant l'aire de Broca et l'aire de Wernicke (figure tirée de [Le cerveau à tous les niveaux, nd]).

Outre les aires de Broca et Wernicke, une troisième région d'importance pour le langage, située dans le cortex pariétal, a été décrite par la suite. Il s'agit du lobule pariétal inférieur (figure 1.4), aussi appelé **territoire de Geschwind**. Grâce à l'imagerie cérébrale, il a été montré que ce lobule pariétal inférieur (comprenant le gyrus angulaire et le gyrus supramarginal) était connecté par d'importants faisceaux de fibres nerveuses à la fois à l'aire de Broca et à celle de Wernicke. L'information pourrait donc transiter entre ces deux régions soit directement par le faisceau arqué, soit par une route parallèle en passant par le territoire de Geschwind. Le lobule pariétal inférieur de l'hémisphère gauche occupe un endroit clé dans le cerveau, à l'intersection des cortex auditif, visuel et somatosensoriel (impliqué dans le traitement des informations liées notamment au toucher, à la température, à la position des membres et aux dommages tissulaires), avec qui il est massivement connecté. De plus, les neurones de cette région ont la particularité d'être multimodaux, c'est-à-dire qu'ils sont capables de traiter simultanément des stimuli de différentes natures (auditif, visuel, sensorimoteur, etc.). Ces caractéristiques font du lobule pariétal inférieur un candidat idéal pour appréhender les multiples propriétés d'un mot : son aspect visuel, sa fonction, son nom, etc. Cette intégration multimodale implique également d'autres structures cérébrales, notamment le thalamus, qui assure la transmission des informations visuelles, auditives, gustatives et somatosensorielles vers le cortex cérébral, tout en participant à la sélection des informations accédant à la conscience.

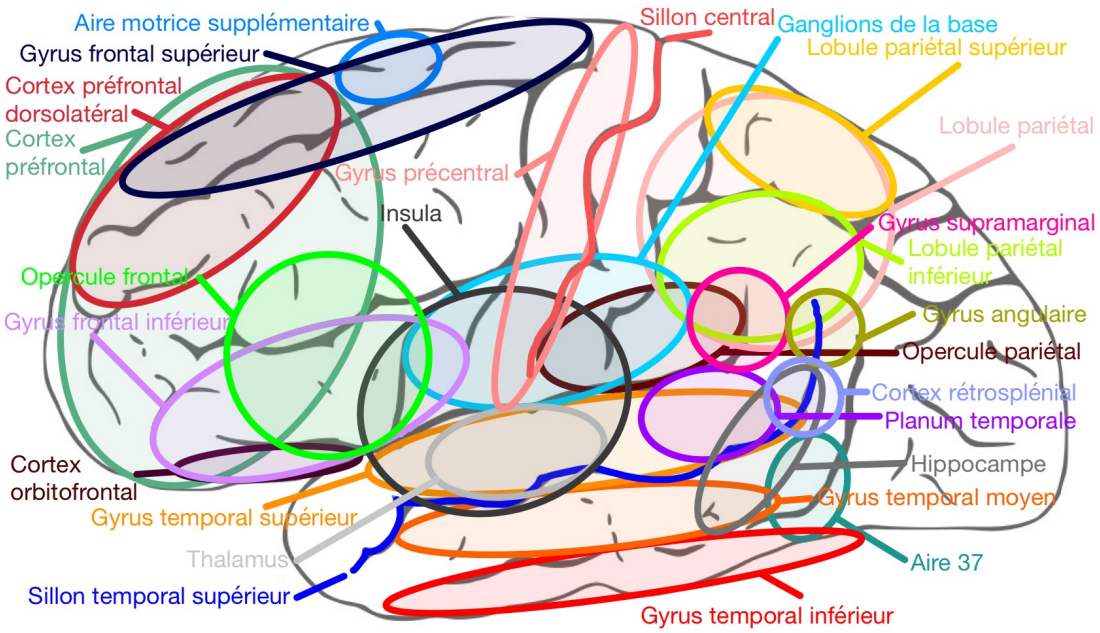


FIGURE 1.4 – Schéma des différentes régions cérébrales pouvant être impliquées dans le traitement des sons.

Selon le modèle proposé par Hickok et Poeppel [2004], les premières étapes du traitement de la parole reposeraient sur un réseau auditif organisé de manière bilatérale, mais asymétrique, dans les régions supérieures du lobe temporal, notamment au niveau du gyrus temporal supérieur. Le traitement se répartirait ensuite entre deux voies principales : une voie ventrale, située dans les régions inférieures du cerveau (figure 1.5) et impliquée dans l'association des sons à leur signification, et une voie dorsale, située dans les régions supérieures, qui participerait davantage à la conversion des signaux acoustiques de la parole en représentations articulatoires [Hickok et Poeppel, 2004, 2007].

La voie ventrale suit une trajectoire ventrolatérale et mobilise d'abord des régions du sillon temporal supérieur, puis des régions situées plus en aval, dans la partie postérieure du lobe temporal inférieur. Dans ces dernières, certaines observations ont montré des réponses plus importantes lors de l'écoute de mots que de pseudomots (i.e. des séquences phonologiques ressemblant à des mots, mais dépourvues de signification). Ce résultat est compatible avec l'hypothèse selon laquelle ces régions contribuent à l'association entre les sons et leur signification [Hickok et Poeppel, 2004]. Alors que les modèles classiques attribuent un rôle prédominant à l'hémisphère gauche dans le traitement de la parole, le modèle de Hickok et Poeppel [2007] propose une organisation bilatérale de la voie ventrale. Cette bilatéralité ne signifie cependant pas que les gyri temporaux supérieurs gauche et droit réalisent des traitements strictement équivalents. Les deux hémisphères pourraient participer conjointement au codage de la parole tout en étant sensibles à des propriétés différentes du signal [Hickok et Poeppel, 2004].

La voie dorsale suit, pour sa part, une trajectoire dorso-postérieure, depuis les régions temporales vers le lobe pariétal, puis vers les régions frontales [Hickok et Poeppel, 2004]. Elle comprend notamment des structures situées dans la partie postérieure du lobe frontal, dans la région dorsale et postérieure du lobe temporal, ainsi que dans l'opercule pariétal. Cette voie auditive dorsale assurerait une interface entre les systèmes perceptifs et moteurs. Elle présente une dominance hémisphérique gauche marquée. Dans l'hémisphère gauche, elle ne mobiliserait toutefois qu'une

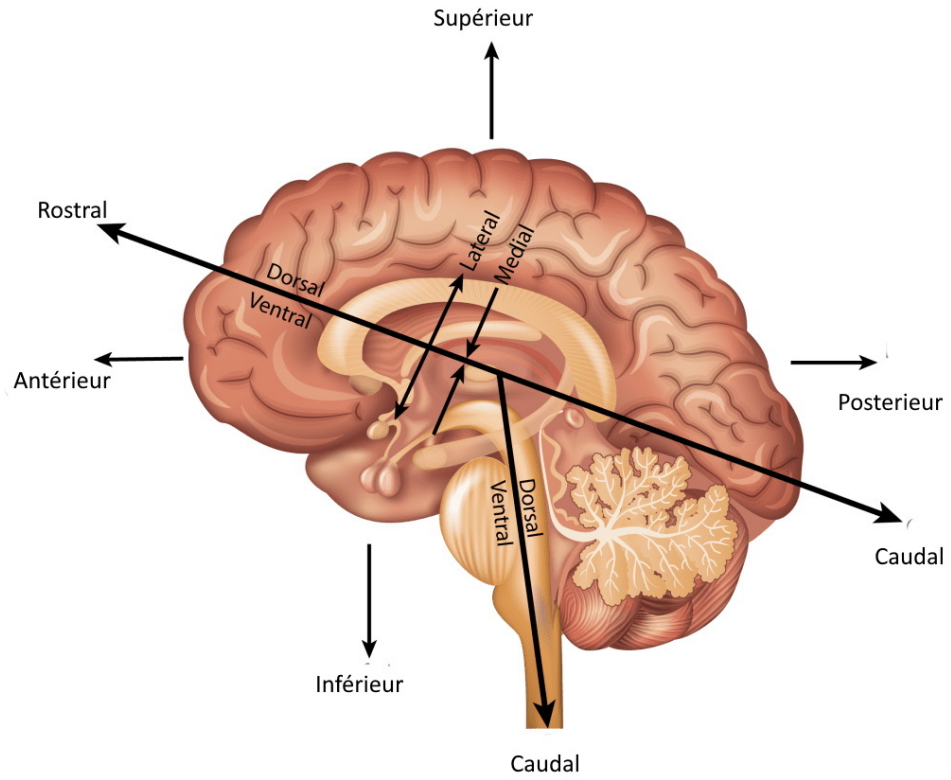


FIGURE 1.5 – Termes directionnels anatomiques utilisés pour décrire la localisation des différentes régions cérébrale (figure tirée de [neuromedia, nd]).

partie du planum temporale [Hickok et Poeppel, 2007]. Bien que cette région ait traditionnellement été associée au traitement de la parole, des études d'imagerie fonctionnelle ont montré qu'elle répond également à différents types de stimuli acoustiques, tels que des séquences de sons purs, des signaux musicaux ou des informations spatiales, ainsi qu'à certains signaux non acoustiques, notamment la parole visuelle, la langue des signes et le mouvement visuel [Hickok et Poeppel, 2004]. Ces résultats ont conduit à envisager le planum temporale non comme une région exclusivement consacrée à la parole, mais comme un carrefour computationnel. Il recevrait des informations provenant du cortex auditif primaire et contribuerait, d'une part, à distinguer les différentes composantes d'un signal sonore complexe et, d'autre part, à analyser leurs caractéristiques spectrales et temporelles. Ces opérations permettraient d'extraire des informations relatives à l'identité des objets sonores, ensuite traitées dans les régions latérales du lobe temporal, ainsi qu'à leur localisation spatiale, prise en charge plus en aval par les structures pariétales [Hickok et Poeppel, 2004].

Enfin, bien que les signaux de parole soient traités de manière bilatérale, les deux hémisphères présenteraient une sensibilité différenciée à certaines propriétés acoustiques. L'hémisphère gauche serait relativement plus sensible aux changements spectrotemporels rapides et privilégierait l'analyse d'informations intégrées sur des fenêtres temporelles courtes, de l'ordre de 25 à 50 ms. À l'inverse, l'hémisphère droit serait davantage impliqué dans le traitement de variations plus lentes, telles que les modulations de hauteur associées à la prosodie. Il opérerait préférentiellement sur des fenêtres d'intégration plus longues, comprises entre 150 et 250 ms.

1.4 Mécanismes cérébraux du traitement de la musique

L'écoute de sons musicaux entraîne une réponse dans le cortex auditif primaire (aire 41) dont la dynamique temporelle diffère significativement entre les deux hémisphères [Izumi et al., 2011]. Plus précisément, le cortex auditif primaire gauche présente, dès l'apparition de la stimulation, une réponse transitoire marquée, caractérisée par une augmentation rapide et brève de l'activité. À l'inverse, le cortex auditif primaire droit ne manifeste pas de composante transitoire initiale, mais une activation soutenue qui s'installe de manière plus progressive et se maintient dans le temps. Ces résultats suggèrent que le cortex auditif gauche est spécialisé dans le traitement temporel rapide et le cortex auditif droit dans le traitement spectral [Zatorre et Belin, 2001; Zatorre et al., 2002]. Ainsi, les aires auditives gauches traitent les stimuli auditifs comme la parole, tandis que les aires auditives droites participent au traitement des stimuli auditifs tels que la musique [Belin et al., 1998].

Au-delà du cortex auditif primaire, l'écoute d'une pièce musicale mobilise un réseau plus étendu (figure 1.6). Une activation significative, ainsi qu'une augmentation de la centralité (i.e. le nombre de connexions associée à une région cérébrale) sont observées, principalement dans l'hémisphère droit, au niveau de plusieurs régions cérébrales impliquées dans la mémoire, l'attention et les processus auditifs, telles que le cortex auditif primaire, l'opercule frontal, les ganglions de la base, l'insula et l'hippocampe [Bonetti et al., 2021]. Plus spécifiquement, l'écoute de mélodies est associée à l'activation de régions du cortex temporal supérieur (aussi appelé **gyrus temporal supérieur**) droit, tandis que la comparaison de la hauteur des sons au sein de ces mélodies implique principalement le cortex préfrontal droit [Bonetti et al., 2021]. Lorsque la musique est inconnue des participants, la centralité du cortex orbitofrontal médian droit est significativement plus forte au cours de l'encodage des sons. Enfin, une étude portant sur la reconnaissance de musiques suggère que la réussite de cette tâche est associée à l'activation des gyri temporaux supérieurs droits, frontaux supérieurs, frontaux inférieurs bilatéraux, précentraux bilatéraux et supramarginaux gauches [Bonetti et al., 2021].

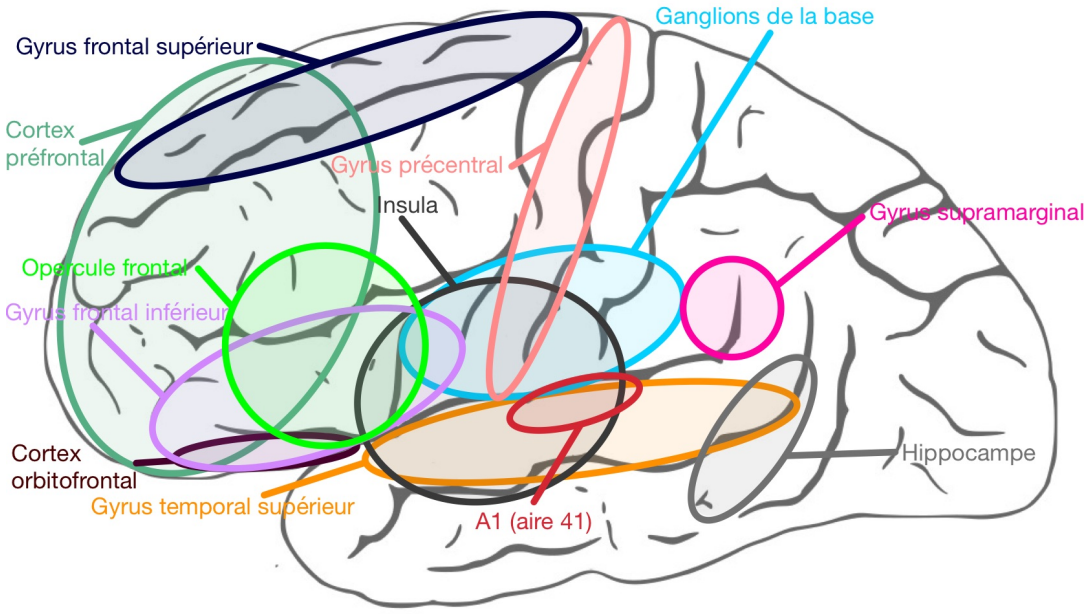


FIGURE 1.6 – Schéma des différentes régions cérébrales impliquées dans le traitement des sons musicaux.

Plus généralement, le traitement de stimuli acoustiques présentant des variations de leurs caractéristiques temporelles et spectrales repose sur l'activation des régions auditives corticales, telles que le gyrus de Heschl et le gyrus temporal supérieur (figure 1.7). Le gyrus de Heschl est une structure transverse, orientée perpendiculairement au gyrus temporal supérieur, sur sa face supérieure. Localisé en profondeur dans le sillon latéral (figure 1.3), il n'est pas directement visible à la surface sans écarter les opercules frontal et pariétal. La modification de la hauteur ou de l'enveloppe spectrale de sons harmoniques (i.e. composés d'une fréquence fondamentale et de plusieurs fréquences multiples de celle-ci) entraîne une activation au sein d'un réseau bilatéral (i.e. aux deux hémisphères) comprenant le gyrus latéral de Heschl et les zones corticales adjacentes dans le lobe temporal supérieur, le planum temporale antérolatéral [Warren et al., 2005]. Selon Patterson et Johnsrude [2008], les neurones de la région du gyrus antérolatéral de Heschl sont impliqués dans l'évaluation de la hauteur et de l'amplitude des sons.

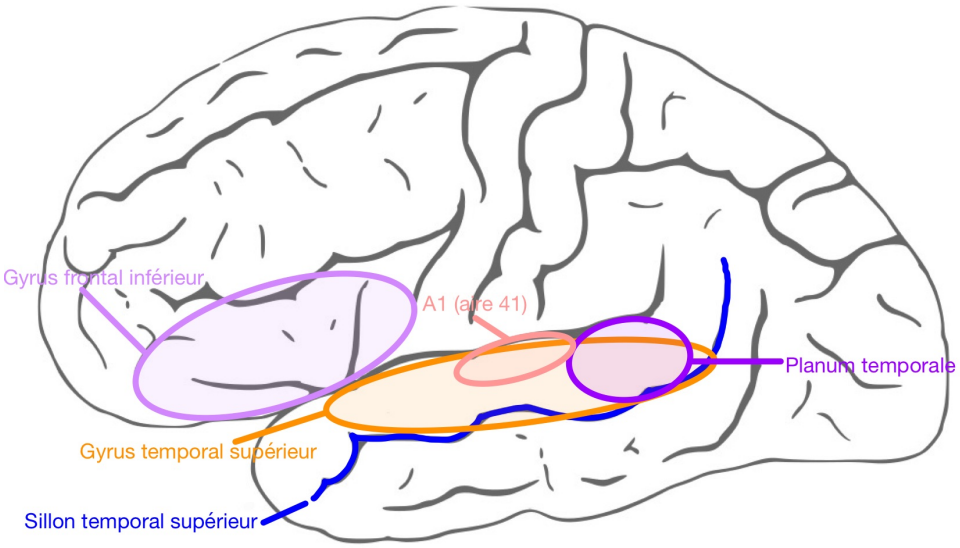


FIGURE 1.7 – Schéma des différentes régions cérébrales impliquées dans le traitement de stimuli acoustiques de différentes natures, présentant des variations de leurs caractéristiques temporelles et spectrales.

Enfin, la nature du stimulus influence l'intensité et l'étendue des activations corticales (figure 1.7). Les stimuli acoustiques simples, tels qu'un sinus pur ou un signal sinusoïdal modulé en amplitude, activent le cortex auditif médian, comprenant le cortex auditif primaire et les régions auditives supérieures environnantes [Hashimoto et al., 2006]. Ces activations demeurent toutefois plus faibles que celles induites par des sons naturels complexes, tels que des vocalisations animales ou des sons d'instruments, qui augmentent notamment l'activation du gyrus de Heschl, en particulier dans l'hémisphère droit [Specht et Reul, 2003]. De plus, les régions corticales entourant bilatéralement le gyrus de Heschl (notamment le planum temporale et le gyrus temporal supérieur dorsolatéral) sont davantage activées par des signaux sinusoïdaux modulés en amplitude que par du bruit, suggérant leur implication dans le traitement de stimuli auditifs structurés dans le temps [Binder et al., 2000]. Par ailleurs, des réponses à divers sons naturels, tels que des outils manipulés par un humain ou des vocalisations animales, ont été observées dans le sillon temporal supérieur [Beauchamp et al., 2004; Lewis et al., 2004]. Les sons d'animaux entraînent aussi des activations dans le gyrus frontal inférieur gauche [Hashimoto et al., 2006]. Des travaux antérieurs ont également montré que, lors d'une écoute passive de stimuli acoustiques simples

(e.g. des sons purs ou modulés en amplitude), l'évolution temporelle de la réponse dans le cortex auditif primaire est symétrique entre les deux hémisphères [Hashimoto et al., 2006].

1.5 Mécanismes cérébraux de la discrimination des catégories sonores

La discrimination des catégories de sons repose sur des réseaux distribués à travers les deux hémisphères, impliquant à la fois les aires auditives primaires et des aires associatives impliquées dans des processus cognitifs complexes (figure 1.8) [Zhang et al., 2020].

Ainsi, la catégorisation de quatre types de sons, tels que des voix humaines, des vocalisations animales, des sons d'outils (e.g. un sèche-cheveux, un marteau) et des bruits de circulation, repose sur des connexions impliquant les régions temporales, frontales, pariétales et occipitales dans les deux hémisphères [Zhang et al., 2020]. Les régions temporales supérieures jouent un rôle central dans cette distinction [Staeren et al., 2009]. Le gyrus temporal supérieur et le gyrus de Heschl présentent notamment une capacité à discriminer les voix humaines, les vocalisations animales et les bruits de circulation. De plus, les gyri temporaux moyen et supérieur, le gyrus de Heschl, ainsi que le sillon temporal supérieur, contribuent de manière significative à la discrimination des voix humaines, des vocalisations animales et des sons d'outils [Zhang et al., 2020]. Enfin, le planum temporale bilatéral et le gyrus temporal supérieur dorsolatéral répondent de la même manière aux sons et aux mots, tandis que le sillon temporal supérieur, les gyri temporaux inférieur et moyen, et le gyrus angulaire, particulièrement dans l'hémisphère gauche, montrent des réponses plus importantes pour les mots que pour les sons [Hashimoto et al., 2006].

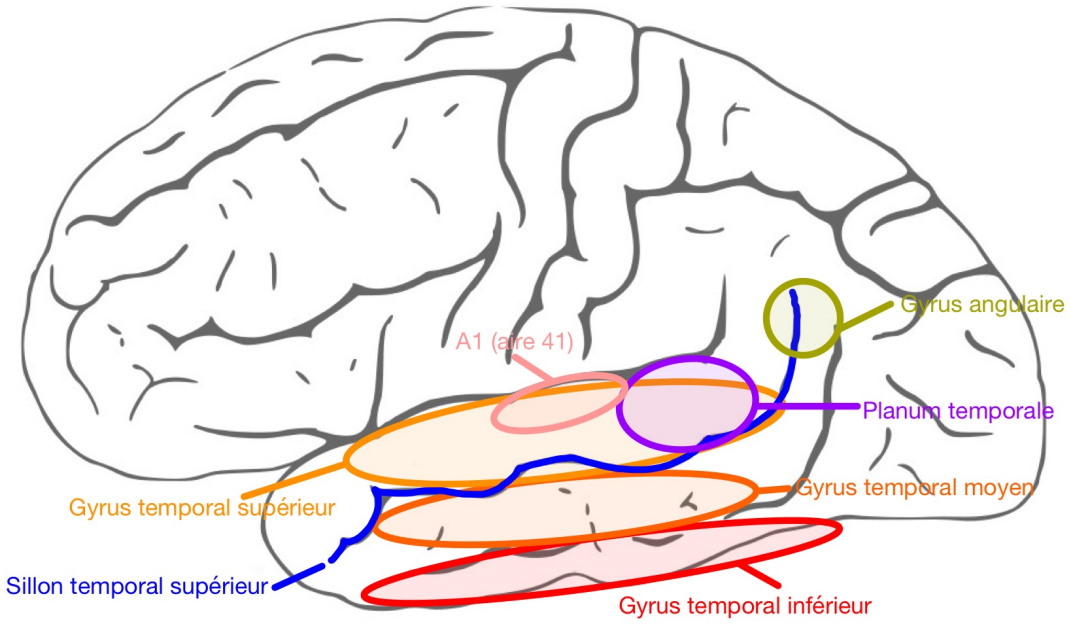


FIGURE 1.8 – Schéma des différentes régions cérébrales impliquées dans la discrimination de catégories de sons.

Par ailleurs, les sons désagréables (e.g. des pleurs de bébé, une alarme de voiture ou une sirène de police) activent des régions cérébrales plus étendues que les sons naturels (e.g. une rivière, une forêt ou la pluie). En revanche, l'intensité des réponses cérébrales est plus élevée pour les

sons naturels que pour les sons désagréables [Hong et Santosa, 2016].

1.6 Mécanismes cérébraux de la localisation sonore

Le traitement auditif spatial repose sur un réseau complexe (figure 1.9) impliquant des régions antérieures et postérieures du lobe temporal, le cortex frontal inférieur, le cortex préfrontal dorsolatéral, ainsi que le cortex pariétal postérieur [Lewald et al., 2008]. En effet, des déficits spécifiques du traitement auditif spatial ont été rapportés chez des patients présentant des lésions du cortex pariétal.

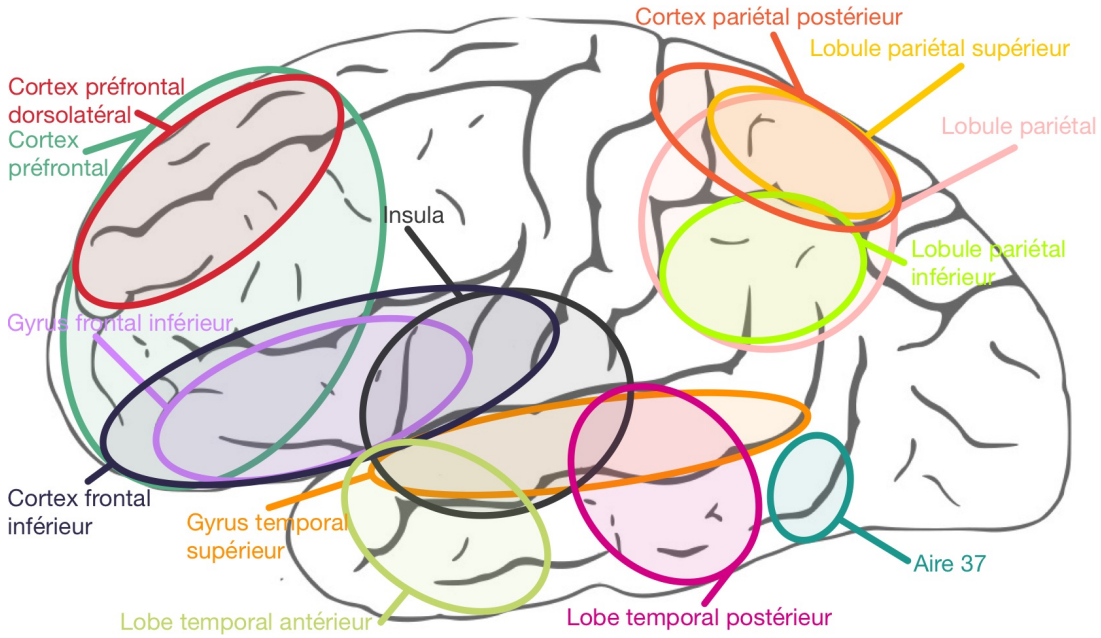


FIGURE 1.9 – Schéma des différentes régions cérébrales impliquées dans le traitement auditif spatial.

Toutefois, différents niveaux de complexité peuvent être distingués dans ce traitement. Les processus les plus simples, tels que la discrimination de l'hémichamp (gauche/droit), impliquent des régions situées à proximité du cortex auditif primaire [Viceic et al., 2006], notamment la jonction temporo-pariéto-occipitale, incluant le gyrus temporal supérieur postérieur, le lobule pariétal inférieur et l'aire 37 de Brodmann [Lewald et al., 2008]. En revanche, l'intégration d'informations spatiales plus complexes, telles que l'excentricité des sources sonores (i.e. leur position angulaire par rapport à l'axe frontal), ferait intervenir des régions plus antérieures du lobe temporal, ainsi que des régions dorsales du lobule pariétal postérieur, mais également des régions plus distantes, telles que le gyrus frontal inférieur et le cortex préfrontal dorsolatéral.

Plus généralement, ces observations suggèrent que le traitement spatial auditif est étroitement lié à d'autres fonctions perceptives, notamment la perception de la hauteur sonore.

Enfin, lorsque l'attention est portée sur le mouvement du son, l'insula, située en profondeur dans le gyrus temporal supérieur [Griffiths et al., 1994], est également activée.

1.7 Variabilité interindividuelle du traitement cérébral des sons

Le traitement des sons présente une variabilité interindividuelle marquée, qui se manifeste tant au niveau fonctionnel qu'anatomique. Certaines de ces différences sont intrinsèques, comme en témoigne la forte variabilité morphologique du gyrus de Heschl entre les individus [Golestani, 2014]. D'autres variations sont associées à des facteurs individuels, tels que l'expertise musicale, la latéralité ou encore le sexe, qui influencent les régions cérébrales recrutées lors du traitement auditif.

1.7.1 Influence de l'expertise musicale

La formation musicale et la possession de l'oreille absolue (i.e. la capacité à identifier la hauteur d'un son sans référence auditive préalable) constituent des facteurs individuels responsables de l'asymétrie hémisphérique du cortex auditif [Izumi et al., 2011].

De plus, lors de l'écoute de sons musicaux, les participants ayant une plus grande expertise musicale présentent une plus grande centralité de l'aire motrice supplémentaire gauche, du gyrus temporal supérieur et de l'insula droit [Bonetti et al., 2021] (figure 1.10). L'activation de l'aire motrice supplémentaire pourrait s'expliquer par l'implication de l'apprentissage moteur dans la pratique musicale, suggérant que les musiciens recrutent également des régions motrices lors de l'encodage des sons.

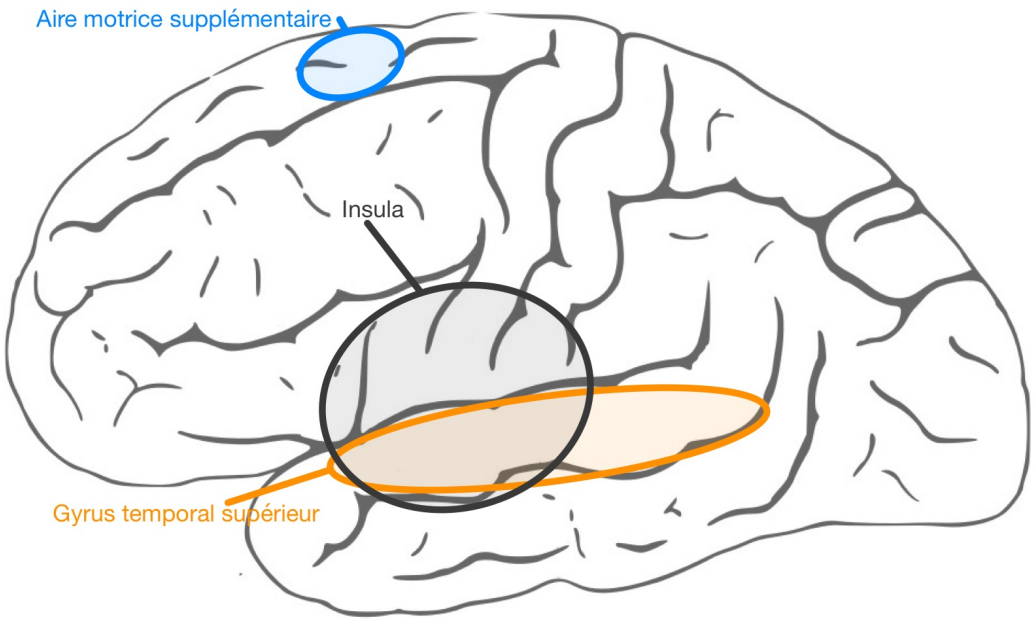


FIGURE 1.10 – Régions cérébrales présentant une plus grande centralité lors de l'écoute de sons musicaux chez des participants ayant une plus grande expertise musicale.

1.7.2 Influence de la latéralité

La préférence manuelle est corrélée à l'asymétrie cérébrale [Golestani, 2014]. En effet, elle influe sur l'hémisphère dominant pour le langage, bien qu'il existe une prédisposition naturelle de l'hémisphère gauche à assurer cette fonction. Ainsi, entre 95 et 99% des droitiers présentent une dominance cérébrale gauche pour le langage, contre 70 à 80% des gauchers [Corballis et al.,

2012], qui représentent environ 10% de la population. Par conséquent, une proportion importante de gauchers présente également une spécialisation du langage dans l'hémisphère gauche, tandis que les autres montrent soit une dominance dans l'hémisphère droit, soit une organisation plus bilatérale.

Par ailleurs, même lorsqu'il n'est pas dominant pour le langage, l'hémisphère droit contribue à certains aspects du traitement linguistique, tels que la compréhension de mots simples et de phrases courtes, le traitement du langage métaphorique ou encore la prosodie (i.e. l'intonation, l'accentuation sonore, le rythme). En cas de lésion de l'hémisphère gauche, son implication peut encore devenir plus importante au niveau du langage [Boatman et al., 1999].

Les différences fonctionnelles associées à la latéralité s'accompagnent également de particularités anatomiques (figure 1.11). Chez les droitiers, le planum temporale gauche, région recouvrant l'aire de Wernicke, est généralement plus grand que le droit, alors qu'aucune asymétrie globale n'est observée chez les gauchers. De même, le sillon central est en moyenne plus profond à gauche chez les droitiers, tandis que l'asymétrie inverse est rapportée chez les gauchers.

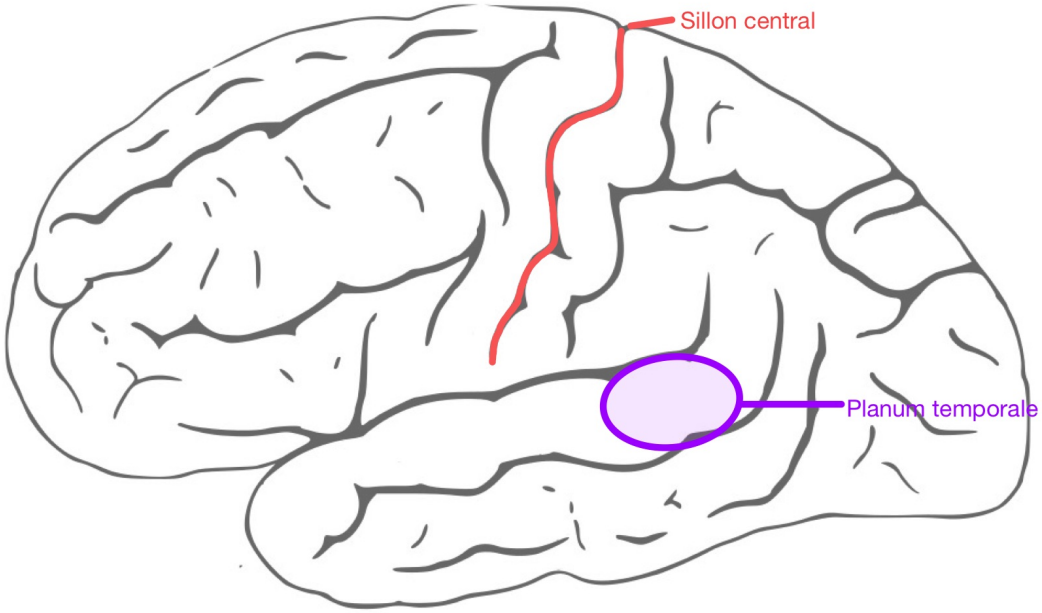


FIGURE 1.11 – Asymétries anatomiques de certaines régions cérébrales associées à la latéralité manuelle.

En outre, la perception musicale diffère entre droitiers, ambidextres et gauchers lorsqu'ils sont exposés à certaines illusions auditives décrites par [Deutsch et al., 2007]. Lorsqu'une note grave et son octave (obtenue par le doublement de la fréquence fondamentale de la note grave initiale) sont présentées simultanément, une dans chaque oreille, en alternant leur latéralisation (e.g. un sol grave à l'oreille gauche et un sol aigu à l'oreille droite, puis l'inverse), la perception diffère statistiquement entre droitiers, ambidextres et gauchers selon deux aspects principaux [Deutsch, 1983]. Premièrement, les droitiers tendent davantage que les ambidextres et les gauchers à entendre un son pur qui alterne entre les deux oreilles, passant de grave à aigu, tout en maintenant la note grave du même côté et la note aiguë de l'autre. Deuxièmement, parmi les droitiers rapportant cette perception d'un son pur, une forte proportion localise la note aiguë dans l'oreille droite et la note grave dans l'oreille gauche, ce qui n'est pas systématiquement observé chez les ambidextres et les gauchers.

Ces illusions étant produites à partir de sons stationnaires, la question se pose de savoir si un phénomène comparable pourrait émerger avec des sons dont la fréquence varie de manière continue. Plus généralement, il semble exister une tendance à percevoir les sons aigus du côté dominant de l'espace (i.e. le côté de l'espace qui est contralatéral à l'hémisphère dominant) et les sons graves du côté non dominant.

1.7.3 Influence du sexe biologique

Enfin, il est suggéré que des différences dans la latéralisation du langage liées au sexe existent. Les hommes semblent présenter une dominance plus marquée de l'hémisphère gauche pour les fonctions liée à la parole, tandis que les femmes montrent une organisation plus bilatérale [McLeun, 1996]. Des différences anatomiques ont également été rapportées, notamment au niveau du planum temporale [Hodgetts et Hausmann, 2022]. Certaines études [Kulynych et al., 1994] indiquent que cette région pourrait être, en moyenne de plus grande taille chez les hommes. De plus, une asymétrie interhémisphérique est observée chez les hommes, le planum temporale droit serait plus volumineux que le gauche, alors qu'elle n'est pas observée chez les femmes. Finalement, des variations pourraient également exister dans la proportion de substance grise et de substance blanche (qui contient les fibres nerveuses assurant les connexions cérébrales), indépendamment de la taille cérébrale [Hodgetts et Hausmann, 2022].

1.8 Influence des substances psychoactives sur le traitement cérébral

La consommation de certaines substances, telles que la caféine, la théine, la nicotine, l'alcool, diverses drogues ou certains traitements médicamenteux, peut moduler l'activité cérébrale, notamment en altérant la puissance des oscillations dans les différentes bandes de fréquences : delta ($[0.5, 4]$ Hz), thêta ($[4, 8]$ Hz), alpha ($[8, 12]$ Hz), bêta ($[12, 35]$ Hz), gamma (> 35 Hz).

1.8.1 Effets de la caféine et de la théine

La caféine, présente notamment dans le café et le thé (où elle est souvent désignée sous le terme de théine), correspond à une même molécule [La Marque et La Micronutrition, nd]. Toutefois, sa concentration est généralement plus élevée dans le café que dans le thé, y compris le thé vert.

La caféine a des effets importants sur la puissance des signaux EEG. En effet, la caféine entraîne une diminution de la puissance des signaux EEG sur l'ensemble des bandes de fréquences, modulée par l'état des yeux (ouverts ou fermés), l'administration concomitante de la nicotine et la région cérébrale considérée [Gilbert et al., 2000]. Lorsque les sujets gardent les yeux ouverts, cette diminution est particulièrement marquée dans les bandes alpha lentes ($[8, 10]$ Hz) et rapides ($[10, 12]$ Hz), ainsi que dans la bande bêta lente ($[12, 20]$ Hz) [Siepmann et Kirch, 2002]. Elle est observée au niveau des régions frontales, centrales, pariétales et occipitales des deux hémisphères. Une tendance similaire est observée lorsque les yeux sont fermés, bien que moins prononcée. Dans cette condition, la puissance dans la bande alpha n'est pas significativement impactée, tandis que la puissance dans la bande bêta lente diminue.

La diminution de la puissance dans les bandes alpha est généralement interprétée comme le reflet d'une augmentation de l'activation corticale, de l'excitabilité et de l'état d'éveil induits par la caféine.

1.8.2 Effets de la nicotine

La nicotine peut entraîner des modifications de la puissance dans certaines bandes de fréquences, dont l'ampleur dépend en grande partie de la condition visuelle (yeux ouverts ou fermés) [Gilbert et al., 2000]. Une diminution significative de la puissance dans la bande delta est toutefois observée indépendamment de cette condition. En revanche, les effets sur les autres bandes de fréquences varient selon l'état des yeux : la nicotine réduit la puissance dans la bande alpha 1 ($[8, 8.9]$ Hz) [Wu et Liu, 1995] lorsque les yeux sont ouverts, tandis qu'elle augmente la puissance dans la bande alpha 2 ($[9, 10.9]$ Hz) lorsque les yeux sont fermés. Par ailleurs, des interactions entre la nicotine et l'état des yeux sont également observées dans la bande bêta. Une augmentation de la puissance est observée dans les bandes bêta 1 ($[12.5, 16]$ Hz) et bêta 2 ($[16.5, 20]$ Hz) avec les yeux fermés, tandis qu'une diminution apparaît avec les yeux ouverts.

1.8.3 Effets de l'alcool

En ce qui concerne la consommation d'alcool, de faibles doses peuvent induire des modifications relativement prévisibles de la fréquence dominante et de la puissance des bandes thêta et alpha lentes [Ehlers et al., 1989]. En revanche, les réponses observées dans les gammes de fréquences plus élevées varient en fonction de facteurs tels que les antécédents de consommation d'alcool et l'activité EEG au repos.

1.8.4 Effets des drogues et des médicaments

Concernant la consommation de drogues ou de médicaments, les effets sur les signaux EEG sont variés et dépendent notamment de la nature du composé. De manière générale, ces substances peuvent induire une large gamme de modifications, incluant l'absence d'effet notable, une augmentation de l'activité bêta, une diminution de l'amplitude et/ou de la fréquence dominante du rythme alpha, ainsi que l'apparition d'une activité thêta et/ou delta superposée. Elles peuvent également moduler l'activité épileptique, soit en la réduisant, soit en abaissant le seuil épileptogène, ce qui peut entraîner une augmentation des décharges de type pointes-ondes [Banoczi, 2005].

Les psychostimulants, tels que les amphétamines, sont notamment associés à une diminution de la puissance totale des signaux EEG, ainsi qu'à une atténuation du rythme alpha [Siepmann et Kirch, 2002]. À l'inverse, les barbituriques et les benzodiazépines induisent généralement un excès d'activité bêta, parfois accompagné d'une augmentation de l'activité thêta [Blume, 2006]. D'autres substances, telles que la cocaïne, le méthylphénidate ou les antidépresseurs tricycliques, peuvent également entraîner une augmentation de l'activité bêta, bien que celle-ci reste de faible amplitude. Certains antiépileptiques, comme la carbamazépine et le valproate, peuvent quant à eux provoquer une augmentation modérée et diffuse (i.e. étendue à plusieurs régions corticales) de l'activité thêta, y compris à des concentrations thérapeutiques.

1.9 Conclusion

Ce chapitre décrit l'organisation anatomique et fonctionnelle du cerveau, en mettant l'accent sur les structures impliquées dans le traitement des sons. Il met en évidence la complexité et la distribution des réseaux neuronaux recrutés en fonction de la nature du stimulus sonore, qu'il s'agisse de parole ou de musique. Ces deux types de stimulations recrutent des régions en partie communes, notamment le cortex auditif primaire, situé sur la face supérieure du lobe temporal.

Néanmoins, une spécialisation hémisphérique est généralement observée : l'hémisphère gauche est associé au traitement temporel rapide, caractéristique de la parole, tandis que l'hémisphère droit est davantage impliqué dans le traitement spectral, plus sollicité dans la musique. Cette organisation reflète une latéralisation fonctionnelle des processus auditifs.

La nature du stimulus acoustique influence également l'intensité et l'étendue des activations corticales. Les stimuli simples, tels que les sons purs ou les signaux sinusoïdaux modulés en amplitude, sont principalement traités par le cortex auditif primaire et les régions auditives adjacentes, mais induisent des activations plus faibles que celles générées par des vocalisations animales ou des sons d'instruments. Par ailleurs, [Hong et Santosa, 2016] montrent que les réponses cérébrales à des sons naturels (e.g. rivière, forêt, pluie) sont plus importantes que celles associées à des sons désagréables (e.g. pleurs de bébé, alarme, sirène). En revanche, ces derniers activent des régions cérébrales plus étendues que les sons naturels.

La discrimination des catégories de sons, ainsi que leur localisation, sollicitent des réseaux cérébraux dépassant le cortex auditif primaire, impliquant notamment des aires associatives distribuées dans les lobes temporaux, frontaux, pariétaux et occipitaux des deux hémisphères.

Enfin, les réponses cérébrales ne dépendent pas uniquement des propriétés acoustiques des stimuli. Des facteurs individuels, tels que l'expertise musicale, la latéralité ou le sexe, peuvent influencer à la fois la perception du son et les mécanismes neurobiologiques sous-jacents. De plus, la consommation de certaines substances (e.g. caféine, théine, nicotine, alcool, drogues ou médicaments) est susceptible de moduler l'activité cérébrale.

Ainsi, l'ensemble de ces éléments souligne la nécessité de considérer conjointement les propriétés acoustiques des stimuli et les caractéristiques individuelles dans l'analyse de l'activité cérébrale. Si les sons naturels tendent à induire des activations plus marquées que les stimuli simples ou désagréables, ces effets peuvent varier selon les individus.

2

Les interfaces cerveau-ordinateur (*Brain-Computer Interfaces* ou BCI)

Sommaire

2.1 Introduction 19
2.2 Classification des BCI 20
2.3 Les méthodes d'imagerie cérébrale 21
2.3.1 Méthodes invasives 21
2.3.2 Méthodes non invasives 22
2.4 Marqueurs neurophysiologiques utilisés dans les BCI actives et réactives 24
2.4.1 Marqueurs neurophysiologiques utilisés dans les BCI actives 25
2.4.2 Marqueurs neurophysiologiques utilisés dans les BCI réactives 26
2.5 Conclusion 34

2.1 Introduction

Les BCI analysent les signaux cérébraux d'un individu afin d'en extraire des informations pertinentes, de les traduire sous forme de commandes permettant d'interagir avec différents dispositifs externes et, enfin, de fournir un retour perceptif à l'utilisateur quant à la transformation de ses intentions en actions [Lotte, 2021].

Il existe plusieurs catégories de BCI, définies notamment par leur mode de fonctionnement, la nature des marqueurs neurophysiologiques exploités, la temporalité du contrôle, ainsi que les modalités d'enregistrement de l'activité cérébrale. Cette dernière peut être mesurée de manière **invasive**, à l'intérieur de la boîte crânienne, ou **non invasive**, à la surface du cuir chevelu. Le choix de la méthode d'enregistrement influence directement les propriétés des signaux acquis, en particulier leurs résolutions temporelle et spatiale. Parmi les marqueurs neurophysiologiques utilisés dans les BCI, les potentiels évoqués stationnaires font partie des approches couramment employées. Ils correspondent à des réponses cérébrales induites par des stimulations sensorielles périodiques. Dans le contexte des BCI, la modalité visuelle est la plus étudiée et la plus largement exploitée [Clerc et al., 2016]. Les travaux consacrés à la modalité auditive étant plus récents et moins nombreux, ce travail s'appuie sur un parallèle entre les approches visuelles et auditives, afin d'identifier les connaissances et les méthodes développées pour les BCI visuelles pouvant être transposées à la modalité auditive.

Ce chapitre présente les différentes catégories de BCI (section 2.2), ainsi que les méthodes d'imagerie cérébrale (section 2.3). Il s'intéresse ensuite aux marqueurs neurophysiologiques exploités dans les BCI, et plus particulièrement, aux potentiels évoqués stationnaires. Une attention particulière est accordée aux BCI dites **réactives** reposant sur ces marqueurs, qu'ils soient d'origine visuelle (section 2.4.2.2) ou auditive (section 2.4.2.3).

2.2 Classification des BCI

Les BCI peuvent être caractérisées selon différentes propriétés comme leur mode de fonctionnement (actif, réactif ou passif), la nature des marqueurs neurophysiologiques exploités (marqueurs spectraux, temporels, spatiaux [Cabestaing et Derambure, 2016; Roy et Frey, 2016]), la méthode d'acquisition des signaux cérébraux (invasive ou non invasive), ainsi que la temporalité du contrôle (synchrone ou asynchrone).

Une BCI est dite **active** lorsque l'utilisateur la contrôle en réalisant volontairement des tâches mentales. Un exemple d'application est l'imagerie motrice en rééducation post-accident vasculaire cérébral [Herrera, 2024] : chez un patient présentant une paralysie, le mouvement qu'il ne peut plus exécuter est simulé mentalement. Les signaux cérébraux associés à cette intention motrice sont alors enregistrés et décodés en temps réel afin de provoquer le mouvement du membre paralysé.

Une BCI **réactive**, quant à elle, exploite la réponse cérébrale de l'utilisateur à des stimulations particulières [Wolpaw, 2013] (figure 2.1). Ce type de BCI peut notamment être utilisé pour contrôler des appareils domestiques (e.g. un téléviseur, une lampe ou une prise connectée) [Si-Mohammed et al., 2022]. Des icônes, correspondant à différentes commandes (e.g. changement de chaîne, réglage du son, allumage d'un appareil, etc.) sont affichées sur un écran, tel qu'un téléviseur ou un casque de réalité étendue. Lorsque l'utilisateur porte son attention sur l'une d'elles, cela se reflète dans son activité cérébrale, ce qui permet de déclencher l'action associée.

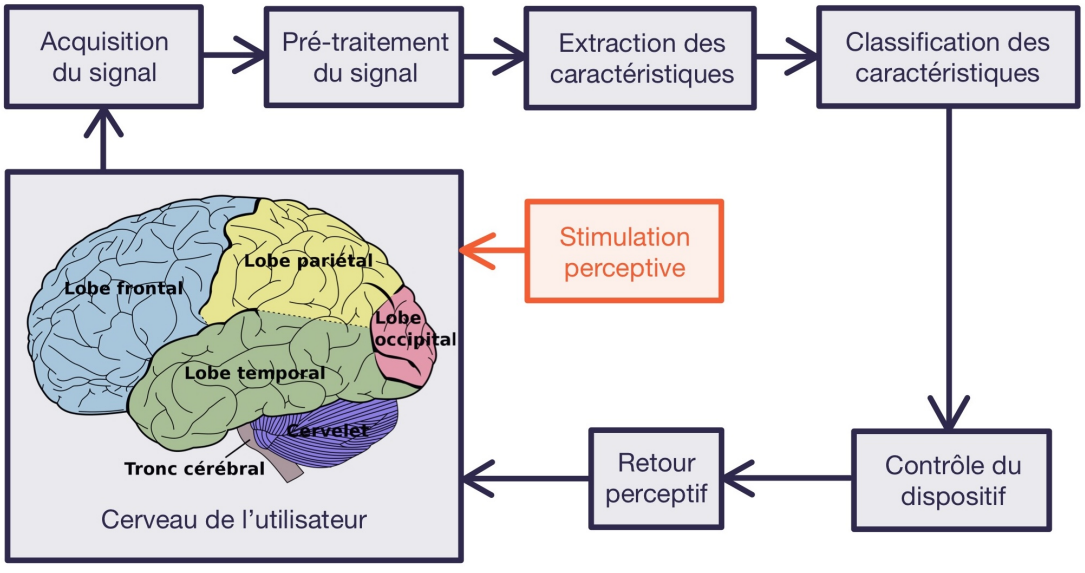


FIGURE 2.1 – Schéma de fonctionnement d'une BCI réactive.

Enfin, le but d'une BCI **passive** est d'observer l'état mental de l'utilisateur en temps réel, sans que ce dernier soit sollicité pour contrôler volontairement un dispositif grâce à des commandes

mentales. Par exemple, dans le domaine de l'aviation, les pilotes peuvent ne pas percevoir certains stimuli auditifs, comme des alarmes, lorsqu'ils sont engagés dans des tâches cognitives exigeantes [Dehais et al., 2019]. Une BCI passive peut alors permettre de surveiller en continu leur état attentionnel afin d'adapter le système en conséquence.

Une BCI composée d'au moins deux types de BCI (e.g. active, réactive ou passive) [Pfurtscheller et al., 2010] est appelée BCI **hybride**. Cette définition s'étend également aux systèmes associant une BCI à un autre dispositif, par exemple un système d'électromyographie (enregistrement de l'activité électrique des muscles) [Amiri et al., 2013]. Selon la manière dont ces systèmes sont combinés, on distingue différents types de BCI hybrides. Dans les BCI hybrides séquentielles, la sortie d'un système est utilisée comme entrée de l'autre système. À l'inverse, dans les BCI hybrides simultanées, les deux systèmes sont traités en parallèle. Une BCI hybride peut ainsi combiner une interface active et une interface réactive, afin d'exploiter simultanément différentes sources d'activité cérébrale et d'améliorer les performances du système. Dans ce cas, l'utilisateur peut réaliser une tâche mentale (e.g. imaginer un mouvement de la main) tout en regardant un stimulus visuel [Amiri et al., 2013].

Les phases d'interaction peuvent être imposées à l'utilisateur par le système, c'est le cas d'une BCI **synchrone**, dans laquelle l'utilisateur ne peut réagir qu'à des moments précis. Au contraire, si l'utilisateur peut la contrôler à tout moment, l'interface est dite **asynchrone**.

2.3 Les méthodes d'imagerie cérébrale

Lorsque l'enregistrement de l'activité cérébrale est réalisé à l'intérieur même de la boîte crânienne, la méthode est dite **invasive**. Les électrodes sont alors placées au plus près des sources électrophysiologiques, et les signaux sont mesurés avant qu'ils ne traversent le crâne, les muscles et la peau. Les mesures reflètent ainsi plus fidèlement l'activité cérébrale que lorsqu'elles sont réalisées de manière **non invasive**, c'est-à-dire à la surface du cuir chevelu.

2.3.1 Méthodes invasives

Dans le domaine des BCI, deux principales technologies d'enregistrement invasif de l'activité cérébrale dominent [Millán et Carmena, 2010] : l'ElectroCorticoGraphie (ECoG) et les ensembles de microélectrodes (*MicroElectrode Arrays* ou MEA). Le principal inconvénient des méthodes invasives réside dans leur caractère chirurgical, nécessitant l'ouverture du crâne et, dans certains cas, de la dure-mère (i.e. la membrane qui tapisse toute la paroi interne du crâne) [Millán et Carmena, 2010]. Dans ce contexte, ce travail porte principalement sur les méthodes non invasives, présentées plus en détail dans la section suivante (section 2.3.2).

2.3.1.1 ElectroCorticoGraphie (ECoG)

Premièrement, l'ECoG mesure les potentiels électriques à la surface du cortex, reflétant l'activité moyenne d'une large population de neurones. Elle enregistre des signaux de même nature que l'EEG, mais avec de meilleures performances en termes de résolution spatiale (de l'échelle du millimètre contre le centimètre), de bande passante (jusqu'à environ 500 Hz contre 50 Hz) et d'amplitude (de l'ordre de 50 à 100 μV contre 10 à 20 μV) [Millán et Carmena, 2010]. La différence de bande passante s'explique notamment par le fait que les tissus biologiques agissent comme un filtre passe-bas, atténuant les composantes hautes fréquences des signaux EEG. Par ailleurs, les signaux EEG sont plus sensibles aux artefacts liés, entre autres, aux contractions musculaires, aux mouvements oculaires ou à des sources externes [Millán et Carmena, 2010].

2.3.1.2 Réseaux de microélectrodes (*MicroElectrode Arrays* ou MEA)

Deuxièmement, les réseaux de microélectrodes permettent l'enregistrement d'activités unitaires (*Single-Unit Activity* ou SUA), c'est-à-dire l'activité électrique de neurones individuels [Waldert, 2016]. Implantés de manière permanente dans les régions corticales, ces réseaux constituent la technique la plus invasive, mais permettent d'enregistrer simultanément l'activité de multiples neurones individuels dans plusieurs régions du cerveau. De plus, en raison de l'atténuation des champs électriques avec la distance, le nombre de neurones devant être activés simultanément dans une zone restreinte pour générer un signal détectable est nettement plus élevé en EEG [Waldert, 2016]. Par conséquent, l'EEG ne permet pas de détecter l'activité de petits groupes neuronaux, ou seulement avec un RSB plus faible.

2.3.2 Méthodes non invasives

Nous nous intéressons ici aux méthodes d'imagerie cérébrale, qui permettent d'observer l'activité du cerveau sans recourir à une intervention chirurgicale. Il existe deux types d'imagerie médicale qui ont pour but de lire l'activité électrique du cerveau [Clerc et al., 2016]. Le premier type est l'imagerie structurelle, qui utilise la résonance magnétique et qui vise à caractériser l'anatomie du cerveau. Elle permet également de visualiser les faisceaux de matière blanche reliant les régions cérébrales entre elles, par une méthode dite de *tractographie*. Le deuxième type est l'imagerie fonctionnelle, qui a pour objectif de localiser les régions impliquées dans une tâche cognitive, ainsi que les interactions entre ces différentes régions. De fait, l'activation d'une région cérébrale entraîne une cascade d'évènements, notamment électromagnétiques et sanguins, qui produisent des signaux mesurables par les différentes méthodes d'imagerie fonctionnelle. Les principales techniques non invasives de mesure de l'activité cérébrale, brièvement présentées par la suite, ainsi que les ordres de grandeur de leurs résolutions temporelle et spatiale, sont illustrées en figure 2.2.

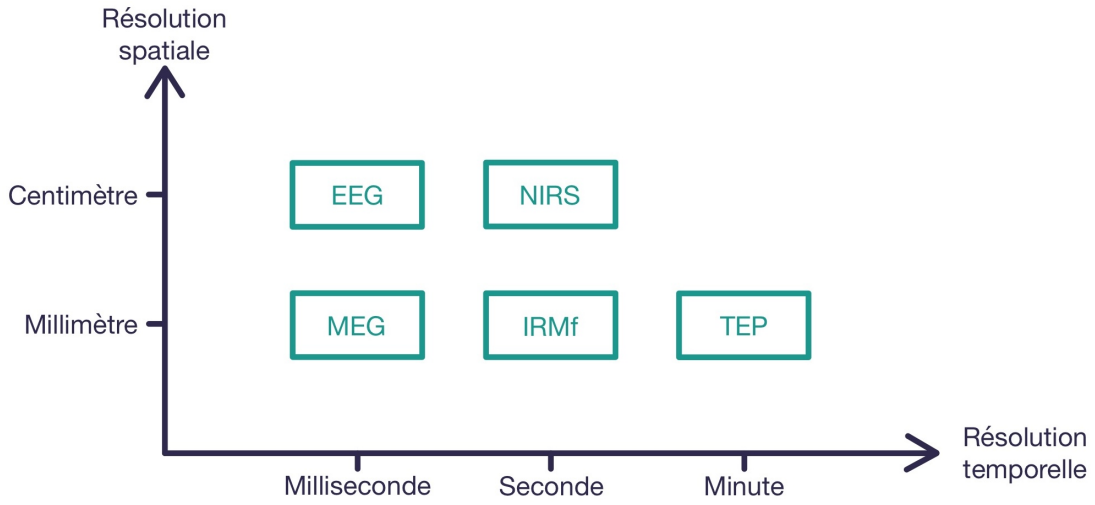


FIGURE 2.2 – Illustration des ordres de grandeur des résolutions temporelle et spatiale de différentes méthodes d'imagerie cérébrale non invasives.

2.3.2.1 ElectroEncéphaloGraphie (EEG)

Dans le domaine des BCI, la méthode non invasive d'enregistrement cérébralw la plus largement utilisée est l'EEG. Elle mesure la différence de potentiel électrique entre une électrode située sur un point du cuir chevelu et une électrode de **référence** [Wolpaw, 2013], sous la forme d'un signal temporel par électrode [Clerc et al., 2016]. Le crâne, étant moins conducteur que le cerveau ou le cuir chevelu, atténue et diffuse les courants, ce qui entraîne un lissage spatial des potentiels électriques. Ainsi, il faut que de nombreux neurones soient actifs en même temps pour que leurs activités élémentaires s'additionnent au point d'atteindre une intensité telle qu'elle sera mesurable à la surface du cuir chevelu. En d'autres termes, l'EEG n'est sensible à l'activité des neurones que s'ils sont assez nombreux à fonctionner en même temps. De fait, observer les variations du potentiel électrique sur le cuir chevelu procure une information grossière sur la provenance de l'activité électrique, dite **représentation topographique**.

L'EEG a une résolution temporelle d'approximativement une milliseconde et une résolution spatiale de quelques centimètres (environ cinq centimètres) [Burle et al., 2015]. Cette technique est principalement sensible à l'activité corticale superficielle et nécessite de bons contacts électriques avec le cuir chevelu, généralement assuré par l'utilisation de gels conducteurs (électrodes dites **humides**). Il est également possible d'utiliser des électrodes **sèches** ou **semi-sèches**. Les électrodes sèches ne nécessitent pas l'application de gel conducteur et permettent une mise en place rapide, mais présentent généralement une qualité de signal plus faible et une sensibilité accrue aux artefacts de mouvements [Li et al., 2020]. Les électrodes semi-sèches constituent un compromis entre électrodes sèches et humides, en utilisant une faible quantité d'électrolyte afin de réduire l'impédance de contact tout en conservant une bonne qualité de signal. Toutefois, leur utilisation peut entraîner une dégradation progressive du signal au cours du temps [Li et al., 2018].

L'EEG est moins efficace pour mesurer l'activité électrique provenant de régions profondes du cerveau et offre une localisation spatiale moins précise que la MagnétoEncéphaloGraphie (MEG) [Carlach et Javaudin, 2020].

2.3.2.2 MagnétoEncéphaloGraphie (MEG)

La MEG mesure les champs magnétiques induits par l'activité électrique des neurones, détectables à l'extérieur de la tête sans contact direct avec le cuir chevelu [Clerc et al., 2016].

Elle possède une bonne résolution temporelle, de quelques millisecondes, ainsi qu'une bonne résolution spatiale (de quelques millimètres à un centimètre). Malgré ses performances, la MEG demeure peu utilisée en pratique en raison de son coût élevé et de son encombrement. Les mesures nécessitent notamment un environnement contrôlé, tel qu'une chambre isolée magnétiquement, afin de limiter les interférences externes.

2.3.2.3 Imagerie par Résonance Magnétique (IRM)

L'IRM permet d'obtenir un signal en chaque point du cerveau en mesurant la réponse des noyaux d'hydrogène soumis à un champ magnétique [Clerc et al., 2016]. L'hydrogène est présent en grande quantité dans les tissus biologiques en raison de leur forte teneur en eau.

Cette méthode d'imagerie cérébrale offre une résolution spatiale élevée, de l'ordre du millimètre. Cependant, son coût d'acquisition est élevé.

2.3.2.4 Imagerie par Résonance Magnétique fonctionnelle (IRMf)

L'IRMf permet de mesurer l'activité cérébrale à travers les modifications hémodynamiques (i.e. les variations de la circulation sanguine) locales [Clerc et al., 2016].

Elle a une faible résolution temporelle, de l'ordre de la seconde, car elle dépend de la réponse hémodynamique (i.e. l'augmentation locale du débit sanguin associée à l'activation d'une région du cerveau). Elle est donc loin d'égaler la MEG. En revanche, sa résolution spatiale, de quelques millimètres, est élevée, ce qui en fait l'une des meilleures techniques d'imagerie fonctionnelle [Krainik et al., 2006]. Contrairement à l'EEG ou à la MEG, elle permet de localiser l'activité cérébrale non seulement en surface, mais également en profondeur [Gosseries et al., 2008]. Toutefois, son coût d'achat et d'entretien demeure élevé.

2.3.2.5 Tomographie par Emission de Positons (TEP)

La TEP permet d'obtenir une cartographie tridimensionnelle d'un organe grâce à l'injection préalable d'un traceur radioactif émetteur de positions [Jouvie, 2013]. Cette cartographie est obtenue en observant la distribution du traceur dans l'organisme au cours du temps [De Dreuille et al., 2002].

Cette technique présente une faible résolution temporelle (de l'ordre de la minute), mais une résolution spatiale de quelques millimètres [Defrise et Trebossen, 2002]. Toutefois, son utilisation est contrainte par l'emploi de traceurs radioactifs, qui nécessitent la présence d'un cyclotron (i.e. un accélérateur de particules) à proximité.

2.3.2.6 Spectroscopie proche infrarouge (*Near-InfraRed Spectroscopy* ou NIRS)

La réponse hémodynamique entraîne une modification de la quantité d'oxyhémoglobine (i.e. hémoglobine oxygénée) et de désoxyhémoglobine (i.e. hémoglobine désoxygénée) [Khoa et Nakagawa, 2008]. La spectroscopie proche infrarouge repose sur les propriétés d'absorption optique spécifiques de ces deux formes d'hémoglobine [Erhart et al., 2023]. Cette méthode permet ainsi de suivre leurs variations de concentration au cours du temps.

La résolution temporelle de cette méthode est limitée, car elle dépend de la réponse hémodynamique, qui intervient plusieurs secondes après l'activation neuronale. Sa résolution spatiale est également limitée (i.e. quelques centimètres) [Pinti et al., 2020; Cui et al., 2011; Erhart et al., 2023] et seule l'activité des couches superficielles du cortex peut être mesurée. Toutefois, cette technique est portable et plus abordable que d'autres méthodes.

2.4 Marqueurs neurophysiologiques utilisés dans les BCI actives et réactives

Le fonctionnement de BCI repose sur l'identification de marqueurs neurophysiologiques associés à une activité cérébrale particulière. Toutefois, la simple définition d'une tâche mentale ne suffit pas à disposer d'un marqueur exploitable de manière robuste par une BCI [Clerc et al., 2016]. Pour qu'un signal cérébral constitue un marqueur pertinent, l'utilisateur doit pouvoir exercer un contrôle volontaire sur une ou plusieurs de ses caractéristiques. Par ailleurs, un marqueur neurophysiologique doit être localisable dans l'espace, i.e. associé à une région cérébrale spécifique, et/ou dans le temps, i.e. pouvant être corrélé à l'instant où l'utilisateur modifie son état mental. Une bonne résolution spatiale permet alors d'identifier précisément les régions cérébrales

impliquées, tandis qu'une bonne résolution temporelle permet de relier précisément les variations de l'activité cérébrale aux intentions ou actions de l'utilisateur.

Les marqueurs utilisés dans les BCI reposent le plus souvent sur l'analyse d'activités neurophysiologiques enregistrées de manière non invasive à l'aide d'un EEG de surface. Ces mêmes activités peuvent également être exploitées dans des interfaces invasives, où les signaux sont enregistrés par ECoG. Dans ce cas, la proximité des électrodes avec les sources neuronales permet généralement d'obtenir une meilleure résolution spatiale, ainsi qu'un rapport signal sur bruit plus élevé.

Dans ce travail, l'étude se concentre sur les BCI actives et réactives, dont l'objectif est de permettre une interaction directe (e.g. commandes, messages) entre un utilisateur et un ordinateur à partir de l'activité cérébrale, contrairement aux BCI passives, principalement utilisées pour le suivi de l'état mental de l'utilisateur [Roy et Frey, 2016]. Dans une BCI active, la modification de l'activité cérébrale résulte d'un événement **endogène** : l'utilisateur décide volontairement de réaliser une tâche mentale spécifique [Padfield et al., 2022]. À l'inverse, dans une BCI réactive, cette modification est induite par un événement **exogène**, c'est-à-dire par une stimulation sensorielle extérieure [Clerc et al., 2016].

2.4.1 Marqueurs neurophysiologiques utilisés dans les BCI actives

Ainsi, les BCI actives reposent sur des modifications volontaires de l'activité cérébrale générées spontanément par l'utilisateur, sans recours à une stimulation sensorielle externe. Elles exploitent différents marqueurs neurophysiologiques associés à la préparation, à l'exécution ou à l'imagination d'une tâche mentale. Parmi les principaux marqueurs utilisés figurent les variations lentes du potentiel cortical moyen (*Slow Cortical Potentials* ou SCP), les potentiels de préparation, ainsi que les phénomènes de synchronisation et de désynchronisation liées à l'événement (*Event-Related Synchronization/Desynchronization* ou ERS/ERD) [Clerc et al., 2016].

2.4.1.1 Variations lentes du potentiel cortical moyen (*Slow Cortical Potentials* ou SCP)

Les SCP correspondent à des variations de l'activité électrique corticale s'étendant sur plusieurs centaines de millisecondes à plusieurs secondes [Gevensleben et al., 2014]. Ces variations sont caractérisées selon le signe de leur évolution, positive ou négative [Hinterberger et al., 2004], par rapport à un niveau de référence correspondant au potentiel électrique moyen mesuré durant une période de repos de l'utilisateur, appelée **ligne de base** (ou *baseline*) [Clerc et al., 2016]. La localisation spatiale des SCP peut varier d'un individu à l'autre, les amplitudes maximales n'étant pas nécessairement observées au niveau des mêmes régions corticales [Clerc et al., 2016].

Malgré cette variabilité inter-sujet, les SCP constituent des marqueurs exploitables dans une BCI. En effet, ils peuvent être mis en évidence relativement simplement en comparant le potentiel électrique moyen observé à un instant donné avec celui de la ligne de base. Les SCP apparaissent naturellement, puisqu'ils traduisent des modifications du niveau d'excitabilité de réseaux corticaux impliqués dans la préparation d'une tâche cognitive ou motrice [Gevensleben et al., 2014]. Par ailleurs, l'utilisateur peut apprendre à moduler volontairement leur amplitude, afin d'améliorer leur détectabilité dans les signaux EEG [Clerc et al., 2016]. Ainsi, les variations positives et négatives des SCP peuvent être interprétées comme des commandes binaires pour sélectionner des éléments affichés sur un écran d'ordinateur [Hinterberger et al., 2004; Pham et al., 2005]. Ce principe est notamment exploité dans le "dispositif de traduction de la pensée" (*Thought Translation Device* ou TTD).

2.4.1.2 Potentiel de préparation (*Readidness Potential* ou RP, également appelé *BereitschaftsPotential* ou BP)

Le BP correspond à une variation de l'activité cérébrale précédant l'exécution d'un mouvement volontaire [Clerc et al., 2016]. Il se caractérise généralement par deux phases successives : une première diminution lente du potentiel cortical, apparaissant entre 2 et 3 secondes avant le mouvement et se poursuivant jusqu'à environ 1 seconde avant son initiation, suivie d'une diminution plus rapide à l'approche du mouvement, qui débute approximativement 1 seconde avant celui-ci [Mahmoodi et al., 2017].

Le BP ne se manifeste pas uniquement lors de l'exécution réelle d'un mouvement [Pineda et al., 2000]. Il peut également être observé lorsqu'un individu regarde une autre personne réaliser un mouvement, anticipe la réalisation d'un mouvement, ou encore, imagine effectuer ce mouvement sans l'exécuter physiquement. Cette dernière propriété rend le BP particulièrement intéressant dans le cadre des BCI actives basées sur l'imagerie motrice [Clerc et al., 2016]. Il peut alors être exploité afin de prédire l'intention de mouvement d'un utilisateur, mais également certaines caractéristiques associées à ce mouvement, telles que sa direction ou le membre concerné.

2.4.1.3 Potentiels évoqués liés à un évènement (*Event-Related Potentials* ou ERP)

La survenue d'un évènement, qu'il soit exogène ou endogène, induit une variation de l'activité cérébrale appelée **potentiel évoqué lié à un évènement** [Clerc et al., 2016]. Toutefois, les réponses associées aux évènements endogènes sont généralement moins marquées que celles provoquées par des stimuli exogènes.

Dans les BCI actives, les marqueurs les plus fréquemment utilisés sont les ERS/ERD [Clerc et al., 2016; Jeon et al., 2011]. Une ERS correspond à une augmentation transitoire de l'amplitude des oscillations cérébrales dans une bande de fréquences donnée, tandis qu'une ERD correspond à une diminution de cette amplitude [Tang et al., 2016]. La localisation spatiale, ainsi que la latence (i.e. le délai séparant l'évènement déclencheur de la réponse cérébrale associée [Galambos et al., 1981]) de ces phénomènes dépendent fortement du type d'évènement endogène considéré, notamment de la complexité de la tâche mentale qui lui est associée [Clerc et al., 2016].

Les phénomènes de ERS/ERD induits par des tâches d'imagerie motrice constituent les marqueurs les plus étudiés dans la littérature sur les BCI actives [Clerc et al., 2016]. Dans ce contexte, les variations d'activité sont principalement observées dans les bandes de fréquences mu ($[10, 13]$ Hz) et bêta, qui peuvent être relativement facilement extraites à partir des signaux d'un EEG de surface [Tang et al., 2016]. De plus, les ERS/ERD associées à l'imagerie motrice présentent généralement une localisation spatiale marquée [Clerc et al., 2016]. Les interfaces utilisant ces marqueurs permettent à un utilisateur d'interagir avec un ordinateur uniquement en imaginant des mouvements, par exemple des mouvements de la main. Cependant, l'imagerie motrice ne constitue pas une tâche mentale triviale. Les stratégies cognitives employées pour imaginer un mouvement sans l'exécuter réellement peuvent varier fortement d'un individu à l'autre, ce qui entraîne une importante variabilité inter-sujet des réponses cérébrales enregistrées.

2.4.2 Marqueurs neurophysiologiques utilisés dans les BCI réactives

Dans les BCI réactives, l'interface présente à l'utilisateur des stimulations sensorielles destinées à provoquer une réponse cérébrale spécifique. La perception de ces stimuli induit des potentiels évoqués sensoriels (*Sensory Evoked Potentials* ou SEP), qui constituent un sous-ensemble des ERP, associé aux stimulations sensorielles [Lu et Hu, 2019]. Les SEP sont en phase avec les stimuli et leur latence peut être déterminée assez précisément. Ainsi, il est possible de les mettre en évidence

en répétant la stimulation plusieurs fois et en moyennant la réponse cérébrale. Toutefois, pour être exploitables dans une BCI, les SEP doivent pouvoir être détectés à partir d'un nombre limité de répétitions, afin de permettre une communication suffisamment rapide [Clerc et al., 2016]. Par ailleurs, lorsque la perception du stimulus entraîne également une réaction cognitive spécifique chez l'utilisateur, d'autres potentiels évoqués plus tardifs peuvent apparaître, comme le P300 [Clerc et al., 2016].

2.4.2.1 P300

La présentation à un individu d'un stimulus sensoriel inattendu, ou différent de celui auquel il s'attendait, peut induire l'apparition de potentiels évoqués spécifiques venant s'ajouter aux potentiels évoqués sensoriels et apparaissant à des latences plus élevées dans ses signaux cérébraux [Clerc et al., 2016].

Parmi eux, le **P3a**, également appelé **potentiel de nouveauté**, est observé lorsque la perception du stimulus capte l'attention de l'individu. Ce potentiel évoqué apparaît généralement entre 250 et 300 millisecondes après la présentation du stimulus.

Par ailleurs, lorsqu'un individu doit détecter un stimulus particulier connu à l'avance parmi une séquence d'autres stimuli, un autre potentiel évoqué, appelé **P3b** et parfois qualifié de **potentiel cible**, apparaît dans ses signaux cérébraux [Picton, 1992]. Sa latence est généralement comprise entre 250 et 500 millisecondes, mais peut être plus élevée lorsque la détection du stimulus cible nécessite un traitement cognitif plus important (e.g. compter le nombre d'occurrences du stimulus cible). Dans les BCI réactives, ce potentiel, alors appelé **P300**, constitue l'un des marqueurs neurophysiologiques les plus exploités pour, par exemple, identifier le stimulus sur lequel l'utilisateur se concentre et ainsi, déclencher l'action associée. Son exploitation est facilitée par le fait qu'il peut généralement être mesuré dès la première session d'utilisation, puisqu'il est associé à des mécanismes attentionnels et de reconnaissance du stimulus naturellement présents chez l'utilisateur. En outre, sa détection est souvent possible en essai unique, et les paradigmes P300 permettent la présentation rapide des stimuli, pouvant atteindre une dizaine de stimuli par seconde, ce qui autorise des débits de communication satisfaisants.

2.4.2.2 Potentiels évoqués stationnaires visuels (Steady-State Visually Evoked Potentials ou SSVEP)

Lorsqu'un individu est exposé à un stimulus visuel bref, celui-ci induit une réponse cérébrale appelée **potentiel évoqué visuel** (*Visual Evoked Potential* ou VEP). Cependant, considérant la réponse à une unique présentation, il est extrêmement difficile de mettre en évidence ce potentiel dans les signaux EEG. Les VEP obtenus par une stimulation brève sont généralement des réponses transitoires (i.e. instable et/ou non-stationnaires) du système visuel [Allison et al., 2008]. En raison de leur faible amplitude, ils sont difficiles à distinguer de l'activité cérébrale de fond [Gaume et al., 2014]. De ce fait, leur utilisation en tant que marqueurs dans une BCI est difficilement envisageable. En revanche, quand les stimulations sont répétées, il en est de même pour la réponse du système visuel [Vialatte et al., 2010; Higashi et al., 2011]. Un "régime de fonctionnement" permanent s'établit alors dans la chaîne de perception visuelle, aboutissant à l'apparition de potentiels évoqués stationnaires visuels. Comme ils reposent sur des stimulations externes, ces potentiels évoqués sont généralement plus robustes et plus facilement détectables que ceux générés intérieurement (e.g. en effectuant des tâches mentales volontaires) [Lotte, 2012].

Stimulations visuelles induisant des SSVEP Les SSVEP peuvent être obtenus via différents types de stimuli visuels. L'idée sous-jacente est toujours la même : un stimulus visuel clignotant ou se déplaçant à une fréquence constante (la **fréquence de stimulation**) provoque une réponse dans le cerveau à la même fréquence, ainsi qu'à ses harmoniques [Vialatte et al., 2010; Lécuyer et Si-Mohammed, 2019] (figure 2.3). Les auteurs de [Odom et al., 2004] considèrent que les fréquences de stimulation permettant d'évoquer des SSVEP sont comprises entre 3 et 50 Hz, mais des études ultérieures ont montré que les SSVEP, ainsi que leurs harmoniques, peuvent être observés jusqu'à 80 Hz [Allison et al., 2008]. Néanmoins, les amplitudes des SSVEP sont généralement plus grandes pour des fréquences de stimulation proches de 10 Hz, ainsi que dans la bande $[16, 18]$ Hz. Pour cette raison, les stimulations dans les bandes de basses ($[1, 12]$ Hz) et moyennes ($[12, 30]$ Hz) fréquences ont été plus largement utilisées que celles dans les hautes fréquences (> 30 Hz) [Yijun et al., 2005]. Toutefois, des évaluations subjectives ont montré que les fréquences comprises entre 5 et 25 Hz sont plus gênantes que les fréquences plus hautes, entraînant de la fatigue visuelle. Par ailleurs, les stimulations visuelles peuvent provoquer des crises épileptiques spécifiquement dans cette gamme de fréquence. De plus, le rythme alpha ($[8, 13]$ Hz), caractéristique de l'activité cérébrale spontanée (i.e. en l'absence de stimulation ou de tâche spécifique [Fox et Raichle, 2007]), peut se superposer aux réponses évoquées et entraîner des faux positifs. Enfin, il est préférable d'éviter la fréquence de 25 Hz, car sa première harmonique, située à 50 Hz, correspond à la fréquence du réseau électrique en Europe, en Asie et en Afrique. La réponse associée à cette harmonique risque alors d'être supprimée lors de traitements visant à réduire le bruit du réseau dans les signaux enregistrés. Ainsi, tous ces désagréments peuvent être évités en utilisant les hautes-fréquences, offrant de plus hauts niveaux de confort et de sûreté.

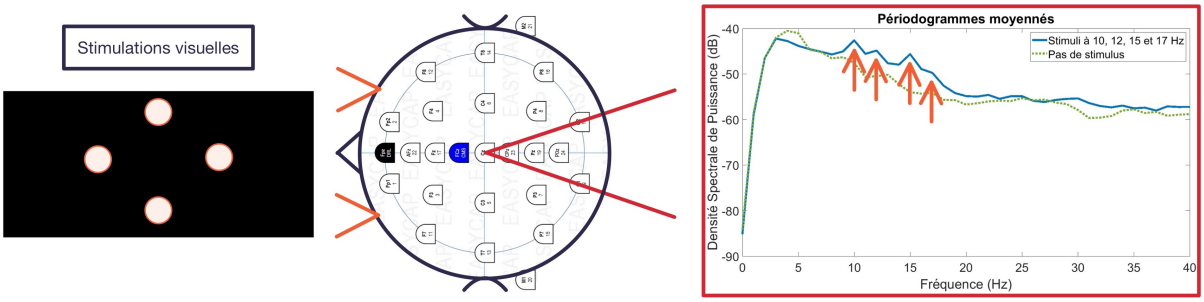


FIGURE 2.3 – Illustration d'un paradigme SSVEP. Quatre stimuli visuels (ronds) clignotent à des fréquences distinctes (10, 12, 15 et 17 Hz) [Si-Mohammed et al., 2022]. L'activité cérébrale est enregistrée à l'aide d'un système EEG à 24 canaux, équipé d'électrodes passives positionnées selon le système international 10–20. La figure présente la réponse cérébrale d'un sujet induite par ces stimulations visuelles (courbe bleue), ainsi que l'activité enregistrée en l'absence de stimulation (courbe pointillée verte). Les flèches oranges indiquent les fréquences des stimulations visuelles.

BCI basées sur les SSVEP Dans la majorité des BCI exploitant les SSVEP, des stimulations visuelles clignotantes à des fréquences spécifiques sont affichées sur un écran, le sujet en regardant un des stimuli peut déclencher l'action correspondante. Ce type de BCI est qualifié de **dépendant**, car le sujet doit au moins disposer d'un mouvement résiduel lui permettant d'orienter son regard [Lotte, 2021]. Les SSVEP sont des marqueurs intéressants pour la mise en œuvre d'une BCI car ils sont relativement facilement mis en évidence par un traitement basique des signaux EEG, tels qu'un filtrage passe-bande étroit centré sur la fréquence de stimulation [Clerc et al., 2016]. De fait, la présence et l'absence de SSVEP peut être estimée grâce aux caractéristiques

spectrales du signal seulement. De plus, leurs propriétés sont relativement peu dépendantes du sujet, à condition de sélectionner des électrodes positionnées au niveau de l'aire occipitale, où l'amplitude des réponses est maximale. Enfin, les SSVEP sont généralement peu sensibles aux artefacts musculaires, notamment aux clignements des paupières et à l'activité des muscles faciaux [Lotte, 2012].

2.4.2.3 Potentiels évoqués stationnaires auditifs (Steady-State Auditory Evoked Potentials ou SSAEP)

La stimulation répétitive d'une autre modalité sensorielle que la vision permet également de susciter des SEP stationnaires. Ainsi, une stimulation tactile périodique peut induire des potentiels évoqués somesthésiques stationnaires (*Steady-State Somatosensory Evoked Potentials* ou SSMEP). Par ailleurs, la modulation en amplitude périodique ou l'allumage et l'extinction répétés d'un son entraîne des potentiels évoqués stationnaires auditifs (figure 2.4) [Higashi et al., 2011; Kim et al., 2011a; Heo et al., 2017].

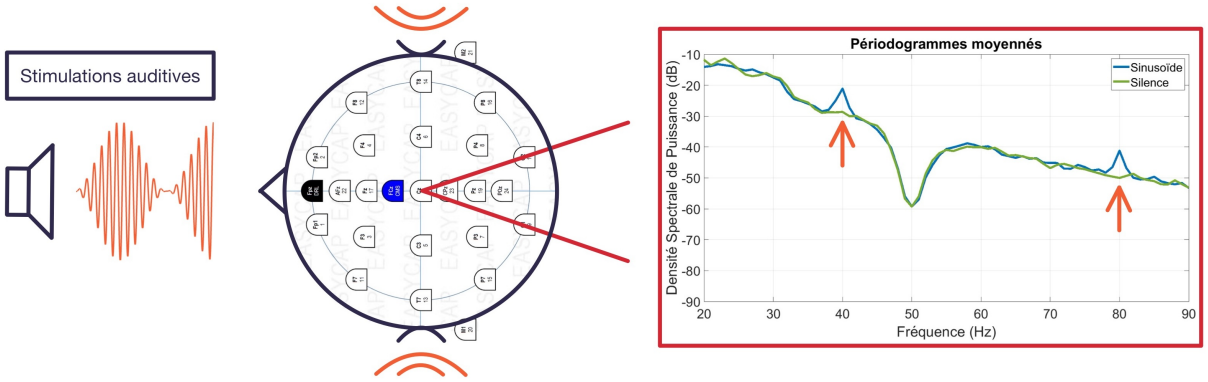


FIGURE 2.4 – Un son modulé en amplitude à 40 Hz est présenté à un sujet. Son activité cérébrale est enregistrée à l'aide d'un système EEG à 24 canaux, équipé d'électrodes passives positionnées selon le système international 10–20. La figure illustre la réponse cérébrale induite par cette stimulation auditive (courbe bleue), comparée à l'activité enregistrée en l'absence de stimulation (courbe verte). Les flèches oranges indiquent la fréquence de modulation du stimulus, ainsi que sa première harmonique.

Les stimulations auditives comme alternative aux stimulations visuelles Les stimuli visuels ont été les premiers à être étudiés et mènent généralement à de meilleures performances de classification que les autres modalités (e.g. stimulations sonores ou tactiles) [Punsawad et Wongsawat, 2011; Smith et al., 2014]. Néanmoins, ils ne sont pas adaptés à l'ensemble des utilisateurs. En effet, ils ne peuvent pas être utilisés lorsque la vision de l'utilisateur est altérée [Nambu et al., 2013], mais aussi lorsque l'utilisateur a des difficultés à contrôler ses mouvements extra-oculaires ou à fixer son regard sur un stimulus visuel spécifique, comme cela peut être le cas pour des sujets atteints de désordres neurologiques sévères [Kim et al., 2011a]. De plus, même pour les personnes ayant une fonction visuelle normale, regarder un stimulus un long moment peut causer de la fatigue ou une perte de concentration vis-à-vis d'autres tâches visuelles. Les stimulations visuelles sont également déconseillées chez les individus souffrant d'épilepsie ou de migraines [Vialatte et al., 2010]. Un autre inconvénient est le manque de confidentialité, car l'écran sur lequel sont affichés les stimuli est visible par d'autres personnes à moins d'utiliser un casque de

réalité étendue, ce qui impose des contraintes additionnelles. Certaines études montrent que les BCI auditives sont une alternative possible afin de pallier ces problèmes, mais les performances sont inférieures [Ferracuti et al., 2013]. Toutefois, l'utilisation de la modalité auditive pourrait permettre d'améliorer la réactivité des BCI, les réponses cérébrales aux stimulations sonores étant généralement plus rapides que celles induites par des stimulations visuelles [Pascual-Leone et al., 1992; Shelton et Kumar, 2010].

Stimulations sonores induisant des SSAEP Les stimulations auditives qui permettent l'apparition de SSAEP peuvent être créées de deux manières différentes. La première méthode consiste en l'allumage et l'extinction répété d'un signal [Kim et al., 2011a]. En ce cas, la durée de l'extinction du signal, qui correspond à une pause, est précisément calculée pour qu'il se répète à une fréquence spécifique, la **fréquence de modulation (f_m)**. Cette approche reste moins utilisée que la seconde méthode, qui repose sur une modulation d'amplitude sinusoïdale, bien que les performances soient similaires [Guého, 2022; Xie et al., 2021]. La modulation d'amplitude sinusoïdale est obtenue en combinant deux signaux [Lopez et al., 2009] (figure 2.5) : un signal porteur (c, d'amplitude A_c), correspondant au son perçu par les utilisateurs, et un signal modulant (m), qui est un signal sinusoïdal à la fréquence de modulation (f_m),

$$m(t) = A_m \cdot cos(2\pi f_m t), \tag{2.1}$$

où A_m définit l'amplitude du signal modulant et t, le temps. La modulation en amplitude (y) est réalisée en multipliant ces deux signaux et en ajoutant le signal porteur :

$$y(t) = [1 + m(t)] \cdot c(t). \tag{2.2}$$

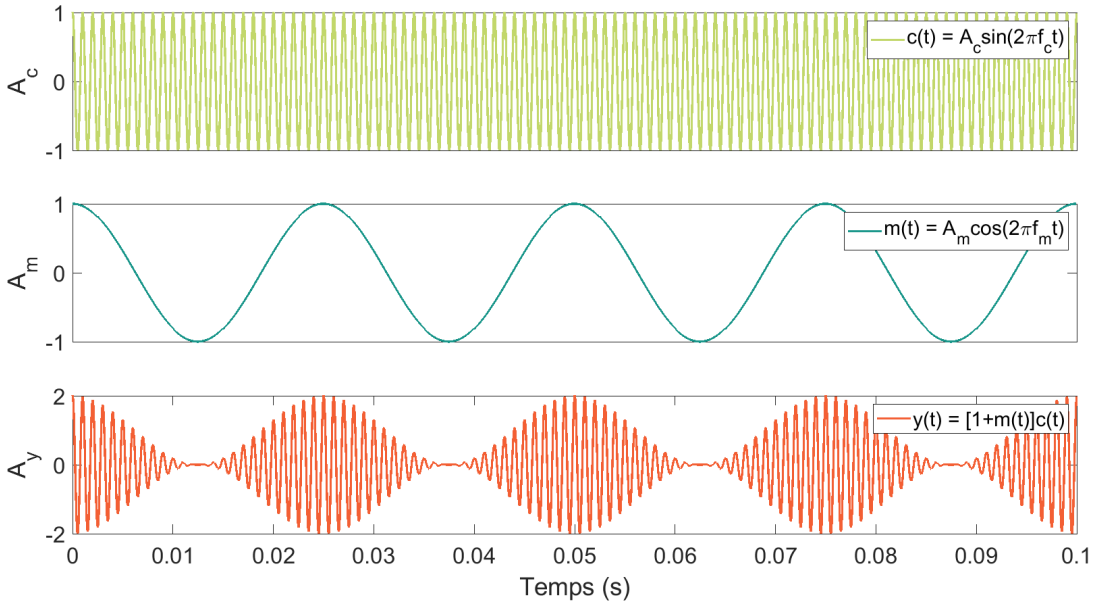


FIGURE 2.5 – Création d'une modulation en amplitude (en orange) à partir de deux signaux sinusoïdaux : le signal porteur (en vert) et le signal modulant (en bleu), avec f_c la fréquence du signal porteur et A_y, l'amplitude du signal modulé.

Le rapport entre l'amplitude du signal modulant (A_m) et celle du signal porteur (A_c) définit l'indice de modulation du signal y [Heo et al., 2017]. Cette modulation en amplitude est analogue

au clignotement des stimuli visuels utilisé dans les BCI visuelles. Ainsi, les SSAEP sont caractérisés par une activité cérébrale oscillant à la fréquence de modulation (et ses harmoniques) [Kim et al., 2011a; Heo et al., 2017]. Dans ce contexte, la fréquence de modulation constitue la fréquence de stimulation utilisée pour évoquer la réponse cérébrale.

BCI basées sur les SSAEP La plupart des BCI basées sur les SSAEP reposent sur la présentation simultanée de deux stimulations sonores distinctes dans l'oreille gauche et l'oreille droite du participant, de façon **dichotique** [Lopez et al., 2009; Kim et al., 2011a; Heo et al., 2017]. Les sujets peuvent alors déclencher une action en concentrant leur attention auditive sur l'un des stimuli, dont la fréquence de modulation est ensuite détectée dans leurs signaux cérébraux [Zhou et al., 2016].

Par ailleurs, les SSAEP peuvent être combinés à d'autres réponses cérébrales dans des systèmes de BCI hybrides, tels que ceux proposés par Kaongoen et Jo [2017], et Xie et al. [2021]. En particulier, Kaongoen et Jo [2017] associent des potentiels évoqués stationnaires auditifs à des potentiels évoqués de type P300 (cf. section 2.4.2.1). De leur côté, Xie et al. [2021] proposent une approche reposant sur la combinaison de stimulations visuelles et auditives.

Signaux porteurs Afin de faciliter la distinction et l'écoute des stimulations auditives, les signaux porteurs doivent être distincts, que ce soit par leur hauteur ou leur nature [McNamara et Ziarani, 2004; Heo et al., 2017].

Les signaux porteurs utilisés dans la majorité des études sont des sinusoïdes. En théorie, leur fréquence doit appartenir à la plage audible par l'humain (i.e. 20 à 20000 Hz). Selon Galambos et al. [1981], le potentiel auditif à 40 Hz (correspondant à la fréquence de modulation) peut être observé dans les signaux cérébraux pour des fréquences porteuses aussi bien basses (i.e. 250 et 500 Hz), que plus élevées (5000 Hz). Lorsque la fréquence porteuse augmente, l'amplitude et la latence de la réponse diminuent [Galambos et al., 1981]. Toutefois, afin d'assurer une perception optimale par la majorité des sujets, une plage plus restreinte est généralement privilégiée, typiquement entre 450 et 2500 Hz [Lopez et al., 2009; Xie et al., 2021; Kim et al., 2011a; Kaongoen et Jo, 2017; Higashi et al., 2011; Heo et al., 2017; Harada et al., 2022]. Par ailleurs, Xie et al. [2021] indiquent que la fréquence porteuse doit être un multiple entier de la fréquence de modulation afin d'optimiser l'efficacité du stimulus. Cette contrainte reste toutefois peu prise en compte dans la littérature, cette étude étant, à notre connaissance, la seule à l'avoir explicitement considérée et mise en œuvre.

Enfin, certaines études, telles que [Heo et al., 2017] et [Harada et al., 2022], utilisent des stimuli auditifs plus complexes, comme des sons musicaux ou naturels (e.g. le bruit d'une cascade d'eau, le chant des cigales, un ruisseau ou le chant d'un rouge-gorge). D'ailleurs, de meilleures performances ont ainsi été rapportées avec ces sons naturels par rapport à des sinusoïdes.

Fréquences de modulation Afin de permettre la discrimination de deux stimuli ou plus dans les signaux cérébraux, leurs fréquences de modulation doivent également être distinctes. D'après les auteurs de [Picton et al., 2003], les SSAEP peuvent être enregistrés en réponse à des stimuli dont la fréquence de modulation est comprise entre 1 et 200 Hz. Toutefois, la présentation de sons modulés à 40 Hz entraîne une réponse sinusoïdale dans le cerveau de plus grande amplitude [Stapells et al., 1984]. Les mécanismes physiologiques à l'origine de ce phénomène ne sont cependant pas encore clairement établis [Lopez et al., 2009]. Selon Picton et al. [1995], ces réponses pourraient résulter d'interactions entre le thalamus et le cortex auditif. Ainsi, l'amplitude de la réponse à des sons modulés entre 40 et 45 Hz est environ deux fois supérieure à celle observée

pour des modulations à 10 et 60 Hz [Stapells et al., 1984]. Par ailleurs, Harada et al. [2022] ont mené une étude visant à identifier les fréquences de modulation optimales pour l'utilisation d'une BCI, c'est-à-dire celles permettant d'induire les réponses de plus grande amplitude. Des sons naturels ont été modulés dans une plage de fréquences allant de 35 à 45 Hz, avec un pas de 2 Hz. Les fréquences associées aux réponses les plus marquées ont notamment été identifiées à 35, 39, 41, 43 et 45 Hz.

Les réponses autour de 40 Hz sont également particulièrement robustes et facilement enregistrables chez les sujets, sans effet significatif de l'âge ou du sexe [Galambos et al., 1981]. Contrairement aux SSVEP, les composantes harmoniques (paires et impaires) les plus hautes sont enregistrées, tandis qu'aucune sous-harmonique n'est observée [McNamara et Ziarani, 2004]. Cependant, lorsque le niveau sonore du stimulus est diminué de 90 à 20 dB HL (i.e. décibels relatifs au seuil d'audition normalisé, voir section 4.4.2.2), l'amplitude de la réponse à 40 Hz décroît linéairement, tandis que la latence (comprise entre 8 et 80 ms) augmente également de façon linéaire [Galambos et al., 1981]. De plus, le potentiel évoqué auditif à 40 Hz présente une amplitude plus élevée lorsqu'un même stimulus est présenté simultanément et en phase dans les deux oreilles (présentation **diotique**) que lorsqu'il est présenté dans une seule oreille (présentation **monaurale**). En revanche, lorsque ce même stimulus est présenté de manière déphasée entre les deux oreilles, l'amplitude observée devient plus faible que celle obtenue en présentation monaurale.

Ainsi, la plupart des BCI basées sur les SSAEP utilisent des fréquences de modulation proches de 40 Hz [Lopez et al., 2009; Kim et al., 2011a; Kaongoen et Jo, 2017; Heo et al., 2017]. Certaines études font toutefois exception, comme celle de Higashi et al. [2011], qui utilise une fréquence de modulation égale à 60 Hz, ou encore celle de Xie et al. [2021], dans laquelle des fréquences plus basses, notamment 9 et 13 Hz, sont employées.

Durées de stimulation Dans la littérature [Xie et al., 2021; Kim et al., 2011a; Higashi et al., 2011; Heo et al., 2017; Harada et al., 2022], les durées de stimulation sont généralement comprises entre 5 et 20 secondes, ces stimuli étant utilisés pour analyser leur influence sur les signaux cérébraux. En pratique, quelques dizaines à centaines de millisecondes suffisent pour qu'une réponse soit détectable dans les signaux cérébraux (e.g. 100 millisecondes [Kaongoen et Jo, 2017]). Cependant, l'augmentation de la durée de stimulation, jusqu'à 44 secondes, permet d'améliorer l'amplitude des SSAEP [Lopez et al., 2009].

Niveau sonore et dispositifs de restitution audio Le niveau sonore de reproduction des stimulations auditives constitue un paramètre essentiel dans les expérimentations basées sur les SSAEP [Galambos et al., 1981]. Toutefois, il est très rarement spécifié de manière précise dans la littérature, ce qui limite fortement la reproductibilité des protocoles expérimentaux. Dans plusieurs cas, les auteurs se contentent d'indications qualitatives, telles qu'un "niveau sonore confortable" [Lopez et al., 2009; Kaongoen et Jo, 2017], sans fournir de mesure objective. Même lorsque des valeurs sont rapportées, elles ne correspondent pas toujours à des niveaux de pression acoustique (en dB SPL), rendant leur interprétation difficile [Xie et al., 2021; Higashi et al., 2011].

Par ailleurs, il n'est jamais précisé si les stimuli sont égalisés en sonie, ce qui peut induire des différences d'intensité perçue entre les stimuli en fonction de leur contenu fréquentiel [Fletcher et Munson, 1933].

Concernant le matériel de reproduction, la plupart des études utilisent des écouteurs intra-auriculaires, tandis que d'autres recourent à des haut-parleurs [Kim et al., 2011a]. L'utilisation d'écouteurs permet un contrôle plus précis de la perception des stimuli, en particulier en ce qui

concerne l'intensité sonore et la spatialisation.

Dispositifs d'enregistrement de l'activité cérébrale Dans la plupart des études portant sur les SSAEP, un système EEG est utilisé pour l'enregistrement de l'activité cérébrale.

Le nombre de canaux employés reste généralement limité, souvent inférieur à une quinzaine [Kaongoen et Jo, 2017; Harada et al., 2022]. En pratique, de nombreuses études utilisent même un nombre très restreint d'électrodes, typiquement entre 2 et 6 [Lopez et al., 2009; Xie et al., 2021; Kim et al., 2011a; Higashi et al., 2011; Heo et al., 2017].

Les électrodes sont principalement positionnées au niveau des régions centrales et temporales, notamment en Fz, Cz, Pz, Oz, T3, T4, T7, T8, TP7 et TP8, conformément au système international 10–20 (figure 2.6).

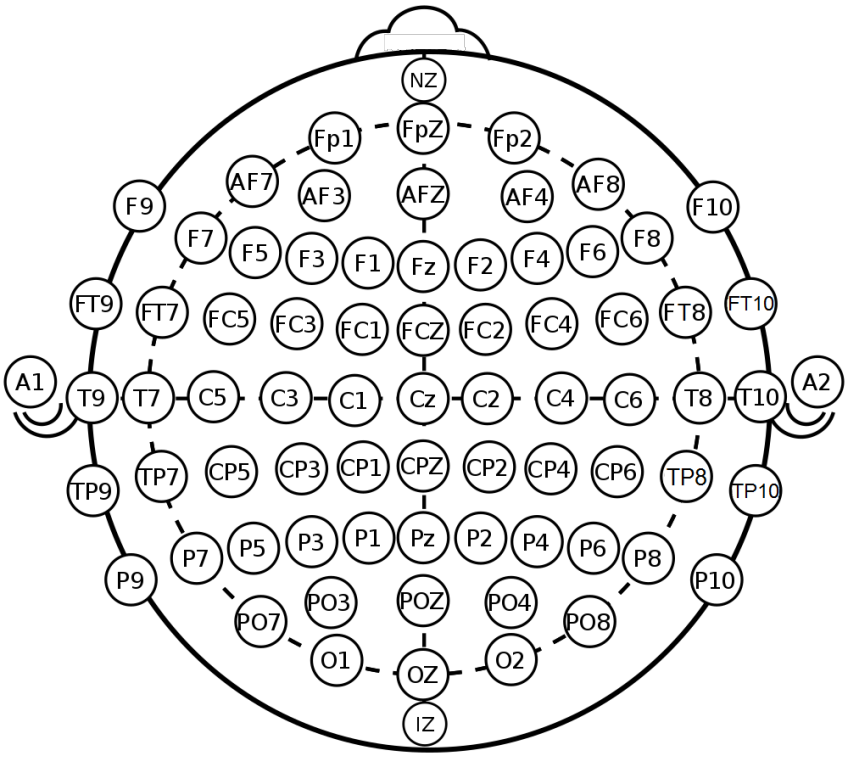


FIGURE 2.6 – Système international 10–20 utilisé pour le positionnement des électrodes EEG (figure tirée de [Wikimedia commons, 2017]).

Acceptabilité des stimuli Le confort associé aux stimuli sonores est rarement évalué dans la littérature. Toutefois, Harada et al. [2022] soulignent que les sons purs utilisés comme signaux porteurs peuvent difficilement être considérés comme des stimuli confortables et non stressants.

Dans cette optique, certaines études, telles que [Harada et al., 2022] et [Heo et al., 2017], proposent des alternatives aux sons purs. Afin d'évaluer si ces alternatives offrent un meilleur confort, Heo et al. [2017] ont mis en place un questionnaire visant à mesurer le confort perçu par les utilisateurs pour différents types signaux porteurs (i.e. sons purs, sons musicaux, sons naturels). Les participants devaient attribuer une note comprise entre 0 (très inconfortable) et 10 (très confortable) à chaque stimulus. Les résultats indiquent que les stimuli basés sur des sons purs sont systématiquement jugés comme les plus inconfortables. À l'inverse, les sons

musicaux et naturels sont perçus comme plus confortables. Bien que les sons naturels obtiennent en moyenne des scores plus élevés que les sons musicaux, cette différence n'est pas statistiquement significative.

2.4.2.4 Comparaison entre les SSVEP et les SSAEP

Les potentiels évoqués stationnaires, qu'ils soit visuel (SSVEP) ou auditifs (SSAEP), reposent sur le même principe neurophysiologique : la capacité du cerveau à se synchroniser avec un stimulus périodique. Dans les deux cas, la présentation d'un stimulus externe périodique, typiquement une lumière clignotante ou un son modulé en amplitude, ou encore un son présenté de manière intermittente par allumage et extinction répétés, induit une activité oscillatoire cérébrale dont la fréquence correspond à celle du stimulus.

La différence principale entre SSVEP et SSAEP réside dans la modalité sensorielle sollicitée (vision vs. audition), mais d'autres distinctions importantes existent. Les plages de fréquences de stimulation diffèrent notamment : elles s'étendent typiquement de 3 à 80 Hz pour la modalité visuelle [Odom et al., 2004; Allison et al., 2008], contre environ 1 à 200 Hz pour la modalité auditive [Picton et al., 2003]. En pratique, les réponses SSVEP présentent des amplitudes maximales pour des fréquences proches de 10 Hz, ainsi que dans la bande $[16, 18]$ Hz, tandis que les SSAEP montrent généralement une réponse optimale autour de 40 Hz [Galambos et al., 1981]. Ces deux types de réponses diffèrent également par leurs modes de reproduction. Les SSVEP sont généralement induits par des dispositifs visuels, tels que des diodes électroluminescentes (*Light-Emitting Diodes* ou LED) [Zhu et al., 2010], des écrans ou des dispositifs de réalité virtuelle/augmentée. À l'inverse, les SSAEP reposent sur des stimuli sonores délivrés via des casques audio, des écouteurs ou des haut-parleurs. De plus, les SSVEP sont principalement générés dans le cortex visuel, situé dans le lobe occipital [Clerc et al., 2016; Lécuyer et Si-Mohammed, 2019], tandis que les SSAEP prendraient leur origine dans le cortex auditif du lobe temporal. Enfin, les SSVEP ont fait l'objet d'un usage beaucoup plus répandu, notamment dans les BCI, contrairement aux SSAEP, qui demeurent encore relativement peu exploités dans ce contexte.

2.5 Conclusion

Parmi les différentes catégories de BCI, les BCI réactives reposent sur l'interprétation des réponses cérébrales évoquées par des stimuli externes [Wolpaw, 2013]. Une approche consiste à exploiter les potentiels évoqués stationnaires, correspondant aux réponses cérébrales induites par des stimulations sensorielles périodiques [Higashi et al., 2011]. Si les stimulations visuelles sont majoritairement utilisées dans ce contexte, elles présentent néanmoins certaines limitations : elles ne conviennent pas à l'ensemble des utilisateurs [Kim et al., 2007] et peuvent engendrer une fatigue visuelle. Dans cette perspective, les BCI auditives apparaissent comme un complément, voire une alternative pertinente.

Lorsque les utilisateurs sont exposés à des sons modulés en amplitude, analogues auditifs des stimuli visuels clignotants, des SSAEP apparaissent dans leurs signaux cérébraux [Kim et al., 2007; Heo et al., 2017]. Ces stimuli résultent de la combinaison de deux composantes : un signal porteur, correspondant au contenu sonore perçu, et un signal modulant, généralement une sinusoïde caractérisée par une fréquence spécifique, appelée **fréquence de modulation**. Bien que des sinusoïdes soient fréquemment utilisées comme signaux porteurs [Kim et al., 2007; Xie et al., 2021], leur caractère monotone et potentiellement désagréable lors d'écoutes prolongées constitue une limite importante à leur utilisation dans des contextes applicatifs réalistes.

Par conséquent, un enjeu central consiste à identifier des signaux porteurs plus confortables, tout en conservant la détectabilité de la fréquence de modulation dans l'activité cérébrale. Ainsi, les sons naturels (e.g. le bruit d'une cascade d'eau, le chant des cigales, un ruisseau ou le chant d'un rouge-gorge) apparaissent comme une alternative prometteuse. Plusieurs études [Heo et al., 2017; Harada et al., 2022] rapportent d'ailleurs de meilleures performances pour la détection de la fréquence de modulation à partir des signaux cérébraux lorsque des sons naturels sont utilisés comme signaux porteurs, comparativement aux signaux sinusoïdaux. Cependant, la dimension perceptive, notamment le confort d'écoute, reste encore largement sous-explorée dans la littérature, alors même qu'elle conditionne l'acceptabilité et l'usage à long terme des BCI. De plus, ces travaux ne précisent pas si les stimuli sont égalisés en sonie, ni, le cas échéant, quel niveau de reproduction est utilisé. Ce manque de précision est également observé dans plusieurs autres études [Kim et al., 2011a; Lopez et al., 2009; Kaongoen et Jo, 2017; Xie et al., 2021; Higashi et al., 2011] et limite la comparabilité des résultats.

En outre, la majorité des travaux reposent sur l'utilisation de systèmes EEG pour l'enregistrement de l'activité cérébrale, comportant un nombre généralement restreint de canaux. Celui-ci est le plus souvent inférieur à quinze [Kaongoen et Jo, 2017; Harada et al., 2022] et peut même se limiter à seulement deux à six électrodes dans de nombreuses études [Lopez et al., 2009; Xie et al., 2021; Kim et al., 2011a; Higashi et al., 2011; Heo et al., 2017]. Les électrodes sont principalement positionnées au niveau des régions centrales et temporales. Or, comme discuté dans le chapitre précédent (chapitre 1), le traitement des sons mobilise également d'autres lobes cérébraux. Cette configuration expérimentale pourrait ainsi conduire à une sous-exploitation de l'information disponible.

Enfin, l'analyse des signaux cérébraux apparaît comme un élément fondamental, en particulier pour l'identification de la fréquence de modulation, et fera l'objet du chapitre suivant (chapitre 3).

3

Analyse des signaux ElectroEncéphaloGraphiques (EEG)

Sommaire

3.1 Introduction 37
3.2 Signaux EEG dans un contexte BCI 38
3.2.1 Acquisition des signaux EEG 38
3.2.2 Organisation des données EEG 39
3.2.3 Propriétés des signaux EEG 40
3.2.4 Classification des signaux EEG 42
3.2.5 Stratégie d'analyse des signaux EEG 42
3.3 Prétraitements des signaux EEG 43
3.3.1 Filtrage passe-bande 44
3.3.2 Extraction temporelle 45
3.4 Réduction de dimension 45
3.4.1 Notations des *epochs* 45
3.4.2 Analyse en composantes principales (*Principal Component Analysis* ou PCA) 46
3.4.3 *Common Spatial Pattern* (CSP) 47
3.4.4 Analyse discriminante linéaire (*Linear Discriminant Analysis* ou LDA) 48
3.5 Détection de la fréquence de modulation 49
3.5.1 Rapport Signal sur Bruit (RSB) 49
3.5.2 Méthodes de classification 50
3.6 Conclusion 66

3.1 Introduction

Dans les BCI réactives basées sur des SSVEP ou des SSAEP, l'utilisateur est exposé à une ou plusieurs stimulations caractérisées par des fréquences distinctes (les fréquences de stimulation). Le principe consiste à identifier la stimulation sur laquelle l'utilisateur porte son attention en détectant, dans son activité cérébrale, la fréquence associée à la stimulation ciblée par ce dernier. La détectabilité de cette fréquence peut être estimée selon deux approches complémentaires. La première repose sur une analyse spectrale, via le calcul du RSB. Cette mesure consiste à comparer,

aux fréquences d'intérêt (i.e. aux fréquences de stimulation, ainsi que leurs harmoniques [Higashi et al., 2011]), la puissance de la réponse cérébrale en présence de la stimulation à celle observée en l'absence de cette dernière. La seconde approche s'appuie sur des méthodes de classification visant à discriminer les réponses EEG associées aux différentes conditions de stimulation. Les performances de classification fournissent alors une estimation de la détectabilité de la fréquence de stimulation, dépendante toutefois du modèle utilisé.

Toutefois, l'interprétation et la classification des signaux EEG demeurent complexes. En effet, ces signaux présentent une variabilité importante, tant intra- qu'inter-sujet [Clerc et al., 2016], et se caractérisent par un faible RSB. De plus, certains utilisateurs présentent des réponses aux stimulations très faibles, voire non détectables au niveau du cuir chevelu, ce qui complique davantage l'identification des fréquences d'intérêt [Vidaurre et Blankertz, 2010]. Afin de pallier ces limitations, de nombreuses méthodes de traitement du signal ont été proposées, notamment des techniques de filtrage temporel et spatial [Lotte et al., 2018], d'optimisation des fenêtres d'analyse [Kalunga et al., 2016], ainsi que des approches de réduction de dimension [Lotte et al., 2018].

Dans ce contexte, ce chapitre débute par une présentation des caractéristiques générales des signaux EEG (section 3.2), avant de détailler les chaînes de traitement mises en œuvre pour la détection des fréquences d'intérêt. Les étapes de prétraitement sont décrites (section 3.3), suivies des méthodes de réduction de dimension (section 3.4). Enfin, l'estimation du RSB, les approches de classification (section 3.5), ainsi que les métriques d'évaluation des performances (section 3.5.2.6), sont présentées.

3.2 Signaux EEG dans un contexte BCI

3.2.1 Acquisition des signaux EEG

En BCI, l'acquisition des signaux cérébraux suit généralement une organisation hiérarchique structurée en essais (ou *trials*), séquences (ou *runs*[2]) et sessions (figure 3.1). Un essai correspond à l'unité élémentaire d'acquisition, associée à l'exécution d'une tâche mentale ou à la présentation d'une stimulation. Un *run* regroupe un ensemble d'essais réalisés de manière consécutive, sans interruption prolongée [Chevallier et al., 2024]. Enfin, une session comprend plusieurs *runs* séparés par des périodes de pause durant lesquelles le dispositif d'enregistrement reste en place [Lotte, 2021].

2. Par souci de cohérence terminologique, et en l'absence d'équivalents français systématiquement établis dans la littérature, un certain nombre de termes anglais seront employés dans la suite de ce manuscrit.

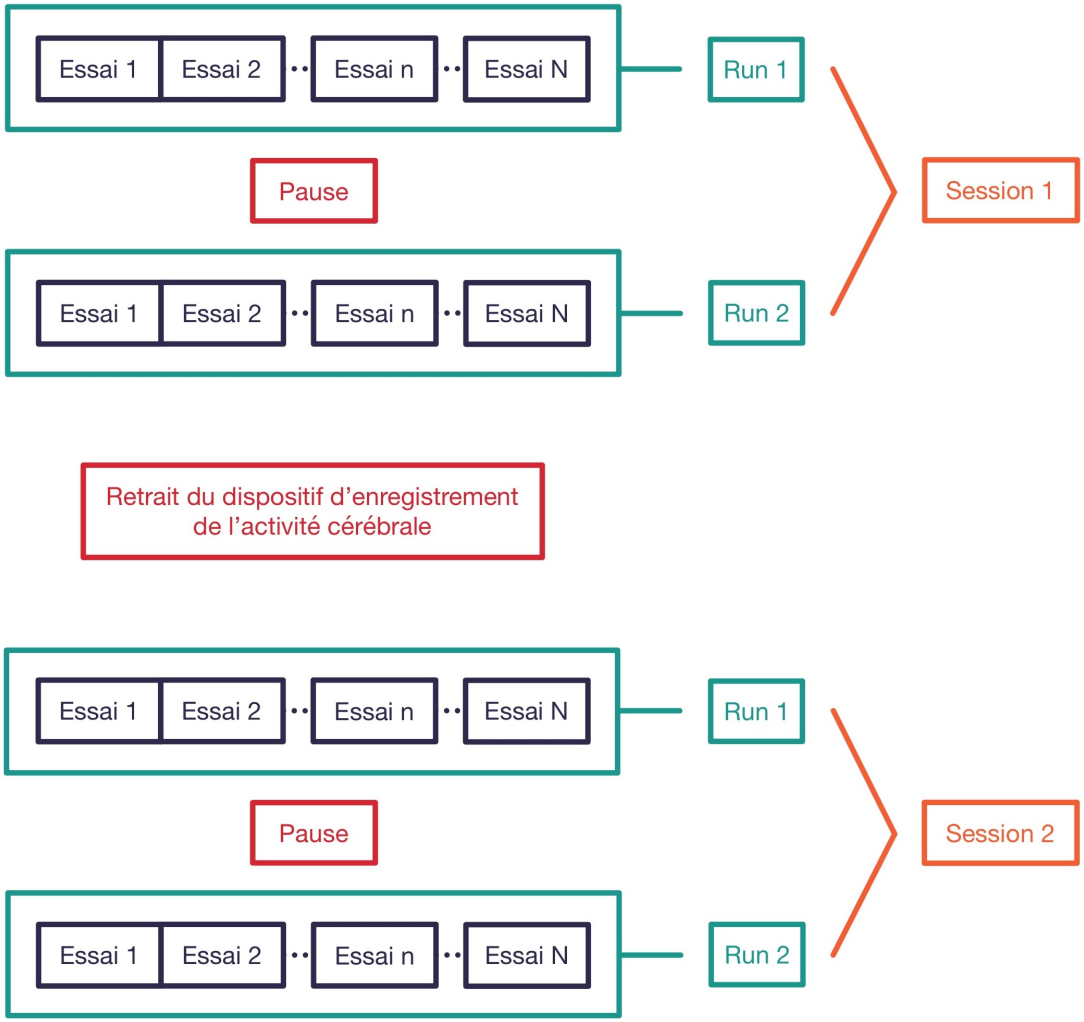


FIGURE 3.1 – Exemple d'organisation hiérarchique d'une acquisition de l'activité EEG dans un contexte BCI : 2 sessions, chacune constituée de 2 *runs*, chaque *run* regroupant N essais correspondant à des répétitions d'une ou de plusieurs tâches (ou stimulations).

3.2.2 Organisation des données EEG

Comme décrit précédemment (section 2.3.2.1), l'EEG mesure des différences de potentiel électrique entre des électrodes réparties sur le cuir chevelu et une électrode de référence [Wolpaw, 2013]. L'amplitude des signaux EEG étant très faible, généralement comprise entre 10 et 100 μV, leur acquisition nécessite l'utilisation d'un amplificateur [Clerc et al., 2016]. Celui-ci fournit, pour chaque canal de mesure, un signal temporel numérisé, obtenu après échantillonnage et quantification (figure 3.2).

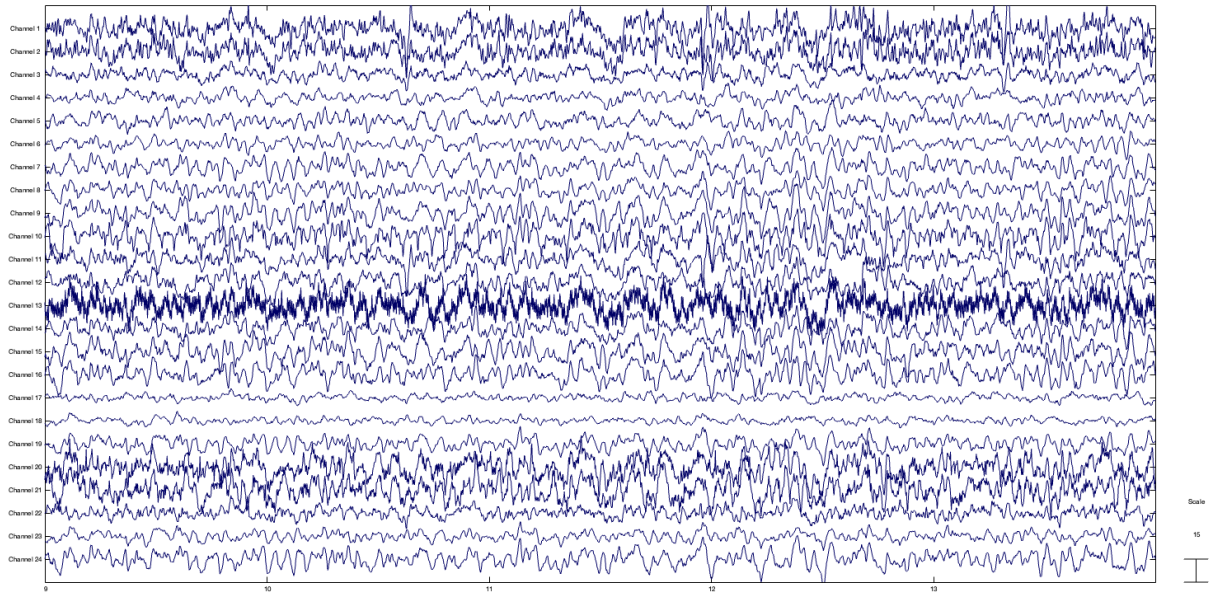


FIGURE 3.2 – Capture d'écran de la bibliothèque EEGLAB montrant les signaux EEG obtenus après échantillonnage et quantification sur 24 canaux (*channels*), pour une durée de 5 secondes.

Pour un enregistrement de N échantillons temporels avec C électrodes, les données peuvent être représentées par une matrice $\mathbf{X} \in \mathbb{R}^{C \times N}$, définie comme :

$$\mathbf{X} = [\mathbf{x}_1 \ \dots \ \mathbf{x}_i \ \dots \ \mathbf{x}_N], \tag{3.1}$$

où chaque vecteur colonne $\mathbf{x}_i \in \mathbb{R}^C$ correspond aux mesures simultanées sur l'ensemble des C électrodes à l'instant i,

$$\mathbf{x}_i = [\mathbf{x}_i^{(1)} \ \dots \ \mathbf{x}_i^{(j)} \ \dots \ \mathbf{x}_i^{(C)}]^\top. \tag{3.2}$$

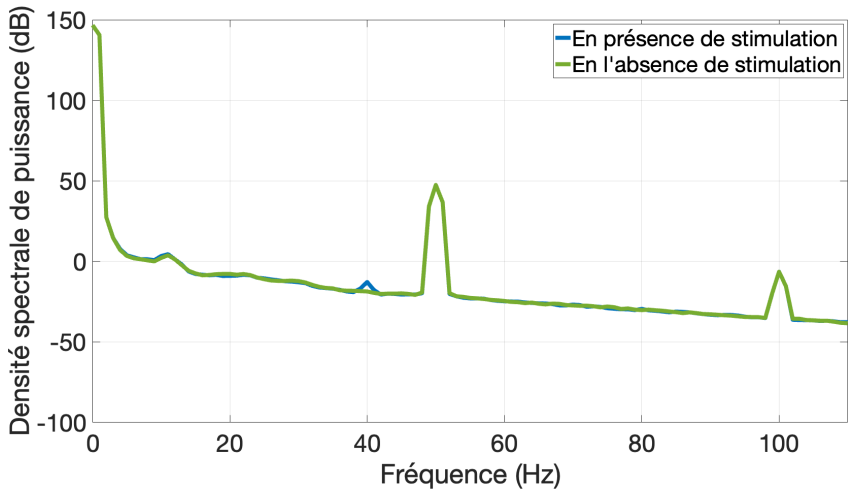
L'exposant $\top$ désigne l'opérateur de transposition appliqué à une matrice. Ainsi, la j-ème ligne de $\mathbf{X}$ contient le signal temporel de l'électrode j, et la i-ème colonne contient les mesures de toutes les électrodes à l'instant i.

3.2.3 Propriétés des signaux EEG

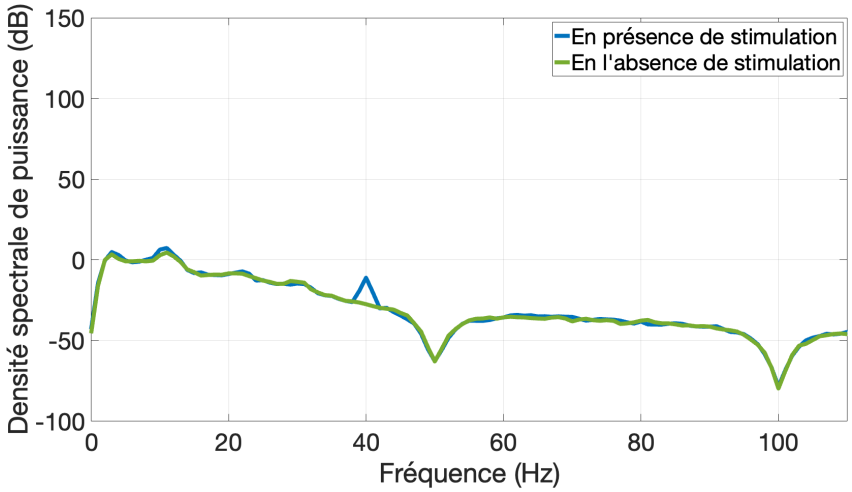
Ainsi, les signaux EEG sont généralement acquis de manière multicanale. Ils possèdent donc à la fois des caractéristiques temporelles et spatiales [Clerc et al., 2016]. L'information temporelle correspond à l'évolution du signal au cours du temps et peut être représentée dans le domaine fréquentiel. Les informations fréquentielles décrivent alors la répartition de la puissance du signal dans différentes bandes de fréquences. L'information spatiale renseigne, quant à elle, sur la distribution des potentiels entre les électrodes et, indirectement, sur la contribution des sources corticales au signal mesuré.

Cependant, les signaux EEG présentent des propriétés complexes : ils sont non linéaires et non gaussiens [Vaid et al., 2015] (i.e. la distribution statistique de leurs amplitudes ne suit par une loi normale, mais présente, au contraire, des asymétries et des valeurs extrêmes fréquentes). Ils se caractérisent également par une faible résolution spatiale, ainsi qu'un RSB généralement réduit. En effet, lors de l'acquisition, le signal d'intérêt est contaminé par de nombreuses sources

d'interférences, notamment les clignements oculaires, l'activité musculaire, les mouvements du sujet, ainsi que le bruit environnemental, en particulier le bruit de réseau électrique (figure 3.4a). Dans ces conditions, les signaux EEG bruts sont rarement directement interprétables, ils doivent être soumis à des traitements visant à atténuer le bruit et à mettre en évidence les composantes pertinentes [Clerc et al., 2016] (figure 3.3b). Dans ce cadre, l'extraction de caractéristiques constitue une étape centrale de l'analyse des signaux EEG. Elle repose généralement sur l'application de filtres temporels et spatiaux permettant d'isoler l'information d'intérêt. L'objectif est alors de représenter les signaux EEG par un ensemble restreint de variables décrivant de manière compacte l'information pertinente qu'ils contiennent, afin de faciliter leur interprétation, ainsi que leur exploitation par des méthodes de classification.



(a) Signaux EEG non filtrés moyennés sur l'ensemble des canaux.



(b) Première composante d'un filtre spatial (CSP, voir section 3.4.3) appliqué à des signaux EEG préalablement filtrés temporellement autour de 40 Hz, dans une bande de ±2 Hz, au moyen d'un filtre IIR (*Infinite Impulse Response* ou à réponse impulsionnelle infinie) implémenté sous forme de sections du second ordre (*Second-Order Sections* ou SOS), voir section 5.2.1.

FIGURE 3.3 – Comparaison de signaux EEG non filtrés et des mêmes signaux filtrés temporellement et spatialement, en présence et en absence de son modulé à 40 Hz. Les signaux sont segmentés en fenêtres temporelles correspondant aux périodes avec et sans stimulation, puis représentés dans le domaine fréquentiel.

3.2.4 Classification des signaux EEG

Les caractéristiques extraites des signaux EEG peuvent être utilisées en entrée d'algorithmes de classification, visant à identifier l'état mental de l'utilisateur ou à détecter la stimulation ciblée dans un paradigme SSVEP/SSAEP. L'objectif est alors d'associer chaque signal à une classe, c'est-à-dire de déterminer, à partir de l'activité cérébrale enregistrée, la condition à laquelle il correspond (e.g. "attention portée sur la stimulation" ou "absence d'attention portée sur la stimulation", dans un paradigme SSVEP/SSAEP binaire). Dans un problème multi-classe (i.e. plus de deux classes), les classes peuvent correspondre à l'attention portée sur différentes stimulations (e.g. stimulation 1, stimulation 2, etc.), ainsi qu'à une condition supplémentaire correspondant à l'absence d'attention portée sur les stimulations.

3.2.5 Stratégie d'analyse des signaux EEG

Le fonctionnement d'une BCI peut être envisagé selon trois modalités : le mode **en ligne** (*online*), qui repose sur un traitement continu des données, en temps réel ; le mode **hors ligne** (*offline*), dans lequel les données, préalablement enregistrées, sont analysées ; et le mode **pseudo-en ligne** (*pseudo-online*), où les données sont traitées a posteriori en reproduisant les conditions du temps réel [Carrara et Papadopoulo, 2024]. La principale différence entre ces modes réside dans la disponibilité des données : en mode hors ligne, l'ensemble des signaux est exploitable pour l'analyse, tandis qu'en modes en ligne et pseudo-en ligne, le traitement repose sur des fenêtres temporelles glissantes. Les deux derniers modes se distinguent par la présence d'une contrainte de temps réel en mode en ligne, absente en mode pseudo-en ligne.

Ces approches répondent à des objectifs complémentaires dans le développement des BCI. L'analyse hors ligne permet d'explorer et de comparer différentes conditions expérimentales et méthodes de traitement des signaux mesurés, sans contrainte de temps réel. Elle est ainsi privilégiée pour l'évaluation des performances, l'optimisation des paramètres et la validation méthodologique. À l'inverse, le mode en ligne traite les données au fur et à mesure de leur acquisition, ce qui est indispensable pour les applications interactives. En pratique, les analyses hors ligne offrent généralement de meilleures performances, tandis que les systèmes en ligne, soumis à des contraintes de latence, de robustesse et de variabilité inter- et intra-individuelle, sont plus représentatifs d'un usage réel. Le mode pseudo-en ligne constitue un compromis, permettant d'approcher le fonctionnement en temps réel tout en relâchant la contrainte temporelle, de manière à estimer les performances maximales atteignables pour une tâche donnée. De nombreuses études en BCI sont évaluées uniquement en mode hors ligne, ce qui peut conduire à des performances peu représentatives des conditions réelles d'utilisation [Carrara et Papadopoulo, 2024]. Cela introduit un biais important, certains algorithmes optimisés dans ce cadre perdant en efficacité lorsqu'ils sont déployés en situation réelle.

Dans un système en ligne, une **phase de calibration** (aussi appelée **session d'entraînement** [Shenoy et al., 2006]) est généralement réalisée avant l'utilisation du système [Si-Mohammed et al., 2022; Kalunga et al., 2016], afin de l'adapter à l'utilisateur. Durant cette phase, l'utilisateur effectue différentes tâches associées aux conditions expérimentales [Shenoy et al., 2006] (e.g. imagination d'un mouvement, écoute d'une stimulation, etc.). Les signaux cérébraux enregistrés servent alors à entraîner le classifieur, qui est ensuite utilisé en temps réel pour prédire, à partir de nouveaux signaux cérébraux, la condition expérimentale correspondante et ainsi, la commande associée [Si-Mohammed et al., 2022].

En analyse hors-ligne, les données sont généralement séparées en un **ensemble d'entraînement** et un **ensemble de test** [Chevallier et al., 2024]. Le modèle permettant de discriminer

les différentes classes est appris sur le premier ensemble, puis évalué sur le second, afin d'estimer sa capacité à déterminer la classe à laquelle appartiennent de nouvelles données [Nguyen et al., 2018].

3.3 Prétraitements des signaux EEG

Les étapes de prétraitement des signaux EEG comprennent généralement l'application d'un filtre passe-haut, destiné à éliminer les basses fréquences (entre 0 et 1 Hz) parfois observées [Delorme et Makeig, 2004; Clerc et al., 2016], notamment dues à des variations temporaires du contact électrode–cuir chevelu [Shahbakhti et al., 2021]. Les artefacts liés au courant alternatif (50 Hz en France) et son harmonique (100 Hz) peuvent être éliminés à l'aide d'un filtre passe-bas lorsque les fréquences d'intérêt sont inférieures à 50 Hz. Dans ce contexte, un filtre passe-bande entre 1 et 50 Hz peut également être utilisé afin de réaliser simultanément ces deux opérations. Toutefois, lorsque des informations au-delà de 50 Hz doivent être conservées, des filtres coupe-bande centrés sur 50 Hz et 100 Hz sont privilégiés, pas exemple des filtres FIR (*Finite Impulse Response* ou à réponse impulsionnelle finie) ou des filtres de type Butterworth.

Ces étapes sont suivies d'un filtrage passe-bande et d'une extraction temporelle, dont l'ordre d'application peut varier selon les analyses (figure 3.4). Le filtrage passe-bande vise à restreindre le signal aux bandes de fréquences d'intérêt, qui correspondent aux réponses cérébrales associées à la tâche considérée, tout en atténuant au maximum le bruit hors bande. L'extraction temporelle (ou *epoching*), quant à elle, consiste à segmenter le signal continu en fenêtres temporelles synchronisées avec l'apparition de chaque stimulus (*onset*), afin d'isoler temporellement l'activité cérébrale correspondante. Une *epoch* correspond donc à un segment temporel extrait d'un essai, durant lequel se produit un évènement élémentaire, déclenché soit par une intention interne, soit par un stimulus externe. Ainsi, une première stratégie [Delorme et Makeig, 2004] (figure 3.4a) consiste, dans un premier temps, à segmenter temporellement les signaux, puis à examiner les fenêtres obtenues afin d'identifier et de rejeter les artefacts ainsi que, si nécessaire, les canaux bruités. Ces étapes peuvent également être automatisées à l'aide de critères appropriés. Le filtrage passe-bande est alors appliqué aux données conservées. À l'inverse, il est également possible d'appliquer le filtrage temporel directement sur les signaux continus, avant leur segmentation en *epochs* (figure 3.4b) [Guého, 2022]. Cette stratégie permet notamment d'éviter l'introduction de distorsions au début et à la fin de chaque fenêtre temporelle (i.e. effets de bord), susceptibles d'apparaître lorsque le filtre passe-bande est appliqué après la segmentation temporelle des signaux. De plus, elle peut conduire au rejet d'un plus faible nombre d'artefacts, certains pouvant être atténués ou supprimés par le filtrage en amont, ce qui permet de conserver une plus grande quantité de données pour les analyses ultérieures. Toutefois, cette approche n'est applicable qu'en mode *offline* (cf. section 3.2.5), puisqu'elle nécessite de disposer des signaux dans leur intégralité avant leur segmentation en *epochs*. En revanche, celle proposée par Delorme et Makeig [2004] est compatible avec un fonctionnement *online*, le traitement étant réalisé indépendamment sur chaque fenêtre après segmentation temporelle.



(a) Stratégie de prétraitement proposée par Delorme et Makeig [2004].

(b) Stratégie alternative de prétraitement proposée par Guého [2022].

FIGURE 3.4 – Deux stratégies d'application des étapes de prétraitement des signaux EEG : à gauche, la stratégie proposée par Delorme et Makeig [2004] ; à droite, une stratégie alternative différant notamment par l'ordre d'application du filtrage passe-bande et de l'extraction temporelle [Guého, 2022].

Dans tous les cas, les signaux EEG étant intrinsèquement bruités, l'utilisation d'un filtre passe-bande est indispensable. L'ajustement de ses paramètres doit alors être réalisé avec soin afin de minimiser le bruit et de supprimer les artefacts situés en dehors des bandes de fréquences d'intérêt, tels que les clignements oculaires ou les interférences liées au courant alternatif [Sen et al., 2023]. Toutefois, selon les propriétés du filtre utilisé, notamment sa réponse en gain et en phase, le filtrage peut également introduire des distorsions du signal [Sen et al., 2023]. Ainsi, les paramètres du filtre, tels que son type, sa bande passante et sa bande de transition, peuvent avoir une incidence directe sur les performances de classification.

3.3.1 Filtrage passe-bande

Dans les paradigmes SSVEP et SSAEP, le filtrage passe-bande vise à isoler les bandes de fréquences correspondant aux fréquences de stimulation, ainsi qu'à leurs harmoniques. Les données sont alors filtrées à l'aide d'un filtre passe-bande centré sur la fréquence de stimulation, ainsi que d'un second filtre passe-bande appliqué autour de la première harmonique [Higashi et al., 2011]. En effet, la prise en compte de l'activité aux fréquences harmoniques peut potentiellement améliorer les performances de classification.

Les informations relatives à la bande passante des filtres restent rares dans la littérature. Une étude sur les SSVEP [Si-Mohammed et al., 2022] utilise une bande passante de ±0,25 Hz autour des fréquences de stimulation, tandis qu'une étude sur les SSAEP rapporte une bande passante plus large, de l'ordre de ±1 Hz [Krishnan et al., 2009]. Pourtant, ce paramètre influence directement la qualité du signal extrait et, par conséquent, les performances du système. Il est donc étudié plus en détail dans la section 5.2.1.1.

3.3.2 Extraction temporelle

L'extraction temporelle est une étape importante pour la classification, en particulier dans le cadre des méthodes basées sur les matrices de covariance [Kalunga et al., 2016] (voir section 3.5.2.4). De fait, l'augmentation de la durée des *epochs* améliore les performances du classifieur en fournissant une estimation plus fiable des matrices de covariance. Dans cette perspective, il est nécessaire de déterminer précisément l'*onset* de chaque stimulation, ainsi que sa durée. Lorsque la stimulation est suffisamment longue, le signal peut être segmenté en plusieurs *epochs*, afin d'exploiter différemment les données disponibles sans acquisition supplémentaire [Lashgari et al., 2020]. Le chevauchement des *epochs* constitue également une stratégie pour multiplier les segments analysés. Toutefois, les *epochs* qui se chevauchent partageant une partie de l'information, elles n'apportent pas nécessairement de nouvelles données discriminantes [Lashgari et al., 2020]. L'intérêt de cette approche réside donc principalement dans la possibilité d'exploiter au mieux les informations contenues dans les signaux en ajustant la durée et le recouvrement des *epochs*. Elle peut également s'avérer utile pour les méthodes qui requièrent un grand nombre d'échantillons (e.g. les méthodes d'apprentissage profond, voir section 3.5.2.5) ou lorsque l'ensemble de données est limité, la subdivision en *epochs*, éventuellement avec recouvrement, permettant alors d'augmenter la taille effective de l'ensemble d'apprentissage.

3.4 Réduction de dimension

A l'issue des étapes de prétraitement, les signaux EEG restent marqués par une variabilité importante et un faible RSB (proche de 0 et variable en fonction du type de stimulus, voir section 5.2). Dans ce contexte, la réduction de dimension consiste à projeter les données dans un nouvel espace afin de regrouper l'information pertinente sur un ensemble de composantes, puis à en sélectionner un sous-ensemble pertinent pour la classification.

Ces techniques sont particulièrement utiles pour faire face aux problèmes liés à la grande dimension, où le "fléau de la dimension" (*curse of dimensionality*) [Banks et Fienberg, 2003] peut entraîner une dégradation des performances, du fait de la dispersion des données et de la complexité accrue du modèle. Ce problème est particulièrement marqué lorsque les évaluations reposent uniquement sur les données d'un seul sujet, dans le but de construire un modèle spécifique à celui-ci. À titre d'exemple, la représentation dans l'espace tangent (*Tangent Space* ou TS, voir section 3.5.2.4) peut conduire à environ 300 caractéristiques à partir de 24 canaux. Dans ce cas, la dimension de l'espace des caractéristiques peut largement dépasser celui des échantillons disponibles pour un sujet, selon les stratégies d'augmentation de données mises en œuvre. Afin de pallier cette situation, des techniques de réduction de dimension, telles que l'analyse en composantes principales (*Principal Component Analysis* ou PCA), le *Common Spatial Pattern* (CSP) et l'analyse discriminante linéaire (*Linear Discriminant Analysis* ou LDA) peuvent être appliquées. En concentrant l'information pertinente sur un nombre réduit de composantes, ces méthodes permettent également d'augmenter le RSB.

3.4.1 Notations des *epochs*

Dans la suite de ce chapitre, les méthodes supervisées utilisent, pour chaque *epoch*, l'information relative à la condition expérimentale dont elle est issue, appelée **étiquette de classe**. Les *epochs* extraites pour une même condition sont ainsi regroupées au sein d'une même classe k. La i-ème *epoch* appartenant à cette classe est notée $\mathbf{X}_i^{(k)}$, avec $i = 1, \ldots, I_k$, où I_k désigne le nombre total d'*epochs* de la classe k. Dans le cas binaire, deux classes A et B sont considérées,

correspondant à deux conditions expérimentales (e.g. "présence de la fréquence de stimulation" et "absence de la fréquence de stimulation"). Les *epochs* appartenant à ces classes sont notées $\mathbf{X}_i^{(A)}$ et $\mathbf{X}_i^{(B)}$, avec $i = 1, \ldots, I_A$ pour la classe A et $i = 1, \ldots, I_B$ pour la classe B, où I_A et I_B correspondent au nombre d'*epochs* disponibles pour chaque classe.

3.4.2 Analyse en composantes principales (*Principal Component Analysis* ou PCA)

La PCA a pour objectif de conserver le maximum d'information, en l'occurrence de variance des données, tout en réduisant la dimension de celles-ci [Abdi et Williams, 2010]. Cette méthode est largement utilisée dans de nombreux domaines, au-delà de l'analyse de signaux EEG, notamment en reconnaissance d'images, en traitement de la parole, ainsi qu'en analyse de données environnementales et climatologiques.

La PCA est une transformation linéaire orthogonale qui projette les données dans un nouvel espace de dimension réduite [Jolliffe, 2002]. Elle construit donc de nouvelles variables, appelées **composantes principales**, qui correspondent à des combinaisons linéaires des variables initiales. Ces composantes sont déterminées de manière à maximiser successivement la variance des données projetées : la première composante principale est calculée de façon à présenter la variance maximale, la deuxième composante est calculée sous la contrainte d'être orthogonale à la première composante et capture la plus grande part de variance restante, et ainsi de suite. Elles sont ainsi, par construction, non corrélées.

Une approche classique pour calculer la PCA consiste à utiliser la décomposition en valeurs singulières (*Singular Value Decomposition* ou SVD). Etant donnée une matrice de données centrée $\mathbf{X} \in \mathbb{R}^{C \times N}$ (i.e. telle que chaque ligne possède une moyenne nulle), sa SVD est donnée par :

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^\top, \tag{3.3}$$

avec $\mathbf{S}$, une matrice diagonale contenant les valeurs singulières de $\mathbf{X}$. Les valeurs singulières sont égales aux racines carrés positives des valeurs propres de $\mathbf{X}^\top\mathbf{X}$ (ou de $\mathbf{X}\mathbf{X}^\top$). La matrice $\mathbf{U}$ contient les vecteurs singuliers gauches de $\mathbf{X}$, tandis que $\mathbf{V}$ contient les vecteurs singuliers droits. Ces derniers correspondent aux vecteurs propres de $\mathbf{X}^\top\mathbf{X}$, tandis que les vecteurs singuliers gauches correspondent aux vecteurs propres de $\mathbf{X}\mathbf{X}^\top$. Les composantes principales sont obtenues en projetant les données sur les vecteurs singuliers droits, ce qui donne :

$$\mathbf{F} = \mathbf{X}\mathbf{V} = \mathbf{U}\mathbf{S}. \tag{3.4}$$

Les colonnes de $\mathbf{F}$ correspondent ainsi aux composantes principales. En sélectionnant les v composantes principales de $\mathbf{F}$ associées aux plus grandes valeurs singulières, c'est-à-dire celles qui capturent le plus de variance, la PCA permet de réduire efficacement la dimension des données. Il est important de noter que cette réduction de la dimension est réalisée de manière non supervisée, donc indépendamment de la pertinence des informations initiales.

3.4.2.1 *Kernel Principal Component Analysis* (KPCA)

La PCA peut être étendue à des transformations non linéaires grâce à l'utilisation de fonctions noyaux comme dans la KPCA [Schölkopf et al., 1998].

L'algorithme permet de projeter les vecteurs de caractéristiques $\mathbf{x}_i$ dans un espace de plus grande dimension (potentiellement infini) à l'aide d'une fonction noyau $\kappa(\mathbf{x}_i, \mathbf{x}_j)$, définie par :

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle_{\mathcal{H}}, \tag{3.5}$$

où $\mathcal{H}$ désigne un espace de caractéristiques de dimension plus élevée (souvent un espace de Hilbert à noyau reproduisant, *Reproducing Kernel Hilbert Space* ou RKHS) [Briscik et al., 2023]. Un RKHS est un espace de Hilbert associé à une fonction noyau, dans lequel le noyau permet de "reproduire", via un produit scalaire, la valeur d'une fonction en un point [Yuan et Cai, 2010]. Cette projection est réalisée à l'aide d'une transformation non linéaire ϕ, qui dans la pratique n'est pas calculée explicitement. Seule la fonction noyau est utilisée, ce qui permet d'effectuer la projection implicite des données dans l'espace de grande dimension.

Les performances de l'algorithme dépendent fortement du choix de la fonction noyau. Les noyaux les plus couramment utilisés sont le noyau linéaire, le noyau polynomial [Pan et al., 2025] et le noyau gaussien [Tayeb et al., 2014]. Le noyau linéaire correspond à un cas particulier pour lequel la KPCA se réduit à la PCA classique, tandis que l'utilisation de noyaux non linéaires permet d'obtenir une meilleure séparation des données lorsque celles-ci ne sont pas linéairement séparables dans l'espace d'origine. Dans ce contexte, la KPCA offre une meilleure capacité que la PCA à capturer la structure non linéaire des données EEG [Pan et al., 2025]. Grâce à la projection non linéaire induite par la fonction noyau, elle permet une réduction de dimension plus efficace tout en préservant les informations clés et les caractéristiques discriminantes. Cette transformation contribue également à réduire les corrélations entre les caractéristiques, ce qui améliore la séparabilité des variables. Les résultats rapportés dans [Pan et al., 2025] et [Tayeb et al., 2014] suggèrent que les caractéristiques extraites par KPCA permettent d'obtenir des performances de classification plus élevées que celles obtenues avec la PCA. De plus, la KPCA se caractérise par une variabilité plus faible des performances de classification, suggérant une meilleure stabilité et une plus grande robustesse.

3.4.3 *Common Spatial Pattern* (CSP)

L'idée principale du CSP est de déterminer une matrice de projection ($\mathbf{W}_{\text{CSP}}$), de manière à projeter les données EEG enregistrées pour plusieurs conditions [Blankertz et al., 2007]. Cette projection linéaire, définie par :

$$\mathbf{X}_{\text{CSP}} = \mathbf{W}_{\text{CSP}}^{\top} \mathbf{X}, \tag{3.6}$$

vise à maximiser la variance d'une condition tout en minimisant simultanément la variance d'une autre condition [Koles et al., 1990; Li et al., 2013]. Les colonnes de la matrice de projection correspondent à des filtres spatiaux linéaires ($\{\mathbf{w}_i\}_{i=1}^{C}$), qui sont des ensembles de poids attribués à chaque canal EEG afin de les pondérer [Wang et al., 2006]. Une réduction de dimension peut être effectuée en sélectionnant un sous-ensemble de l filtres spatiaux, ce qui permet de projeter les données dans un sous-espace de dimension réduite.

Initialement introduite pour des problèmes de classification binaire, la méthode est basée sur la PCA. Le problème peut être formulé sous la forme d'un quotient de Rayleigh [Grosse-Wentrup et Buss, 2008] :

$$J(\mathbf{w}) = \underset{\mathbf{w} \in \mathbb{R}^C}{\arg\max} \frac{\mathbf{w}^{\top} \mathbf{P}_{\mathbf{A}} \mathbf{w}}{\mathbf{w}^{\top} \mathbf{P}_{\mathbf{B}} \mathbf{w}}, \tag{3.7}$$

où J est la fonction objectif, $\mathbf{P}_{\mathbf{A}}$ et $\mathbf{P}_{\mathbf{B}}$ sont les matrices de covariance pour les deux conditions $\mathbf{X}^{(A)}$ et $\mathbf{X}^{(B)}$ (cf. section 3.4.1) [Li et al., 2013]. La solution à l'équation 3.7 est donnée en résolvant l'équation aux valeurs propres :

$$\mathbf{P_A w} = \lambda \mathbf{P_B w}, \tag{3.8}$$

où λ représente la valeur propre et $\mathbf{w}$, le vecteur propre associé, correspondant à un filtre spatial. Afin d'obtenir plusieurs filtres spatiaux, ce problème (équation 3.8) est étendu à une décomposition matricielle, en résolvant :

$$\mathbf{P_B}^{-1}\mathbf{P_A}\mathbf{W}_{\text{CSP}} = \mathbf{\Lambda}\mathbf{W}_{\text{CSP}}, \tag{3.9}$$

où $\mathbf{W}_{\text{CSP}} \in \mathbb{R}^{C\times C}$ est la matrice dont les colonnes sont les vecteurs propres $\mathbf{w}_i$ de $\mathbf{P_B}^{-1}\mathbf{P_A}$, et $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_C)$, la matrice diagonale des valeurs propres correspondantes [Wang et al., 1999]. La décomposition CSP fournit donc un ensemble complet de C filtres spatiaux.

Les premiers et derniers vecteurs de $\mathbf{W}_{\text{CSP}}$ sont les filtres spatiaux les plus importants, qui peuvent expliquer la plus grande variance d'une condition et la plus petite d'une autre, comme ils maximisent et minimisent celle-ci respectivement [Lotte et Guan, 2010]. Le filtre spatial $\mathbf{w}_1$ correspond à la plus grande valeur propre λ_1, qui maximise la variance de la condition A tout en la minimisant pour la condition B, et $\mathbf{w}_C$, correspond à la plus petite valeur propre, qui maximise la variance pour la condition B tout en minimisant celle de la condition A. Les filtres spatiaux associés à des valeurs propres intermédiaires produisent des signaux projetés (composantes) pour lesquels le rapport de variance entre les deux conditions est moins marqué que pour les filtres extrêmes, et sont donc généralement moins discriminants. Le pouvoir discriminant diminue ainsi à mesure que l'on se rapproche des indices centraux (autour de $\frac{C}{2}$). La réduction de dimension consiste alors généralement à ne conserver qu'un sous-ensemble de l filtres spatiaux $\{\mathbf{w}_i\}_{i=1}^{l}$, avec $l < C$, sélectionnés parmi les filtres associés aux valeurs propres extrêmes (figure 3.3).

En inversant la transposée de la matrice de projection complète ($\mathbf{W}_{\text{CSP}}$), les motifs spatiaux $\mathbf{A} = (\mathbf{W}_{\text{CSP}}^{\top})^{-1}$ [Barachant et al., 2010a] sont obtenus. Ils permettent de visualiser de manière significative les poids attribués à chaque électrode. Les filtres spatiaux contenus dans $\mathbf{W}_{\text{CSP}}$ décrivent uniquement la manière de combiner les signaux des différents canaux, tandis que les motifs spatiaux contenus dans les colonnes de $\mathbf{A}$ peuvent être interprétés directement comme la cartographie des contributions relatives des électrodes à l'information utile [Haufe et al., 2014; Pan et al., 2022] (voir section 5.4).

Malgré sa popularité et son efficacité, le CSP est connu pour être très sensible aux valeurs aberrantes, qui existent largement dans les applications de BCI, dues, entre autres, au serrage de la mâchoire ou à la perte de contact des électrodes [Li et al., 2013].

3.4.4 Analyse discriminante linéaire (*Linear Discriminant Analysis* ou LDA)

La LDA vise à maximiser la séparabilité entre les classes [Tang et al., 2014], en maximisant la variance inter-classe tout en minimisant la variance intra-classe. Afin d'atteindre son objectif, la LDA repose sur la détermination d'une matrice de projection linéaire ($\mathbf{W}_{\text{LDA}}$), dont les colonnes correspondent aux directions maximisant la séparation entre les conditions [Alberti et al., 2017]. Les signaux EEG mesurés sont alors projetés dans un espace de dimension réduite $(K-1)$, où K est le nombre de conditions :

$$\mathbf{X}_{\text{LDA}} = \mathbf{W}_{\text{LDA}}^{\top}\mathbf{X}. \tag{3.10}$$

Soient μ_A et μ_B les moyennes des données associées aux conditions A et B respectivement, et $\mathbf{P}_A$ et $\mathbf{P}_B$ les matrices de covariances, l'objectif est de trouver le vecteur de projection $\mathbf{v} \in \mathbb{R}^C$ qui maximise la dispersion entre les deux conditions tout en minimisant la dispersion au sein de

chaque condition [Kim et al., 2007]. La matrice de dispersion inter-classe ($\mathbf{S}_B$) et la matrice de dispersion intra-classe ($\mathbf{S}_W$) sont définies [Hastie et al., 2008] telles que :

$$\mathbf{S}_B = (\mu_A - \mu_B)(\mu_A - \mu_B)^\top \tag{3.11}$$

et

$$\mathbf{S}_W = N_A \mathbf{P}_A + N_B \mathbf{P}_B, \tag{3.12}$$

avec N_A et N_B, le nombre d'échantillons des matrices $\mathbf{X}^{(A)}$ et $\mathbf{X}^{(B)}$. La dispersion inter-classe identifie la direction qui maximise la distance entre les moyennes des classes, mais elle ne tient pas compte du chevauchement entre les distributions. Par conséquent, la dispersion intra-classe est aussi intégrée pour minimiser le chevauchement, ce qui permet d'obtenir la direction de projection optimale, par maximisation de la fonction objectif :

$$J(\mathbf{v}) = \frac{\mathbf{v}^\top \mathbf{S}_B \mathbf{v}}{\mathbf{v}^\top \mathbf{S}_W \mathbf{v}}. \tag{3.13}$$

Comme précédemment (cf. section 3.4.3), ce problème peut également être formulé sous la forme d'un quotient de Rayleigh. La direction de projection optimale est alors donnée par :

$$\mathbf{v} = \mathbf{S}_W^{-1}(\mu_A - \mu_B). \tag{3.14}$$

Les données peuvent alors être projetées sur cette direction afin d'améliorer la séparabilité des classes. Ainsi, lorsque la classification est binaire ($K = 2$), cette projection se réduit à une seule dimension, ce qui évite d'avoir à sélectionner le nombre de composantes. Cependant, cela peut simplifier à l'excès le problème, car une seule dimension peut ne pas suffire à capturer suffisamment d'informations pour une classification précise, en particulier avec des données complexes, telles que les signaux EEG.

La LDA est généralement utilisée pour la réduction de dimension, mais elle peut également être employée comme classifieur linéaire [Hastie et al., 2008], puisqu'elle incorpore explicitement les étiquettes de classe. La classification peut alors être réalisée directement dans l'espace projeté, en déterminant un seuil de décision approprié.

3.5 Détection de la fréquence de modulation

3.5.1 Rapport Signal sur Bruit (RSB)

Une façon d'estimer la détectabilité des fréquences d'intérêt consiste à calculer le RSB. Le signal étant alors défini comme les enregistrements EEG en présence du stimulus, tandis que le bruit correspond aux signaux EEG en l'absence de stimulus. Le RSB est ensuite défini à partir de la puissance spectrale de ces signaux à la fréquence d'intérêt, notées respectivement P_{signal} et P_{bruit} :

$$RSB = 10 \log_{10} \frac{P_{\text{signal}}}{P_{\text{bruit}}}. \tag{3.15}$$

Ce RSB permet ainsi de caractériser l'émergence de la fréquence de stimulation dans l'activité cérébrale. Plus sa valeur est élevée, plus la fréquence de stimulation est détectable dans les signaux EEG, suggérant un potentiel accru de discrimination par des méthodes de classification automatique.

3.5.2 Méthodes de classification

La détectabilité des fréquences d'intérêt peut également être évaluée à partir des performances de classification obtenues à l'aide de modèles d'apprentissage automatique appliqués aux réponses cérébrales. Plus précisément, le problème est formulé comme une tâche de classification visant à discriminer des signaux EEG contenant la fréquence de stimulation de signaux ne la contenant pas. Les performances obtenues constituent alors une mesure de la détectabilité de cette fréquence, bien qu'elles dépendent également des capacités du modèle utilisé à extraire et discriminer les caractéristiques pertinentes des signaux EEG.

L'utilisation des réponses cérébrales induites par des stimulations auditives, et plus particulièrement des SSAEP, a été peu étudiée. Par conséquent, aucune méthode de classification n'a été développée spécifiquement pour ce type de réponses. En revanche, les SSVEP ont fait l'objet de nombreuses recherches ayant conduit au développement de plusieurs méthodes de classification adaptées à leur détection. Les SSVEP et les SSAEP partageant des caractéristiques similaires (cf. section 2.4.2.4), il apparaît raisonnable d'appliquer dans un premier temps aux SSAEP les techniques développées pour les SSVEP.

L'approche la plus simple pour qualifier des SSVEP repose sur l'analyse spectrale des signaux EEG. En effet, puisque le spectre des signaux EEG est censé présenter un pic à la fréquence de stimulation, ainsi qu'à ses harmoniques, un simple détecteur de pic peut donc permettre d'identifier la fréquence de stimulation et de classer le signal en conséquence [Cheng et al., 2002]. Toutefois, cette méthode nécessite un RSB élevé. Afin d'améliorer les performances de classification, des méthodes d'apprentissage supervisé classiques peuvent également être utilisés, tels que la LDA [Kanagaluru et Sasikala, 2024; Carvalho et al., 2015]. Pour des tâches de classification binaire, la régression logistique (*Logistic Regression* ou LR) [Acampora et al., 2019; Khan et al., 2023] et la *Support Vector Machine* (SVM) [Kanagaluru et Sasikala, 2024; Carvalho et al., 2015] peuvent être employées. Ces méthodes sont couramment utilisées pour la classification de signaux EEG [Lotte et al., 2018; Aggarwal et Chugh, 2022]. Au-delà de ces approches classiques, d'autres classifieurs peuvent également être envisagés. Parmi les plus fréquemment utilisés figurent notamment [Lotte et al., 2018] :

1. l'analyse canonique des corrélations (*Canonical Correlation Analysis* ou CCA),
2. les classifieurs basés sur la géométrie riemannienne,
3. les modèles d'apprentissage profond (*Deep Learning* ou DL), y compris les réseaux neuronaux convolutifs (*Convolutional Neural Network* ou CNN).

3.5.2.1 Régression logistique (*Logistic Regression* ou LR)

La régression logistique repose sur une fonction logistique, appelée sigmoïde, qui transforme une combinaison linéaire de variables indépendantes en une probabilité, $p_{\boldsymbol{\omega}}(\mathbf{x}_i)$, comprise entre 0 et 1, représentant la probabilité a posteriori d'appartenance à la classe positive [Prasad et al., 2013; Gaur et al., 2021]. La probabilité prédite par le modèle est donnée par :

$$p_{\boldsymbol{\omega}}(\mathbf{x}_i) = \frac{1}{1 + e^{-\boldsymbol{\omega}\mathbf{x}_i}}, \tag{3.16}$$

avec $\mathbf{x}_i^{(0)} = 1$ correspondant au terme de biais, permettant de décaler la frontière de décision du modèle. Les coefficients de régression $\boldsymbol{\omega} = [\boldsymbol{\omega}_0 \ \dots \ \boldsymbol{\omega}_j \ \dots \ \boldsymbol{\omega}_C]$ sont estimés par la méthode du maximum de vraisemblance [Hosmer Jr et al., 2013], à partir d'un ensemble de données $\{(\mathbf{x}_i, y_i)\}_{i=\{1,\dots,N\}}$, où y_i désigne l'étiquette associée à l'observation $\mathbf{x}_i$. Cette méthode conduit

à privilégier les valeurs des paramètres pour lesquelles les probabilités prédites sont les plus cohérentes avec les étiquettes observées.

Lorsque $y_i \in \{0, 1\}$ est une variable binaire, le modèle définit les probabilités conditionnelles ($\mathbb{P}$) des deux classes sachant $\mathbf{x}_i$:

$$\mathbb{P}(y_i = 1 \mid \mathbf{x}_i) = p_{\boldsymbol{\omega}}(\mathbf{x}_i), \quad \mathbb{P}(y_i = 0 \mid \mathbf{x}_i) = 1 - p_{\boldsymbol{\omega}}(\mathbf{x}_i).$$

Ainsi, pour une observation $(\mathbf{x}_i, y_i)$, la probabilité associée est donnée par :

$$p_{\boldsymbol{\omega}}(\mathbf{x}_i)^{y_i}(1 - p_{\boldsymbol{\omega}}(\mathbf{x}_i))^{1-y_i}. \tag{3.17}$$

La fonction de vraisemblance sur l'ensemble des données s'écrit comme le produit des probabilités associées à chaque observation :

$$\mathcal{L}(\boldsymbol{\omega}) = \prod_{i=1}^{N} p_{\boldsymbol{\omega}}(\mathbf{x}_i)^{y_i}(1 - p_{\boldsymbol{\omega}}(\mathbf{x}_i))^{1-y_i}. \tag{3.18}$$

Le principe du maximum de vraisemblance consiste à choisir les paramètres $\boldsymbol{\omega}$ qui maximisent la fonction de vraisemblance (équation 3.18). En pratique, on maximise plutôt la log-vraisemblance pour des raisons de stabilité numérique et de simplicité analytique :

$$\ell(\boldsymbol{\omega}) = \ln[\mathcal{L}\boldsymbol{\omega})] = \sum_{i=1}^{N} y_i \ln[p_{\boldsymbol{\omega}}(\mathbf{x}_i)] + (1 - y_i) \ln[(1 - p_{\boldsymbol{\omega}}(\mathbf{x}_i))], \tag{3.19}$$

ce qui revient à annuler son gradient, c'est-à-dire à vérifier pour tout j,

$$\frac{d\ell}{d\boldsymbol{\omega}_j} = 0. \tag{3.20}$$

La recherche des paramètres optimaux nécessite donc le recours à des méthodes d'optimisation numérique.

3.5.2.2 *Support Vector Machine* (SVM)

La SVM vise à déterminer un hyperplan séparant au mieux deux conditions en maximisant la marge, définie comme la distance entre l'hyperplan et les échantillons les plus proches de celui-ci appartenant à chaque condition, appelés **vecteurs de support** (figure 3.5) [Aguilar et al., 2015].

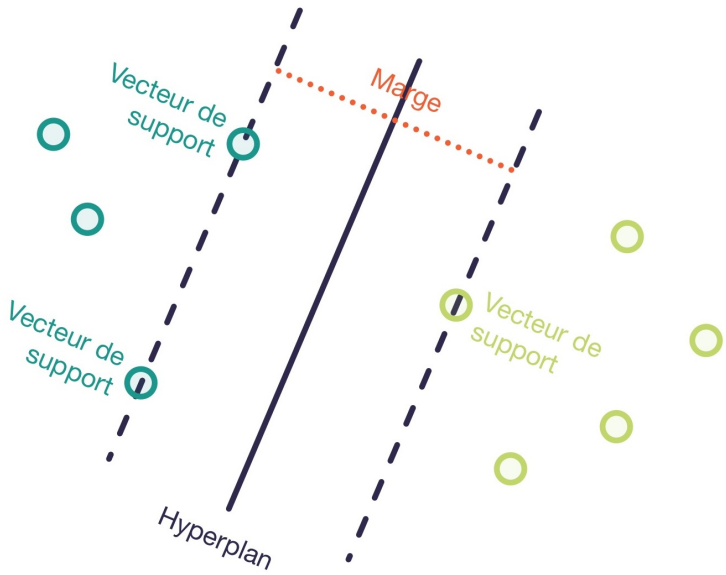


FIGURE 3.5 – Principe de la SVM : l'hyperplan séparateur maximise la marge, définie comme la distance entre l'hyperplan et les échantillons les plus proches appartenant à chaque classe, appelés **vecteurs de support**.

L'objectif est de construire une frontière optimale qui permette une classification précise des nouvelles observations [Vishwakarma et al., 2020]. Dans le cas linéaire, l'hyperplan séparateur est défini par :

$$h(\mathbf{x}_i) = \boldsymbol{\psi}\mathbf{x}_i + \psi_0, \tag{3.21}$$

où $\boldsymbol{\psi}$ est le vecteur de poids normal à l'hyperplan et ψ_0, le terme de biais [Jakkula, 2006].

La SVM impose non seulement que les données d'entraînement soient correctement séparées par l'hyperplan, mais également qu'elles respectent une marge de sécurité afin d'améliorer la capacité de généralisation [Quitadamo et al., 2017]. Le but est donc de trouver les paramètres $\boldsymbol{\psi}$ et ψ_0 maximisant la marge. Ainsi, pour tout exemple i, la contrainte suivante doit être satisfaite :

$$y_i(\boldsymbol{\psi}\mathbf{x}_i + \psi_0) \geq 1. \tag{3.22}$$

Cette contrainte impose donc à la fois une classification correcte des données et le respect de la marge. En effet, dans l'équation 3.22, le terme $(\boldsymbol{\psi}\mathbf{x}_i + \psi_0)$ correspond à la position du point $\mathbf{x}_i$ par rapport à l'hyperplan séparateur, tandis que $y_i \in \{-1, +1\}$ est son étiquette et indique sa classe. Ainsi, le produit $y_i(\boldsymbol{\psi}\mathbf{x}_i + \psi_0)$ est positif lorsque l'échantillon est correctement classé. L'imposition d'une borne supérieure ou égale à 1 garantit donc que les observations se situent à une distance minimale de l'hyperplan, définissant ainsi la marge.

La largeur de la marge est inversement proportionnelle à la norme du vecteur $\boldsymbol{\psi}$, la distance entre les deux hyperplans marginaux (i.e. les deux hyperplans parallèles à l'hyperplan séparateur qui délimitent la marge) étant égale à $2/\|\boldsymbol{\psi}\|$. Maximiser la marge revient alors à minimiser la norme du vecteur poids sous la contrainte de l'équation 3.22. Le problème d'optimisation s'écrit alors :

$$J(\boldsymbol{\psi}) = \frac{1}{2}\|\boldsymbol{\psi}\|^2, \tag{3.23}$$

où $J(\boldsymbol{\psi})$ est la fonction objectif, correspondant au critère à minimiser afin de déterminer l'hyperplan séparateur optimal.

Lors de la phase de test, une nouvelle observation ($\mathbf{z}$) est classée selon la règle de décision :

$$\hat{y} = \text{sign}[h(\mathbf{z})] = \text{sign}[\boldsymbol{\psi}\mathbf{z} + \psi_0], \tag{3.24}$$

ce qui signifie que $\mathbf{z}$ est assigné à la classe positive (correspondant à une condition) si $h(\mathbf{z}) \geq 0$, et à la classe négative (correspondant à l'autre condition) sinon, en fonction du côté de l'hyperplan dans lequel il se situe [Lal et al., 2004].

3.5.2.3 Analyse canonique des corrélations (*Canonical Correlation Analysis* ou CCA)

La CCA est l'une des méthodes les plus populaires dans le domaine des BCI basées sur les SSVEP [Talevi et al., 2017]. Elle permet de classer les réponses cérébrales associées à S stimulations de fréquences distinctes (f_s, avec $s = 1, 2, \ldots, S$). Le principe de la CCA consiste à rechercher la corrélation maximale entre deux ensembles de données [Zhang et al., 2014] : d'une part, les signaux EEG multi-canaux enregistrés pendant une stimulation ($\mathbf{X}_i$), et d'autre part, une matrice de signaux de référence ($\mathbf{R}_s \in \mathbb{R}^{2H \times N}$) construite à partir d'une fréquence de stimulation (f_s). Le facteur 2 provient de l'utilisation d'une sinusoïde et d'une cosinusoïde, pour chaque fréquence $i \cdot f_s$ avec $i = 1, \ldots, H$ (où H désigne le nombre d'harmoniques considérées), permettant de prendre en compte les décalages de phase. Ces signaux de référence constituent des modèles optimaux des réponses attendues dans l'activité cérébrale lorsque le sujet est stimulé à la fréquence f_s :

$$\mathbf{R}_s = \begin{pmatrix} \sin(2\pi \cdot 1 \cdot f_s\mathbf{t}) \\ \cos(2\pi \cdot 1 \cdot f_s\mathbf{t}) \\ \vdots \\ \sin(2\pi \cdot H \cdot f_s\mathbf{t}) \\ \cos(2\pi \cdot H \cdot f_s\mathbf{t}) \end{pmatrix}, \mathbf{t} = \frac{1}{f_e}, \frac{2}{f_e}, \ldots, \frac{N}{f_e}, \tag{3.25}$$

où f_e est la fréquence d'échantillonnage.

La CCA consiste à rechercher une paire de transformations linéaires, appelées **variables canoniques**, $\mathbf{q}_x \in \mathbb{R}^C$ et $\mathbf{q}_r \in \mathbb{R}^{2H}$, maximisant la corrélation entre les combinaisons linéaires $\tilde{\mathbf{x}} = \mathbf{q}_x^\top \mathbf{X}_i$ et $\tilde{\mathbf{r}} = \mathbf{q}_r^\top \mathbf{R}_s$. Les matrices $\mathbf{X}_i$ et $\mathbf{R}_s$ sont normalisées pour avoir une moyenne nulle et une variance unitaire. Le coefficient de corrélation est donné par :

$$\rho = \frac{\mathbf{q}_x^\top \mathbf{X}_i \mathbf{R}_s^\top \mathbf{q}_r}{\sqrt{\mathbf{q}_x^\top \mathbf{X}_i \mathbf{X}_i^\top \mathbf{q}_x \mathbf{q}_r^\top \mathbf{R}_s \mathbf{R}_s^\top \mathbf{q}_r}}, \tag{3.26}$$

qui mesure la similarité entre les deux ensembles de données [Talevi et al., 2017]. Le coefficient de corrélation maximal (ρ_{max}) par rapport à $\mathbf{q}_x$ et $\mathbf{q}_r$ est la corrélation canonique maximale ($\max_{\mathbf{q}_x, \mathbf{q}_r} \rho$).

Pour chaque nouvelle observation ($\mathbf{Z}$), le coefficient de corrélation (ρ) est calculé entre $\mathbf{Z}$ et chacune des matrices de signaux de référence ($\mathbf{R}_s$). La fréquence de stimulation est alors identifiée comme celle associée à la matrice de référence ($\mathbf{R}_s$) produisant le coefficient de corrélation maximal (ρ_{max}) avec $\mathbf{Z}$,

$$\hat{f} = \arg\max_{f_s} \rho_{max}. \tag{3.27}$$

Cette méthode a montré des performances de reconnaissance significativement supérieures à celles de l'analyse traditionnelle de la densité spectrale de puissance [Zhang et al., 2014]. Toutefois, la CCA ne permet pas de détecter un **état de repos** (également appelé **classe de rejet** ou **absence de stimulation**) [Kalunga et al., 2016]. En effet, la CCA sélectionne systématiquement la fréquence de stimulation pour laquelle la corrélation est maximale avec les signaux EEG, même lorsque l'activité cérébrale est mesurée en l'absence de stimulation. Elle ne dispose donc pas d'un mécanisme permettant de rejeter une observation ne correspondant à aucune des fréquences considérées.

Avant l'introduction des approches riemanniennes, les méthodes basées sur la CCA obtenaient généralement les meilleures performances de classification dans les BCI [Chevallier et al., 2021].

3.5.2.4 Modèles riemanniens

L'idée principale des approches riemanniennes consiste à utiliser des matrices de covariance pour représenter les signaux EEG et à exploiter une géométrie adaptée à leur structure afin de classifier directement ces matrices [Barachant et al., 2011], qui sont symétriques et définies positives (*Symmetric and Positive-Definite* ou SPD) [Chevallier et al., 2021]. L'espace des matrices SPD, noté $\mathcal{P}(C)$, est une variété (*manifold*) riemannienne $\mathcal{M}$ [Barachant et al., 2011]. Une variété est un espace géométrique qui peut être courbe, mais qui peut être localement assimilé à un espace euclidien [Fillard, 2004]. Par exemple, une sphère est une variété : bien qu'elle soit courbe globalement, elle apparaît localement comme un plan. Lorsqu'un espace tangent est défini en chaque point de la variété, correspondant à une approximation locale euclidienne de la variété (figure 3.6), celle-ci est dite **différentielle** [Kalunga et al., 2016]. Une variété riemannienne est alors une variété différentielle munie d'une métrique riemannienne, c'est-à-dire d'un produit scalaire défini sur chaque espace tangent.

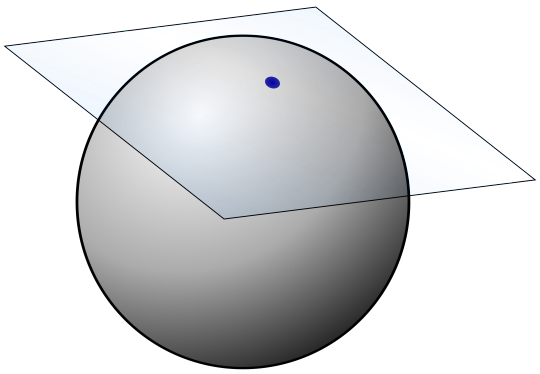


FIGURE 3.6 – Représentation du plan tangent à une sphère, en un point. Ce plan constitue une approximation euclidienne locale de la surface, illustrant la différence entre géométrie euclidienne (locale) et géométrie riemannienne (globale) (figure tirée de [Wikipédia, 2007]).

Le cadre riemannien offre des avantages en surmontant les limites des espaces euclidiens et en introduisant de nouvelles métriques et distances [Barachant et al., 2011]. Par exemple, les métriques affines invariantes, c'est-à-dire des métriques riemanniennes invariantes par transformations affines, ont montré de très bonnes performances dans les paradigmes SSVEP [Chevallier et al., 2021]. Toutefois, l'un des inconvénients des approches riemanniennes est que les matrices SPD ne peuvent pas être directement utilisées par les classifieurs conçus pour des espaces euclidiens, tels que la LR et la SVM.

Sur la variété riemannienne, chaque matrice de covariance correspond à un point $\mathbf{P}_i$. La matrice de covariance peut être estimée à l'aide de l'estimateur le plus couramment utilisé, la **matrice de covariance de l'échantillon** (en anglais *Sample Covariance Matrix* ou SCM) [Barachant et al., 2011],

$$\mathbf{P}_i = \frac{1}{N-1}\mathbf{X}_i(\mathbf{X}_i)^\top, \tag{3.28}$$

où $\mathbf{X}_i \in \mathbb{R}^{C\times N_i}$ correspond à l'*epoch* de la i-ième répétition. Cette estimation est simple et rapide à calculer [Kalunga et al., 2016]. La SCM constitue un estimateur non biaisé de la matrice de covariance à condition que le nombre d'échantillons (N_i) soit significativement supérieur au nombre de canaux (C) [Barachant et al., 2011]. Lorsque $C \approx N_i$, la SCM peut fournir une estimation peu fiable de la covariance [Kalunga et al., 2016].

Etant donné un point de référence $\mathbf{P} \in \mathcal{P}(C)$, généralement la moyenne des matrices de covariance [Tibermacine et al., 2024], chaque point $\mathbf{P}_i \in \mathcal{P}(C)$ peut être associé à un vecteur tangent $\mathbf{S}_i$ appartenant à l'espace tangent en $\mathbf{P}$, noté $\mathcal{T}_p$ (figure 3.7). Ce vecteur tangent correspond à la dérivée en $t = 0$ de la géodésique, $\gamma(t)$, correspondant au plus court chemin et reliant $\mathbf{P}$ à $\mathbf{P}_i$, tel que $\mathbf{S}_i = \dot{\gamma}(0)$.

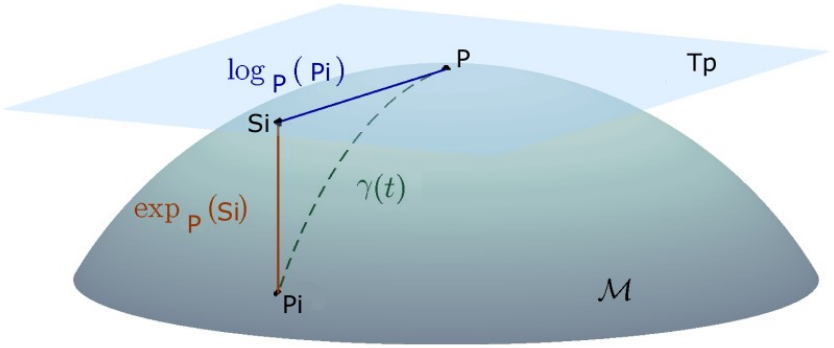


FIGURE 3.7 – Espace tangent $\mathcal{T}_p$ au point $\mathbf{P}$. $\mathbf{S}_i$ correspond à la projection de $\mathbf{P}_i$ sur l'espace tangent via l'application logarithmique riemannienne ($\text{Log}_{\mathbf{P}}$) (figure tirée de [Barachant et al., 2010b]).

Ainsi, la géodésique est définie, pour $t \in [0, 1]$ [Barachant et al., 2010b],

$$\gamma(t) = \mathbf{P}^{\frac{1}{2}}(\mathbf{P}^{-\frac{1}{2}}\mathbf{P}_i\mathbf{P}^{-\frac{1}{2}})^\top\mathbf{P}^{\frac{1}{2}}. \tag{3.29}$$

La distance riemannienne entre ces deux points correspond à la longueur de cette géodésique et s'écrit :

$$\delta(\mathbf{P}, \mathbf{P}_i) = \left[\sum_{j=1}^{C} \log^2 \lambda_j\right]^{\frac{1}{2}}, \tag{3.30}$$

où λ_j sont les valeurs propres réelles de $\mathbf{P}^{-1}\mathbf{P}_i$.

L'espace tangent ($\mathcal{T}_p$) est un espace vectoriel constitué de l'ensemble des vecteurs tangents ($\mathbf{S}_i$). La variété et l'espace tangent sont de même dimension, $\zeta = \frac{C(C+1)}{2}$. L'application logarithmique riemannienne $\log_{\mathbf{P}} : \mathcal{P}(C) \to \mathcal{T}_{\mathbf{P}}$ permet d'associer à chaque matrice $\mathbf{P}_i \in \mathcal{P}(C)$ un vecteur tangent donné par $\mathbf{S}_i = \log_{\mathbf{P}}(\mathbf{P}_i)$. Cette application s'écrit :

$$\mathbf{S}_i = \log_{\mathbf{P}}(\mathbf{P}_i) = \mathbf{P}^{\frac{1}{2}} \log(\mathbf{P}^{-\frac{1}{2}} \mathbf{P}_i \mathbf{P}^{-\frac{1}{2}}) \mathbf{P}^{\frac{1}{2}}. \tag{3.31}$$

L'ensemble des données est alors projeté dans l'espace tangent ($\mathcal{T}_p$), dans lequel la métrique est plate (euclidienne), ce qui permet d'utiliser la moyenne arithmétique, ainsi que d'autres outils classiques de la géométrie euclidienne [Barachant et al., 2010b]. Les données peuvent ensuite être projetées dans la variété riemannienne d'origine des matrices SPD ($\mathcal{P}(C)$), à l'aide de l'application exponentielle riemannienne $\exp_{\mathbf{P}} : T_{\mathbf{P}} \to \mathcal{P}(C)$ définie par :

$$\mathbf{P}_i = \exp_{\mathbf{P}}(\mathbf{S}_i) = \mathbf{P}^{\frac{1}{2}} \exp(\mathbf{P}^{-\frac{1}{2}} \mathbf{S}_i \mathbf{P}^{-\frac{1}{2}}) \mathbf{P}^{\frac{1}{2}}. \tag{3.32}$$

***Riemannian Support Vector Machine* (RiSVM)** Lorsque les données ne sont pas linéairement séparables, la SVM peut utiliser des fonctions noyaux afin de projeter les données dans un espace de dimension plus élevée dans lequel une séparation linéaire devient possible [Barachant et al., 2013b]. Cette approche repose sur l'astuce du noyau (*kernel trick*), qui consiste à appliquer une transformation non linéaire (ϕ) aux vecteurs de caractéristiques $\mathbf{x}_i$ [Yun et al., 2013]. Dans ce cas, la fonction de décision peut s'écrire :

$$h(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i y_i \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}) \rangle_{\mathcal{H}} + \psi_0, \tag{3.33}$$

où α sont les multiplicateurs de Lagrange associés au problème d'optimisation dual et $\mathcal{H}$, un RKHS (cf. section 3.4.2.1).

Dans le cas des matrices de covariance SPD, les données appartenant à une variété riemannienne plutôt qu'à un espace euclidien classique, la RiSVM utilise un noyau riemannien permettant de mesurer la similarité entre matrices de covariance tout en respectant leur géométrie intrinsèque [Barachant et al., 2013b]. Pour cela, les matrices de covariance $\mathbf{P}_i$ sont projetées dans l'espace tangent en un point de référence $\mathbf{P}$ à l'aide de l'application logarithmique riemannienne : $\phi(\mathbf{P}_i) = \log_{\mathbf{P}}(\mathbf{P}_i)$. Cette transformation permet de définir un noyau basé sur le produit scalaire dans l'espace tangent :

$$\begin{aligned} \kappa(\mathbf{P}_i, \mathbf{P}_j) &= \langle \phi(\mathbf{P}_i), \phi(\mathbf{P}_j) \rangle_{\mathbf{P}} \\ &= \mathrm{Tr}[\log(\mathbf{P}^{-\frac{1}{2}} \mathbf{P}_i \mathbf{P}^{-\frac{1}{2}}) \log(\mathbf{P}^{-\frac{1}{2}} \mathbf{P}_j \mathbf{P}^{-\frac{1}{2}})], \end{aligned} \tag{3.34}$$

où Tr désigne la trace de la matrice.

***Minimum-Distance-to-Mean* (MDM)** Un classifieur tel que le MDM est construit en deux étapes.

Dans un premier temps, la phase d'entraînement consiste à estimer la matrice de covariance $\mathbf{P}_i$ pour chacune des *epochs* $\mathbf{X}_i$. Pour chaque fréquence de stimulation, correspondant à une classe k, la moyenne riemannienne des matrices de covariance est calculée. Contrairement à la moyenne euclidienne, qui consiste à effectuer une moyenne terme à terme des matrices de covariance, la moyenne riemannienne est définit comme la matrice $\mathbf{P}_k$ minimisant la somme des distances riemanniennes aux matrices observées $\mathbf{P}_i$. Elle prend ainsi en compte la géométrie non euclidienne de l'espace des matrices SPD et garantit que la matrice moyenne reste définie positive,

$$\mathbf{P}_k = \arg\min_{\mathbf{P}} \sum_{i=1}^{I_k} \delta^2(\mathbf{P}, \mathbf{P}_i), \tag{3.35}$$

où I_k désigne le nombre de matrices de covariance appartenant à la classe k et $\delta(.,.)$ la distance riemannienne (équation 3.30). Estimer une matrice de covariance pour chaque *epoch*, puis calculer la moyenne riemannienne par classe, permet d'obtenir une représentation plus robuste que l'estimation d'une unique matrice de covariance à partir de l'ensemble des données.

Pour classer une nouvelle *epoch* $\mathbf{X}_x$, où x désigne l'indice de l'*epoch* considérée, sa matrice de covariance $\mathbf{P}_x$ est d'abord estimée. Elle est ensuite comparée aux moyennes riemanniennes de chaque classe ($\mathbf{P}_k$, avec k allant de 1 à 4 par exemple, figure 3.8), en utilisant la distance riemannienne (équation 3.30) [Yger et al., 2016]. L'*epoch* est alors affectée à la classe pour laquelle cette distance est minimale. La classe estimée est ainsi définie par :

$$\hat{k} = \arg\min_k \delta^2(\mathbf{P}_x, \mathbf{P}_k). \tag{3.36}$$

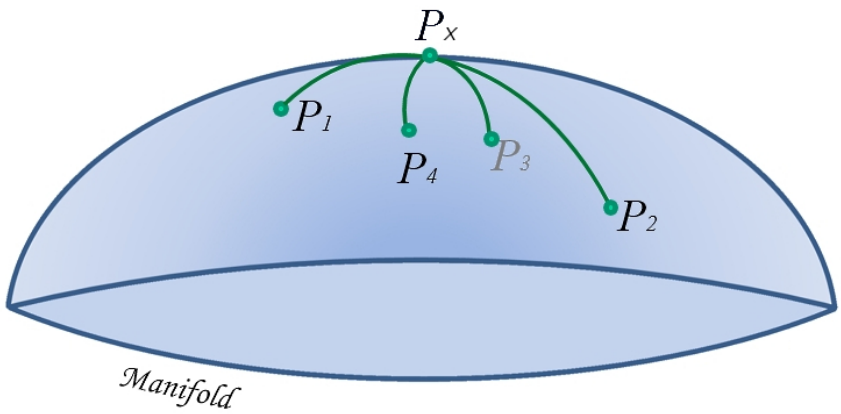


FIGURE 3.8 – La matrice de covariance $\mathbf{P}_x$ de la nouvelle *epoch* $\mathbf{X}_x$ est comparée aux moyennes riemanniennes $\mathbf{P}_k$ de chaque classe $k = 1, \ldots, 4$ à l'aide de la distance riemannienne (figure tirée de [Bleuzé et al., 2021]).

***Fisher Geodesic Minimum-Distance-to-Mean* (FGMDM)** L'un des principaux inconvénients du MDM est qu'il ne prend pas en compte la séparabilité des classes [Kumar et al., 2019]. Afin de pallier cette limitation, le FGMDM repose sur le même principe que le MDM, mais introduit une étape supplémentaire visant à améliorer cette séparabilité à l'aide d'une projection dans l'espace tangent [Barachant et al., 2010b].

Durant la phase d'entraînement, toutes les matrices de covariance d'entraînement sont projetées dans l'espace tangent défini au point de la moyenne riemannienne globale P, calculée sur l'ensemble des données d'entraînement, toutes classes confondues. Une matrice de projection ($\mathbf{W}_{\text{LDA}}$) est ensuite estimée à l'aide d'une LDA, afin de maximiser la séparation entre classes dans cet espace.

Lors de la phase de test (figure 3.9), les matrices de covariance moyennes pour chaque classe ($\mathbf{P}_k$), ainsi que la matrice de covariance d'une nouvelle *epoch* ($\mathbf{P}_x$), sont projetées dans l'espace tangent, puis transformées par la matrice $\mathbf{W}_{\text{LDA}}$, afin d'améliorer la séparabilité inter-classe. Les matrices ainsi filtrées sont ensuite reprojetées sur la variété riemannienne ($\mathcal{M}$) à l'aide de l'application exponentielle (équation 3.32). Enfin, l'algorithme MDM est appliqué à ces matrices de covariance transformées pour effectuer la classification.

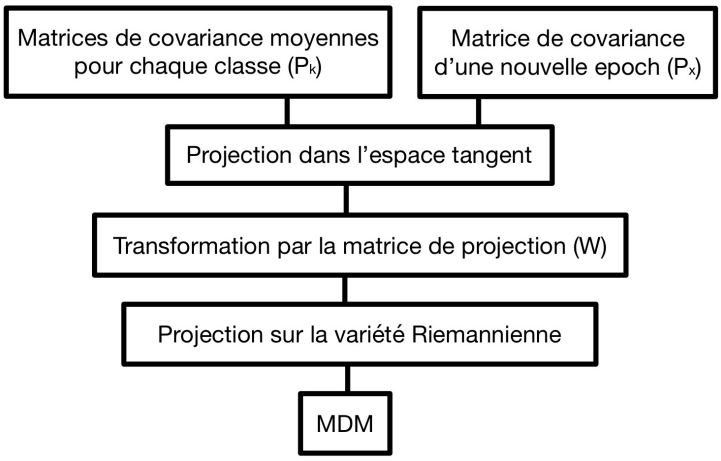


FIGURE 3.9 – Phase de test de la méthode FgMDM.

Rejet des artefacts par la méthode *Riemannian potato* Afin d'obtenir une estimation plus robuste de la moyenne des matrices de covariance pour chaque classe ($\mathbf{P}_k$), les artefacts peuvent être supprimés à l'aide de la méthode dite de la **patate riemannienne** (*Riemannian potato*) [Barachant et al., 2013a; Chevallier et al., 2021].

Le principe consiste à estimer une matrice de covariance de référence ($\mathbf{P}_r$), puis à rejeter chaque *epoch* ($\mathbf{X}_y$) dont la matrice de covariance ($\mathbf{P}_y$) est trop éloignée de cette référence au sens de la distance riemannienne [Barachant et al., 2010b]. La matrice de référence ($\mathbf{P}_r$) est généralement définie comme la moyenne riemannienne des matrices de covariance d'apprentissage, calculée soit globalement, soit pour une classe donnée k (notée alors $\mathbf{P}_k$). Une *epoch* $\mathbf{X}_y$ est considérée comme un artefact si la distance riemannienne $\delta(\mathbf{P}_y, \mathbf{P}_r)$, dépasse un seuil ($th$) prédéfini. Ce seuil est estimé à partir de la moyenne (μ) et de l'écart type (σ) des distances riemanniennes observées :

$$th = \mu + 2{,}5\sigma. \tag{3.37}$$

Les paramètres μ et σ définissent ainsi une région d'intérêt sur la variété riemannienne, centrée sur la matrice de référence ($\mathbf{P}_r$) et délimitée par un seuil sur la distance riemannienne (th). En raison de la géométrie non euclidienne de l'espace des matrices SPD, cette région ne correspond pas à une sphère euclidienne, mais adopte une forme irrégulière de "patate". Par conséquent, si une *epoch* se situe en dehors de cette région, elle est considérée comme un artefact et est rejetée [Barthélemy et al., 2019].

3.5.2.5 Méthodes d'apprentissage profond (*Deep Learning* ou DL)

Les méthodes de DL reposent sur les réseaux de neurones artificiels, qui sont des modèles computationnels inspirés du fonctionnement du système nerveux humain [Islam et al., 2019]. Un réseau de neurones artificiel est constitué d'un ensemble de nœuds de calcul interconnectés, appelés **neurones**, organisés en couches successives [O'shea et Nash, 2015] (figure 3.10).

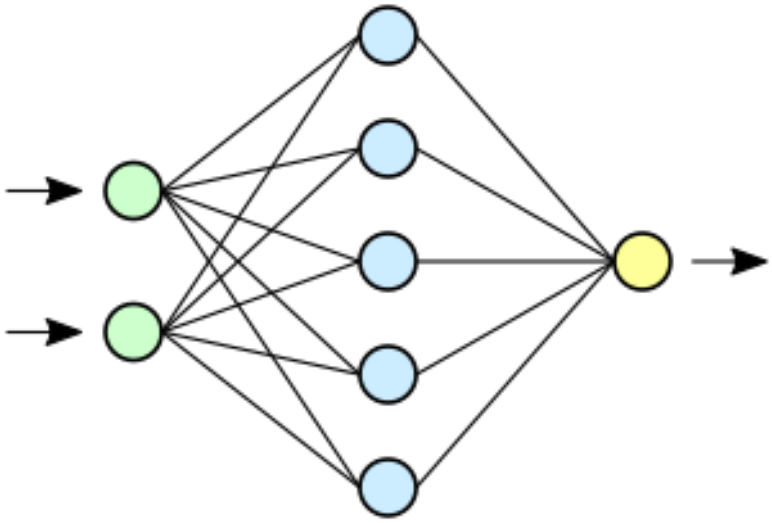


FIGURE 3.10 – Exemple d'un réseau de neurones artificiel à trois couches : une couche d'entrée (verte), une couche cachée (bleue) et une couche de sortie (jaune). Les cercles représentent les neurones, et les connexions entre eux correspondent aux poids du modèle (figure tirée de [Wikipédia, 2006b]).

Les données d'entrée, généralement représentées sous forme de tenseurs (i.e. vecteurs, matrices ou structures de dimension supérieure), sont injectées dans une couche d'entrée, puis propagées à travers une ou plusieurs couches cachées jusqu'à une couche de sortie. Chaque connexion entre neurones est associée à un poids ajustable, tandis que chaque neurone est caractérisé par un biais. Chaque neurone calcule une somme pondérée de ses entrées, à laquelle est ajouté le biais, puis applique une fonction d'activation au résultat obtenu [Damien et al., 2023]. Les fonctions d'activation introduisent de la non-linéarité dans le modèle [André, 2025] et déterminent, par exemple avec la fonction d'unité linéaire rectifiée (*Rectified Linear Unit* ou ReLU, nulle lorsque la valeur d'entrée est négative et égale à l'identité sinon), si l'information doit être transmise aux neurones des couches suivantes. Le processus d'apprentissage consiste ensuite à ajuster de manière itérative les poids et les biais du réseau, à l'aide de méthodes d'optimisation, afin de minimiser une fonction de coût qui mesure l'écart entre les prédictions du modèle et les valeurs réelles attendues [Damien et al., 2023]. Ainsi, les réseaux de neurones permettent d'apprendre automatiquement des représentations des données (i.e. des transformations des données d'entrée en caractéristiques plus pertinentes pour la tâche considérée), réduisant le besoin de définir manuellement ces caractéristiques [Schirrmeister et al., 2017]. Cela ne dispense toutefois pas d'une étape de préparation des données, qui demeure essentielle.

Le DL correspond à une extension de ces modèles fondée sur l'utilisation de multiples couches cachées [O'shea et Nash, 2015]. Cette profondeur permet d'apprendre progressivement des représentations de plus en plus abstraites, chaque couche construisant des caractéristiques plus complexes à partir de celles de la précédente. En empilant un nombre suffisant de transformations non linéaires, ces modèles peuvent approximer des fonctions hautement complexes. Dans les tâches de classification, les couches profondes tendent à mettre en évidence les caractéristiques discriminantes tout en atténuant l'influence des variations non pertinentes [LeCun et al., 2015].

Parmi ces modèles, les CNN occupent une place centrale. Inspirés en partie des mécanismes de la perception visuelle [Gu et al., 2018], ils sont spécifiquement conçus pour exploiter les dépen-

dances locales présentes dans des données structurées, telles que les images ou les signaux, où les observations voisines dans l'espace ou le temps sont fortement corrélées [LeCun et al., 2015]. Les CNN sont principalement constitués de trois types de couches (figure 3.11) : les couches de convolution, les couches de sous-échantillonnage (*pooling*) et les couches entièrement connectées (*Fully Connected* ou FC) [O'shea et Nash, 2015]. Les couches de convolution consistent à appliquer des filtres (ou noyaux) à des sous-parties de taille limitée de l'entrée, considérées successivement sur l'ensemble des données. Les paramètres de ces filtres étant partagés sur l'ensemble des données, les CNN peuvent détecter des structures caractéristiques présentes dans des portions localisées des données, tout en limitant le nombre de paramètres à apprendre. Les couches de sous échantillonnage effectuent ensuite une réduction de la dimension spatiale des représentations obtenues, ce qui permet de diminuer le nombre de paramètres et la complexité du modèle. Enfin, les couches entièrement connectées remplissent un rôle similaire à celui des réseaux de neurones classiques, en exploitant les caractéristiques extraites afin de déterminer la classe associée aux données d'entrée. Ces couches principales sont généralement associées à des fonctions d'activation, telles que ReLU ou l'unité linéaire exponentielle (*Exponential Linear Unit* ou ELU). Selon les architectures, elles peuvent également être complétées par une couche de *BatchNormalization*, destinée à stabiliser l'apprentissage, ainsi que par une couche de *Dropout*, qui désactive aléatoirement une fraction des neurones à chaque passe d'entraînement, forçant le réseau à apprendre des représentations redondantes, à des fins de régularisation. Ainsi, les CNN permettent de combiner, dans un même cadre d'optimisation, le filtrage spatial et temporel, ainsi que l'approximation de relations non linéaires [Lawhern et al., 2018]. Cette capacité rend les CNN particulièrement adaptés à l'analyse de signaux, notamment pour la classification des signaux EEG.

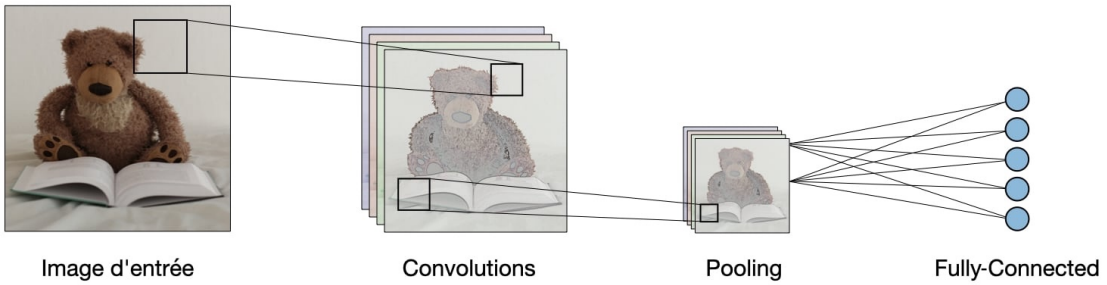


FIGURE 3.11 – Illustration des principaux types de couches d'un réseau de neurones convolutif : convolution, *pooling* (sous-échantillonnage) et *fully connected* (entièrement connectée) (figure tirée de [Shervine Amidi, nd]).

Toutefois, les modèles de DL requièrent davantage de données que les méthodes de filtrage spatial et les méthodes riemanniennes, ce qui peut constituer une contrainte dans le domaine de l'EEG [Lashgari et al., 2020]. Cette limitation s'explique notamment en raison du nombre restreint de sujets, de la forte variabilité inter-sujet, du nombre limité d'enregistrements disponibles, du faible RSB et de la non-stationnarité des signaux EEG. Dans ce contexte, ces modèles sont également plus coûteux en calcul et plus difficiles à entraîner. Ils comportent en outre un grand nombre de paramètres à optimiser, souvent de l'ordre de plusieurs milliers. La sélection des hyperparamètres revêt ainsi une importance particulière afin de garantir la stabilité de l'apprentissage et de limiter le surapprentissage, phénomène particulièrement marqué dans le contexte des données EEG. Dans ce contexte, différentes méthodes de DL appliquées aux signaux EEG dans le cadre des BCI sont présentées ci-après.

EEGNet L'EEGNet est un CNN compact largement utilisé pour la classification et l'interprétation de signaux EEG pour les BCI [Lawhern et al., 2018]. Il s'agit de l'une des architectures les plus populaires dans ce domaine (figure 3.12). Ce modèle peut être entraîné efficacement même avec des ensembles de données relativement limités [Saideepthi et al., 2023]. L'architecture EEGNet comporte trois étapes principales d'extraction de caractéristiques en amont de la phase de classification.

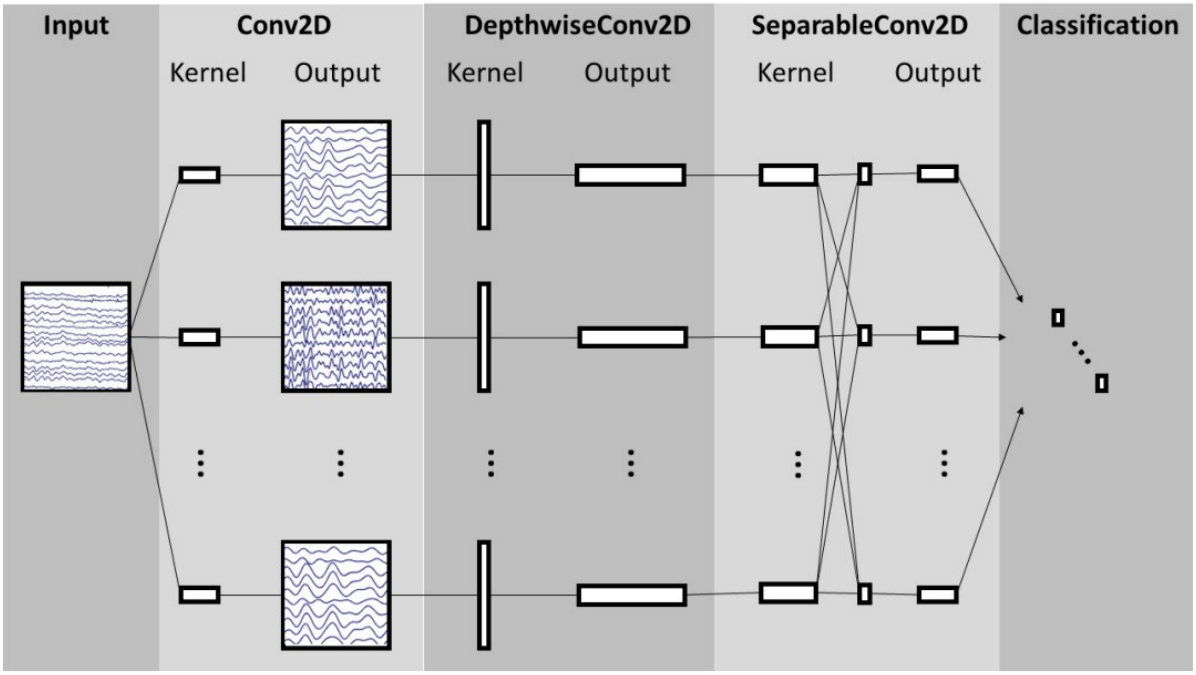


FIGURE 3.12 – Architecture du modèle EEGNet (figure tirée de [Lawhern et al., 2018]).

La première étape consiste en une convolution temporelle agissant comme un banc de filtres passe-bande, permettant d'extraire les composantes des signaux EEG dans différentes bandes de fréquences.

Elle est suivie d'une deuxième étape reposant sur une convolution spatiale en profondeur (*depthwise convolution*) [Lawhern et al., 2018], où un filtre convolutionnel distinct est appliqué aux canaux EEG, indépendamment pour chaque carte de caractéristiques (*feature maps*, correspondant aux sorties des filtres convolutifs) issue de la convolution temporelle, sans mélange d'information entre ces cartes [Guo et al., 2019]. Cette opération permet ainsi, pour chaque filtre temporel, d'apprendre des filtres spatiaux et d'extraire des caractéristiques spatiales spécifiques à chaque bande de fréquence. Chaque filtre spatial correspond à une transformation linéaire combinant l'ensemble des canaux en une seule caractéristique [Deng et al., 2021]. Cette séquence convolutive en deux étapes s'inspire en partie de l'algorithme *Filter-Bank Common Spatial Pattern* (FBCSP) [Ang et al., 2012], qui utilise un banc de filtres pour décomposer les signaux EEG en plusieurs bandes de fréquences avant d'appliquer l'algorithme CSP, afin de calculer les filtres spatiaux [Deng et al., 2021].

La troisième étape consiste en une convolution séparable, qui combine une convolution en profondeur et une convolution en point (*pointwise convolution*) [Chollet, 2017]. La convolution en profondeur permet d'extraire l'information temporelle pour chaque carte de caractéristiques séparément, tandis que la convolution en point combine ensuite ces différentes cartes de carac-

téristiques de manière optimale. Les convolutions séparables offrent plusieurs avantages. Elles permettent notamment de réduire le nombre de paramètres à ajuster, rendant le modèle plus simple et plus rapide à entraîner [Saideepthi et al., 2023]. Elles permettent également de distinguer deux types de relations : celles qui existent au sein d'une même carte de caractéristiques et celles qui existent entre les différentes cartes de caractéristiques. Dans le cas des signaux EEG, cette opération permet ainsi de traiter séparément l'évolution temporelle de chaque carte de caractéristiques, ainsi que la manière dont ces cartes se combinent pour représenter l'information [Lawhern et al., 2018].

Trois types de caractéristiques sont utilisés en entrée du classifieur [Saideepthi et al., 2023], permettant une classification précise : les caractéristiques temporelles, les caractéristiques fréquentielles et enfin, les caractéristiques spatiales. Une fois ces différentes caractéristiques extraites, le modèle les combine afin de produire des probabilités correspondant aux différentes classes. La décision finale est ensuite obtenue à l'aide d'une couche entièrement connectée utilisant la fonction *softmax*, qui transforme ces probabilités en une prédiction de classe.

Les idées principales d'EEGNet sont similaires à celles de l'algorithme FBCSP. Ainsi, l'EEGNet peut être considéré comme une évolution, dans le cadre du DL, d'un algorithme classique [Deng et al., 2021]. L'EEGNet utilise des convolutions séparables et en profondeur, afin de construire un modèle EEG explicite intégrant efficacement des concepts bien établis d'extraction de caractéristiques pour les BCI [Saideepthi et al., 2023]. Ce modèle est suffisamment robuste pour apprendre une large gamme de caractéristiques interprétables dans différentes tâches de BCI. Toutefois, son architecture compacte et flexible [Wang et al., 2020] peut avoir pour contrepartie une précision, définie comme la proportion de prédictions correctement classées parmi l'ensemble des prédictions réalisées, inférieure à celle du ShallowConvNet (SCN) pour les paradigmes oscillatoires, tels que les SSVEP.

ShallowConvNet (SCN) L'architecture SCN a été spécialement conçue pour la classification de signaux oscillatoires, notamment en extrayant des caractéristiques liées à la puissance dans certaines bandes de fréquences [Schirrmeister et al., 2017], d'où son intérêt pour les signaux EEG en réponse à des stimulations périodiques. Comme l'EEGNet, elle s'inspire de l'architecture du FBCSP, mais intègre l'ensemble des étapes de calcul au sein d'un seul réseau, ce qui permet d'optimiser le processus de manière conjointe.

Ainsi, les deux premières couches du SCN effectuent une convolution temporelle et un filtrage spatial, comme dans la méthode FBCSP [Schirrmeister et al., 2017] (figure 3.13). Ces couches sont ensuite suivies d'une mise au carré (*squaring nonlinearity*), afin de calculer la puissance des signaux, d'un moyennage (*mean pooling*) et d'une fonction d'activation logarithmique (*logarithmic activation function*). Ensemble, ces étapes sont analogues au calcul du logarithme de la variance des signaux filtrés spatialement dans la méthode FBCSP. L'application du logarithme permet d'obtenir des caractéristiques dont la distribution est plus proche d'une distribution normale. La classification est finalement réalisée à l'aide d'une couche entièrement connectée (*dense layer*), suivie d'une fonction *softmax* permettant d'estimer la probabilité d'appartenance des données d'entrée aux différentes classes.

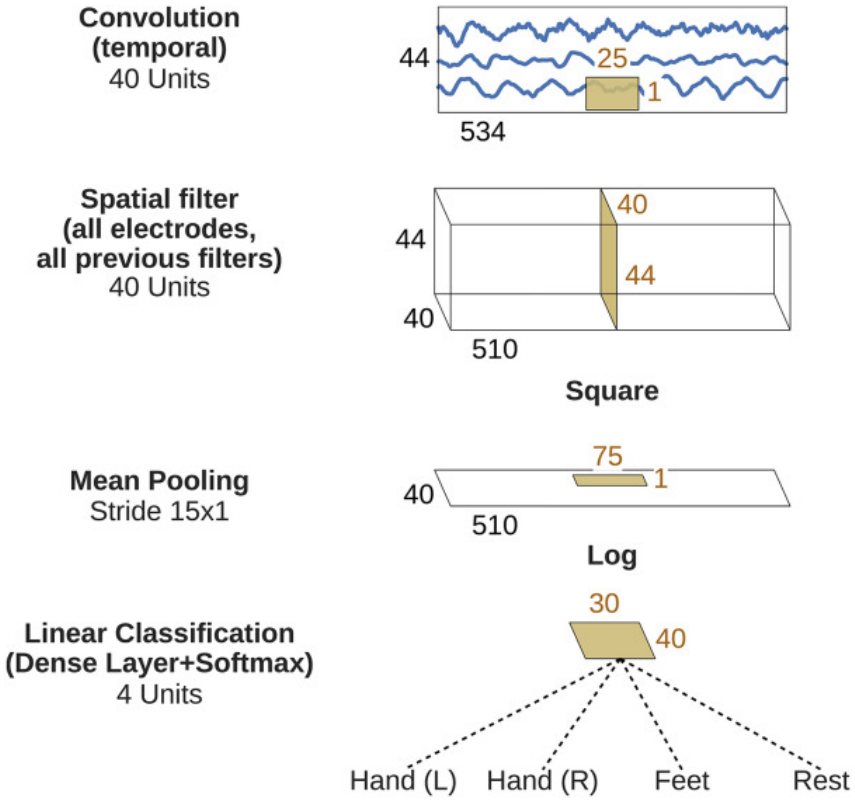


FIGURE 3.13 – Architecture du modèle SCN (figure tirée de [Schirrmeister et al., 2017]).

***Gabor Riemannian* EEGNet (GREEN)** Le GREEN combine l'extraction de caractéristiques temps-fréquence à l'aide d'un banc de filtres de Gabor, de transformations géométriques et de couches entièrement connectées, au sein d'un réseau de neurones (figure 3.14) [Paillard et al., 2024, 2025]. Cette architecture vise à surmonter les limites des applications actuelles du DL à l'EEG.

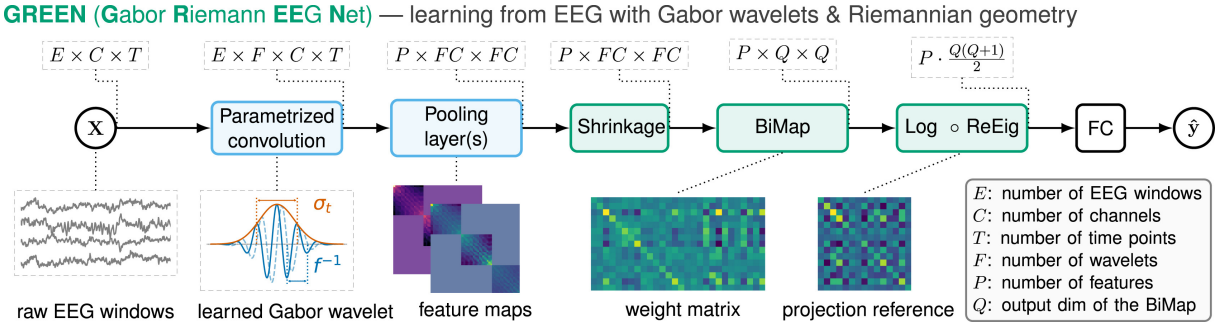


FIGURE 3.14 – Architecture du modèle GREEN (figure tirée de [Paillard et al., 2024]).

La première étape de cette architecture correspond à une couche de convolution paramétrée (*parametrized convolution*). Elle consiste à extraire des composantes temporelles et fréquentielles similaires à une transformée en ondelettes, réalisée par l'application d'une banque de filtres de Gabor (i.e. des filtres passe-bande à réponse impulsionnelle sinusoïdale modulée par une fonction gaussienne).

Le signal transformé par les filtres de Gabor est ensuite transmis à une couche de sous-échantillonnage, de manière à réduire la dimension et à résumer l'information contenue dans les cartes de caractéristiques. La matrice de covariance d'échantillon est calculée à partir de ces signaux transformés. Lorsque le nombre d'échantillons est limité par rapport au nombre de canaux, les matrices de covariance d'échantillon sont connues pour surestimer la dispersion des valeurs propres, ce qui peut entraîner des instabilités numériques et dégrader les performances des algorithmes de classification [Ledoit et Wolf, 2004]. Ainsi, une opération de régularisation (*shrinkage*) est appliquée. L'intensité de cette régularisation est contrôlée par un paramètre appris automatiquement par le réseau.

Les couches suivantes de projection bilinéaire et de rectification des valeurs propres permettent d'effectuer des calculs sur la variété riemannienne, incluant le filtrage spatial, la réduction de dimension et l'introduction de non-linéarités sur les valeurs propres. Ces couches sont issues du réseau SPDNet [Huang et Van Gool, 2017]. La couche de transformation bilinéaire (*Bi-linear Mapping layer* ou *BiMap layer*) est utilisée afin de projeter les données dans un sous-espace de dimension réduite tout en préservant les propriétés des matrices SPD. Cette étape est similaire aux approches classiques de filtrage spatial. La couche de rectification des valeurs propres (*Rectified Eigenvalues layer* ou *ReEig layer*) introduit une non-linéarité via un seuil minimal sur les valeurs propres, garantissant ainsi la stabilité numérique pendant l'entraînement. Enfin, l'application de la carte logarithmique (*Logarithmic Map* ou *LogMap*) projette les matrices SPD de leur variété riemannienne vers un espace tangent euclidien, permettant ainsi d'effectuer des calculs euclidiens dans le domaine logarithmique [Arsigny et al., 2005]. Cette approche, fondée sur la métrique log-euclidienne, fournit des résultats proches de ceux obtenus avec la métrique affine invariante, tout en présentant un coût de calcul plus faible [Thanwerdas et Pennec, 2019].

Finalement, une ou plusieurs couches entièrement connectées peuvent être ajoutées en sortie du réseau. Pour les tâches de régression et de classification, les fonctions de perte d'erreur quadratique moyenne et d'entropie croisée sont respectivement utilisées. Ces fonctions permettent de mesurer l'écart entre les prédictions du modèle et les valeurs attendues : l'erreur quadratique moyenne évalue l'écart moyen entre les valeurs prédites et réelles, tandis que l'entropie croisée mesure la différence entre les probabilités prédites par le modèle et celles des classes réelles.

3.5.2.6 Méthodes d'évaluation des performances de classification

Dans le cadre des BCI, différentes méthodes d'évaluation existent pour répartir les données en ensembles d'apprentissage et de test, et évaluer les performances de la BCI [Chevallier et al., 2024] : l'évaluation intra-session (*Within-Session* ou WS), l'évaluation inter-session (*Cross-Session*) et l'évaluation inter-sujet (*Cross-Subject* ou CS).

L'évaluation WS vise à mesurer les performances du modèle pour un même sujet (figure 3.15). Cette stratégie peut prendre la forme d'une validation croisée en β blocs, dans laquelle les *epochs* sont divisées en β sous-ensembles (i.e. blocs) de taille similaire. À chaque itération, $\beta - 1$ blocs sont utilisés pour l'apprentissage tandis que le bloc restant est utilisé pour le test. L'opération est répétée β fois, chaque bloc étant utilisé une fois comme ensemble de test. La performance finale du modèle correspond alors à la moyenne des performances obtenues sur l'ensemble de ces partitions. Il s'agit de la stratégie d'évaluation la plus couramment utilisée dans la littérature [Chevallier et al., 2024]. Toutefois, elle présente certaines limitations, car elle ne prend que partiellement en compte les variations entre sujets et entre sessions, telles que de légers déplacements des électrodes entre sessions ou des changements de l'état cognitif de l'utilisateur.

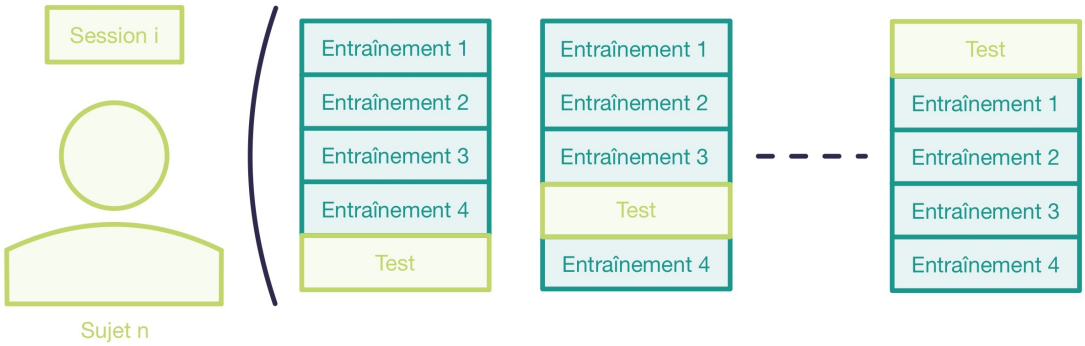


FIGURE 3.15 – Evaluation WS par validation croisée à $\beta = 5$ blocs, pour chaque sujet (n) et chaque session (i). À chaque itération, 4 blocs sont utilisés pour l'apprentissage, tandis que le bloc restant est utilisé pour le test. La performance finale du modèle correspond à la moyenne des performances obtenues sur l'ensemble des itérations.

L'évaluation inter-sujet (*cross-subject* ou CS) ou inter-session (*cross-session*) correspond à une approche *Leave-One-Out* (LOO) ou plus précisément, LOSO pour *Leave-One-Subject-Out* et *Leave-One-Session-Out*. Pour l'évaluation inter-sujet, les données de tous les sujets, à l'exception d'un seul, sont utilisées pour l'entraînement, tandis que les données du sujet exclu servent exclusivement pour le test (figure 3.16). Le principe est similaire pour l'évaluation inter-session, dans laquelle une session est utilisée comme ensemble de test, les autres servant à l'apprentissage. Les évaluations *cross-subject* et *cross-session* mettent davantage l'accent sur les approches de *transfert learning*, qui consistent à exploiter les données issues d'autres sessions ou d'autres sujets, afin de faire face à la variabilité inter-sujet et inter-session. En particulier, l'évaluation CS permet d'évaluer la capacité du modèle à généraliser à un individu non inclus dans l'ensemble d'entraînement. Elle se place ainsi dans un contexte sans phase de calibration propre à un individu, aucune donnée spécifique au sujet testé n'étant disponible pour adapter le modèle à ses caractéristiques propres [Chevallier et al., 2024].

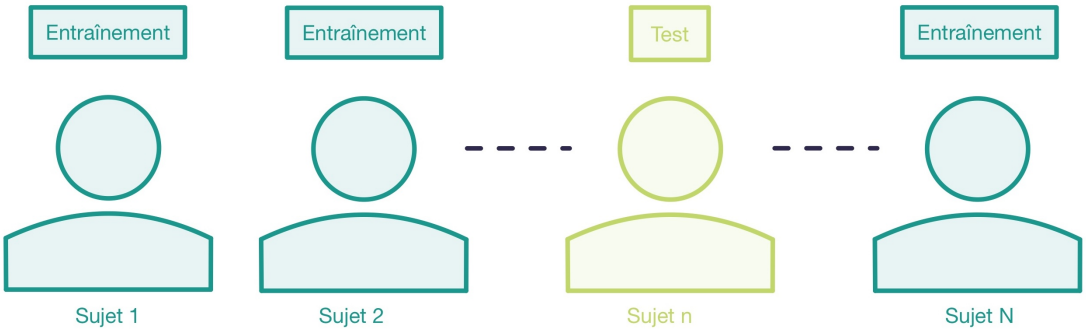


FIGURE 3.16 – Evaluation CS. Les données de tous les sujets, à l'exception d'un seul (n), sont utilisées pour l'apprentissage, tandis que les données du sujet exclu servent exclusivement au test.

3.6 Conclusion

Ce chapitre propose un panorama des principales stratégies de l'état de l'art pour l'analyse et la classification des signaux EEG dans le cadre des BCI.

Les étapes de pré-traitement, comprenant notamment le filtrage temporel et l'extraction des *epochs*, jouent un rôle essentiel dans l'amélioration de la qualité des signaux et dans l'isolation de l'activité cérébrale associée aux stimuli. Toutefois, les paramètres associés à ces traitements doivent être étudiés avec soin, afin d'identifier ceux favorisant la détectabilité de la fréquence de stimulation dans les signaux EEG enregistrés en réponse à des stimulations auditives périodiques.

Les techniques de réduction de dimension, telles que la PCA, le CSP et la LDA, visent quant à elles à extraire des caractéristiques pertinentes des données EEG, tout en réduisant la dimension de leur espace de représentation. Elles permettent ainsi de limiter les problèmes liés à la grande dimension de cet espace, tout en améliorant les performances de classification.

Le calcul du RSB constitue une première mesure de détectabilité des fréquences de stimulation et fournit une indication sur la qualité du signal, susceptible d'influencer les performances de classification. Cette mesure peut être réalisée aux fréquences de stimulation, ainsi qu'à leurs harmoniques, afin d'analyser leur présence dans les signaux cérébraux et d'évaluer la pertinence de leur prise en compte pour la classification. Les approches de classification habituellement utilisées sur des données EEG incluent des méthodes classiques d'apprentissage automatique, des approches basées sur la géométrie riemannienne, ainsi que des architectures d'apprentissage profond adaptées aux signaux EEG. Bien que ces méthodes aient été largement étudiées dans le cadre de paradigmes SSVEP, leur application aux données issues de paradigmes SSAEP demeure encore peu explorée.

Enfin, dans le contexte des BCI, trois stratégies d'évaluation des performances de classification sont généralement considérées : les approches intra-session, inter-session et inter-sujet. Ces différentes approches permettent d'évaluer la robustesse et la capacité de généralisation des méthodes proposées face aux variations pouvant apparaître entre sessions ou entre sujets.

Deuxième partie

Contributions

4

A la recherche d'alternatives aux sons purs : mise en place d'une expérience

Sommaire

4.1 Introduction 69
4.1.1 Deux préoccupations originales : les facteurs acoustiques et les facteurs humains 70
4.2 Travaux antérieurs 70
4.2.1 Paradigme expérimental 73
4.2.2 Stimuli auditifs 73
4.2.3 Dispositifs matériels 74
4.2.4 Participants 75
4.2.5 Classification 75
4.2.6 Performances 75
4.3 Protocole expérimental 76
4.3.1 Objectifs et hypothèses 76
4.3.2 Paradigme expérimental 76
4.3.3 Stimuli auditifs 77
4.3.4 Collecte de métadonnées 81
4.3.5 Dispositifs matériels 82
4.3.6 Participants 83
4.4 Facteurs acoustiques 83
4.4.1 Casque d'écoute 84
4.4.2 Audiomètre 88
4.5 Facteurs individuels 93
4.5.1 Questionnaires 93
4.6 Conclusion 99

4.1 Introduction

La partie précédente (partie I) met en évidence les enjeux associés à la conception d'une interface cerveau-ordinateur basée sur les SSAEP, notamment en termes de confort des stimulations sonores et de détectabilité de la fréquence de modulation.

Dans une perspective d'applications industrielles, en particulier au sein de l'entreprise Orange, l'un des cas d'usage envisagé est celui du technicien augmenté. Une BCI pourrait ainsi permettre à un technicien de contrôler certains dispositifs sans mobilisation des mains, par exemple lors d'interventions en hauteur sur des infrastructures télécom. Toutefois, de telles situations imposent plusieurs contraintes : le système doit être peu intrusif, reposer sur des dispositifs simples, et offrir un niveau de performance et d'acceptabilité compatibles avec un usage en conditions réelles. En pratique, le technicien porte déjà un casque de protection et est soumis à une charge visuelle importante, liée à son environnement de travail. Dans ce contexte, l'ajout d'une modalité visuelle, notamment via un dispositif de réalité étendue, apparaît dès lors peu adapté, voire incompatible avec ces contraintes opérationnelles. Le recours à des stimuli auditifs constitue alors une alternative pertinente. Une expérimentation a donc été mise en place afin d'évaluer l'influence du signal porteur sur les réponses cérébrales et les performances de classification, dans l'objectif d'identifier des compromis entre confort auditif et efficacité pour les BCI auditives.

4.1.1 Deux préoccupations originales : les facteurs acoustiques et les facteurs humains

Lors de la conception de cette expérimentation, deux aspects ont fait l'objet d'une attention particulière. Premièrement, il nous a semblé essentiel de garantir une reproduction fidèle des stimuli auditifs, ainsi qu'une perception correcte de leur niveau sonore par les participants, indépendamment de leur contenu fréquentiel. L'expérimentation a également été conçue de manière à en assurer la reproductibilité. Ainsi, chaque composant de la chaîne de reproduction sonore a été soigneusement sélectionné et caractérisé, notamment par la mesure de sa réponse en fréquence à l'aide d'un HATS [Brüel & Kjaer, 2021]. Par ailleurs, les capacités auditives des participants ont été évaluées, de manière à détecter d'éventuelles déficiences auditives majeures. À cette fin, un test auditif spécifique a été développé, afin de conserver une procédure expérimentale la plus simple possible, sans recourir à une audiométrie dans des structures médicales spécialisées. Deuxièmement, divers facteurs individuels, incluant les caractéristiques démographiques, la latéralité, la pratique musicale ou encore la consommation de substances psychoactives, sont susceptibles d'influencer les processus cognitifs, et par conséquent, les signaux EEG collectés. Afin de prendre en compte cette variabilité interindividuelle, des questionnaires ont été élaborés pour recueillir des informations détaillées sur chaque participant. Ces données visent à mieux caractériser les profils individuels et à faciliter l'interprétation des résultats expérimentaux.

Ce chapitre présente l'ensemble des éléments relatifs au protocole expérimental construit selon ces deux principales préoccupations. La section 4.2 synthétise l'état de l'art sur les BCI basées sur les SSAEP, en détaillant les paramètres expérimentaux utilisés. Le protocole et les paramètres expérimentaux retenus sont ensuite présentés en section 4.3. Les procédures liées aux facteurs acoustiques sont décrites dans la section 4.4, incluant les mesures électroacoustiques et la correction du matériel de reproduction, la construction des stimuli auditifs, la mesure de la sonie, ainsi que la procédure audiométrique. Enfin, la section 4.5 détaille les questionnaires élaborés pour recenser les facteurs humains susceptibles d'influencer l'activité cérébrale.

4.2 Travaux antérieurs

Les études portant sur les BCI basées sur les SSAEP se caractérisent par une grande disparité des paramètres expérimentaux utilisés (tableaux 4.1 et 4.2), ce qui limite considérablement la possibilité de comparer directement les performances rapportées. Cette diversité concerne notamment la nature des signaux porteurs, les fréquences de modulation, la durée des stimulations,

ainsi que les dispositifs de restitution des stimuli et d'enregistrement des signaux cérébraux. Elle s'étend également aux protocoles expérimentaux, avec des différences marquées dans leur mise en œuvre, notamment dans le nombre d'essais, de *runs* et de sessions.

Référence	**Paradigme expérimental**	**Protocole expérimental**	**Signaux porteurs**	**Fréquences de modulation (Hz)**	**Durée (s)**
[Lopez et al., 2009]	Présentation dichotique de deux stimuli auditifs avec tâche d'attention spatiale (G/D/aucune)	• 2 sessions • 10 essais/session	• sons purs : 1 kHz (G), 2.5 kHz (D)	38 (G), 42 (D)	44
[Xie et al., 2021]	Paradigme hybride (SSAEP/SSVEP) : présentation dichotique de deux stimuli auditifs avec tâche d'attention spatiale (G/D) et quatre stimuli visuels affichés à des positions distinctes sur un écran	• 24 *runs* • 10 essais/*run*	• sons purs : 450 Hz (G), 650 Hz (D) • visuel : 8,5, 9, 11, 13 Hz	9 (G), 13 (D)	5, intervalle interstimulus : 2 s
[Kim et al., 2011a]	Présentation dichotique de deux stimuli auditifs avec tâche d'attention spatiale (G/D)	• 2 sessions (10 minutes) • 25 essais/session	• sons purs : 2,5 kHz (G), 1 kHz (D)	37 (G), 43 (D)	20, intervalle interstimulus : $[6, 10]$ s
[Kaongoen et Jo, 2017]	Paradigme hybride (SSAEP/P300) : présentation dichotique de deux stimuli auditifs avec attention spatiale (G/D) et dix bips (brève augmentation du volume)	• 2 sessions • 10 *runs*/session • 2 essais/*run* (57 secondes)	• sons purs : 1 kHz (G), 2 kHz (D)	37 (G), 43 (D)	21, intervalle interstimulus : 5 s
[Higashi et al., 2011]	Présentation monaurale séquentielle aléatoire de deux stimuli auditifs sur l'une ou l'autre oreille avec tâche d'attention spatiale (G/D/aucune)	• 100 essais	• sons purs : 500 Hz (G/D)	35 (G), 60 (D)	10
[Heo et al., 2017]	Présentation dichotique de deux stimuli auditifs (sons purs ou musicaux ou naturels)	• 7 sessions • 50 essais/session	• sons purs : 1 kHz (G), 2 kHz (D) • sons musicaux : violon (G), piano (D) • sons naturels : cascade d'eau (G), chant des cigales (D)	38 (G), 42 (D)	20
[Harada et al., 2022]	Présentation dichotique de deux stimulations partageant le même signal porteur avec tâche d'attention spatiale (G/D/aucune)	• 180 essais	• son pur : 1 kHz (G/D) • sons naturels : ruisseau (G/D), rouge-gorge (G/D)	$[35, 45]$, par pas de 2 (adaptation individuelle), une fréquence à chaque oreille	10

TABLE 4.1 – Synthèse des études sur les BCI basées sur les SSAEP. Le terme **dichotique** désigne un mode de présentation des stimuli auditifs dans lequel les deux oreilles (gauche : G/droite : D) reçoivent des stimulations différentes, par opposition au mode **diotique**, où le même stimulus est présenté aux deux oreilles. Les nombres de sessions, de *runs* et d'essais correspondent aux totaux rapportés dans chaque étude, toutes conditions confondues. Suite dans le tableau 4.2.

Référence	**Niveau sonore**	**Matériel de reproduction audio**	**Participants**	**Mesure EEG (voir figure 2.6)**	**Classification**	**Performances (précision en %)**
[Lopez et al., 2009]	"niveau confortable"	écouteurs intra-auriculaires, modèle non précisé	• 8 : 6 hommes, 2 femmes • normo-entendants (absence d'information concernant la réalisation d'un test audiométrique)	• Cz, Oz (2 canaux) • référence : mastoïde droite	SOM	–
[Xie et al., 2021]	"30% du volume maximal du casque"	casque, modèle non précisé	• 6	• SSAEP : T7, TP7, P7, T8, TP8, P8 (6 canaux) • référence : lobe de l'oreille (unilatéral) • masse : FPz	CCA	73, 3
[Kim et al., 2011a]	–	haut-parleurs latéraux (à maximum 60 cm, position exacte par rapport à l'oreille non précisée)	• 6 : 5 hommes, 1 femme	• Cz, Oz, T7, T8 (4 canaux) • référence : derrière l'oreille droite • masse : derrière l'oreille gauche	distance euclidienne	64,7
[Kaongoen et Jo, 2017]	"volume confortable ajusté par les participants"	écouteurs intra-auriculaires, modèle non précisé	• 10 : 6 hommes, 4 femmes • normo-entendants (absence d'information concernant la réalisation d'un test audiométrique)	• SSAEP : Fz, Cz, Pz, Oz, T3, T4 (6 canaux) • référence : lobe de l'oreille droite • masse : lobe de l'oreille gauche	CCA	85, 4
[Higashi et al., 2011]	"niveau de pression acoustique : −14 dB"	écouteurs intra-auriculaires, modèle non précisé	• 10 hommes • normo-entendants (absence d'information concernant la réalisation d'un test audiométrique)	• T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CPz, TP8 (12 canaux) • référence : lobe de l'oreille • masse : front	PCA + LDA, SVM	62 à 100
[Heo et al., 2017]	–	écouteurs intra-auriculaires, modèle non précisé	• 6 : 4 hommes, 2 femmes • normo-entendants (absence d'information concernant la réalisation d'un test audiométrique)	• Cz, Oz, T7, T8 (4 canaux) • référence : oreille droite • masse : front	LDA	• sons purs : 74 • sons musicaux : 89,7 • sons naturels : 87,7
[Harada et al., 2022]	–	écouteurs intra-auriculaires, modèle non précisé	• 8 hommes • normo-entendants (absence d'information concernant la réalisation d'un test audiométrique)	• F3, Fz, F4, T3, C3, Cz, C4, T4, P3, Pz, P4, O1, O2, A1, A2 (15 canaux)	SVM	• son pur : 42,9 • ruisseau : 41,9 • rouge-gorge : 43,6

TABLE 4.2 – Suite du tableau 4.1. Synthèse des études sur les BCI basées sur les SSAEP. Dans la mesure EEG, l'activité cérébrale est captée sous forme de différences de potentiel entre chaque électrode et une électrode de référence. Une électrode de masse est également utilisée afin de stabiliser les mesures et de réduire les interférences électriques, notamment le bruit du réseau.

4.2.1 Paradigme expérimental

La plupart des BCI recensées dans les tableaux 4.1 et 4.2 reposent sur des paradigmes dichotiques, impliquant la présentation simultanée de deux stimulations associées à des fréquences de modulation distinctes [Lopez et al., 2009; Xie et al., 2021; Kim et al., 2011a; Kaongoen et Jo, 2017; Heo et al., 2017]. Ces paradigmes s'appuient sur des tâches d'attention spatiale, dans lesquelles l'utilisateur est contraint de focaliser son attention sur un stimulus cible. Cette approche garantit l'engagement attentionnel des participants vis-à-vis de la stimulation. Certaines études [Xie et al., 2021; Kaongoen et Jo, 2017] ne reposent pas exclusivement sur les SSAEP, mais intègrent également d'autres marqueurs neurophysiologiques. Cette hétérogénéité méthodologique rend les comparaisons directes délicates avec les travaux fondés uniquement sur des paradigmes SSAEP.

Le nombre d'essais par condition expérimentale (i.e. le nombre de répétitions d'une même tâche) varie fortement, de 10 [Lopez et al., 2009] à 50 [Higashi et al., 2011], bien que la majorité des études en comprennent entre 20 et 25 [Kaongoen et Jo, 2017; Harada et al., 2022; Kim et al., 2011a; Heo et al., 2017]. Par ailleurs, le nombre de conditions reste généralement limité, le plus souvent à deux (attention portée sur le stimulus gauche ou droit), voire trois (attention portée sur le stimulus gauche, droit ou aucun), ce qui permet d'augmenter le nombre de répétitions sans allonger excessivement la durée totale de l'expérimentation. Toutefois, un nombre d'essais compris entre 20 et 25 apparaît suffisant.

4.2.2 Stimuli auditifs

Les signaux porteurs utilisés sont majoritairement des sinusoïdes, dans une gamme de fréquences comprise entre 450 et 2500 Hz [Lopez et al., 2009; Xie et al., 2021; Kim et al., 2011a; Kaongoen et Jo, 2017; Higashi et al., 2011]. Toutefois, une seule de ces études évalue explicitement le confort associé à ces stimuli [Harada et al., 2022]. Seules deux études [Heo et al., 2017; Harada et al., 2022] explorent des alternatives, en utilisant des sons naturels (e.g. bruit d'une cascade d'eau, chant des cigales, ruisseau ou chant d'un rouge-gorge) ou musicaux (e.g. violon, piano). D'ailleurs, Heo et al. [2017] rapportent de meilleures performances de classification avec des sons naturels qu'avec des sinusoïdes. Ces résultats suggèrent que ce type de stimuli pourrait améliorer les performances et potentiellement, l'acceptabilité. Néanmoins, il apparaît nécessaire d'évaluer la perception utilisateur vis-à-vis de ces stimulations(e.g. confort, fatigue, préférence), puisqu'elle constitue un facteur déterminant pour l'utilisation en conditions réelles.

La modulation d'amplitude constitue la méthode la plus couramment utilisée dans la littérature pour la génération de stimuli [Lopez et al., 2009; Xie et al., 2021; Kaongoen et Jo, 2017; Higashi et al., 2011; Harada et al., 2022]. Elle offre une certaine flexibilité de conception, notamment à travers le contrôle de l'indice de modulation, paramètre pourtant rarement discuté dans les travaux existants. Ce manque constitue une limite importante, dans la mesure où l'indice de modulation est susceptible de jouer un rôle déterminant dans la génération et la détectabilité des réponses SSAEP. Seuls Heo et al. [2017] mettent en évidence l'influence de cet indice sur les performances. Il apparaît donc pertinent, dans notre étude, de s'appuyer sur les valeurs pour lesquelles les meilleures performances sont obtenues dans leurs travaux.

4.2.2.1 Fréquence de modulation

Les fréquences de modulation considérées se situent majoritairement autour de 40 Hz, sauf dans les travaux de Xie et al. [2021], où des fréquences plus basses, dépendantes des fréquences porteuses, sont proposées. Certaines approches adaptent également les fréquences de modulation

à chaque individu dans une plage restreinte ([35, 45] Hz) [Harada et al., 2022]. Ces différents résultats semblent ainsi converger vers l'existence d'une fréquence de modulation optimale autour de 40 Hz, bien que cette hypothèse demeure encore peu étudiée de manière systématique (cf. section 2.4.2.3).

4.2.2.2 Durée

La durée des stimulations varie considérablement, de 5 à 44 secondes. Les durées les plus longues [Lopez et al., 2009; Kim et al., 2011a; Heo et al., 2017] apparaissent difficilement compatibles avec des protocoles comportant un nombre considérable de conditions expérimentales et un nombre suffisant d'essais, puisqu'elles allongent significativement la durée totale de l'expérimentation. Une durée intermédiaire, de l'ordre de 10 secondes, semble ainsi constituer un compromis raisonnable entre qualité des acquisitions et contraintes pratiques. Par ailleurs, le temps inter-stimulus demeure peu documenté dans la littérature, hormis dans quelques études [Xie et al., 2021; Kim et al., 2011a; Kaongoen et Jo, 2017]. Les auteurs de [Kim et al., 2011a] proposent de le faire varier entre 6 et 10 secondes au cours de l'expérimentation, une approche pouvant contribuer à limiter les phénomènes d'habituation chez les participants.

4.2.2.3 Niveau sonore de présentation

En ce qui concerne le niveau sonore de présentation des stimuli, aucune étude ne fournit d'informations précises. Celui-ci n'est généralement pas exprimé à l'aide d'unités physiques standardisées et demeure, le plus souvent, décrit de manière qualitative ou subjective. Ces imprécisions méthodologiques constituent une limite importante, dans la mesure où le niveau sonore influence directement la perception des stimuli, compromettant ainsi fortement la reproductibilité des résultats. Afin de garantir la fidélité et la reproductibilité des expériences, il apparaît donc essentiel de définir et de rapporter des paramètres physiques mesurables.

4.2.3 Dispositifs matériels

4.2.3.1 Reproduction des stimulations audio

Les dispositifs de restitution sonore ne sont jamais considérés comme des paramètres expérimentaux à part entière dans les études existantes. Le matériel est décrit de manière générique dans la plupart des travaux (e.g. "écouteurs intra-auriculaires" ou "haut-parleurs"), sans précision quant au modèle utilisé ni aux caractéristiques techniques associées. Cette absence de caractérisation empêche non seulement la reproductibilité des expériences, mais introduit également une incertitude quant aux conditions réelles de perception des stimulations. En particulier, l'utilisation de haut-parleurs, pour lesquels la position relative au participant n'est pas rigoureusement contrôlée, constitue une source supplémentaire de variabilité, notamment en raison de la distance, de l'orientation ou encore de l'environnement acoustique [Kim et al., 2011a]. Dans ce contexte, une caractérisation rigoureuse du matériel, ainsi que des conditions de diffusion, apparaît indispensable pour maîtriser les conditions de stimulation acoustique et la reproductibilité des résultats.

4.2.3.2 Acquisition des signaux EEG

En outre, la majorité des études repose sur l'utilisation de systèmes EEG pour l'enregistrement de l'activité cérébrale, avec un nombre de canaux relativement restreint. Celui-ci est, le plus

souvent, inférieur à 15 [Kaongoen et Jo, 2017; Harada et al., 2022] et se limite même, dans de nombreuses études, à seulement 2 à 6 électrodes [Lopez et al., 2009; Xie et al., 2021; Kim et al., 2011a; Higashi et al., 2011; Heo et al., 2017]. Les électrodes sont principalement positionnées au niveau des régions centrales et temporales. Toutefois, comme discuté dans le chapitre 1, le traitement auditif mobilise un réseau cérébral étendu incluant des régions frontales, pariétales, temporales et occipitales. Le nombre limité d'électrodes et leur répartition spatiale peuvent ainsi restreindre l'observation complète des SSAEP.

4.2.4 Participants

L'ensemble des études considérées repose sur un nombre limité de participants, généralement compris entre 6 et 10, ce qui limite la portée statistique des résultats. En outre, les échantillons apparaissent souvent déséquilibrés en termes de répartition hommes/femmes. Par exemple, certaines études présentent un déséquilibre marqué en faveur des participants masculins [Higashi et al., 2011; Harada et al., 2022]. Or, des différences liées au sexe peuvent influencer les réponses EEG (cf. section 1.7.3), ce qui souligne la nécessité de constituer des échantillons équilibrés.

Par ailleurs, bien que plusieurs travaux indiquent que les participants sont normo-entendants [Lopez et al., 2009; Kaongoen et Jo, 2017; Higashi et al., 2011; Heo et al., 2017; Harada et al., 2022], il n'est généralement pas précisé si cette caractéristique a été vérifiée à l'aide de tests audiométriques standardisés. D'autres études [Xie et al., 2021; Kim et al., 2011b] ne fournissent, quant à elles, aucune information à ce sujet. Pourtant, le contrôle de l'audition apparaît essentiel dans des paradigmes auditifs, dans la mesure où les capacités auditives des participants peuvent directement influencer la perception des stimuli et, par conséquent, les réponses cérébrales mesurées.

4.2.5 Classification

Contrairement aux études basées sur les SSVEP, qui s'appuient fréquemment sur des méthodes avancées de détection des fréquences d'intérêt, telles que les approches riemanniennes ou les algorithmes d'apprentissage profond (cf. chapitre 3), les études portant sur les SSAEP utilisent principalement des méthodes issues de l'apprentissage automatique classique. Parmi celles-ci figurent le SOM (*Self-Organizing Map* ou carte auto-organisatrice) [Lopez et al., 2009], un algorithme d'apprentissage non-supervisé qui projette les données sur une carte en regroupant les observations qui se ressemblent [Van Hulle, 2012], la CCA [Xie et al., 2021; Kaongoen et Jo, 2017], la LDA [Higashi et al., 2011; Heo et al., 2017] et le SVM [Higashi et al., 2011; Harada et al., 2022]. Kim et al. [2011b] utilisent également la distance euclidienne, qui mesure la proximité entre les caractéristiques extraites des différentes conditions.

Cette différence méthodologique pourrait s'expliquer par le nombre limité d'études consacrées aux SSAEP, mais également par la faible quantité de données actuellement disponibles, qui restreint l'utilisation de méthodes nécessitant d'importants volumes de données d'entraînement, en particulier les approches de DL.

4.2.6 Performances

Enfin, bien que les études reposent sur des algorithmes d'apprentissage automatique classiques, les performances de classification rapportées sont globalement satisfaisantes. Les précisions obtenues restent toutefois difficiles à comparer d'une étude à l'autre en raison de l'importante variabilité des paramètres expérimentaux employés comme les signaux porteurs, les fréquences de modulation, les durées de stimulations, l'enregistrement des signaux cérébraux ou

encore les protocoles expérimentaux. En outre, la reproductibilité de ces travaux est compromise par l'absence d'informations concernant certains paramètres essentiels, tels que le niveau sonore de présentation des stimuli ou le matériel utilisé pour leur reproduction.

Par ailleurs, bien que Heo et al. [2017] rapportent de meilleures performances avec des sons musicaux et naturels utilisés comme signaux porteurs qu'avec des sons purs, cette conclusion doit être interprétée avec prudence. En effet, l'étude ne précise ni le niveau sonore de restitution des stimuli, ni si ceux-ci ont été égalisés afin de présenter une intensité sonore perçue similaire, indépendamment de leur contenu spectral. Or, ces paramètres sont susceptibles d'influencer les réponses cérébrales et, par conséquent, les performances de classification.

4.3 Protocole expérimental

4.3.1 Objectifs et hypothèses

L'objectif de notre étude est de mesurer l'activité EEG d'un panel d'individus en réponse à quatre stimulations sonores, ainsi qu'à une condition de silence correspondant à l'absence de stimulation auditive. Cette condition permet notamment de définir une *baseline* pour la classification et d'estimer l'activité EEG spontanée, non liée aux stimulations. Les stimuli sonores reposent sur des signaux porteurs modulés en amplitude : un son pur, utilisé comme référence, ainsi que trois alternatives destinées à être évaluées.

Le principal facteur expérimental est le signal porteur. Toutefois, le niveau sonore de présentation des stimulations constitue également un paramètre expérimental et fait l'objet d'une variation contrôlée.

Les hypothèses formulées sont les suivantes :

1. les trois stimuli alternatifs présentent un niveau de confort supérieur à celui du son pur, sans dégradation significative de la détectabilité de leur fréquence de modulation ;
2. une augmentation du niveau sonore améliore la détectabilité de la fréquence de modulation dans l'activité cérébrale enregistrée.

4.3.2 Paradigme expérimental

La phase de test (figure 4.1) consiste à installer le dispositif d'enregistrement de l'activité cérébrale, puis à mesurer cette activité chez les participants en réponse aux cinq stimulations. Elle se compose d'une session comprenant deux *runs* de 10 minutes chacun, séparés par une pause de 5 minutes, pour une durée totale d'acquisition de 25 minutes. Chaque *run* comprend 50 essais, correspondant à 10 présentations de chacun des cinq stimuli. Ainsi, chaque stimulus est présenté 20 fois par session [Kaongoen et Jo, 2017; Harada et al., 2022]. Les essais sont séparés par un intervalle compris entre 6 et 10 secondes [Kim et al., 2011a], afin d'introduire une variabilité temporelle entre les essais et de limiter les effets d'anticipation. Les stimuli sont présentés dans un ordre pseudo-aléatoire, de manière à éviter la répétition immédiate d'un même stimulus et à empêcher qu'un stimulus clôturant une séquence de cinq essais ne débute la séquence suivante. Cette contrainte est implémentée directement dans le logiciel OpenViBE [Renard et al., 2010], qui assure le déclenchement des stimulations et l'acquisition des signaux cérébraux. Il est demandé aux participants de rester immobiles, les yeux fermés, pendant les deux *runs*, afin de limiter le bruit et les artefacts liés aux mouvements oculaires, aux mouvements de la tête ou à la perte de contact des électrodes [Li et al., 2013]. Les participants sont également invités à porter leur attention uniquement sur les stimuli. Au final, la phase de test dure environ 40 minutes, en

ajoutant les 15 minutes d'installation du dispositif d'enregistrement de l'activité cérébrale aux 25 minutes d'acquisition.



FIGURE 4.1 – Participant équipé d'un casque EEG et d'écouteurs intra-auriculaires lors de la phase de test.

4.3.3 Stimuli auditifs

Des sons alternatifs aux sons purs, supposés être plus confortables à l'écoute, sont évalués en tant que signaux porteurs. À cette fin, deux sons naturels sont retenus : le chant des cigales [Heo et al., 2017] et le ronronnement de chat. Un bruit Brownien est également inclus. Il s'agit d'un signal artificiel dont les propriétés, contrôlées par des paramètres définis, permettent une reproductibilité parfaite, à l'instar d'un son pur. Son contenu spectral est proche de celui de certains sons présents dans la nature, tels que les bruits d'eau, souvent perçus comme agréables. Il évoque ainsi le son d'une cascade, un son naturel également utilisé par Heo et al. [2017]. Ces différents signaux porteurs sont comparés à un son pur de 1000 Hz, utilisé comme référence et largement employé dans les études sur les BCI basées sur les SSAEP [Lopez et al., 2009; Kim et al., 2011a; Kaongoen et Jo, 2017; Heo et al., 2017; Harada et al., 2022].

Dans un premier temps, chaque signal porteur voit sa bande de fréquences limitée au moyen d'un filtre passe-haut à 300 Hz et d'un filtre passe-bas à 10000 Hz. Ce choix est motivé par des contraintes liées à la reproduction sonore, en particulier à la correction du casque d'écoute (voir section 4.4.1.5). D'une part, les casques présentent des performances limitées aux basses fréquences, notamment en dessous de 200 Hz, leur correction peut alors entraîner des distorsions affectant également les fréquences plus élevées. D'autre part, au-delà de 10 kHz, la réponse en fréquence devient fortement dépendante du positionnement du casque, rendant toute correction délicate et susceptible d'introduire des artefacts. La limitation de la bande fréquentielle permet ainsi de garantir une reproduction plus stable et contrôlée des stimuli. Une modulation d'amplitude sinusoïdale est ensuite appliquée à chaque signal porteur (voir section 4.3.3.1). Cette

approche correspond à celle employée dans les travaux de Heo et al. [2017], qui a motivé le choix des signaux porteurs considérés dans notre étude. Ces auteurs évaluent également plusieurs indices de modulation. Pour les signaux musicaux, les valeurs testées sont 1, 0.5 et 0.3, tandis que pour les sons naturels, les valeurs considérées sont 1, 0.7, 0.5. Les meilleurs résultats sont obtenus pour un indice de modulation égal à 1, cette valeur est ainsi retenue dans notre travail. En effet, l'amplitude des SSAEP tend à diminuer lorsque l'indice de modulation décroît. Enfin, le filtre de correction du casque est appliqué [3], avant de procéder à l'égalisation de la sonie (voir section 4.3.3.3). La figure 4.2 regroupe les représentations temporelles et les spectrogrammes des signaux porteurs, ainsi que des stimuli auditifs correspondants, représentés tels qu'ils sont perçus par les participants, sans application du filtre de correction du casque.

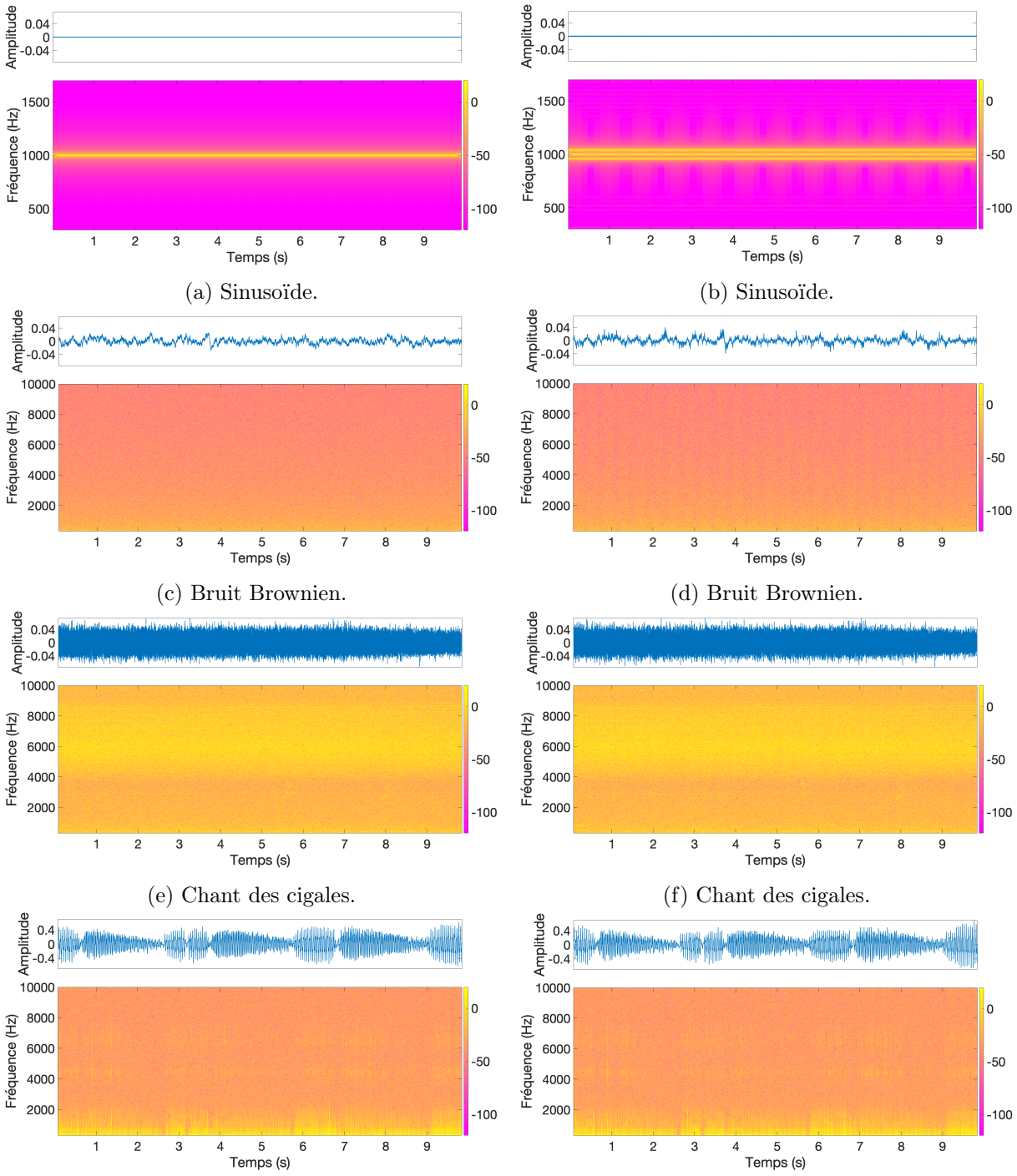
4.3.3.1 Fréquence de modulation

Compte tenu de la présentation séquentielle des stimulations sonores et de l'objectif de l'expérimentation, centré sur l'étude de l'influence des signaux porteurs plutôt que des fréquences de modulation, une seule fréquence de modulation est utilisée. La valeur de 40 Hz est choisie puisqu'elle fait consensus concernant la maximisation des SSAEP chez les humains (cf. section 2.4.2.3).

4.3.3.2 Durée

Afin de limiter la durée de l'expérimentation, une durée de stimulation de 10 secondes est privilégiée. Bien que les réponses cérébrales apparaissent rapidement après le début de la stimulation [Kaongoen et Jo, 2017], des durées plus longues permettent d'améliorer le RSB par moyennage temporel. De plus, une durée similaire est utilisée dans plusieurs études [Higashi et al., 2011; Harada et al., 2022].

3. Chaque signal modulé est normalisé dans le domaine temporel en le divisant par sa valeur absolue maximale (normalisation en amplitude), afin d'éviter tout écrêtage et de préserver une dynamique comparable.



(a) Sinusoïde.

(b) Sinusoïde.

(c) Bruit Brownien.

(d) Bruit Brownien.

(e) Chant des cigales.

(f) Chant des cigales.

(g) Ronronnement de chat.

(h) Ronronnement de chat.

FIGURE 4.2 – Représentations temporelles et spectrogrammes (échelle en dB) des signaux porteurs (à gauche) et des stimuli auditifs correspondants (à droite), tels qu'ils sont perçus par les participants (i.e. sans application du filtre de correction du casque). Les gains utilisés pour égaliser l'intensité perçue des stimuli sont également appliqués aux signaux porteurs, permettant une comparaison cohérente entre les différentes conditions. Pour les représentations temporelles, une échelle d'amplitude différente est utilisée pour le ronronnement de chat, dont l'amplitude plus élevée réduisait la visibilité des autres stimuli. Pour le son pur, l'échelle fréquentielle est limitée à 2000 Hz, l'énergie du signal étant concentrée autour de 1000 Hz. Dans les autres cas, elle est restreinte à 10000 Hz, y compris pour les signaux porteurs présentant des composantes au-delà de cette fréquence, afin de faciliter la comparaison avec les stimuli correspondants.

4.3.3.3 Niveau sonore de présentation

Afin de garantir que les stimuli sonores soient perçus avec la même sensation d'intensité, indépendamment de leur contenu fréquentiel, ils doivent être égalisés en sonie [Fletcher et Munson, 1933], c'est-à-dire présentés à **isosonie**. La sonie dépend à la fois des propriétés du son (i.e. sa pression acoustique, son contenu spectral et sa durée) et de l'auditeur [Edjekouane, 2016]. Il existe plusieurs modèles de sonie, fondés entièrement ou en partie sur le modèle de Zwicker [Zwicker, 1958]. L'unité associée aux courbes d'isosonie (figure 4.3) est le phone, et 1 phone correspond à la sonie perçue d'un son de 1 dB SPL à 1000 Hz. Le niveau de sonie dépend de la fréquence : l'intensité sonore perçue diminue lorsque la fréquence diminue ou augmente par rapport à 1000 Hz. À sonie égale, un son présenté aux deux oreilles est perçu comme plus intense que le même son présenté à une seule oreille ; cette différence est appelée **gain binaural** [Edjekouane, 2016]. Selon Whilby et al. [2006], ce gain binaural est proche de 10 dB pour des sons compris entre 50 et 60 dB SPL. Par exemple, une sinusoïde de 1000 Hz à 50 dB SPL (soit 50 phones) est perçue à 60 dB SPL (soit 60 phones) lorsqu'elle est diffusée simultanément dans les deux oreilles.

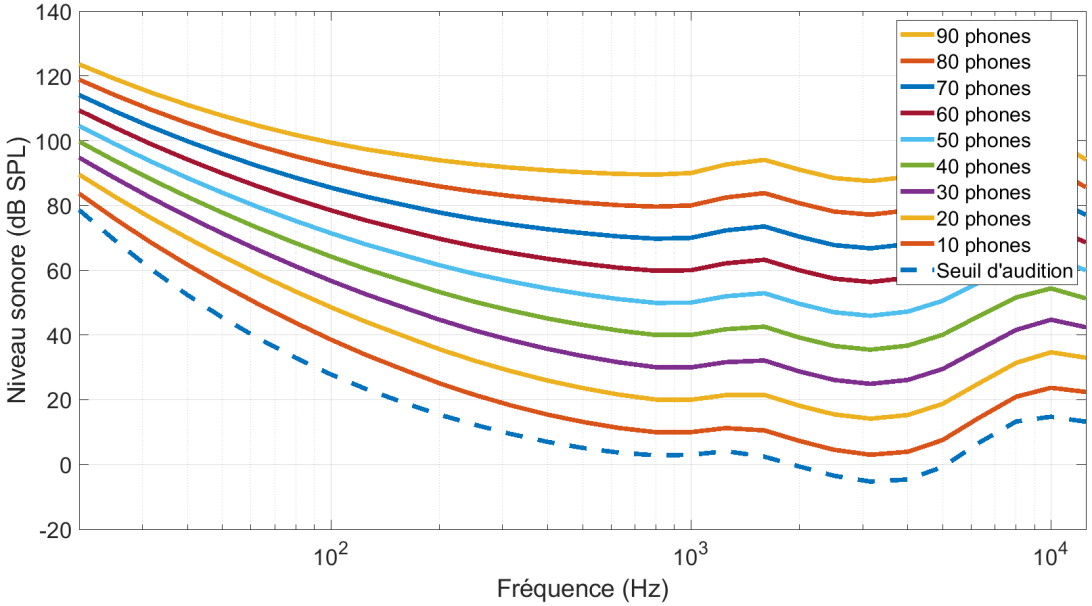


FIGURE 4.3 – Courbes isosoniques définies par la norme ISO 226 :2023 (*International Organization for Standardization*) [ISO, 2023].

Pour la parole, un niveau sonore d'environ 65 dB SPL reproduit aux deux oreilles est généralement considéré comme offrant un niveau d'écoute confortable [Edjekouane, 2016]. Afin de tester notre hypothèse, selon laquelle le niveau sonore peut influencer les réponses EEG, et qu'une augmentation de l'intensité est susceptible d'améliorer la détectabilité de la fréquence de modulation, deux niveaux de sonie distincts sont comparés dans notre expérimentation : 60 et 66 phones, pour une présentation diotique des stimuli (figure 4.4). Le premier niveau de 60 phones correspond à 60 dB SPL pour un son pur de 1000 Hz. Ce niveau est inférieur au seuil de confort rapporté dans les travaux de Edjekouane [2016], mais permet de s'assurer que l'écoute du signal sinusoïdal de 1 kHz, et par conséquent, celle de l'ensemble des stimuli puisqu'ils sont égalisés en sonie, reste confortable. Dans cette configuration, chaque oreille reçoit un stimulus de 50 phones, le niveau perçu global étant augmenté par le gain binaural d'environ 10 dB. Le second niveau sonore est défini de manière à représenter une augmentation notable sans être excessive. Une

élévation de 6 dB pour la sinusoïde est retenue, conduisant à une sonie globale de 66 phones. De fait, une augmentation de 10 phones correspond approximativement à un doublement de la sonie perçue, un tel doublement à partir de 60 phones (soit 70 phones) est jugé trop élevé pour une session d'environ 25 minutes.

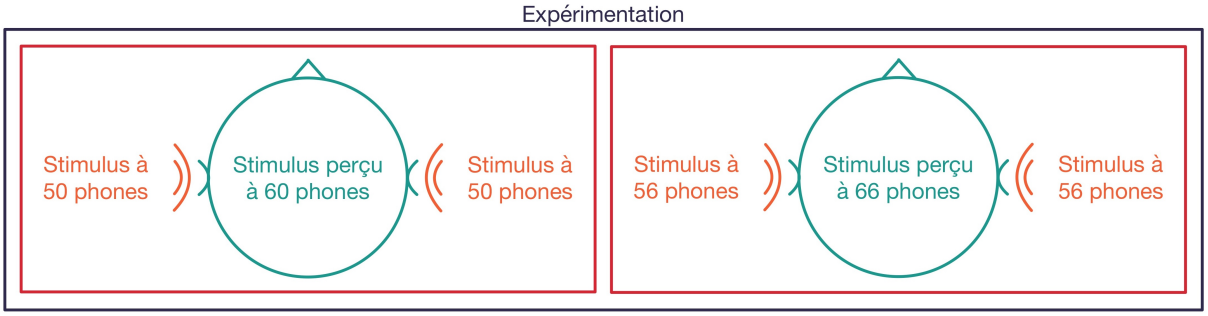


FIGURE 4.4 – Illustration des deux conditions expérimentales : à gauche, chaque oreille reçoit un stimulus de 50 phones, conduisant à une sonie globale perçue de 60 phones grâce au gain binaural ; à droite, chaque oreille reçoit un stimulus de 56 phones, conduisant à une sonie globale perçue de 66 phones.

4.3.4 Collecte de métadonnées

En complément des données EEG, des métadonnées décrivant le profil de chaque participant sont collectées, afin de prendre en compte les facteurs individuels susceptibles d'influencer les réponses cérébrales, ainsi que les retours subjectifs relatifs à l'expérimentation et notamment, aux stimulations. Des phases pré-test et post-test sont intégrées, respectivement, avant et après la phase de test (figure 4.5).

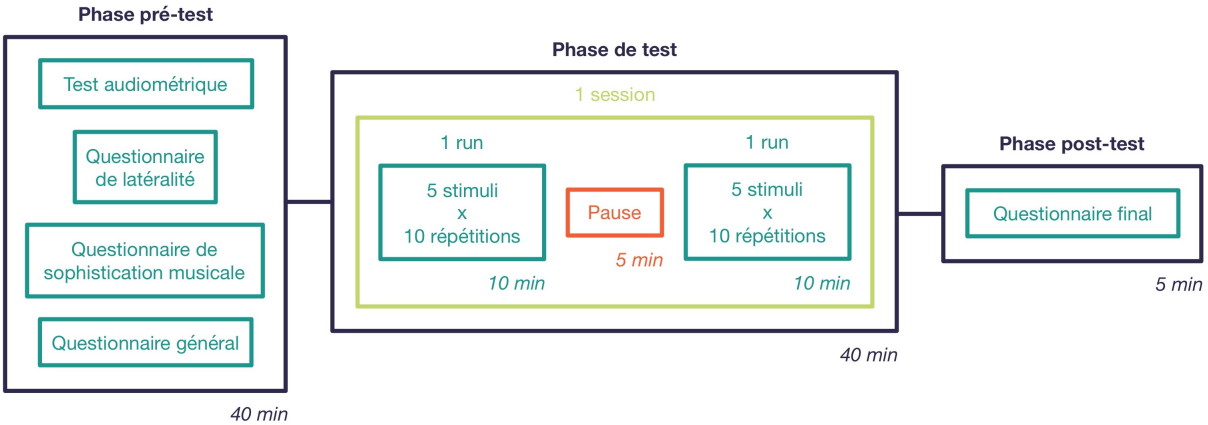


FIGURE 4.5 – Vue d'ensemble du déroulement de l'expérimentation, comprenant une phase pré-test, une phase de test et une phase post-test. Les phases pré- et post-test sont dédiées à la collecte de métadonnées sur les participants, ainsi qu'à la collecte de leurs retours subjectifs relatifs aux stimulations et à l'expérimentation. La phase de test consiste à mesurer l'activité cérébrale des participants en réponse aux cinq stimuli.

La phase pré-test, d'une durée d'environ 40 minutes, comprend un test audiométrique (présenté en section 4.4.2), ainsi que des questionnaires portant sur la latéralité, la musicalité et des

informations générales. La phase post-test repose sur un questionnaire final d'une durée d'environ 5 minutes. Le détail de ces questionnaires est présenté en section 4.5.1.

Ainsi, ce protocole respecte la durée maximale d'expérimentation d'1h30 souhaitée par les comités d'éthique.

4.3.5 Dispositifs matériels

L'expérimentation est menée dans une salle de 25 m^2 insonorisée, conforme à la norme ITU-T P800 (*International Telecommunication Union*) [ITU, 1996], au sein d'Orange Innovation à Lannion (France). Cette salle bénéficie d'un traitement acoustique destiné, d'une part, à maîtriser la réverbération, avec un temps de réverbération maintenu inférieur à 500 millisecondes, et, d'autre part, à assurer une isolation vis-à-vis du bruit extérieur. Le niveau de bruit ambiant y demeure ainsi inférieur à 30 dB(A). L'unité dB(A) est utilisée afin de représenter les niveaux sonores en tenant compte de la sensibilité fréquentielle de l'oreille humaine.

4.3.5.1 Reproduction des stimulations audio

Des écouteurs intra-auriculaires sont utilisés pour la diffusion des stimuli sonores, conformément à la configuration employée dans la majorité des études présentées dans le tableau 4.1. Ce dispositif présente plusieurs avantages dans le contexte de l'expérimentation. D'une part, il est compatible avec le dispositif d'enregistrement de l'activité cérébrale décrit en section 4.3.5.2. D'autre part, il permet un contrôle précis de la pression acoustique à l'entrée du conduit auditif des participants, contrairement à des hauts-parleurs. Enfin, le modèle retenu, le SHURE SE215-K [SHURE, 2023], de taille réduite et équipé d'un système de maintien derrière l'oreille, assure une bonne stabilité durant toute la durée de la session et facilite son utilisation auprès du maximum de participants. Toutefois, pour des raisons d'hygiène, il est nécessaire de procéder à son nettoyage entre chaque participant.

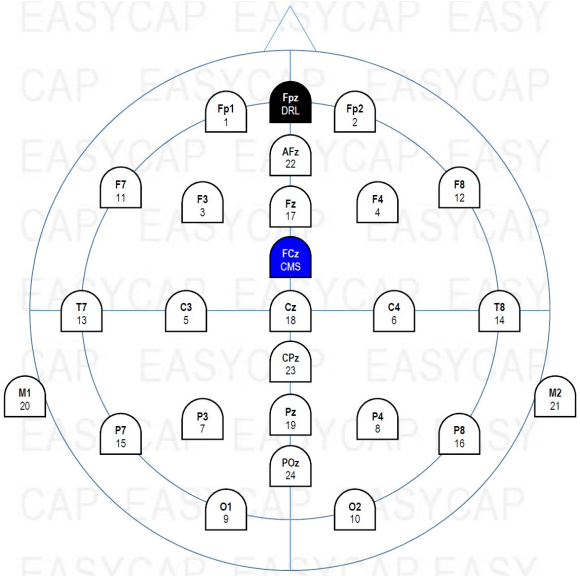
4.3.5.2 Acquisition des signaux EEG

Les participants sont équipés, tout au long de la session, d'un casque EEG de la marque mBrainTrain [EASYCAP, nd] (figure 4.6a) à 24 canaux, comportant des électrodes passives positionnées selon le système international $10-20$ (figure 4.6b). Ce dispositif permet d'enregistrer l'activité cérébrale sur l'ensemble du cuir chevelu avec une résolution spatiale relativement détaillée, permettant d'identifier les régions cérébrales présentant les réponses les plus saillantes lors de l'exposition des participants aux différents stimuli sonores. Par ailleurs, le nombre limité d'électrodes garantit un temps de préparation raisonnable pour l'expérimentation. Le contact entre les électrodes et le cuir chevelu est assuré au moyen d'un gel conducteur. Il est considéré comme satisfaisant lorsque l'impédance est inférieure à 10 kΩ. L'ensemble constitue ainsi une solution d'enregistrement non invasive, relativement peu coûteuse et largement répandue dans la littérature (tableau 4.1).

Le casque est connecté à un boîtier d'acquisition *Smarting* (mBrainTrain, figure 4.6a). Celui-ci transmet les données au logiciel OpenViBE [Renard et al., 2010] via une connexion *Bluetooth*, avec une fréquence d'échantillonnage de 500 Hz. Les données sont enregistrées au format EDF (*European Data Format*).



(a) Casque EEG connecté à un boîtier d'acquisition *Smarting*, de la marque mBrainTrain (figure tirée de [Si-Mohammed et al., 2023]).



(b) Configuration des électrodes selon le système international $10-20$. Les électrodes Fpz et FCz sont utilisées respectivement comme masse et référence.

FIGURE 4.6 – Casque EEG utilisé pour l'expérimentation et configuration de ses électrodes [EASYCAP, nd].

4.3.6 Participants

Au total, 61 participants ont été recrutés pour cette expérimentation. Les participants ont été répartis en deux groupes, chacun associé à un niveau sonore (60 et 66 phones). À l'issue des acquisitions, 24 participants ont été retenus dans chaque groupe. Certains participants ont en effet été exclus en raison de problèmes techniques (e.g. électrodes défectueuses, dysfonctionnements logiciels) ou de facteurs liés à leur comportement (e.g. mouvements excessifs, endormissement). Par ailleurs, 6 participants retenus ont pris part aux deux conditions de niveau sonore et sont ainsi communs aux deux groupes. L'expérience a été menée conformément à la Déclaration d'Helsinki de la *World Medical Association* (WMA) relative aux principes éthiques applicables à la recherche médicale impliquant des sujets humains. Tous les participants ont reçu une lettre d'information et donné leur consentement éclairé.

4.4 Facteurs acoustiques

Compte tenu de l'objectif de cette expérimentation, il est essentiel de garantir une restitution des stimuli auditifs aussi fidèle et reproductible que possible. Par conséquent, les propriétés acoustiques de chaque élément du système de reproduction sonore sont mesurées et, si nécessaire, compensées afin de correspondre à la réponse cible définie dans le cahier des charges (voir section 4.4.1.5). Il est également indispensable de s'assurer que les participants perçoivent correctement les stimuli auditifs, ce qui est vérifié au moyen d'un test audiométrique. Cette section présente l'ensemble des procédures mises en œuvre pour atteindre ces objectifs.

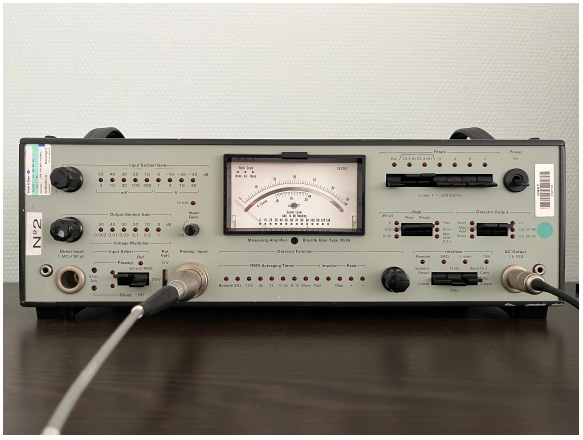
4.4.1 Casque d'écoute

4.4.1.1 Dispositif de mesure

La réponse en fréquence du casque d'écoute est mesurée à l'aide d'un HATS (figure 4.7a), un mannequin acoustique simulant l'interaction entre les ondes acoustiques (notamment les phénomènes de diffraction) et le corps humain, principalement la tête, le torse et le pavillon de l'oreille. Ce dispositif permet d'estimer la pression acoustique à l'entrée du conduit auditif, telle qu'elle serait perçue par un auditeur humain. La sortie du HATS (i.e. les signaux captés par les microphones situés dans ses oreilles) est reliée à un préamplificateur [Brüel & Kjaer, nd] (figure 4.7b), qui convertit l'amplitude du signal en niveau de pression acoustique, exprimé en dB SPL. Pour l'acquisition, le traitement et l'analyse ultérieurs, la sortie du préamplificateur est ensuite connectée à un ordinateur via une carte son Focusrite Scarlett 6i6 2e génération.



(a) HATS Brüel & Kjær type 4128C.



(b) Préamplificateur Brüel & Kjær type 2636.

FIGURE 4.7 – Dispositif de mesure de la chaîne électroacoustique : HATS et préamplificateur.

4.4.1.2 Calibration de la chaîne de mesure

Une calibration du HATS est nécessaire avant toute mesure. Cette étape essentielle permet d'interpréter directement l'amplitude du signal de sortie du HATS en termes de niveau de pression acoustique, exprimé en dB SPL. Pour ce faire, une source étalon, émettant une sinusoïde de 1000 Hz à 94 dB SPL, est couplée à l'une des oreilles du HATS à l'aide d'un adaptateur de calibration (figure 4.8), qui induit une augmentation du niveau sonore de 3,1 dB [Brüel & Kjaer, 2019]. Par conséquent, le niveau attendu à la sortie de la chaîne de mesure est de 97,1 dB SPL. Si cette valeur n'est pas obtenue, le gain du préamplificateur de mesure est ajusté jusqu'à atteindre précisément ce niveau.

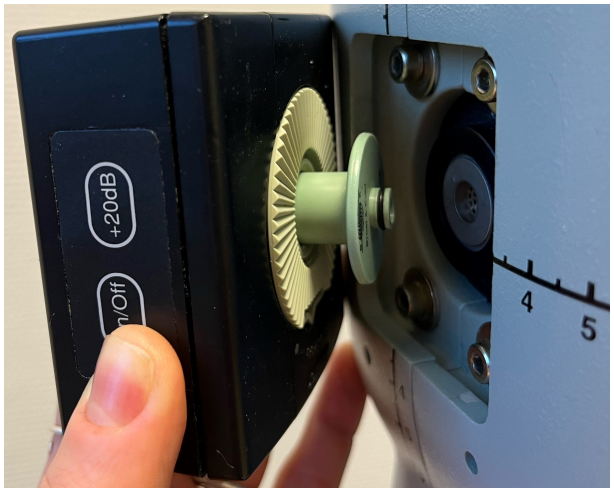


FIGURE 4.8 – Source étalon [Brüel & Kjaer, 2022] utilisée pour la calibration du HATS, équipée d'un adaptateur de calibration permettant le couplage avec l'entrée d'une oreille du HATS.

4.4.1.3 Rodage du casque d'écoute

Lorsque le casque d'écoute n'a jamais été utilisé auparavant, il est indispensable de procéder à son rodage en diffusant un signal audio contenant une gamme riche et étendue de fréquences pendant environ 20 heures. Un bruit rose (i.e. un signal couvrant une large bande de fréquences) remplit parfaitement cette fonction. Ce processus permet d'assouplir les transducteurs (i.e. les moteurs électro-acoustiques), afin d'atteindre leurs propriétés acoustiques optimales et surtout, de stabiliser la réponse en fréquence, celle-ci dépendant de la souplesse des transducteurs. La réponse en fréquence du casque est donc déterminée uniquement après cette phase de rodage.

4.4.1.4 Réponse en fréquence du casque d'écoute

Afin de mesurer sa réponse en fréquence, le casque est positionné sur les oreilles du HATS, en veillant à aligner correctement ses émetteurs avec les microphones de mesure. Un bruit rose est diffusé pendant 30 secondes dans une oreille, et la pression acoustique résultante, captée par le HATS, est enregistrée par un ordinateur. Chaque mesure est répétée cinq fois [4] par oreille, les écouteurs étant repositionnés entre chaque essai, afin de prendre en compte, dans le moyennage des mesures, les variations induites par leur positionnement, particulièrement marquées dans les hautes fréquences [Olive et al., 2013]. La réponse en fréquence est ensuite calculée en divisant la puissance du signal de sortie (P_{sortie}, le signal enregistré par le HATS) par la puissance du signal d'entrée ($P_{\text{entrée}}$, le signal envoyé au HATS) :

$$RF(f) = 10 \log_{10} \frac{P_{\text{sortie}}(f)}{P_{\text{entrée}}(f)}, \tag{4.1}$$

où f désigne la fréquence. Les mesures sont réalisées sur les deux canaux de deux casques (figure 4.10) : les écouteurs intra-auriculaires SHURE SE215-K (figure 4.9a), utilisés lors de la phase de test, et un casque circum-aural SENNHEISER HD650 (figure 4.9b), utilisé pour l'audiométrie.

4. Ce nombre peut être augmenté en fonction des besoins expérimentaux, notamment lorsque les repositionnements entraînent une variabilité importante des mesures.

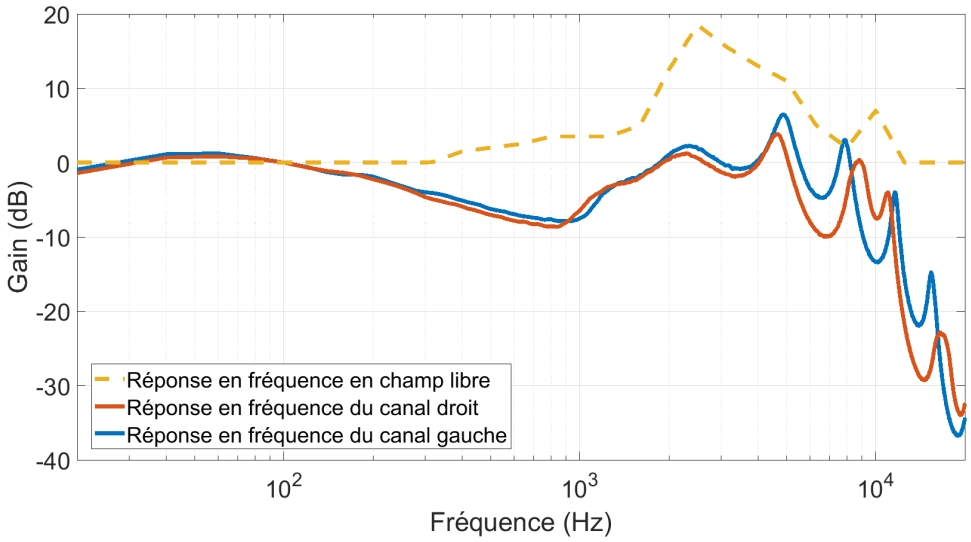


(a) Ecouteurs intra-auriculaires SHURE SE215-K utilisés lors de la phase de test.

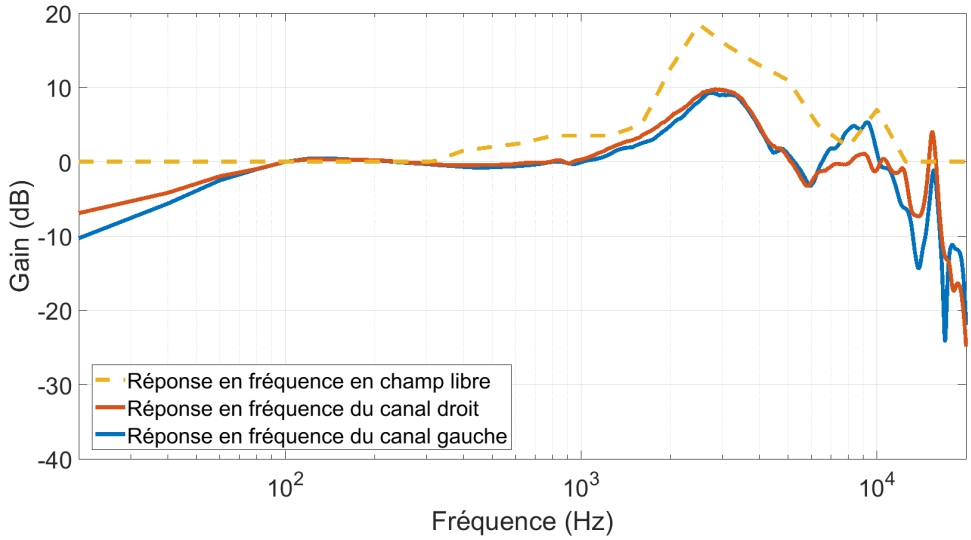


(b) Casque circum-aural SENNHEISER HD650 utilisé pour l'audiométrie.

FIGURE 4.9 – Dispositifs d'écoute utilisés lors de l'expérimentation.



(a) Réponse en fréquence des écouteurs intra-auriculaires SHURE SE215-K.



(b) Réponse en fréquence du casque circum-aural SENNHEISER HD650.

FIGURE 4.10 – Réponses en fréquence des canaux gauche et droit des dispositifs d'écoute utilisés lors de l'expérimentation, comparées à une réponse en fréquence en champ libre.

4.4.1.5 Correction du casque d'écoute

Dans le cadre d'une expérience utilisant un casque pour la reproduction de stimulations sonores, l'objectif est généralement de reproduire une condition d'écoute en champ libre, c'est-à-dire que la pression acoustique reproduite à l'entrée du conduit auditif de l'auditeur soit aussi proche que possible de celle qu'induirait une source acoustique située à la distance de référence (dans notre cas, 1 mètre) face à l'auditeur et émettant le même signal en conditions de champ libre (i.e. en environnement anéchoïque) [Edjekouane, 2016]. Cette écoute en champ libre définit la réponse cible, dite **réponse en champ libre** (*free-field frequency response*, figure 4.11), qui est largement utilisée en téléphonométrie (bien qu'il existe des variantes), afin d'obtenir une reproduction sonore au casque la plus fidèle possible. Cette réponse est établie à partir de la moyenne de mesures réalisées sur un grand nombre de morphologies humaines [ITU, 2013].

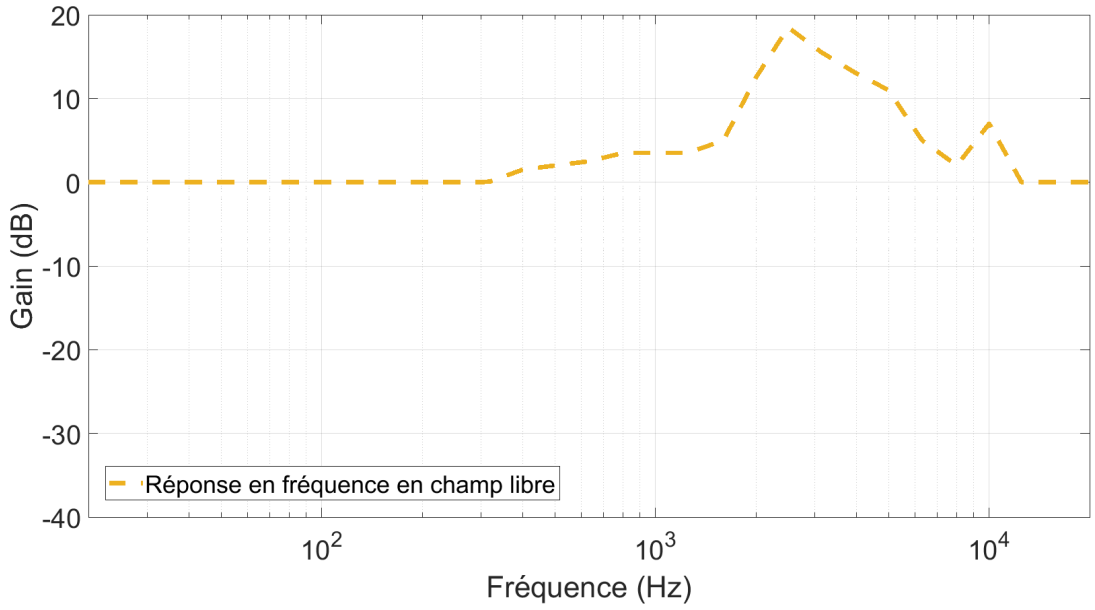
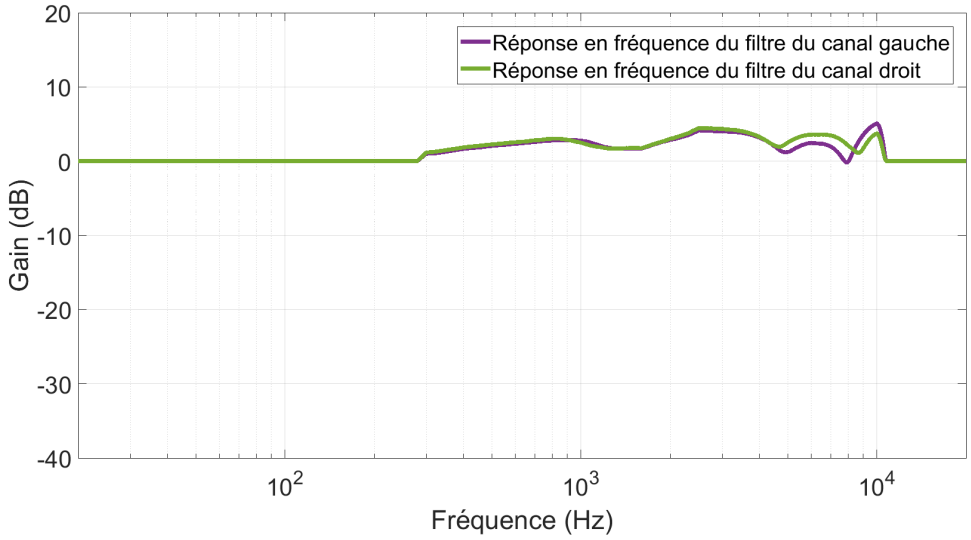


FIGURE 4.11 – Réponse en fréquence cible correspondant à une écoute en champ libre.

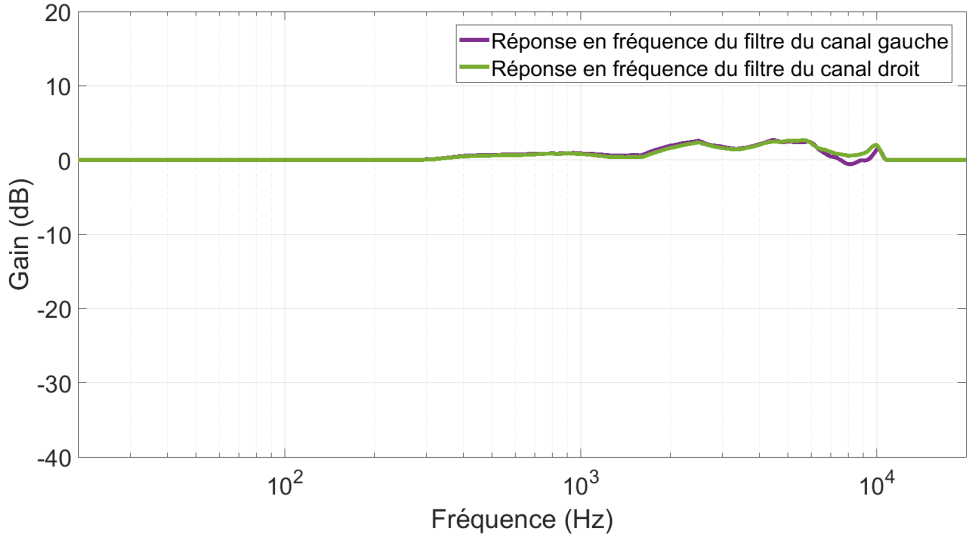
Les écouteurs SHURE SE215-K présentent une réponse en fréquence sensiblement différente de la réponse cible en champ libre (figure 4.10a). Alors que la réponse en champ libre est relativement plate entre 20 et 1000 Hz, celle des écouteurs n'est uniforme que dans la bande $[20, 100]$ Hz, puis décroît progressivement jusqu'à 1 kHz. La réponse en fréquence du casque SENNHEISER HD650 se rapproche davantage de la réponse en champ libre (figure 4.10b). Elle présente toutefois un écart dans les basses fréquences, caractérisé par une augmentation progressive du gain entre 20 et 100 Hz. Entre 100 et 1000 Hz, la réponse demeure globalement similaire à la réponse cible en champ libre. À partir de 1 kHz, la réponse en fréquence en champ libre présente une augmentation notable suivie d'une diminution aux alentours de 3 kHz, tendance également observée pour les deux dispositifs mesurés. Au-delà de 4 kHz, des irrégularités marquées (pics et creux) apparaissent, accompagnées d'écarts importants entre les canaux gauche et droit de chacun des casques. L'ensemble de ces irrégularités doit être compensé afin d'assurer une reproduction fidèle des stimuli auditifs, une attention particulière devant être portée aux fréquences supérieures à 4 kHz. Néanmoins, une correction trop précise des hautes fréquences peut entraîner l'apparition d'artefacts dans les fréquences plus basses.

Ainsi, la réponse en fréquence du filtre de correction associé à chaque casque est obtenue en

divisant la réponse en champ libre par la réponse en fréquence du casque correspondant (figure 4.12). Leurs réponses impulsionnelles sont ensuite calculées par transformée de Fourier inverse. L'ensemble des stimuli est alors convolué, dans le domaine temporel, avec le filtre correspondant au casque utilisé pour leur reproduction, afin de compenser la réponse en fréquence du casque et de tenir compte de la réponse en fréquence en champ libre.



(a) Réponses en fréquence des filtres de correction des écouteurs intra-auriculaires SHURE SE215-K.



(b) Réponses en fréquence des filtres de correction du casque circum-aural SENNHEISER HD650.

FIGURE 4.12 – Réponses en fréquence des filtres de correction des canaux gauche et droit, appliqués aux stimuli audio afin de compenser la réponse en fréquence des dispositifs d'écoute et de se rapprocher de conditions d'écoute en champ libre.

4.4.2 Audiomètre

La procédure standardisée d'évaluation de l'audition nécessite des installations dédiées (e.g. une cabine insonorisée, un audiomètre normalisé ou un casque audio spécifique, tel que le SENN-

HEISER HDA 200), qui ne sont pas toujours disponibles pour des raisons pratiques ou de coût. Son intégration dans un protocole expérimental déjà complexe peut ainsi s'avérer contraignante. Pour l'ensemble de ces raisons, nous avons choisi de développer une procédure audiométrique simplifiée. Celle-ci n'a pas pour but d'établir un diagnostic médical, mais vise uniquement à détecter d'éventuelles déficiences auditives majeures (susceptibles d'entraîner l'exclusion d'un participant), par la mesure du seuil d'audition des participants en fonction de la fréquence. Aucune information quantitative relative aux résultats de l'audiométrie n'a été communiquée aux participants pendant l'expérience. En revanche, en cas de détection d'une déficience auditive importante, il était prévu d'inviter le participant à consulter un spécialiste ORL (Oto-Rhino-Laryngologiste), sans fournir d'interprétation complémentaire. Cette situation ne s'est toutefois pas présentée au cours de l'étude.

4.4.2.1 Dispositif matériel

L'audiomètre proposé repose sur une interface graphique développée sous MATLAB et exécutée sur un Mac mini (figure 4.13). Les stimuli auditifs sont délivrés via une carte son Focusrite Scarlett 6i6 2e génération et un casque SENNHEISER HD650, choisi pour son confort et sa qualité de restitution audio. Afin de limiter les perturbations sonores liées à la présence de l'expérimentateur, seul le participant est présent dans la salle durant cette procédure.



FIGURE 4.13 – Interface graphique de l'audiomètre développé sous MATLAB.

4.4.2.2 Niveau d'audition (*Hearing Level* ou HL) et niveau de pression acoustique (*Sound Pressure Level* ou SPL)

Dans un test d'audiométrie tonale, l'objectif est de déterminer les seuils d'audition, c'est-à-dire les niveaux en dessous desquels l'auditeur ne perçoit plus les sons purs à des fréquences données. Pour ce faire, les écouteurs délivrent des stimuli à des niveaux de pression acoustique contrôlés à l'entrée du conduit auditif. Les seuils d'audition sont d'abord exprimés en dB SPL (mesurés à l'aide d'un HATS), puis doivent être convertis en dB HL (*Hearing Level* ou **niveau d'audition**). Ce niveau d'audition correspond à l'écart entre les seuils auditifs du participant (exprimés en dB SPL) et les seuils auditifs de référence (également exprimés en dB SPL), appelés **zéros audiométriques** (figure 4.14 et première colonne du tableau 4.3). Ces valeurs de référence correspondent à la moyenne des niveaux minimaux de pression acoustique perceptibles par une population de jeunes adultes des deux sexes, âgés de 18 à 25 ans, conformément à la norme ISO 389 [Sterkers-Artières et Vincent, 2014; ISO, 2023]. Ainsi, les niveaux d'audition (en dB HL) tiennent compte de la réponse en fréquence moyenne du système auditif humain. Un niveau d'audition supérieur de plus de 20 dB HL aux zéros audiométriques est révélateur d'une perte auditive à la fréquence considérée [Artières et al., 2006].

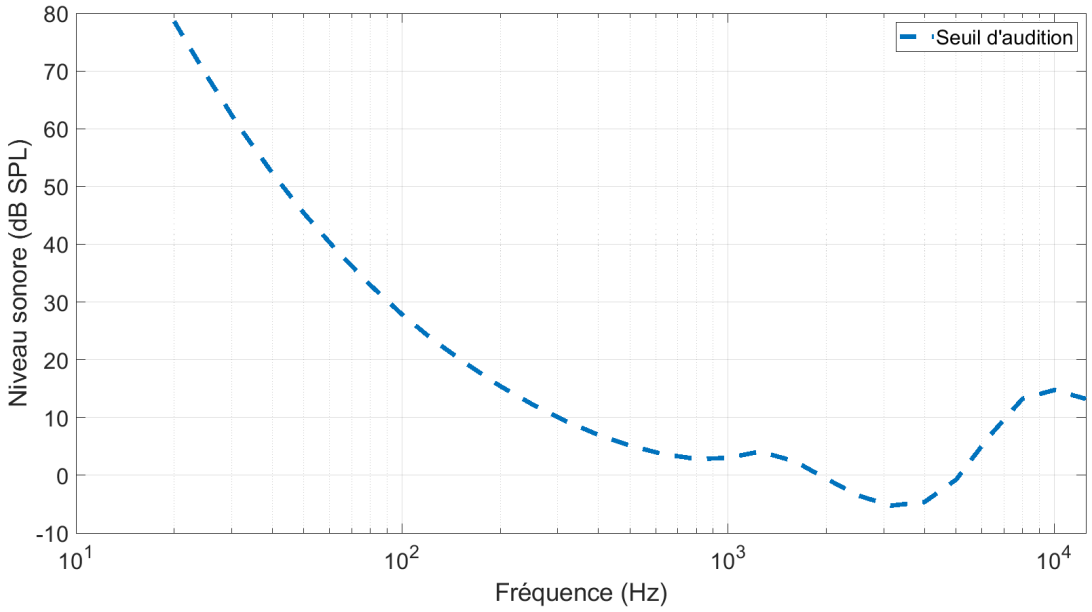


FIGURE 4.14 – Seuils auditifs de référence définis par la norme ISO 389, également appelés **zéros audiométriques** [Sterkers-Artières et Vincent, 2014; ISO, 2023].

Toutefois, avec l'âge, un phénomène de vieillissement physiologique de l'oreille peut entraîner des pertes auditives, principalement dans les hautes fréquences : il s'agit de la **presbyacousie** [Pouchain et al., 2007]. Cet effet est particulièrement marqué au-delà de 2000 Hz et tend à être plus prononcé chez les hommes que chez les femmes [Artières et al., 2006; Kim et al., 2010]. La presbyacousie s'aggrave progressivement avec l'âge (figure 4.15) : elle touche la quasi-totalité de la population après 65 ans, à des degrés variables [Pouchain et al., 2007], et peut se manifester progressivement dès 40 ans [Behtani, 2023]. Elle concerne environ 50 à 60% des personnes de plus de 60 ans et atteint 80 à 90% au-delà de 80 ans [Hull et Kerschen, 2010]. Son apparition et son évolution sont probablement influencées par des facteurs environnementaux et génétiques, ce qui explique la variabilité interindividuelle. La presbyacousie constitue la forme la plus fréquente

de surdité de perception bilatérale progressive chez l'adulte.

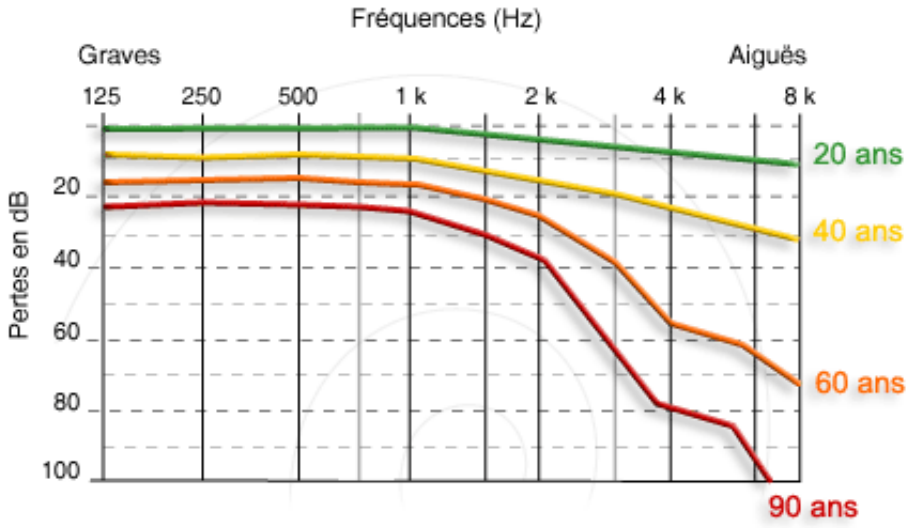


FIGURE 4.15 – Evolution de la perte auditive (en dB HL) en fonction de l'âge, illustrant l'effet de la presbyacousie, particulièrement marquée dans les hautes fréquences (figure tirée de [Fillon, 2004]).

Fréquence (Hz)	**Seuils auditifs de référence (dB SPL)**	**Niveaux sonores de calibration**	
		en dB HL	**en dB SPL**
125	23,4	60	83,4
250	12,3	70	82,3
500	5,1	80	85,1
750	3,1	80	83,1
1000	3	80	83
1500	2,9	80	82,9
2000	-0,6	80	79,4
3000	-4,9	90	85,1
4000	-4,6	90	85,4
6000	5,1	80	85,1
8000	13,3	70	83,3

TABLE 4.3 – Seuils auditifs de référence exprimés en dB SPL et correspondance, pour chaque fréquence, entre les niveaux sonores de calibration en dB HL et en dB SPL.

4.4.2.3 Calibration

Afin de garantir des mesures précises du niveau d'audition en dB HL pour chaque fréquence testée, l'ensemble de la chaîne de mesure (audiomètre et casque) doit être calibré. La procédure de calibration standard nécessite l'utilisation d'un simulateur d'oreille (CEI 60318-1, Commission Electrotechnique Internationale) et d'un coupleur acoustique (CEI 60318-3), conformément aux

normes ISO 389 [Artières et al., 2006]. Pour chaque fréquence et chaque modèle de casque, des niveaux de pression acoustique de référence, exprimés en dB SPL, induits à l'entrée du conduit auditif sont spécifiés. Cependant, en l'absence de cet équipement normalisé, une solution alternative a été développée afin de reproduire ces normes au plus près.

Le casque SENNHEISER HD650 est placé sur un HATS préalablement calibré (cf. section 4.4.1.2), utilisé ici à la place du simulateur d'oreille défini par la norme. Tout d'abord, pour chaque fréquence f testée lors de la procédure audiométrique, une sinusoïde :

$$\Xi(t) = A \cdot \sin(2\pi f t) \tag{4.2}$$

d'amplitude A (choisie de façon arbitraire) est générée. Ce signal est ensuite traité par le filtre de correction spécifique au casque HD650 (cf. section 4.4.1.5). Le niveau sonore résultant à l'entrée du conduit auditif du HATS est mesuré en dB SPL, pour chaque fréquence et pour les deux canaux. L'expérimentateur ajuste alors l'amplitude A afin que le niveau sonore mesuré corresponde au niveau de calibration requis, indiqué dans le tableau 4.3 (dernière colonne). Les basses et hautes fréquences sont produites à des niveaux perçus plus faibles (en dB HL) que les fréquences médianes. Par conséquent, l'écart maximal entre les niveaux sonores en dB SPL n'est que de 6 dB, ce qui permet de maintenir des conditions de fonctionnement similaires du système de reproduction audio, quelle que soit la fréquence. Cette étape de calibration permet de déterminer l'amplitude nominale A du stimulus $\Xi(t)$ pour chaque fréquence. Lors de l'audiométrie, le niveau sonore en dB HL est ensuite ajusté en appliquant un gain g égal à la différence entre le niveau souhaité et le niveau de calibration (i.e. la deuxième colonne du tableau 4.3), soit

$$\Xi(t) = g \cdot A \sin(2\pi f t). \tag{4.3}$$

Par exemple, pour un niveau cible de 40 dB HL à 125 Hz, le gain est $G = 40 - 60 = -20$ dB, ce qui correspond à un gain linéaire en amplitude de $g = 10^{G/20} = 0{,}1$.

4.4.2.4 Procédure expérimentale

La procédure utilisée dans notre expérimentation suit la méthode ascendante (*ascending method*) [Hughson et Westlake, 1944; Carhart et Jerger, 1959] : le seuil d'audibilité minimal est déterminé en augmentant progressivement le niveau sonore du stimulus. Toutefois, une phase initiale descendante est mise en œuvre afin de débuter la phase ascendante au plus près du seuil auditif du participant, réduisant ainsi la durée du test par rapport à un départ à 0 dB HL ou à un niveau inférieur. Le test débute donc à 40 dB HL [Arlinger, 1979; Carhart et Jerger, 1959]. Tant que le participant perçoit la tonalité, le niveau sonore est réduit de 10 dB HL [Aras, 2003; Arlinger, 1979]. Lorsque le participant n'entend plus le stimulus, le niveau sonore est alors augmenté par paliers de 5 dB HL [Aras, 2003; Arlinger, 1979], jusqu'à ce qu'il indique percevoir à nouveau la tonalité. Le stimulus suivant est alors présenté 10 dB HL plus bas, et une nouvelle séquence ascendante par pas de 5 dB HL est initiée. Dans cette procédure, le seuil auditif est défini comme le plus faible niveau auquel le stimulus est perçu dans plus de la moitié des séquences ascendantes [Aras, 2003; Liu et al., 2022]. La procédure est appliquée à l'ensemble des fréquences, dans l'ordre suivant : 1000, 1500, 2000, 3000, 4000, 6000, 8000, 750, 500, 250 et 125 Hz. Conformément aux recommandations de [Artières et al., 2006], la durée de stimulation est fixée à $d = 1$ seconde, avec l'application d'une rampe de 250 ms au début et à la fin du signal. L'intervalle entre deux stimuli est choisi aléatoirement entre 0 et 3 secondes, afin d'éviter que les participants n'anticipent le rythme des présentations. L'ensemble de la procédure est automatisé et piloté par les réponses du participant : le niveau sonore est ajusté de manière adaptative en fonction de la détection du

stimulus, tandis que la fréquence est mise à jour une fois le seuil auditif estimé pour la condition considérée.

4.5 Facteurs individuels

4.5.1 Questionnaires

Comme indiqué précédemment (cf. section 4.3.4), des données spécifiques sont recueillies auprès des participants au moyen de plusieurs questionnaires, afin de prendre en compte l'influence des facteurs individuels (i.e. sexe, âge, consommation de substances psychoactives, état d'esprit, latéralité, musicalité, etc.) sur les réponses EEG. Ces questionnaires permettent également de collecter, à l'issue de l'expérience, des retours subjectifs relatifs aux stimulations et au déroulement de l'expérimentation.

4.5.1.1 Questionnaire de latéralité

L'*Edinburgh Handedness Inventory* (EHI) est la mesure de la latéralité manuelle la plus couramment utilisée [Oldfield, 1971]. Il comprend 41 tâches, pour lesquelles les participants indiquent la main préférentiellement utilisée, l'absence de préférence ou le fait qu'ils ne savent pas (figure 4.16). Les scores Γ et Δ sont initialement fixés à zéro. Pour chaque élément, les scores sont mis à jour selon les règles suivantes :

— $\Gamma = \Gamma + 1$, lorsque la préférence est attribuée à la main gauche,
— $\Delta = \Delta + 1$, lorsque la préférence est attribuée à la main droite,
— $\Gamma = \Gamma + 1$ et $\Delta = \Delta + 1$, en cas d'absence de préférence,
— $\Gamma = \Gamma + 0$ et $\Delta = \Delta + 0$, lorsque le participant ne sait pas.

Le score de latéralité est alors défini par :

$$\Theta = \frac{\Delta - \Gamma}{\Delta + \Gamma} \cdot 100, \tag{4.4}$$

où Θ est compris entre -100 et 100. Un participant est considéré comme gaucher si $\Theta < -40$, droitier si $\Theta > 40$, et ambidextre si $\Theta \in [-40, 40]$.

Dans le but de limiter la durée du questionnaire, nous avons décidé d'utiliser un sous-ensemble des 41 éléments. Plusieurs chercheurs ont proposé des versions abrégées de l'EHI, chacune visant à corriger certaines limites perçues du questionnaire original [Edlin et al., 2015]. Toutefois, aucun consensus ne s'est imposé quant au choix d'une version réduite de référence. Dans ce contexte, une étude pilote, menée auprès de 28 participants, a comparé les scores obtenus avec l'EHI complet et ceux obtenus à partir des rubriques les plus fréquemment utilisées, identifiées par [Edlin et al., 2015] (version courte de l'EHI). Pour les participants présentant un score proche des seuils de classification, c'est-à-dire autour de -40 et 40 (e.g. entre gaucher et ambidextre, ou entre droitier et ambidextre), la catégorisation pouvait différer entre la version complète du questionnaire et sa version courte. Il a donc été décidé de ne pas recourir à la version courte du test, mais de sélectionner 12 éléments issus de la version complète, de manière à préserver au mieux les résultats :

— écrire,
— jeter,
— utiliser un peigne,
— ouvrir une boîte (main qui tient le couvercle),
— échanger des cartes (main qui tient la carte qui est échangée),

— tenir une bouteille,
— tenir un verre,
— mettre la clé dans la serrure,
— jouer à la balle,
— pointer précisément (un endroit par exemple),
— utiliser une télécommande,
— agrafer des feuilles de papier (main qui tient l'agrafeuse).

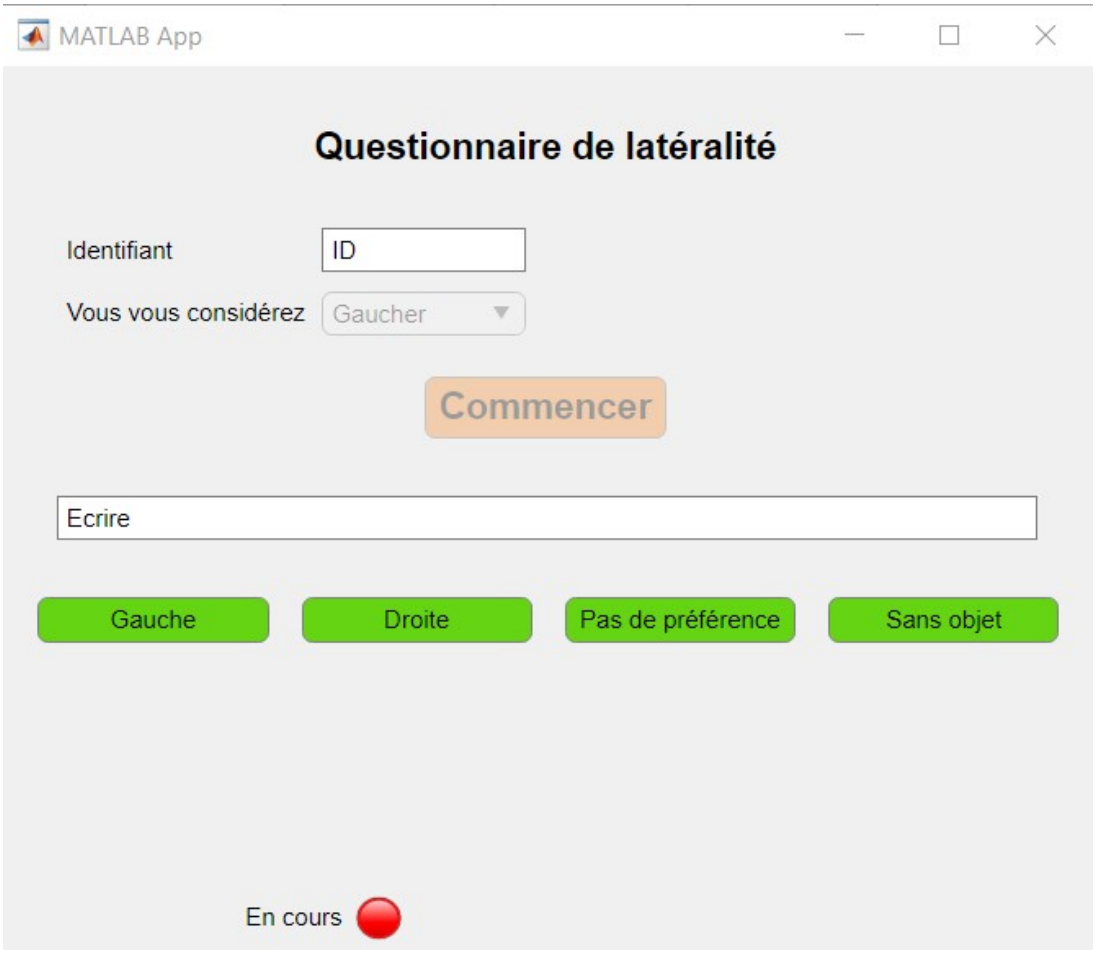


FIGURE 4.16 – Interface graphique du questionnaire de latéralité développé sous MATLAB.

4.5.1.2 Questionnaire de sophistication musicale

La musicalité et l'expérience musicale ne dépendent pas uniquement d'une formation instrumentale formelle. Elles reflètent également l'engagement dans diverses activités musicales [Degrave et Dedonder, 2019]. De nombreuses compétences musicales se développent ainsi par une interaction répétée et attentive avec la musique, plutôt que dans le cadre d'un apprentissage formel [Müllensiefen et al., 2013]. L'indice de sophistication musicale (*Goldsmiths Musical Sophistication Index* ou Gold-MSI), développé par la *Goldsmiths, University of London*, a été conçu pour mesurer la musicalité et l'expérience musicale telles qu'elles s'expriment dans les sociétés occidentales. Il comprend cinq dimensions : l'engagement actif (comportements d'implication dans la musique), les capacités perceptives (aptitudes liées à l'écoute musicale), la formation musicale,

les capacités vocales (compétences et activités liées au chant) et les émotions (comportements associés aux réponses émotionnelles à la musique). Le **facteur général de sophistication musicale** combine des aspects issus de l'ensemble de ces dimensions. Afin de ne pas allonger la durée de l'expérimentation, seul ce facteur général est utilisé ici. Il repose sur 17 éléments et fournit une estimation globale des capacités musicales de chaque participant [Degrave et Dedonder, 2019] :

- Je consacre une partie importante de mon temps libre à des activités liées à la musique.
- J'aime écrire à propos de musique (ex. sur des forums, des blogs, etc.).
- Si quelqu'un chante une chanson que je ne connais pas, je peux me mettre à chanter avec lui.
- Je suis capable de trouver les bonnes notes lorsque je chante accompagné(e) d'un enregistrement.
- Je suis capable de comparer et discuter des différences entre deux interprétations ou versions d'un même morceau de musique.
- Je n'ai jamais été complimenté(e) pour mes talents de chanteur ou de musicien.
- Je lis ou je cherche souvent sur internet des informations en lien avec la musique.
- Je ne suis pas capable de chanter sur une tonalité différente lorsque quelqu'un chante une mélodie que je connais.
- Je suis capable d'identifier ce qu'un morceau de musique a de particulier.
- Quand je chante, je n'ai aucune idée de si je chante juste ou pas.
- La musique est une sorte d'addiction pour moi - je ne pourrais pas vivre sans elle.
- Je n'aime pas chanter en public car j'ai peur de chanter les mauvaises notes.
- Je ne me considérerais pas comme un(e) musicien(ne).
- Après avoir entendu une nouvelle chanson deux ou trois fois, je peux normalement la chanter seul(e).
- Je me suis investi(e) dans une pratique régulière, quotidienne d'un instrument de musique pendant __ an(s).
- Au sommet de mon intérêt, je pratiquais mon instrument principal durant __ heure(s) par jour.
- Je sais jouer __ instrument(s) de musique.

Chaque élément est évalué à l'aide d'une échelle de Likert à 7 niveaux (figure 4.17). Pour les formulations négatives, l'échelle est inversée (e.g. une réponse de 7, "tout à fait d'accord", à un élément formulé négativement est recodée en 1). Le score global est calculé en sommant les réponses obtenues pour l'ensemble des éléments. Afin de permettre la comparaison entre participants, ce score est ensuite normalisé par le score maximal théorique, égal au produit du nombre total d'éléments par 7. Le résultat est exprimé en pourcentage du score maximal, constituant ainsi un indicateur synthétique et comparable de la musicalité et de l'expérience musicale de chaque participant.

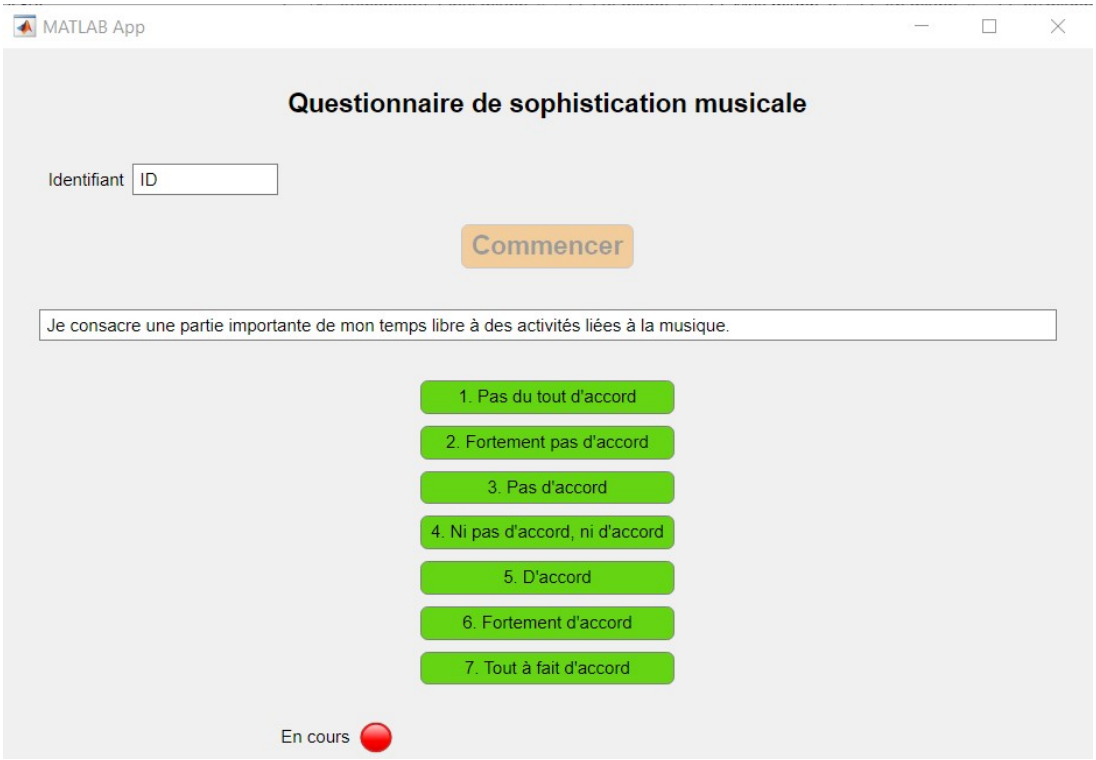


FIGURE 4.17 – Interface graphique du questionnaire de musicalité développé sous MATLAB.

4.5.1.3 Questionnaire général

Les participants remplissent également un questionnaire général (figure 4.18), portant sur différentes informations personnelles, telles que le sexe et l'âge, l'état d'esprit, l'émotion dominante, ainsi que la consommation récente de substances au cours des 24 dernières heures (e.g. caféine, théine, cigarettes, alcool, stupéfiants ou médicaments). L'état d'esprit est évalué en termes de valence et d'excitation, sur une échelle de 1 à 9, où 1 correspond respectivement à "déplaisant" et "somnolent", et 9 à "plaisant" et "énergique" (*Self-Assessment Manikin* ou SAM de [Bradley et Lang, 1994]). Le questionnaire comporte également des questions portant sur d'éventuels troubles auditifs, notamment l'hyperacousie (i.e. une hypersensibilité aux sons [Bizaguet, 2011]) et les acouphènes (i.e. une perception auditive ressentie en l'absence de stimulation sonore [Noreña et al., 2020]), ainsi que sur la familiarité des participants avec les tests auditifs et les BCI. L'ensemble de ces facteurs est susceptible d'influencer l'activité cérébrale. En particulier, l'hyperacousie, les acouphènes ou une perte auditive peuvent altérer la perception des stimulations sonores, tandis que la familiarité avec les tests auditifs et les BCI peut faciliter la concentration de l'attention sur les stimuli et, par conséquent, améliorer les performances [Lotte et al., 2018].

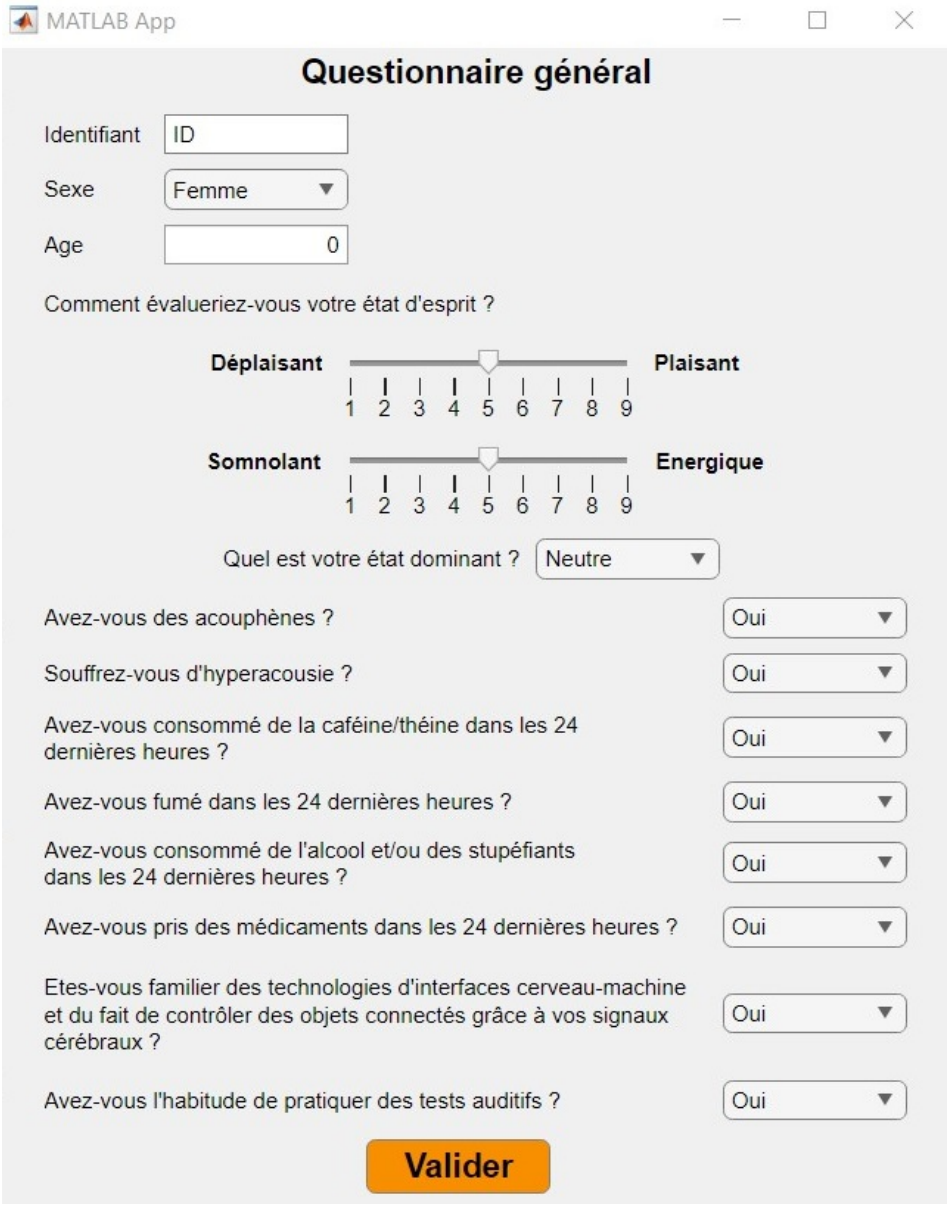


FIGURE 4.18 – Interface graphique du questionnaire général développé sous MATLAB.

4.5.1.4 Questionnaire final

Le questionnaire final (figure 4.19) vise à évaluer l'état d'esprit des participants, ainsi que leurs retours sur l'expérimentation. Il permet notamment de comparer leur état d'esprit avant et après l'expérimentation (cf. section 4.5.1.3), d'en évaluer l'impact, ainsi que de déterminer dans quelle mesure l'expérimentation a été perçue comme agréable ou non, et pour quelles raisons. À ce stade, les participants ne connaissent pas la nature exacte de chacun des stimuli présentés. Il leur est alors demandé de classer les stimuli selon leur préférence, d'indiquer ceux qu'ils jugent désagréables et ceux qui paraissent plus forts que les autres. Ces informations permettent d'identifier les stimulations les plus appréciées par les participants, et donc potentiellement les plus adaptées à une utilisation dans une BCI. Par ailleurs, malgré l'égalisation de la sonie, l'intensité

perçue peut varier d'un individu à l'autre : un stimulus perçu comme plus fort par un participant pourrait induire des réponses neuronales plus marquées, et donc être plus facilement détecté dans l'activité cérébrale.

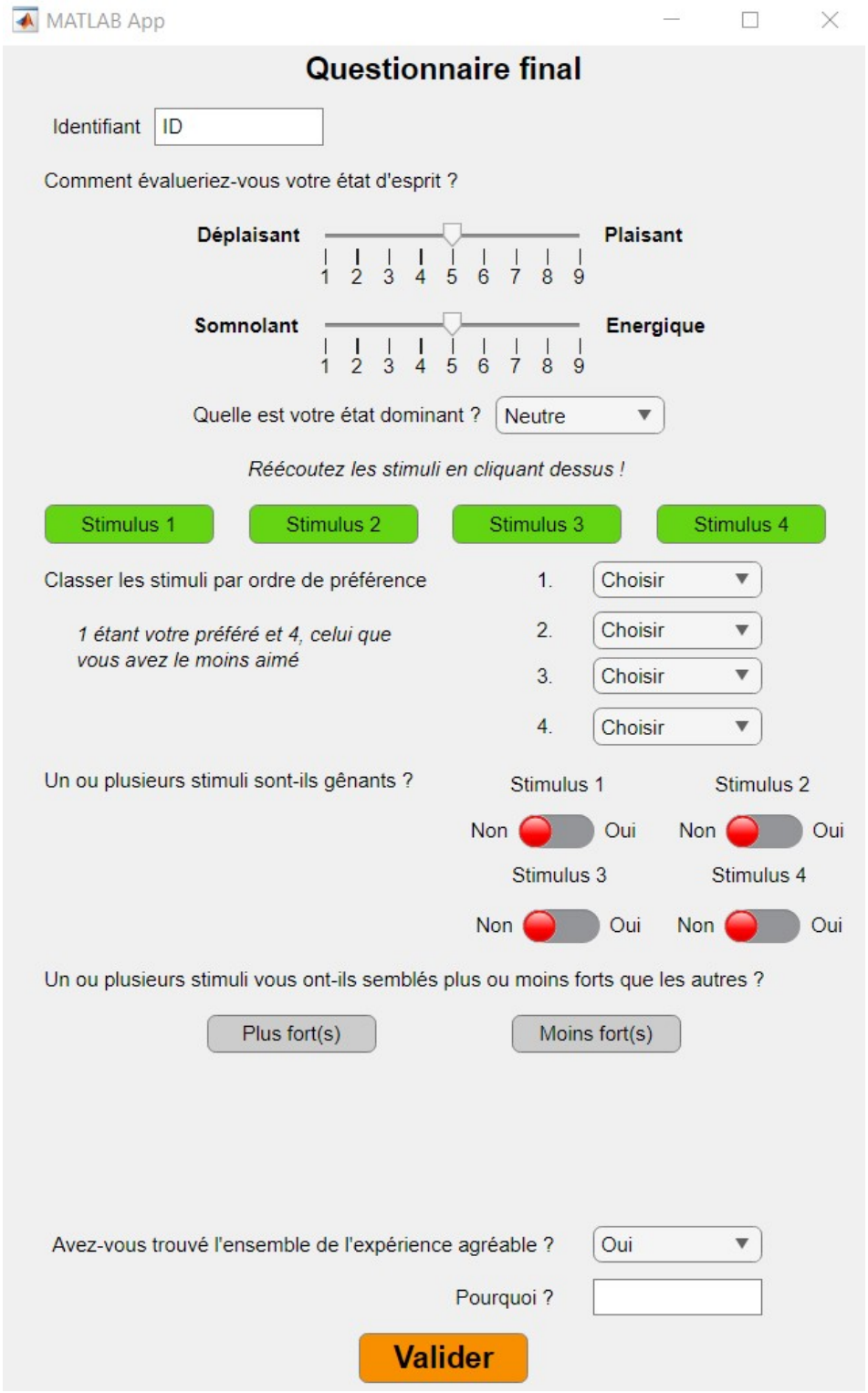


FIGURE 4.19 – Interface graphique du questionnaire final développé sous MATLAB.

4.6 Conclusion

Ce chapitre présente le paradigme expérimental conçu pour évaluer l'utilisation de sons alternatifs aux sinusoïdes dans le cadre d'une BCI basée sur les SSAEP. Il met en particulier l'accent sur les enjeux liés à la fidélité de la reproduction sonore, à la perception des stimuli par les participants, ainsi qu'à la prise en compte de caractéristiques individuelles pouvant impacter l'activité cérébrale.

La conception des stimuli auditifs est décrite en détail, notamment les traitements acoustiques mis en œuvre afin de rendre leur reproduction indépendante du matériel et leur niveau sonore perçu indépendant de leur contenu fréquentiel. Ce travail a également conduit au développement d'une procédure simplifiée d'audiométrie, ne nécessitant pas d'installations médicales dédiées. Cette approche est adaptable et réutilisable, sous réserve de disposer d'un HATS pour la calibration ou de mettre en place une méthode de calibration alternative. Il en va de même pour les questionnaires et interfaces développés afin de recueillir des données individuelles. Ainsi, cette étude propose un ensemble d'outils méthodologiques, réutilisables dans de futures expérimentations en audio ou en BCI.

La question de la détectabilité de la fréquence de modulation dans les signaux EEG, en fonction du signal porteur, constitue désormais la première problématique à étudier. Elle fait l'objet du chapitre suivant (chapitre 5).

5

Evaluation de la détectabilité de la fréquence de modulation selon le signal porteur

Sommaire

5.1 Introduction et présentation du contexte 101
5.2 RSB 103
5.2.1 Prétraitements des signaux EEG préalables au calcul du RSB 104
5.2.2 Détectabilité de la fréquence de modulation 107
5.3 Classification 111
5.3.1 Prétraitements des signaux EEG préalables à l'application de classifieurs 111
5.3.2 Résultats de détection de la fréquence de modulation 114
5.4 Cartographie cérébrale 122
5.5 Analyse critique des propriétés des stimulations 125
5.6 Conclusion 126

5.1 Introduction et présentation du contexte

Les signaux EEG de 48 participants ont été enregistrés lors de leur exposition à quatre stimuli sonores modulés : un son pur (utilisé comme référence), un son artificiel (le bruit Brownien) et deux sons naturels (le chant des cigales et le ronronnement de chat), ainsi que lors d'une condition correspondant à l'absence de stimulation auditive (figure 5.1).

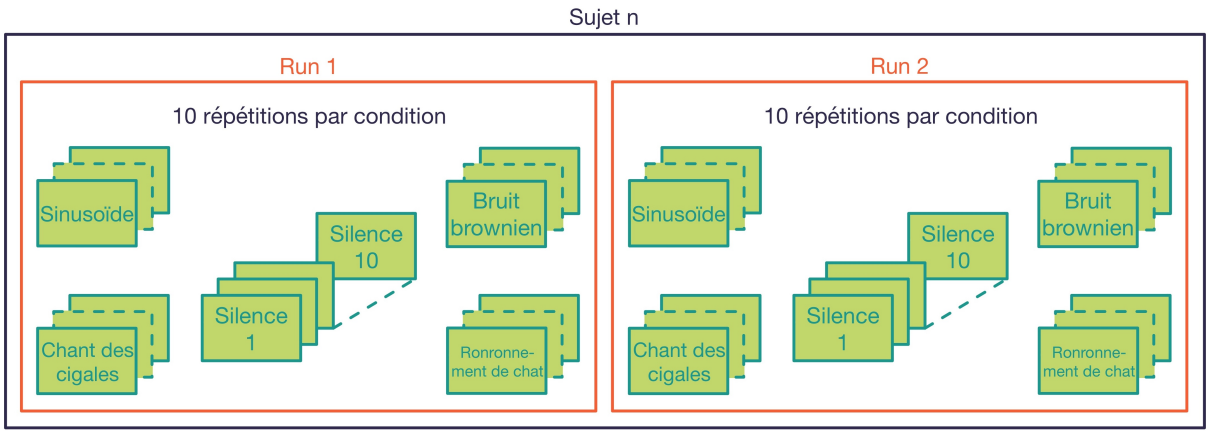


FIGURE 5.1 – Pour chaque participant (n), les réponses cérébrales associées à 10 répétitions de chaque stimulus ont été enregistrées pour chaque *run*, soit un total de 20 répétitions pour les 2 *runs* de chaque session (cf. section 4.3.2).

L'expérimentation (figure 5.2) a été réalisée à deux niveaux de sonie distincts, afin d'évaluer l'influence de l'intensité sonore perçue sur la détectabilité de la fréquence de modulation dans les signaux cérébraux. Les enregistrements ont été effectués à l'aide d'un casque EEG comportant 24 électrodes réparties sur l'ensemble du cuir chevelu.

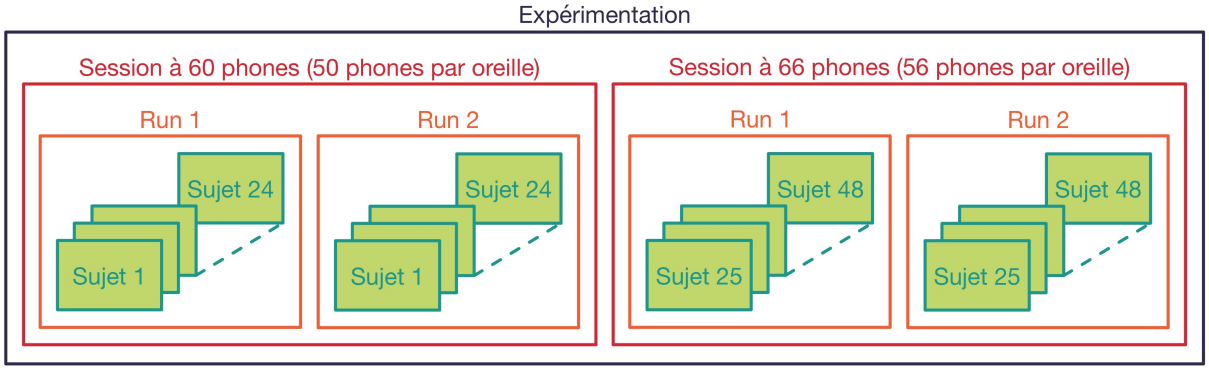


FIGURE 5.2 – Répartition des 48 participants entre les deux conditions expérimentales : 24 participants ont été exposés à des stimuli égalisés à 60 phones et 24 autres, à des stimuli égalisés à 66 phones.

L'objectif principal est d'évaluer si la fréquence de modulation associée à chaque stimulus sonore peut être détectée dans l'activité cérébrale des participants. Plus précisément, il s'agit d'estimer la puissance des signaux EEG à cette fréquence pendant la présentation d'une stimulation auditive (i.e. le signal utile) et en l'absence de stimulation (i.e. le bruit). Le rapport entre ces deux puissances constitue, par définition, un RSB, qui est étudié dans un premier temps (section 5.2). Pour chaque stimulus, le problème de détection de la fréquence de modulation peut également être formulé comme un problème de classification binaire, dans lequel deux classes sont considérées :

— **Classe 1** : signaux EEG contenant la fréquence de modulation,
— **Classe 2** : signaux EEG ne contenant pas la fréquence de modulation.

Cette approche est explorée dans un second temps (section 5.3). Elle est d'ailleurs largement

exploitée pour l'analyse de données EEG, en particulier dans le cadre de paradigmes SSVEP. En revanche, la classification de données issues de paradigmes SSAEP reste encore peu étudiée. L'objectif est donc d'identifier les approches de classification les plus adaptées à la détection et à la discrimination des SSAEP. Dans ce cadre, dix stratégies de classification sont considérées, incluant des méthodes classiques d'apprentissage automatique, des architectures d'apprentissage profond, ainsi que des approches riemannienne (notamment dans l'espace tangent). Ces dernières reposent sur l'utilisation de matrices de covariance calculées à partir des signaux EEG, lesquelles capturent les relations spatiales entre électrodes. Ces matrices étant définies sur une variété riemannienne, les méthodes associées permettent d'exploiter une géométrie adaptée à leur structure (cf. section 3.5.2.4).

Par ailleurs, différentes étapes de prétraitement, telles que le filtrage passe-bande et le fenêtrage temporel, sont évaluées afin d'identifier les paramètres susceptibles d'améliorer la qualité des signaux EEG et la détectabilité des SSAEP (sections 5.2.1.1 et 5.3.1).

La distribution spatiale des réponses cérébrales est également analysée en section 5.4 dans le but d'identifier, pour chaque stimulation sonore, les électrodes contribuant le plus à la détection des SSAEP. À terme, cette analyse pourrait permettre d'optimiser le positionnement et le nombre d'électrodes, de manière à rendre les acquisitions EEG moins contraignantes pour les utilisateurs tout en conservant des performances de détection satisfaisantes.

Enfin, les propriétés acoustiques de chaque stimulus sont étudiées en section 5.5 afin de mieux comprendre les différences de détectabilité observées.

5.2 RSB

Dans ce travail, le RSB (cf. section 3.5.1) est utilisé afin de quantifier l'émergence de la fréquence de modulation dans les signaux EEG. Bien que cet estimateur soit davantage employé en traitement du signal qu'en BCI, il constitue une mesure indicative pertinente de la détectabilité des SSAEP. Le RSB est ici évalué à l'aide de deux approches.

La première consiste à le calculer sur les signaux EEG prétraités, sans application préalable d'un filtrage passe-bande, auquel cas il s'agit du RSB brut, noté RSB_{BRUT}. Les puissances spectrales sont estimées pour chaque stimulus aux fréquences d'intérêt, à savoir la fréquence de modulation et sa première harmonique, puis moyennées sur l'ensemble des répétitions associées, des canaux. Le RSB est ensuite calculé à partir de ces puissances moyennées pour chaque sujet, avant d'être moyenné sur l'ensemble des sujets.

Dans une seconde approche, nous proposons de calculer le RSB à partir des signaux prétraités ayant fait l'objet d'un filtrage passe-bande, puis d'une réduction de dimension par un algorithme de CSP (cf. section 3.4.3). Cette approche permet d'évaluer l'impact de ces étapes de traitement sur la qualité des signaux et sur la détectabilité des fréquences d'intérêt. Deux matrices de projection distinctes ($\mathbf{W}_{\text{CSP}}$) sont alors déterminées : l'une à partir des données filtrées passe-bande autour de la fréquence de modulation, et l'autre à partir des données filtrées autour de la première harmonique. Pour chacune d'elles, le filtre spatial qui maximise la variance des signaux EEG en réponse au stimulus par rapport à l'absence de stimulus est appliqué aux signaux EEG non filtrés passe-bande. Le RSB est ensuite calculé en moyennant les résultats sur l'ensemble des répétitions et des sujets. Ces RSB sont désignés par RSB_{CSP}. Les différentes étapes conduisant au calcul de ces deux types de RSB sont résumées dans la figure 5.3.

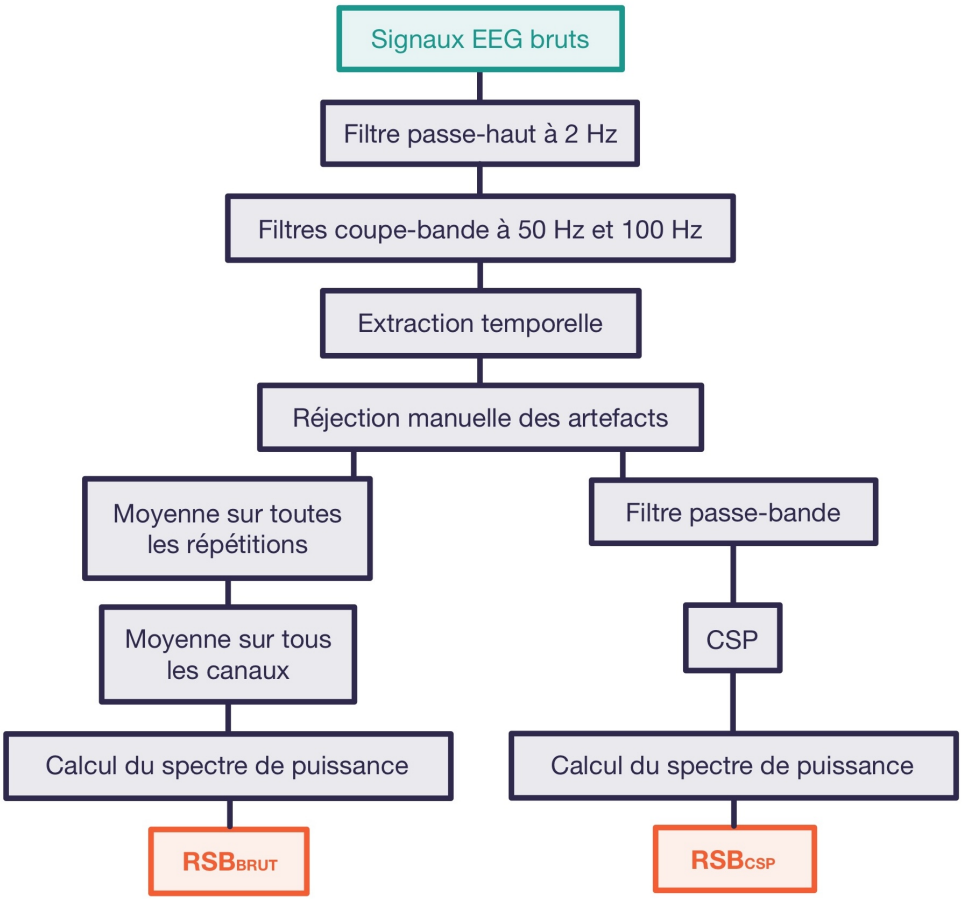


FIGURE 5.3 – Etapes du calcul des deux types de RSB : RSB_{BRUT} et RSB_{CSP}.

5.2.1 Prétraitements des signaux EEG préalables au calcul du RSB

Avant le calcul du RSB, les données font ainsi l'objet d'un prétraitement (cf. section 3.3). Elles sont d'abord filtrées passe-haut avec une fréquence de coupure de 2 Hz, afin d'atténuer les composantes de très basse fréquence et de supprimer les dérives lentes parfois observées dans les signaux EEG [Delorme et Makeig, 2004; Clerc et al., 2016]. Ces dérives se traduisent par des variations progressives de l'amplitude du signal au cours du temps, souvent liées à des fluctuations temporaires du contact entre l'électrode et le cuir chevelu [Shahbakhti et al., 2021]. Deux filtres coupe-bande sont ensuite appliqués à 50 et 100 Hz pour supprimer les interférences liées au réseau électrique. Les *epochs* sont ensuite extraites entre 1 seconde après le début du stimulus et 1 seconde avant sa fin (i.e. 9 secondes après le début du stimulus), de manière à ne conserver qu'une phase d'écoute stable. Ce choix permet d'exclure les phases transitoires liées à l'apparition et à la disparition du stimulus, et de ne conserver que les périodes où l'utilisateur est effectivement engagé dans l'écoute. Les *epochs* sont ensuite soumises à un rejet manuel des artefacts. Pour l'expérimentation menée à 60 phones, 3.75% des *epochs* sont ainsi rejetées, tout sujets confondus, contre 6.45% des *epochs* pour l'expérimentation menée à 66 phones.

Pour le calcul du CSP utilisé dans l'estimation du RSB_{CSP}, les données prétraitées sont ensuite filtrées passe-bande autour de la fréquence de modulation, afin de réduire le bruit et d'atténuer les composantes situées en dehors de la bande d'intérêt [Sen et al., 2023]. Une seule étude relative aux SSAEP [Krishnan et al., 2009] aborde le choix de la bande passante et suggère une largeur

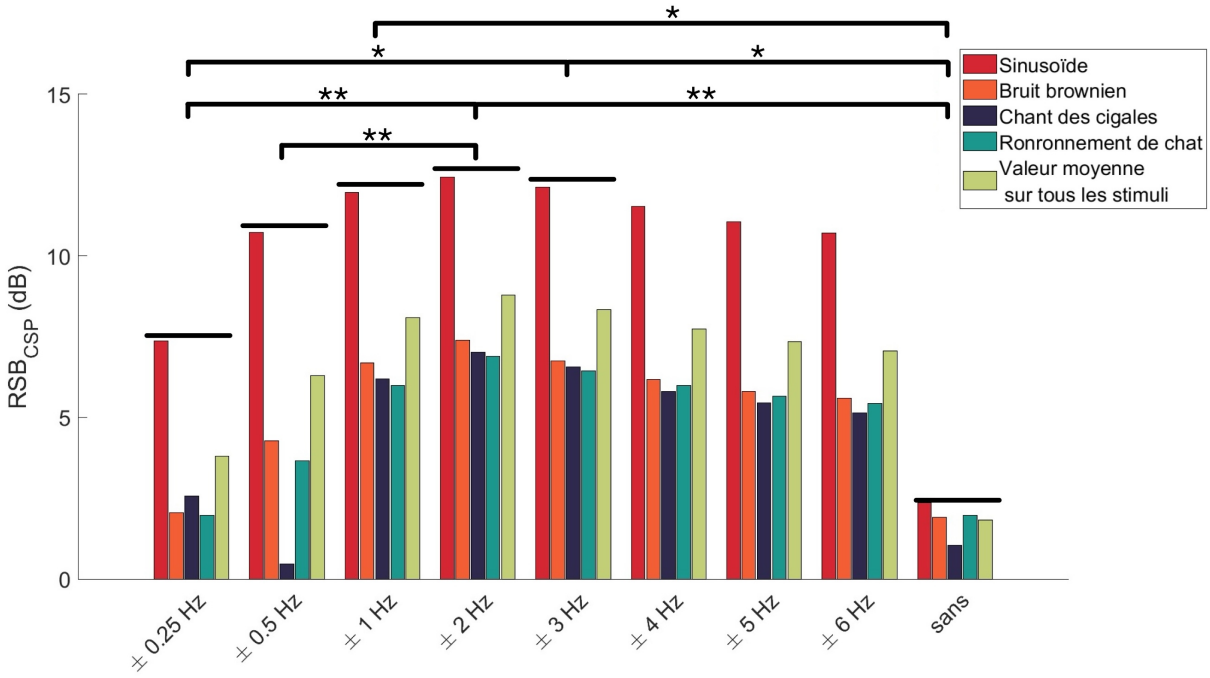
de l'ordre de ±1 Hz (cf. section 3.3.1). Nous avons donc choisi d'étudier l'influence de différentes largeurs de bande ($\pm 0.25, 0.5, 1, 2, 3, 4, 5, 6$ Hz), ainsi que celle de l'absence de filtre passe-bande, sur le RSB_{CSP} pour l'expérimentation menée à 60 phones, compte tenu de la similarité des valeurs de RSB obtenues pour les deux niveaux de sonie. Cette analyse repose sur l'utilisation d'un filtre numérique de type IIR implémenté sous forme de SOS. Cette approche consiste en une cascade de filtres d'ordre 2 (i.e. biquadratiques), appliqués successivement au signal, ce qui permet d'améliorer la stabilité numérique.

L'ensemble des traitements liés au calcul du RSB est réalisé sous MATLAB. Le rejet manuel des artefacts est effectué à l'aide de la bibliothèque EEGLAB [Delorme, 2011].

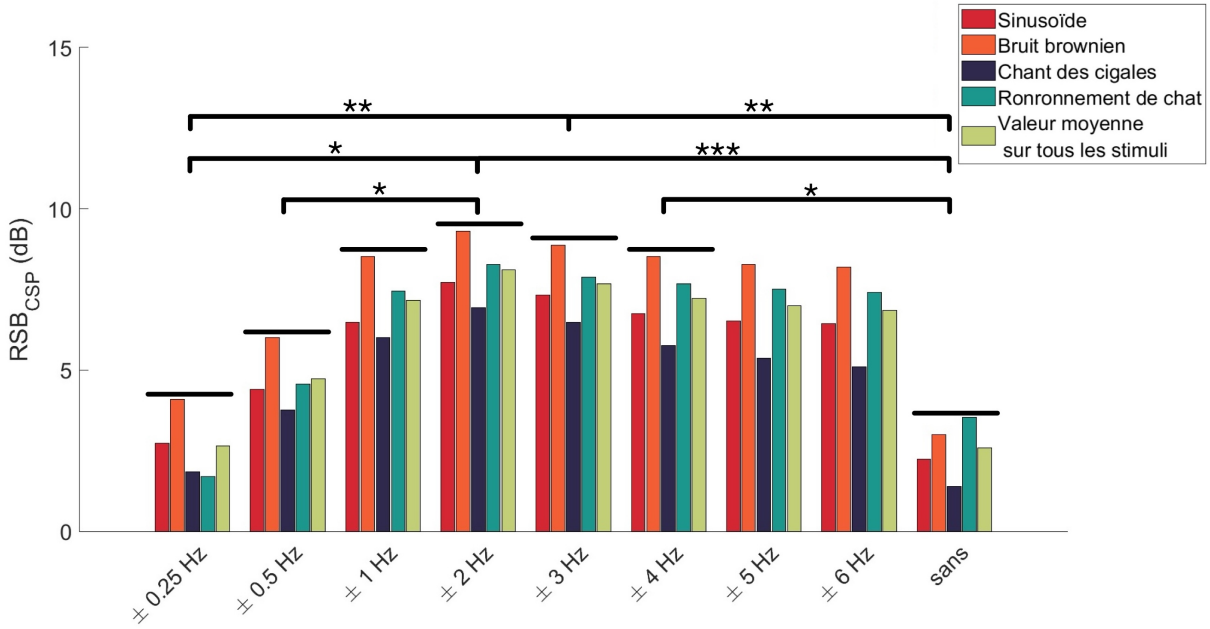
5.2.1.1 Filtrage passe-bande

Le RSB_{CSP} moyenné sur tous les sujets, calculé à la fois à 40 et à 80 Hz sur l'ensemble des conditions, augmente avec la largeur de bande jusqu'à ±3 Hz, ce qui suggère une augmentation du signal utile par rapport au bruit de fond, et donc une meilleure détectabilité de la fréquence de modulation (figure 5.4). Au-delà de cette valeur, il diminue. Des différences significatives sont mises en évidence en fonction de la bande passante, à 40 Hz et à 80 Hz (test de Friedman, $p < 0{,}001$). Les valeurs les plus élevées sont obtenues avec une bande passante de ±2 Hz, significativement supérieures à celles obtenues avec des bandes plus étroites (±0,25 Hz, ±0,5 Hz), ainsi qu'en l'absence de filtrage passe-bande. Ces résultats confirment l'intérêt du filtrage autour de la fréquence de modulation, qui permet d'augmenter significativement le RSB, d'environ 1 dB pour une bande passante de ±0,25 Hz jusqu'à 6 dB pour une bande passante de ±2 Hz, indépendamment du type de stimulus.

Sur la base de ces résultats, la bande passante est fixée à ±2 Hz pour le calcul du RSB_{CSP}.



(a) Filtre passe-bande centré sur 40 Hz appliqué aux signaux EEG lors du prétraitement.



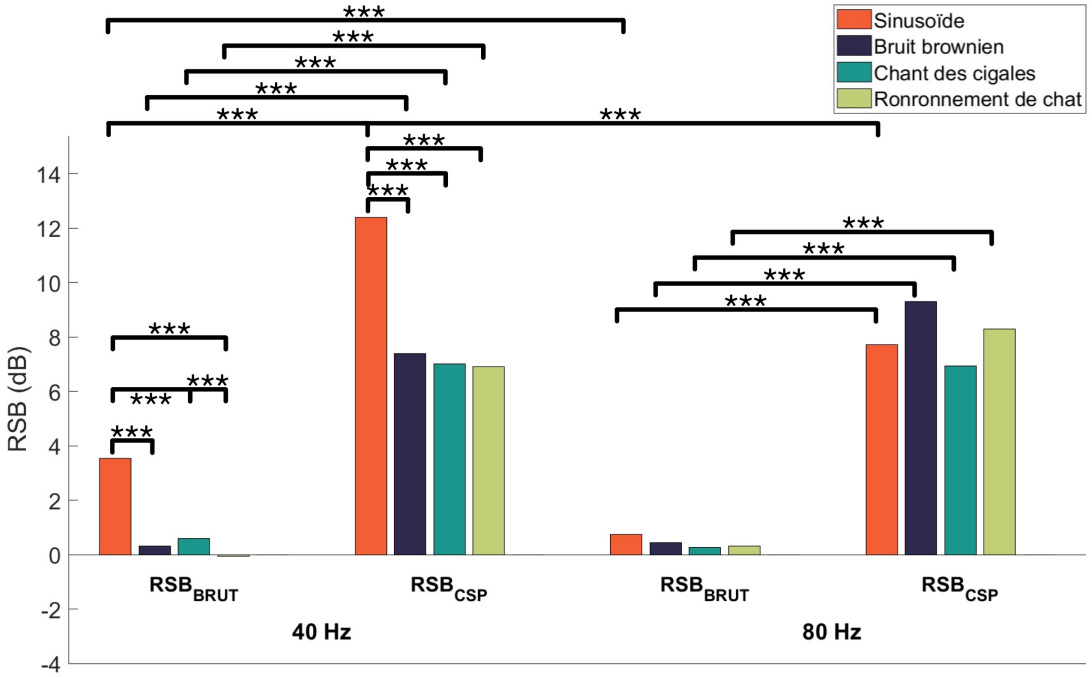
(b) Filtre passe-bande centré sur 80 Hz appliqué aux signaux EEG lors du prétraitement.

FIGURE 5.4 – Expérimentation menée à 60 phones : évolution du RSB_{CSP} (dB) en fonction de la largeur de bande des filtres passe-bande, centrés sur 40 et 80 Hz, appliqués aux signaux EEG lors du prétraitement. * indique $p < 0.05$, ** indique $p < 0.01$ et *** indique $p < 0.001$ (test des rangs signés de Wilcoxon avec correction de Bonferroni).

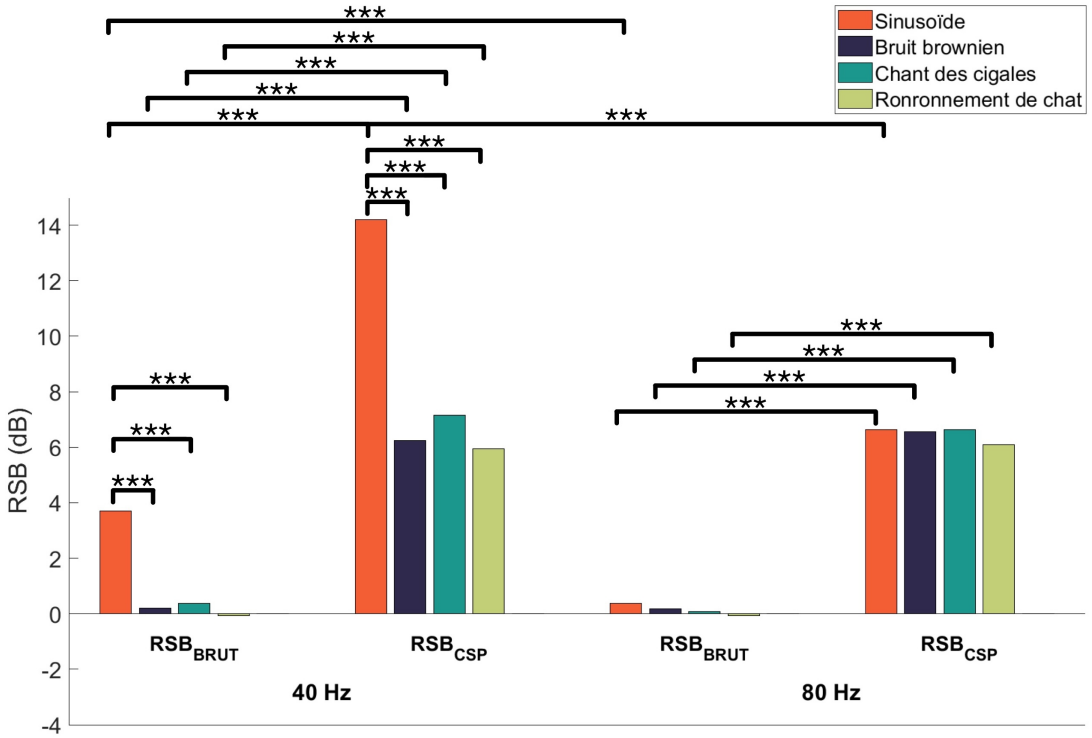
5.2.2 Détectabilité de la fréquence de modulation

Quel que soit le niveau de sonie, les valeurs du $\mathbf{RSB_{BRUT}}$ calculées à 40 Hz diffèrent significativement selon le stimulus (test de Friedman, $p < 0{,}001$). Le son pur présente les valeurs les plus élevées, significativement supérieures à celles des autres stimuli (figure 5.5). À 60 phones, le RSB_{BRUT} du chant des cigales diffère également significativement de celui du ronronnement de chat. D'ailleurs, le RSB_{BRUT} associé à ce dernier est négatif à 40 Hz, à 60 comme à 66 phones, indiquant que la puissance du bruit est supérieure à celle du signal d'intérêt. À 80 Hz, le RSB_{BRUT} est globalement plus faible qu'à 40 Hz. À 60 phones, cette diminution est observée pour la sinusoïde et le chant des cigales, tandis qu'à 66 phones, elle concerne l'ensemble des stimuli. Toutefois, la différence entre 40 et 80 Hz n'est significative que pour la sinusoïde. Enfin, l'augmentation du niveau de sonie de 60 à 66 phones n'a pas d'effet significatif sur le RSB_{BRUT}, indépendamment du type de stimulus considéré. Ce résultat infirme notre hypothèse initiale selon laquelle une augmentation de l'intensité sonore perçue pourrait améliorer le RSB. Il suggère qu'une augmentation de cette intensité pourrait également s'accompagner d'une hausse du bruit dans les signaux EEG, limitant ainsi l'amélioration attendue du RSB. En effet, pour les 6 participants ayant pris part aux deux conditions de niveau sonore, le RSB_{BRUT} est plus faible lors de l'expérimentation menée à 66 phones que lors de celle menée à 60 phones, avec une différence moyenne d'environ 0,4 dB à 40 Hz et de 0,2 dB à 80 Hz. Toutefois, la plage de niveaux sonores explorée dans notre étude demeure limitée, ce qui ne permet pas de tirer de conclusions générales quant à l'effet du niveau sonore.

Le $\mathbf{RSB_{CSP}}$ est significativement supérieur au RSB_{BRUT}, pour l'ensemble des stimuli, aux deux fréquences (40 et 80 Hz) et pour les deux niveaux de sonie considérés (figure 5.5). Le gain observé est d'environ 7 dB pour le bruit Brownien, le chant des cigales et le ronronnement de chat, et d'environ 8 dB pour la sinusoïde. À 40 Hz, le son pur présente les valeurs les plus élevées, significativement supérieures à celles des autres stimuli (plus de 5 dB à 60 phones et plus de 7 dB à 66 phones), comme pour le RSB_{BRUT}. Ainsi, l'application d'un filtre spatial de type CSP, visant à maximiser la variance des réponses EEG à des stimuli modulés en amplitude tout en minimisant la variance des réponses EEG au silence, apparaît déterminante. La comparaison entre 40 et 80 Hz met en évidence des tendances similaires à celles observées pour le RSB_{BRUT}. Pour la sinusoïde, le RSB_{CSP} est significativement plus faible à 80 Hz qu'à 40 Hz. En revanche, pour les autres stimuli, les différences restent faibles et non significatives, avec une légère augmentation observée pour le bruit Brownien et le ronronnement de chat. Enfin, aucune différence significative n'est observée entre les niveaux de sonie (60 et 66 phones), ce qui confirme les résultats obtenus avec le RSB_{BRUT}.



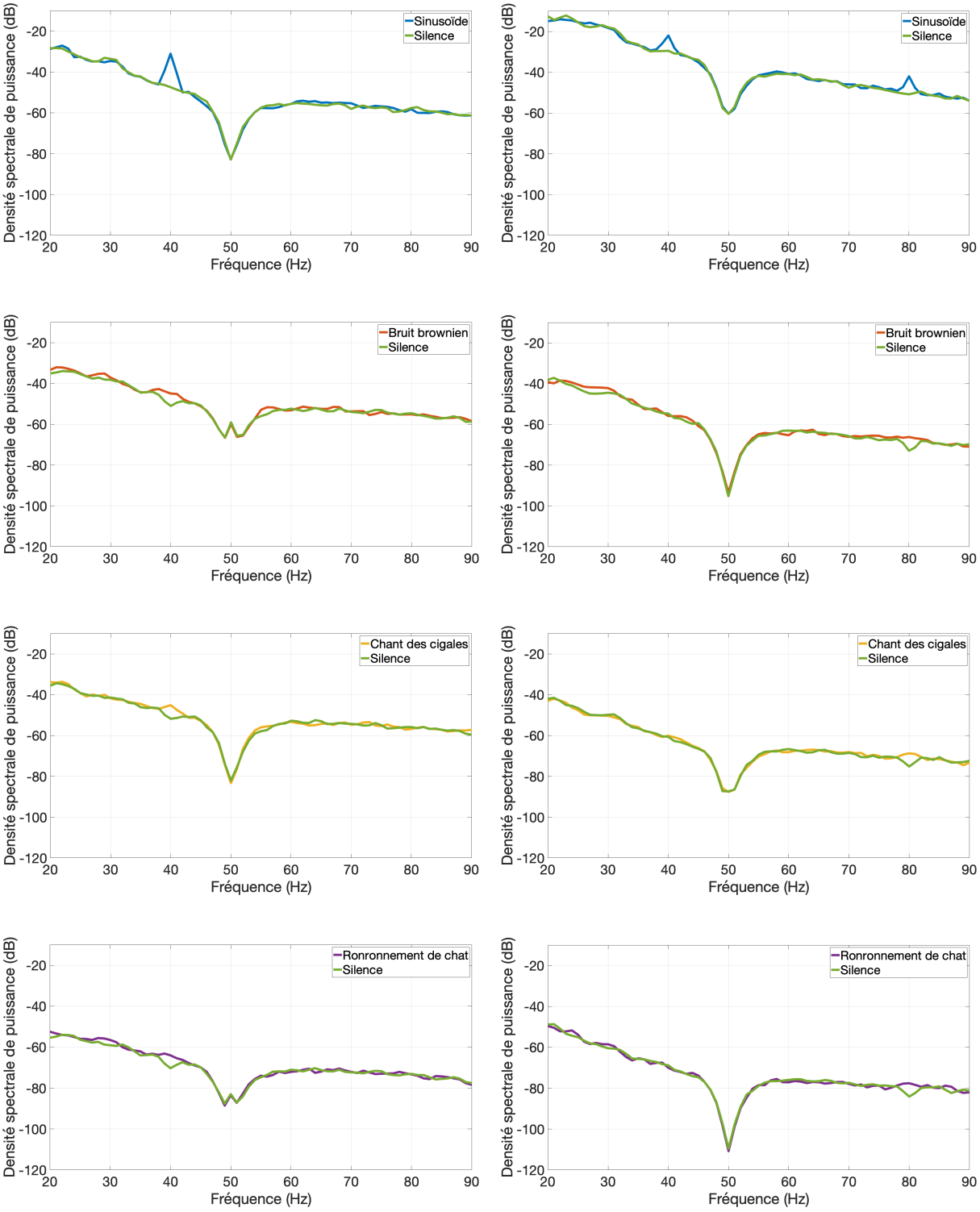
(a) Expérimentation menée à 60 phones.



(b) Expérimentation menée à 66 phones.

FIGURE 5.5 – Valeurs des RSB_{BRUT} et RSB_{CSP} pour chaque stimulus, moyennées sur l'ensemble des sujets. *** indique $p < 0{,}001$ (test des rangs signés de Wilcoxon avec correction de Bonferroni).

Bien que le RSB_{CSP} calculé à 80 Hz soit inférieur à celui obtenu à 40 Hz pour le son pur, il atteint néanmoins près de 10 dB, ce qui signifie que la puissance du signal utile est environ dix fois supérieure à celle du bruit. Pour les autres stimuli, en particulier le bruit Brownien et le ronronnement du chat, le filtrage passe-bande des signaux EEG à la fréquence harmonique conduit à une augmentation du RSB_{CSP}. Ces résultats soulignent l'intérêt de prendre en compte les fréquences harmoniques afin d'améliorer la détection de la fréquence de modulation. La figure 5.6 présente la première composante du CSP, calculée, pour un *run* et un sujet donnés, à partir de l'ensemble des *epochs* associés à une stimulation donnée, ainsi qu'à la condition de silence. Les filtres spatiaux sont estimés à partir de signaux EEG filtrés passe-bande autour de 40 et 80 Hz, puis appliqués aux signaux EEG sans filtrage passe-bande. Les périodogrammes obtenus conservent ainsi une énergie répartie sur l'ensemble du spectre fréquentiel, bien que les filtres spatiaux aient été optimisés à partir des bandes fréquentielles d'intérêt. Cependant, pour certains participants et certaines conditions expérimentales, l'analyse de la première composante CSP met en évidence des cas de défaillance du calcul. Dans ces situations, la composante présente une amplitude anormalement élevée. L'examen des signaux EEG associés révèle que certaines électrodes sont défectueuses, c'est-à-dire qu'elles sont trop bruitées pour fournir une information exploitable. Le filtrage spatial opéré par le CSP vise en principe à atténuer ces canaux bruités. Néanmoins, lorsque leur amplitude domine largement celle des autres électrodes, ils peuvent être interprétés à tort comme des sources informatives et être amplifiés, conduisant à une estimation erronée des composantes CSP, souvent anormalement élevée. Dans ces cas particuliers, l'utilisation d'un filtre spatial estimé pour le même sujet dans une autre condition (e.g. sinusoïde versus silence) permet une estimation plus fiable des composantes. Par ailleurs, même en l'absence de filtrage spatial, une moyenne des canaux permet d'obtenir des périodogrammes exploitables. Ces observations suggèrent que la robustesse du CSP face aux canaux bruités constitue un point critique, qui nécessite des investigations complémentaires.



(a) Filtrage passe-bande centré sur 40 Hz, appliqué lors du prétraitement précédant le calcul du CSP.

(b) Filtrage passe-bande centré sur 80 Hz, appliqué lors du prétraitement précédant le calcul du CSP.

FIGURE 5.6 – Périodogramme du premier filtre spatial CSP appliqué aux signaux EEG prétraités, sans filtrage passe-bande. Les filtres CSP sont calculés, pour un sujet et un *run* donnés, à partir des *epochs* associées à chaque stimulation, ainsi qu'à la condition de silence (courbes vertes).

5.3 Classification

5.3.1 Prétraitements des signaux EEG préalables à l'application de classifieurs

Avant l'application des méthodes de classification, seulement trois étapes de prétraitement sont appliquées : un filtrage passe-bande, un filtrage coupe-bande centré à 50 Hz et un fenêtrage temporel. Aucun filtre coupe-bande supplémentaire à 100 Hz n'est utilisé, le filtre passe-bande retenu étant considéré comme suffisamment étroit pour atténuer efficacement les composantes fréquentielles éloignées de la fréquence de modulation ou de son harmonique. En revanche, le filtre coupe-bande centré à 50 Hz est conservé, cette composante étant plus proche de la fréquence de modulation et donc davantage susceptible d'altérer les performances de classification. De plus, l'interférence liée au courant alternatif présente généralement une amplitude plus importante à 50 Hz qu'à son harmonique, à 100 Hz, ce qui motive l'application de ce filtre, afin d'assurer la suppression de cette composante. Des paramètres tels que le type de filtre, sa largeur de bande et sa bande de transition peuvent influencer directement les performances de classification, de même que la durée des fenêtres temporelles. Il est donc nécessaire d'en évaluer l'impact de manière à identifier des paramètres de filtrage et de fenêtrage adaptés. Il est à noter qu'aucune suppression d'artefacts n'est effectuée et que l'ensemble des canaux est conservé, afin de préserver la dimension des matrices de covariance (et de l'espace tangent associé), laquelle dépend du nombre de canaux. La suppression d'un canal dans l'ensemble d'entraînement nécessiterait sa suppression dans l'ensemble de test, de manière à garantir la cohérence. Par ailleurs, une correction de *baseline* est appliquée afin de recentrer les signaux EEG. Pour cela, la moyenne du signal calculée sur l'intervalle d'une seconde précédant l'apparition du stimulus est soustraite à chaque *epoch* [Sasatake et Matsushita, 2025; Marissens Cueva et al., 2025].

L'influence du filtrage passe-bande et celle de l'extraction temporelle sont évaluées en analysant les performances du classifieur MDM (cf. section 3.5.2.4). Ce classifieur est retenu pour sa simplicité, son faible nombre d'hyperparamètres à ajuster, mais également parce qu'il repose sur la géométrie riemannienne, comme une grande partie des méthodes évaluées dans ce travail. Ainsi, une amélioration des performances du MDM traduit une meilleure séparabilité des classes. Cette séparabilité se traduisant par une augmentation des distances entre matrices de covariance, à la fois sur la variété riemannienne et dans l'espace tangent, les améliorations observées avec le MDM sont susceptibles de bénéficier également aux autres approches riemanniennes évaluées. Une évaluation intra-session (cf. section 3.5.2.6) est réalisée, cette stratégie étant la plus couramment utilisée dans la littérature [Chevallier et al., 2024]. La précision est ensuite moyennée sur l'ensemble des sujets. La tâche de classification consiste à détecter la fréquence de modulation portée par des sinusoïdes, en comparaison avec une condition de silence, sans modulation. Les données utilisées pour cette analyse proviennent de l'expérimentation réalisée à 60 phones.

5.3.1.1 Filtrage passe-bande

Trois filtres couramment utilisés pour le traitement des signaux EEG sont considérés : deux filtres IIR, de types Butterworth et Bessel, ainsi qu'un filtre FIR [Sen et al., 2023]. La bande passante (variable) de chaque filtre est centrée sur la fréquence de modulation. Les filtres IIR sont d'ordre 4, tandis que le filtre FIR dispose de 16501 coefficients (pour une fréquence d'échantillonnage de 500 Hz). Cette longueur élevée s'explique par le choix d'une bande de transition vers la bande atténuée très étroite, fixée à 0,1 Hz, afin d'obtenir un filtrage sélectif autour de la fréquence de modulation et de limiter ainsi la contribution des composantes fréquentielles voisines. L'évaluation est réalisée à l'aide de fenêtres temporelles de 2 secondes sans recouvrement.

Les résultats (figure 5.7) mettent en évidence que pour des bandes passantes (B) strictement supérieures à 0,08 Hz, le filtre de Bessel offre les meilleures performances, devant le filtre de Butterworth et le FIR, avec un optimum obtenu pour une bande passante d'environ 0,1 Hz. Pour des bandes passantes plus larges, les performances des deux filtres IIR diminuent d'environ 2% par augmentation de 0,1 Hz, tandis que pour des bandes passantes plus étroites, la diminution est plus marquée (environ 10% par réduction de 0,01 Hz). Cette dégradation, observée pour les plus petites largeurs de bande, peut s'expliquer par la difficulté des filtres IIR à obtenir une sélectivité fréquentielle élevée avec un ordre limité. La très faible bande passante requise lors de la conception du filtre peut aussi entraîner de trop faibles atténuations dans la zone de transition, voire dans la zone atténuée (autrement dit, le filtre fait ce qu'il peut avec les contraintes données). En revanche, pour le filtre FIR, les performances augmentent légèrement lorsque la bande passante se réduit, probablement en raison du nombre très élevé de coefficients utilisés (16501) par rapport aux filtres IIR, permettant d'obtenir une meilleure sélectivité fréquentielle sans contrepartie. Toutefois, malgré cette amélioration, les performances restent similaires à celles obtenues avec un filtre de Bessel pour $B = 0{,}1$ Hz, alors que le filtre FIR implique une complexité calculatoire plus importante et l'introduction d'un délai de filtrage conséquent qu'il faut alors absolument compenser. Ainsi, un filtre de Bessel d'ordre 4, avec une bande passante de 0,1 Hz centrée sur la fréquence de modulation est retenu pour le prétraitement des données en amont des méthodes de classification.

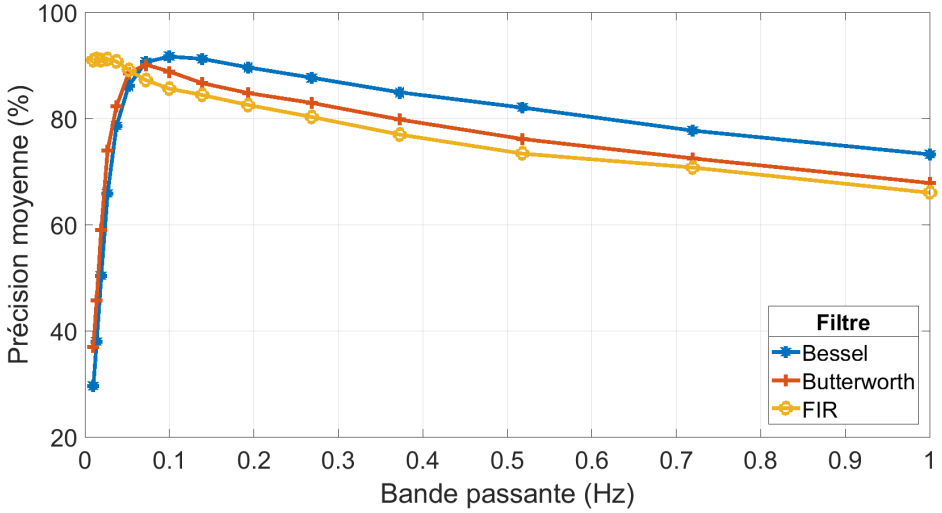


FIGURE 5.7 – Expérimentation menée à 60 phones : influence du type de filtre et de la largeur de bande sur les performances de classification, en utilisant des fenêtres de 2 secondes sans recouvrement. Les résultats sont moyennés sur l'ensemble des sujets.

Ces résultats soulignent l'importance du choix du filtre et de ses paramètres. En particulier, une isolation précise de la fréquence de modulation apparaît déterminante, avec une amélioration des performances de classification pouvant dépasser 15%.

5.3.1.2 Extraction temporelle

La figure 5.8 présente l'évolution des performances du classifieur en fonction de la taille des fenêtres temporelles, pour différentes configurations de recouvrement avec le filtre de Bessel optimal (ordre 4, $B = 0{,}1$ Hz), ainsi que sans recouvrement pour différentes largeurs de bande

avec un filtre de Bessel d'ordre 4. Quelle que soit la valeur du recouvrement, les courbes obtenues se superposent, indiquant que ce paramètre n'a pas d'impact significatif sur les performances du classifieur, contrairement à la taille des fenêtres et à la bande passante du filtre. De fait, plus la bande passante du filtre diminue, plus la précision augmente. Concernant la taille des fenêtres, l'augmentation à 2000 échantillons (soit 4 secondes) conduit aux meilleurs résultats. Toutefois, une interaction entre ces deux paramètres est observée : pour des bandes passantes supérieures ou égales à 0,5 Hz, l'augmentation de la durée des fenêtres améliore les performances d'environ 8%. En revanche, pour des bandes plus étroites, les performances diminuent lorsque la taille de la fenêtre augmente, les meilleurs résultats étant obtenus pour des tailles comprises entre 1000 et 2500 échantillons (2 à 5 secondes). Bien qu'un optimum soit observé autour de 2000 échantillons (4 secondes), l'écart de performance avec des fenêtres plus courtes reste limité. La réduction de la durée des fenêtres permet en outre de segmenter les enregistrements en un plus grand nombre d'*epochs*, augmentant ainsi le nombre d'observations disponibles pour l'apprentissage et l'évaluation. Des fenêtres comprises entre 750 et 1500 échantillons (1,5 à 3 secondes) offrent ainsi un compromis entre le nombre d'*epochs* disponibles et la détectabilité des classes. Par conséquent, les données sont finalement segmentées en *epochs* de 1000 échantillons (2 secondes), sans recouvrement, avant l'application des classifieurs.

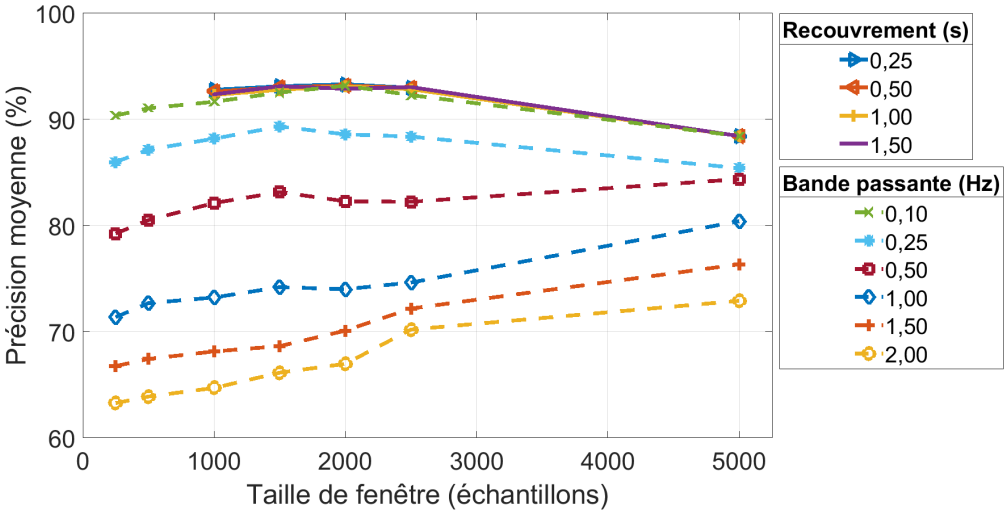


FIGURE 5.8 – Expérimentation menée à 60 phones : influence de la taille des fenêtres pour différentes configurations de recouvrement avec le filtre de Bessel optimal (ordre 4, $B = 0{,}1$ Hz), ainsi que sans recouvrement pour différentes largeurs de bande avec un filtre de Bessel d'ordre 4. Les résultats sont moyennés sur l'ensemble des sujets.

Les résultats suggèrent que l'influence de la taille des fenêtres diminue à mesure que la qualité du filtrage s'améliore, ce qui est cohérent avec des travaux antérieurs montrant que l'allongement des fenêtres favorise une meilleure estimation des matrices de covariance, en accumulant davantage d'informations pertinentes [Chevallier et al., 2018; Kalunga et al., 2016]. Ainsi, une bande passante plus large peut être compensée par l'utilisation de fenêtres plus longues, en particulier dans des scénarios de classification en ligne où les contraintes liées au filtrage sont plus marquées. Toutefois, des fenêtres excessivement longues peuvent introduire des délais incompatibles avec une utilisation en temps réel et augmenter la charge de calcul.

5.3.2 Résultats de détection de la fréquence de modulation

Le prétraitement et la classification des signaux EEG sont effectués hors ligne. Par conséquent, les contraintes habituellement associées à un traitement en temps réel ou pseudo-temps réel ne s'appliquent pas (cf. section 3.2.5).

Comme tous les stimuli non silencieux partagent la même fréquence de modulation (40 Hz), il n'est pas pertinent de comparer directement les réponses cérébrales associées à ces différentes stimulations (e.g. le chant des cigales et le ronronnement de chat), puisque la fréquence détectable dans les signaux cérébraux demeure identique dans tous les cas. Une telle comparaison ne serait justifiée que dans un problème multi-classe, où les stimuli seraient modulés à des fréquences distinctes, permettant ainsi de déclencher différentes actions. Le problème doit alors être formulé comme une classification binaire opposant, pour chaque stimulus, une stimulation sonore à une condition de silence.

Pour cette tâche de classification, dix chaînes de traitement (*pipelines*) sont retenues (figure 5.9).

- La méthode CSP+LDA (*Common Spatial Pattern* + *Linear Discriminant Analysis*, cf. sections 3.4.3 et 3.4.4) est utilisée comme approche de référence, en raison de son emploi répandu dans la littérature, et notamment par Heo et al. [2017], de sa simplicité de mise en œuvre et de son efficacité reconnue pour l'analyse de signaux EEG.
- Les approches TS+SVM et TS+LR (*Tangent Space* + *Support Vector Machine* et *Logistic Regression*, cf. sections 3.5.2.4, 3.5.2.2 et 3.5.2.1), employées dans des paradigmes SSVEP [Chevallier et al., 2024], sont également incluses.
- Afin de comparer l'utilisation de la SVM dans l'espace tangent (TS+SVM) et dans l'espace riemannien, une version appliquée directement aux matrices de covariance, notée RiSVM (*Riemannian Support Vector Machine*, cf. section 3.5.2.4), est également considérée.
- Par ailleurs, une étape de réduction de dimension est intégrée à TS+LR, donnant lieu à la variante TS+KPCA+LR (*Tangent Space* + *Kernel Principal Component Analysis* + *Logistic Regression*, cf. sections 3.5.2.4, 3.4.2.1 et 3.5.2.1). La méthode KPCA est privilégiée à la PCA, les travaux existants montrant que les caractéristiques qu'elle extrait permettent d'obtenir de meilleures performances de classification [Pan et al., 2025; Tayeb et al., 2014].
- Enfin, compte tenu de la quantité limitée de données EEG [Lashgari et al., 2020] et de la complexité des modèles DL comportant des milliers de paramètres entraînables, seules des architectures simples sont retenues (cf. section 3.5.2.5) : EEGNet, SCN (ShallowConvNet) et GREEN (*Gabor Riemannian* EEGNet).

La CCA n'est pas considérée, en raison de son incapacité à détecter un état de repos (en l'absence de stimulation), ce qui ne lui permet pas de discriminer les réponses cérébrales associées à une stimulation sonore de celles observées en condition de silence (cf. section 3.5.2.3).

L'ensemble des méthodes d'apprentissage automatique est implémenté en Python. Les approches classiques reposent sur la bibliothèque scikit-learn [Pedregosa et al., 2011], tandis que les méthodes riemanniennes s'appuient sur pyRiemann [Barachant et al., nd]. Les architectures de DL, quant à elles, sont mises en œuvre à l'aide de la bibliothèque Braindecode [Schirrmeister et al., 2017].

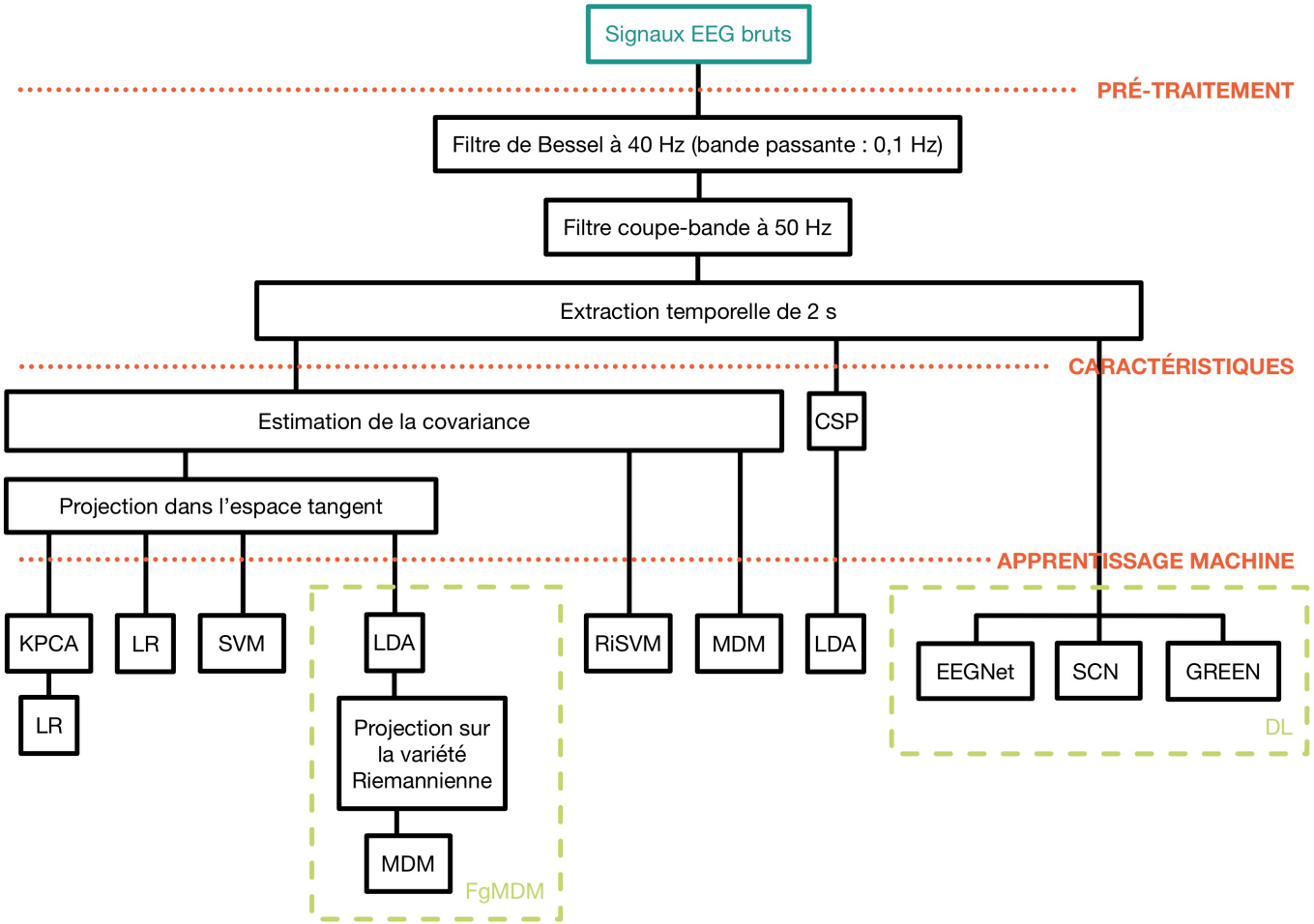


FIGURE 5.9 – Vue d'ensemble des *pipelines* utilisés dans cette étude. KPCA : *Kernel Principal Component Analysis*, LR : *Logistic Regression*, SVM : *Support Vector Machine*, LDA : *Linear Discriminant Analysis*, MDM : *Minimum-Distance-to-Mean*, FgMDM : *Fisher Geodesic Minimum-Distance-to-Mean*, RiSVM : *Riemannian Support Vector Machine*, CSP : *Common Spatial Pattern*, SCN : *ShallowConvNet*, GREEN : *Gabor Riemannian* EEGNet, DL : *Deep Learning*.

Etant donné que les données ont été acquises au cours d'une seule session, il n'est pas possible de réaliser une évaluation inter-session, laquelle nécessite au moins deux sessions. L'évaluation repose donc sur deux stratégies (cf. section 3.5.2.6) : une approche intra-session et une approche inter-sujet.

Les hyperparamètres jouent un rôle clé dans l'obtention de performances optimales. Toutefois, il n'existe pas de règle permettant de déterminer directement les valeurs les plus adaptées pour des paramètres tels que le facteur de régularisation ou le nombre de filtres spatiaux. L'ajustement des classifieurs, à l'exception des modèles de DL, est donc effectué à l'aide d'une recherche par grille (*grid search*, tableau 5.1), où un ensemble de valeurs est défini pour chaque paramètre et toutes les combinaisons possibles sont testées systématiquement. Concernant les modèles de DL, les hyperparamètres retenus s'appuient sur des configurations déjà optimisées lors de travaux préliminaires, au cours desquels une procédure de recherche par grille avait été réalisée. Ces configurations ont été conservées dans cette étude afin de limiter le nombre de paramètres à ajuster, tout en maintenant une cohérence avec les résultats obtenus pour les approches classiques.

Etape	**Hyperparamètres**	**Valeurs**
CSP	*n_filters*	[2, 4, 8, 12, 16]
LDA	-	-
MDM	*metric*	[riemann, logeuclid]
FgMDM	*metric*	[riemann, logeuclid]
TS	-	-
LR	*C*	[0.001, 0.01, 0.1, 1, 10]
KPCA	*n_components*	[2, 8, 16, 24, 32, 40, 80]
	kernel	[linear, rbf]
SVM	*C*	[0.001, 0.01, 0.1, 1, 10]
	kernel	[linear, rbf]
RiSVM	*C*	[0.001, 0.01, 0.1, 1, 10]

TABLE 5.1 – Hyperparamètres utilisés pour la recherche par grille lors de l'évaluation WS.

Pour l'évaluation WS, le choix est fait d'utiliser une validation croisée imbriquée (*nested cross-validation*) [Parvandeh et al., 2020] dans le cadre de la recherche par grille. Cette approche permet de sélectionner les hyperparamètres du modèle, tout en fournissant une estimation non biaisée de ses performances après optimisation des hyperparamètres [Wainer et Cawley, 2021].

En revanche, pour l'évaluation CS, une validation croisée de type *leave-one-out* est mise en œuvre.

Les hyperparamètres pour les modèles de DL, en conditions WS et CS, sont présentés dans le tableau 5.2.

Hyperparamètres	Valeurs WS	Valeurs CS
Général		
Batch Size	16	16
Loss	BCEWithLogits	BCEWithLogits
Optimizer	Adam	Adam
Epochs	20	30
Learning Rate	0,0005	0,001
EEGNet		
F1	2	2
D	12	12
F2	24	24
kernel_length	32	32
Batch Size	16	16
SCN		
n_filters_time	2	2
n_filters_spat	8	12
pool_time_length	100	100
pool_time_stride	25	25
pool_mode	mean	mean
Batch Size	16	32
GREEN		
n_freqs	5	5
kernel_width_s	0,1	0,1
hidden_dim	32	100
bi_out	24	24
Batch Size	16	32

TABLE 5.2 – Hyperparamètres des modèles de DL.

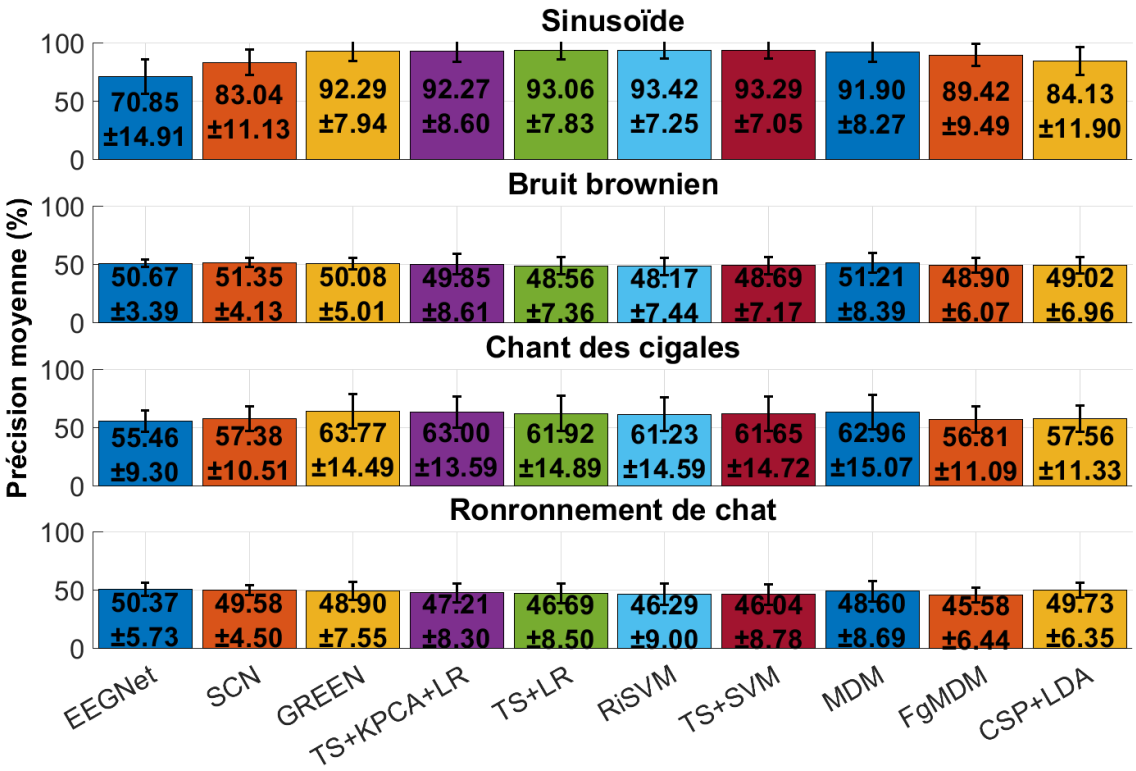
5.3.2.1 Evaluation intra-session (Within-Session ou WS)

Indépendamment du niveau sonore de présentation des stimuli, la sinusoïde obtient les performances de classification les plus élevées, avec une précision moyenne supérieure à 88% (figure 5.10). Les résultats sont plus faibles pour le chant des cigales, mais demeurent supérieurs au niveau de chance (50%), autour de 60%, indiquant que la fréquence de modulation reste détectable. Les précisions obtenues pour le son pur et le chant des cigales diffèrent significativement selon les *pipelines* (test de Friedman avec correction de Bonferroni, $p < 0{,}001$ dans tous les cas). À l'inverse, pour le bruit Brownien et le ronronnement de chat, les performances des algorithmes ne diffèrent pas du niveau du hasard, suggérant que les réponses cérébrales associées à ces stimuli modulés deviennent indiscernables de celles observées en condition de repos. Ainsi, parmi les stimulations évaluées, le chant des cigales constitue la seule alternative au son pur qui fait ses preuves. Par ailleurs, les performances associées à la sinusoïde et au chant des cigales tendent à augmenter avec le niveau sonore, tandis que celles obtenues avec le bruit Brownien et le ronronnement de chat restent faibles. Toutefois, les performances globales de classification restent similaires entre les deux niveaux de sonie, indépendamment des stimuli ou des classifieurs.

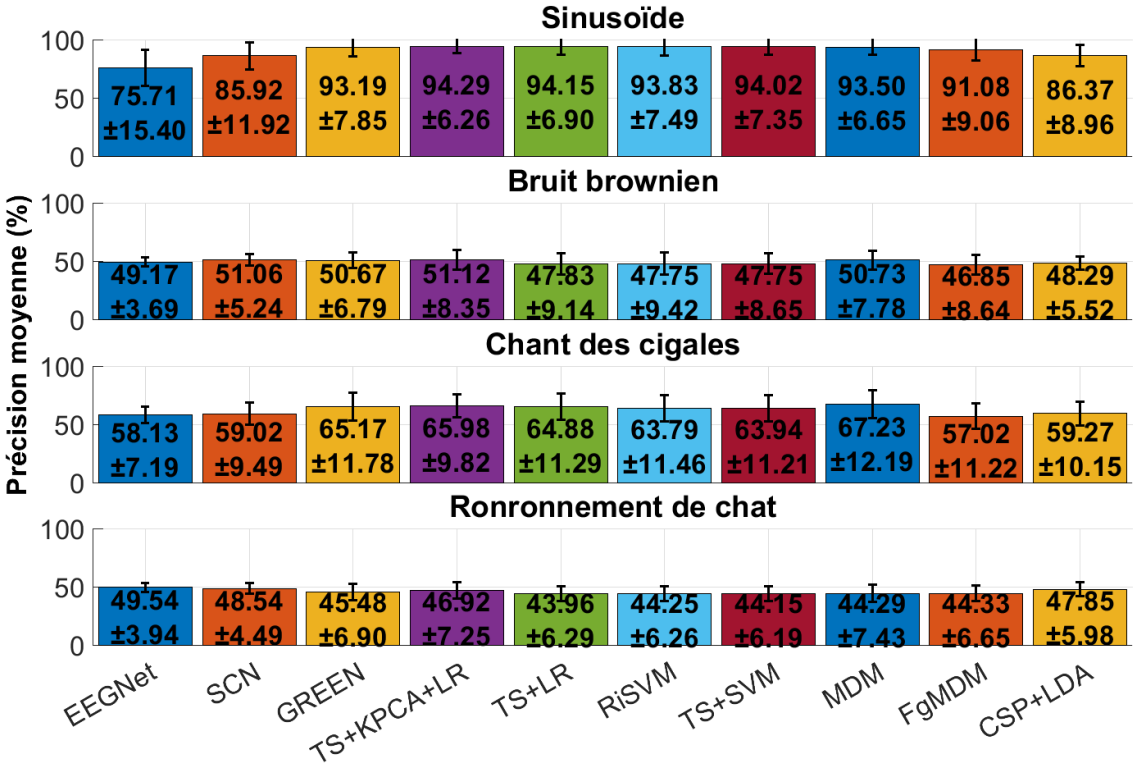
Les résultats obtenus diffèrent de ceux rapportés par Heo et al. [2017], ce qui peut s'expliquer par plusieurs différences méthodologiques. Premièrement, la durée de stimulation utilisée par

Heo et al. est deux fois plus longue (20 secondes), ce qui peut favoriser la détectabilité des SSAEP. Deuxièmement, la tâche expérimentale diffère : dans l'étude de Heo et al., des stimuli distincts sont présentés simultanément à chaque oreille (écoute dichotique) et les participants doivent porter leur attention sur une source spécifique (à gauche ou à droite), tandis que, dans notre protocole, un stimulus est présenté de manière diotique sans consigne d'attention sélective. Troisièmement, les problèmes de classification ne sont pas identiques. Notre approche repose sur une classification de type "un-contre-un", opposant un stimulus modulé à 40 Hz à une condition de repos, tandis que Heo et al. distinguent des stimuli modulés à 38 Hz et à 42 Hz. Enfin, il faut noter que les informations concernant l'intensité sonore et l'égalisation des stimuli ne sont pas décrites dans cette étude [Heo et al., 2017]. Or, ces paramètres peuvent également contribuer aux différences de performance observées. Faute d'informations plus précises, il est difficile d'affiner la comparaison.

En ce qui concerne les performances des différentes chaînes de traitement, la précision varie significativement en fonction du stimulus (test de Friedman avec correction de Bonferroni, $p < 0{,}001$ dans tous les cas). Pour chacune d'elles, les résultats obtenus avec la sinusoïde sont systématiquement et significativement supérieures à ceux des autres stimuli (test des rangs signés de Wilcoxon avec correction de Bonferroni, $p < 0{,}001$ pour chaque comparaison). Les meilleures performances sont obtenues par les algorithmes basés sur le TS, en particulier la RiSVM lorsque l'expérimentation est menée à 60 phones et le TS + KPCA + LR à 66 phones, ainsi que par les modèles GREEN et MDM. Ces modèles surpassent les algorithmes EEGNet, SCN, FgMDM et CSP+LDA (test des rangs signés de Wilcoxon avec correction de Bonferroni, $p < 0{,}001$ pour chaque comparaison). Concernant le chant des cigales, les résultats diffèrent significativement de ceux des autres stimuli selon le *pipeline* considéré : lorsque l'évaluation est menée à 60 phones, cette différence est observée pour les approches basées sur l'espace tangent, ainsi qu'avec GREEN et le MDM ; à 66 phones, elle est observée sur l'ensemble des algorithmes (test des rangs signés de Wilcoxon avec correction de Bonferroni, $p < 0{,}001$ pour chaque comparaison). Des tendances similaires à celles observées avec la sinusoïde sont retrouvées : les méthodes fondées sur le TS, ainsi que GREEN et le MDM, obtiennent également les meilleures performances de détection, tandis que l'EEGNet présente les résultats les plus faibles. Plus généralement, les modèles d'apprentissage profond obtiennent les performances les plus faibles, aussi bien pour la sinusoïde que pour le chant des cigales. Cette observation s'explique vraisemblablement par la quantité limitée de données d'entraînement disponibles.



Expérimentation menée à 60 phones



Expérimentation menée à 66 phones

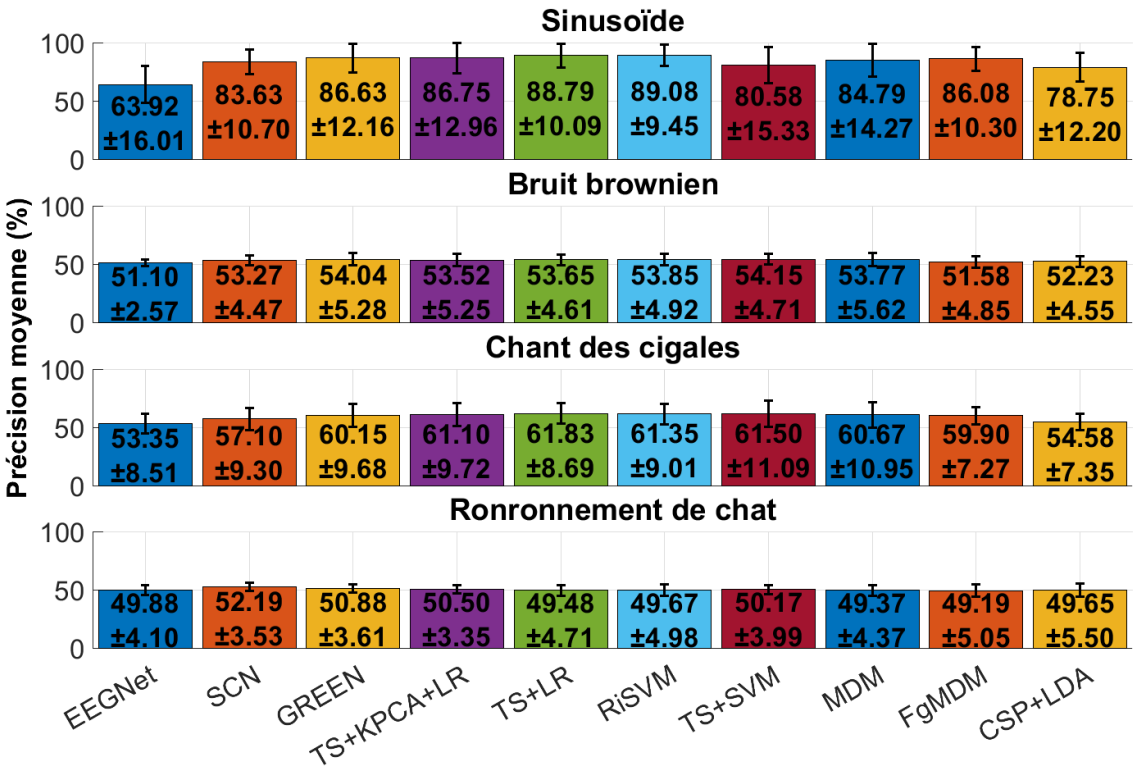
FIGURE 5.10 – Précision moyenne, en WS, calculée sur l'ensemble des sujets pour chaque *pipeline*. Les barres représentent la précision moyenne ± l'écart type.

5.3.2.2 Evaluation inter-sujet (Cross-Subject ou CS)

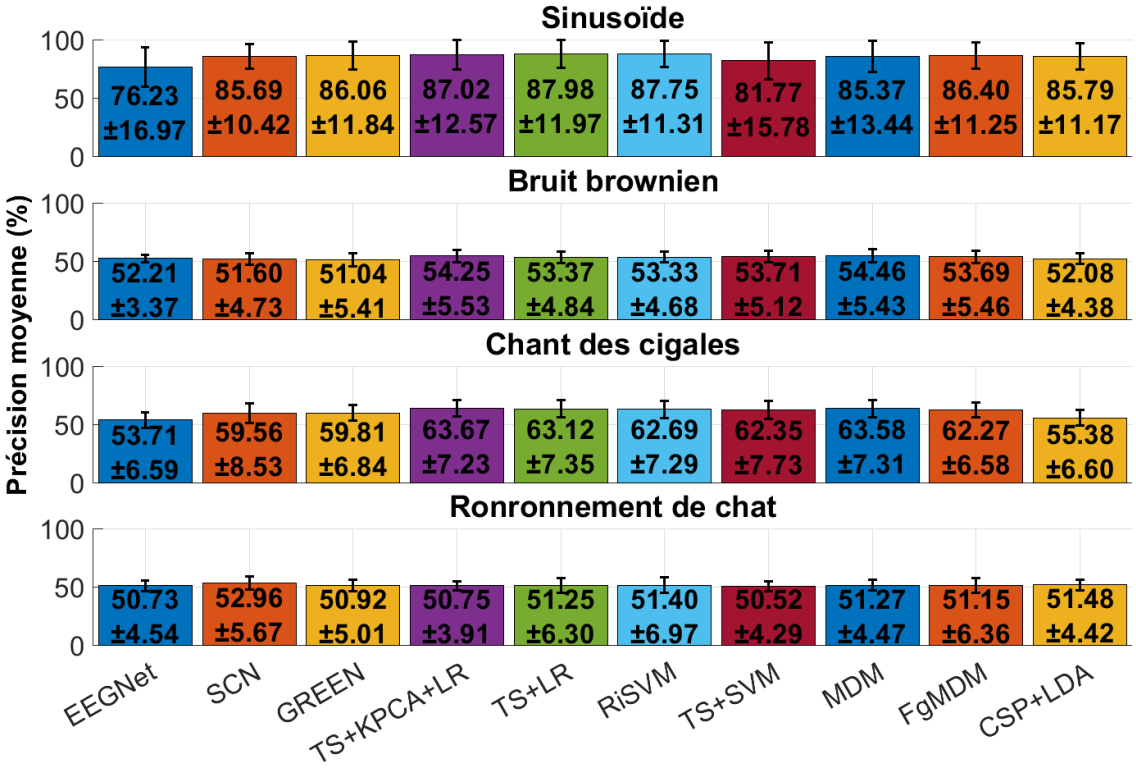
Les résultats pour l'évaluation CS sont globalement similaires à ceux obtenus lors de l'évaluation WS, pour tous les sons porteurs. Indépendamment du niveau sonore, la sinusoïde présente la meilleure détectabilité, avec des précisions supérieures à 83% (figure 5.11), suivie du chant des cigales, dont les performances se situent autour de 60%. Toutefois, les résultats dépendent significativement des *pipelines* considérés (test de Friedman avec correction de Bonferroni, $p < 0{,}001$ dans tous les cas). En revanche, pour les deux autres stimuli, les performances restent au niveau du hasard (50%), pour l'ensemble des algorithmes. Ainsi, pour chaque *pipeline*, les précisions obtenues varient significativement en fonction du stimulus (test de Friedman avec correction de Bonferroni, $p < 0{,}001$ dans tous les cas). Les meilleures performances de détection sont obtenues avec les méthodes basées sur le TS, suivies de près par GREEN et le MDM, à la fois pour le son pur (en particulier la RiSVM à 60 phones et le TS+LR à 66 phones) et pour le chant des cigales (notamment le TS+LR à 60 phones et le TS+KPCA+LR à 66 phones). Enfin, l'EEGNet présente systématiquement les précisions les plus faibles sur l'ensemble des conditions.

Les méthodes basées sur l'espace tangent, ainsi que GREEN et le MDM, obtiennent ainsi les meilleures performances, pour les deux stratégies d'évaluation, aussi bien pour la sinusoïde que pour le chant des cigales, surpassant les autres approches, en accord avec les observations rapportées dans le *Mother Of All BCI Benchmarks* (MOABB) [Chevallier et al., 2024]. Parmi ces méthodes, l'algorithme MDM présente des avantages notables, notamment un besoin limité en réglage d'hyperparamètres, un temps d'entraînement réduit et une bonne robustesse. Il est généralement préféré à GREEN, bien que ce dernier repose également sur les matrices de covariance et leur projection dans l'espace tangent, en raison de sa plus faible complexité computationnelle et de sa simplicité de mise en œuvre. À l'inverse, les modèles d'apprentissage profond (EEGNet, SCN), ainsi que la chaîne de traitement de référence, le CSP+LDA, présentent systématiquement les performances les plus faibles. L'évaluation des approches de DL a volontairement été simplifiée, afin de privilégier des méthodes légères et faciles à paramétrer, comme les approches riemanniennes, tout en garantissant un apprentissage stable et en limitant le surapprentissage compte tenu du nombre restreint de fenêtres temporelles disponibles par sujet. Ce choix méthodologique a toutefois pu désavantager les modèles de DL.

En général, la classification inter-sujet introduit des difficultés supplémentaires liées à la variabilité interindividuelle, entraînant une diminution des performances par rapport à l'approche intra-session. Cependant, pour le son pur et le chant des cigales, les taux de détection supérieurs à 83% et 60%, respectivement, indiquent que les modèles conservent une bonne capacité de généralisation, non seulement entre les sessions, mais également entre les sujets. Ces résultats suggèrent que les réponses cérébrales induites par une sinusoïde modulée à 40 Hz sont robustes d'un individu à l'autre et globalement détectables.



Expérimentation menée à 60 phones



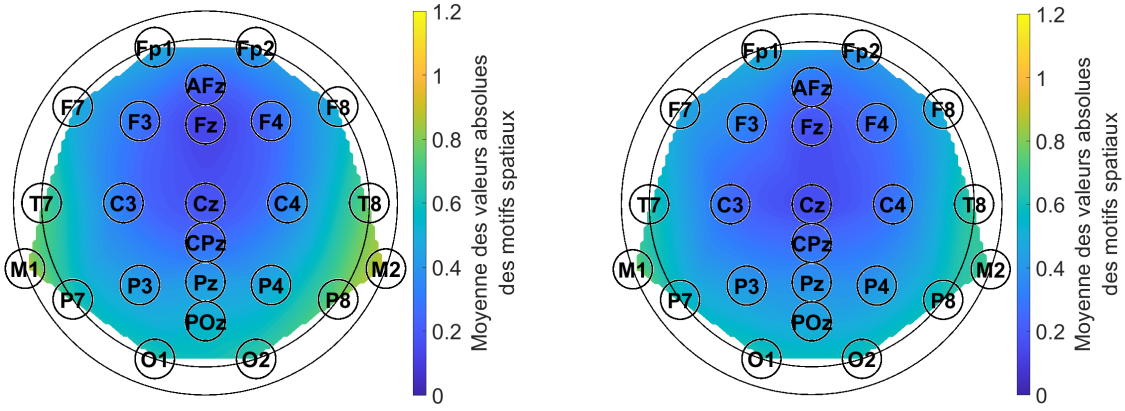
Expérimentation menée à 66 phones

FIGURE 5.11 – Précision moyenne, en CS, calculée sur l'ensemble des sujets pour chaque *pipeline*. Les barres représentent la précision moyenne ± l'écart type.

5.4 Cartographie cérébrale

La cartographie cérébrale permet de représenter la distribution spatiale de l'activité cérébrale. Dans cette étude, elle est utilisée de manière à identifier les régions cérébrales les plus impactées lors de l'écoute des sons modulés en amplitude. Pour ce faire, des motifs spatiaux sont calculés, afin de déterminer les électrodes contribuant le plus fortement à la discrimination entre les réponses cérébrales évoquées par les stimulations sonores et celles enregistrées en condition de silence. Les motifs spatiaux ($\mathbf{A}$), correspondant à des poids attribués à chaque électrode, sont dérivés des filtres spatiaux contenus dans la matrice de projection ($\mathbf{W}_{\text{CSP}}$) obtenue par la méthode du CSP (cf. section 3.4.3). Les colonnes de cette matrice maximisent la variance d'une condition tout en minimisant celle d'une autre condition. Ainsi, elle est calculée, pour chaque stimulus en opposition au silence, sur les deux *runs* réalisés par chaque sujet, puis le premier motif spatial en est extrait. Deux motifs spatiaux sont donc obtenus par stimulus et par sujet, un pour chaque *run*. Les valeurs absolues des poids associés à chaque électrode sont ensuite calculées, puis moyennées entre les sujets, de manière à identifier les électrodes présentant les contributions les plus importantes à l'activité cérébrale observée, à la fois pour chaque stimulus et sur l'ensemble des stimuli.

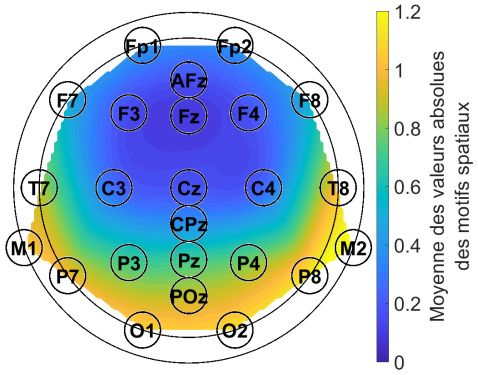
La figure 5.12 représente, pour chaque niveau de sonie, la moyenne des valeurs absolues des poids associés à chaque électrode, calculée sur l'ensemble des sujets et des stimulations. Plus la valeur associée à une électrode est élevée, plus sa contribution à l'activité discriminante (i.e. la distinction entre les réponses cérébrales évoquées par les stimuli auditifs et celles enregistrées en condition de silence) est importante. Ainsi, dans les deux conditions de niveau sonore, l'électrode M1, située à proximité de l'électrode TP9 selon le système international 10–20 (figure 2.6), présente la valeur la plus élevée. Des différences significatives sont observées entre les poids associés aux électrodes (test de Friedman, $p < 0.001$ et $p < 0.01$, respectivement pour l'expérimentation menée à 60 et à 66 phones). Cependant, pour plus de la moitié des électrodes, les poids ne diffèrent pas significativement de celui associé à M1. À 60 phones, onze électrodes diffèrent significativement de M1 : Fp1, Fp2, F3, F4, C3, C4, F7, Fz, AFz, Cz, CPz (test des rangs signés de Wilcoxon avec correction de Holm-Bonferroni, $p < 0{,}001$ pour toutes comparaisons), tandis qu'à 66 phones, seule l'électrode F4 diffère significativement de M1 (test des rangs signés de Wilcoxon avec correction de Holm-Bonferroni, $p < 0{,}01$). L'étendue des zones impliquées tend donc à augmenter avec le niveau sonore.



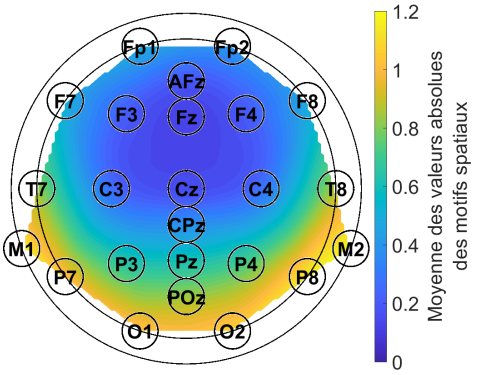
Expérimentation menée à 60 phones | Expérimentation menée à 66 phones

FIGURE 5.12 – Moyenne des motifs spatiaux (valeurs absolues des poids associées à chaque électrode) sur l'ensemble des sujets, pour tous les stimuli confondus en opposition à la condition de silence.

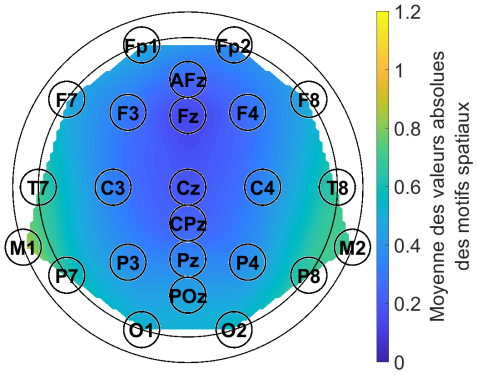
Toutefois, le moyennage des valeurs absolues des poids associés à chaque électrode sur l'ensemble des stimuli peut masquer des informations pertinentes. Il est donc intéressant d'analyser séparément l'activité cérébrale associée à chaque stimulus (figure 5.13). Ainsi, l'électrode présentant la valeur la plus élevée est M2, électrode voisine de TP10 dans le système international 10-20 (figure **??**), pour la sinusoïde et le chant des cigales, et M1 pour le bruit Brownien. Pour le ronronnement de chat, la valeur maximale est observée sur les électrodes T8 et M1, respectivement pour l'expérimentation menée à 60 et 66 phones. Par ailleurs, des différences significatives entre électrodes apparaissent principalement pour la sinusoïde. Pour les autres stimuli, les moyennes des valeurs absolues des poids associés aux électrodes sont globalement homogènes, sans distinction marquée entre électrodes. Néanmoins, cette cartographie de l'activité cérébrale dépend directement du calcul des filtres spatiaux de type CSP. Les erreurs d'estimation décrites dans la section 5.2.2 influencent donc également l'interprétation de ces cartographies. Cependant, ce phénomène n'affecte pas la sinusoïde, permettant ainsi d'interpréter plus fiablement les régions cérébrales impliquées lors de la tâche d'écoute, pendant la phase de test. Dans ce cas, à 60 phones, toutes les électrodes diffèrent significativement de M2 (test des rangs signés de Wilcoxon avec correction de Holm-Bonferroni, $p < 0{,}01$). À 66 phones, quatre électrodes ne diffèrent pas significativement de M2 : O2, P7, P8 et M1 (test des rangs signés de Wilcoxon avec correction de Holm-Bonferroni, $p = 0.016$, $p = 0.017$, $p = 0.291$ et $p = 0.854$, respectivement). Ces résultats suggèrent que, outre les électrodes situées à proximité des oreilles (M1, M2), des électrodes localisées dans les régions occipitales et pariétales (O2, P7, P8) contribuent également à la détection de la fréquence de modulation. Cette observation est cohérente avec des travaux antérieurs [Zhang et al., 2006] et pourrait s'expliquer par des mécanismes de plasticité intermodale. Bien que ce phénomène soit principalement documenté chez les personnes présentant une déficience visuelle, il a également été observé, dans une moindre mesure, au sein du reste de la population. Cette interprétation demeure toutefois spéculative, aucune méthode de localisation de source, permettant d'estimer l'origine de l'activité électrique cérébrale, n'ayant été appliquée.



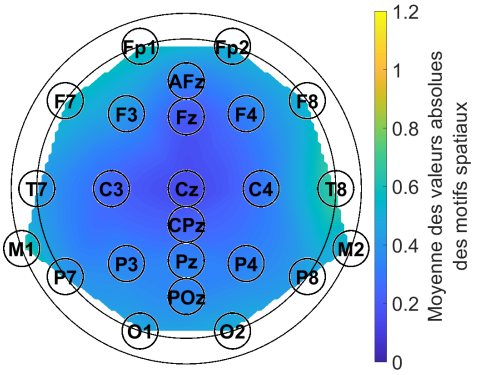
(a) Sinusoïde.



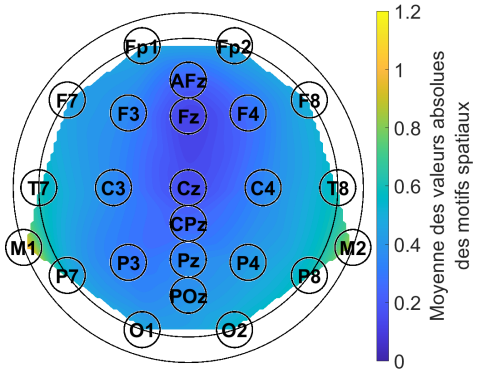
(e) Sinusoïde.



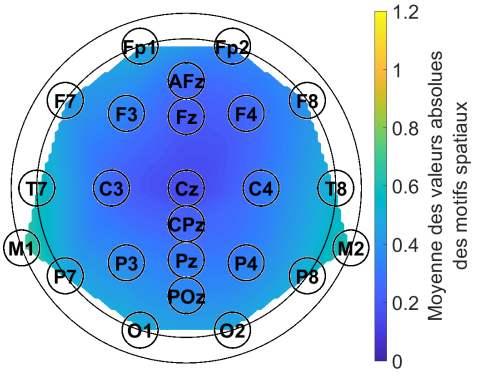
(b) Bruit Brownien.



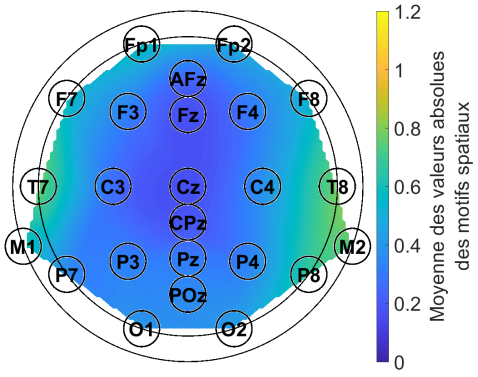
(f) Bruit Brownien.



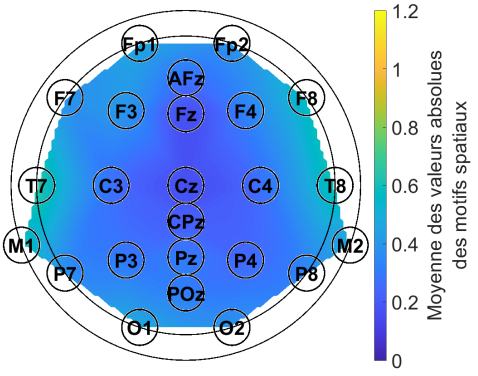
(c) Chant des cigales.



(g) Chant des cigales



(d) Ronronnement de chat.



(h) Ronronnement de chat.

FIGURE 5.13 – A gauche, expérimentation menée à 60 phones ; à droite, expérimentation menée à 66 phones. Moyenne des motifs spatiaux (valeurs absolues des poids associées à chaque électrode) sur l'ensemble des sujets, pour chaque stimulus en opposition à la condition de silence.

5.5 Analyse critique des propriétés des stimulations

Le bruit Brownien et le ronronnement de chat ne constituent pas des signaux porteurs efficaces. Une première hypothèse est que cette inefficacité soit liée à certaines propriétés, fréquentielles ou temporelles, de ces signaux, susceptibles d'empêcher la modulation d'être portée de manière suffisamment distincte ou stable, et ainsi, de limiter sa détectabilité dans les signaux cérébraux.

Dans le cas d'un son pur modulé en amplitude, la modulation se traduit par la présence d'énergie de part et d'autre de la fréquence porteuse, aux fréquences correspondant à celle-ci plus ou moins la fréquence de modulation ($f_c \pm f_m$). Pour le son pur à 1 kHz, ces composantes sont nettement visibles sur le spectrogramme de la figure 4.2, à 960 Hz et 1040 Hz. En revanche, lorsque la modulation est portée par le bruit Brownien, le chant des cigales ou le ronronnement de chat, le contenu spectral de ces signaux masque fortement ces composantes, ce qui empêche leur identification claire.

Dans ce contexte, une analyse quantitative du contenu spectral des stimuli apparaît nécessaire, afin de mieux caractériser leurs propriétés acoustiques. À cette fin, deux descripteurs spectraux sont considérés. Le premier descripteur est le centroïde spectral (i.e. le barycentre du spectre), défini par :

$$\Phi = \frac{\sum_{b=b_1}^{b_2-1} f_b \cdot s_b}{\sum_{b=b_1}^{b_2-1} s_b}, \tag{5.1}$$

où f_b désigne la fréquence associée au *bin* b, s_b la puissance du spectre pour ce *bin*, et b_1 et b_2, les indices délimitant la bande fréquentielle considérée. Le second descripteur correspond à l'étalement spectral moyen (i.e. la largeur de bande moyenne du spectre autour de sa valeur moyenne) [Peeters et al., 2004]. L'étalement spectral est d'abord calculé à chaque instant, puis moyenné sur toute la durée du stimulus. Pour un instant donné, il s'écrit :

$$\Omega = \sqrt{\frac{\sum_{b=b_1}^{b_2-1} (f_b - \Phi)^2 \cdot s_b}{\sum_{b=b_1}^{b_2-1} s_b}}. \tag{5.2}$$

Par ailleurs, dans le but de caractériser le comportement temporel des signaux, la moyenne est calculée par :

$$\mu = \frac{1}{N} \sum_{n=1}^{N} x_n, \tag{5.3}$$

où x_n est la valeur du signal à l'indice n ; la variance par :

$$\sigma^2 = \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu)^2, \tag{5.4}$$

et l'autocovariance au cours du temps par :

$$\gamma(\tau) = \frac{1}{N-\tau} \sum_{n=1}^{N-\tau} (x_n - \mu)(x_{n+\tau} - \mu), \tag{5.5}$$

où τ représente le décalage (i.e. le nombre de pas séparant deux échantillons comparés). Ainsi, si ces trois grandeurs demeurent constantes au cours du temps, le signal est considéré comme stationnaire [Zhivomirov et Nedelchev, 2020; Zhivomirov, 2024].

Le tableau 5.3 synthétise l'ensemble de ces caractéristiques, fréquentielles et temporelles, pour chaque stimulus. Le son pur, le bruit Brownien et le chant des cigales présentent un centroïde spectral élevé, nettement supérieur à la fréquence de modulation, et présentent un comportement stationnaire. Le ronronnement de chat se distingue, quant à lui, par un centroïde spectral faible et une variance qui évolue au cours du temps, traduisant le caractère non stationnaire de ce signal. Par ailleurs, le son pur possède la largeur spectrale la plus étroite, tandis que le bruit Brownien présente la plus large.

	Sinusoïde	**Bruit Brownien**	**Chant des cigales**	**Ronronnement de chat**
Centroïde spectral (Hz)	1007	1082	6241	534
Etalement spectral moyen (Hz)	208	1391	863	492
Moyenne	stable	stable	stable	stable
Variance				instable
Auto-covariance				stable

TABLE 5.3 – Critères utilisés pour comparer les propriétés spectrales et temporelles des signaux porteurs.

Ces observations suggèrent que plusieurs propriétés des signaux porteurs pourraient favoriser la détectabilité de la fréquence de modulation : un centroïde spectral élevé par rapport à la fréquence de modulation, une largeur spectrale relativement étroite et un caractère stationnaire du signal. Toutefois, ces hypothèses reposent uniquement sur des analyses temporelles et spectrales préliminaires, et ne couvrent probablement pas l'ensemble des facteurs pertinents.

5.6 Conclusion

Des alternatives aux sons purs sont évaluées comme stimuli auditifs pour des BCI basées sur les SSAEP. L'analyse porte sur la détectabilité de la fréquence de modulation dans les réponses EEG en fonction du signal porteur, en comparaison avec un son pur à 1 kHz. Cette étude repose sur deux approches complémentaires : d'une part, le calcul du RSB, qui constitue une mesure indicative de l'émergence de la fréquence de stimulation dans les signaux EEG, et, d'autre part, la discrimination des réponses EEG en présence et en absence de stimulation, à l'aide de différentes méthodes de classification. Cette seconde approche permet en outre de comparer plusieurs chaînes de traitement, d'évaluer leurs performances de classification et d'étudier l'impact des étapes de prétraitement sur celles-ci.

Concernant le prétraitement, les résultats mettent en évidence que le filtrage constitue le facteur le plus déterminant, le filtre de Bessel à bande étroite apparaissant comme le plus performant. Quant à l'effet du fenêtrage temporel, il est plus modéré lorsque le filtrage est optimal, mais peut compenser l'utilisation de bandes passantes plus larges.

Les résultats mettent en évidence une influence marquée du signal porteur sur la détectabilité de la fréquence de modulation. La sinusoïde à 1 kHz présente les niveaux de détection les plus élevés, avec des précisions intra-session supérieures à 88%. Le chant des cigales demeure détectable au-delà du niveau du hasard, bien que de manière nettement plus limitée (environ 60%). En revanche, les autres stimuli ne peuvent pas être discriminés de manière fiable avec les méthodes considérées. L'étude des propriétés fréquentielles et temporelles de chacun des stimuli suggère

que la détectabilité de la fréquence de modulation dépend des caractéristiques des signaux porteurs. Les meilleures performances semblent notamment associées à des signaux présentant un centroïde spectral élevé par rapport à la fréquence de modulation, une largeur spectrale relativement étroite et un caractère stationnaire. Par ailleurs, l'augmentation du niveau sonore n'entraîne pas d'amélioration significative des performances, ce qui suggère qu'une hausse de l'intensité sonore pourrait également s'accompagner d'une augmentation du bruit dans les signaux EEG. Ces résultats sont cohérents avec ceux obtenus à partir du RSB, ce qui confirme que cette mesure, relativement simple à mettre en œuvre, constitue un indicateur pertinent de la détectabilité de la fréquence de modulation.

Les classifieurs basés sur la géométrie riemannienne, en particulier ceux reposant sur l'espace tangent, ainsi que GREEN et le MDM, surpassent les autres modèles pour la sinusoïde comme pour le chant des cigales. En revanche, les modèles d'apprentissage profond (EEGNet, SCN) présentent des performances systématiquement plus faibles. Toutefois, leur évaluation ayant été volontairement simplifiée, les résultats obtenus pourraient ne pas refléter pleinement leurs capacités. Ainsi, bien que les méthodes issues des paradigmes SSVEP apparaissent globalement transposables aux SSAEP, des travaux complémentaires restent nécessaires pour exploiter davantage le potentiel des approches d'apprentissage profond. En particulier, une optimisation systématique des hyperparamètres, par exemple au moyen de recherches par grille, pourrait contribuer à améliorer les performances de classification.

L'interprétation de la cartographie de l'activité cérébrale en fonction du signal porteur demeure délicate, en raison des limites associées à l'estimation des filtres spatiaux de type CSP (cf. section 5.2.2). Néanmoins, pour le stimulus sinusoïdal, pour lequel ces filtres sont estimés de manière plus fiable, la détection de la fréquence de modulation implique non seulement les électrodes situées à proximité des oreilles, mais également des régions localisées dans les lobes occipitaux et pariétaux. De plus, l'augmentation du niveau sonore semble s'accompagner d'une extension des zones cérébrales impliquées.

Ces résultats mettent en évidence la nécessité d'étudier la perception des stimuli, ainsi que le profil des participants, afin d'explorer, dans de futurs travaux, d'éventuelles différences interindividuelles en termes de performances de détection de la fréquence de modulation et de régions cérébrales impliquées. L'analyse de la perception des stimuli peut également permettre d'évaluer dans quelle mesure certains stimuli, notamment le chant des cigales, constituent des alternatives pertinentes aux sons purs dans des conditions d'utilisation réelles des BCI. Ces aspects feront l'objet du chapitre suivant (chapitre 6).

6

Evaluation perceptive et caractérisation des participants

Sommaire

6.1 Introduction 129
6.2 Evaluation perceptive des stimuli auditifs et de l'expérimentation 130
6.2.1 Classement des stimuli par ordre de préférence 130
6.2.2 Acceptabilité des stimuli 131
6.2.3 Perception de la sonie des stimuli 132
6.2.4 Acceptabilité et impact émotionnel de l'expérimentation 133
6.3 Profils des participants 133
6.4 Conclusion 134

6.1 Introduction

Dans le chapitre précédent (chapitre 5), il a été mis en évidence que le son pur constitue le signal porteur le plus performant, suivi par le chant des cigales, tandis que le bruit Brownien et le ronronnement de chat ne permettent pas de porter efficacement la fréquence de modulation, de sorte à ce qu'elle soit détectable dans les signaux cérébraux. Ces résultats soulèvent une question fondamentale concernant la sinusoïde : comment celle-ci est-elle perçue par les participants ? En effet, malgré son efficacité, l'utilisation du son pur peut être remise en question dans l'hypothèse où il serait perçu comme désagréable par les participants. Une telle perspective interroge directement la pertinence de poursuivre le développement de BCI reposant sur ce type de stimulus. À l'inverse, il convient d'examiner dans quelle mesure les autres stimuli pourraient constituer des alternatives viables, sous réserve d'une adaptation de leurs caractéristiques permettant de rendre la fréquence de modulation détectable. Par ailleurs, bien que les stimuli aient été égalisés en sonie, l'analyse de la perception subjective de leur intensité demeure nécessaire. De fait, si le signal sinusoïdal est systématiquement perçu comme plus intense au cours des deux expérimentations, cette différence perceptive pourrait contribuer, au moins en partie, à expliquer sa meilleure détectabilité. Enfin, l'évaluation du caractère agréable de l'expérimentation constitue un enjeu central, en particulier dans la perspective d'un déploiement en ligne de BCI basées sur les SSAEP, où l'acceptabilité par les utilisateurs représente un facteur déterminant.

Au-delà des propriétés des stimuli, l'analyse du profil des participants apparaît également essentielle. Les données recueillies, auprès de 48 participants, ne se limitent pas aux enregis-

trements EEG, mais incluent aussi des informations complémentaires, telles que le sexe, l'âge, les substances consommées, la latéralité, la musicalité, ainsi que les capacités auditives. Collectées à l'aide de questionnaires pré- et post-tests, ces données offrent un cadre particulièrement riche pour explorer les sources de variabilité interindividuelle, notamment en ce qui concerne l'amplitude des SSAEP, ainsi que la distribution spatiale des régions cérébrales impliquées.

Ainsi, l'ensemble de ces analyses vise à mieux comprendre les facteurs perceptifs et individuels susceptibles d'influencer les performances observées. Dans un premier temps, les retours perceptifs des participants concernant l'expérimentation, et plus particulièrement les stimuli, sont examinés (section 6.2). Les profils des participants, établis à partir des réponses aux questionnaires, sont ensuite présentés dans la section 6.3. Toutefois, faute de temps, une analyse approfondie des signaux EEG au regard des caractéristiques individuelles n'a pas pu être menée dans le cadre de cette thèse.

6.2 Evaluation perceptive des stimuli auditifs et de l'expérimentation

Les réponses aux questionnaires pré- et post-test, recueillies lors des expérimentations menées à 60 et 66 phones, sont exploitées afin d'identifier les préférences des participants concernant les différents stimuli, de mesurer leur caractère potentiellement désagréable et d'examiner les différences de perception de leur intensité, de manière à vérifier l'efficacité de l'égalisation en sonie. Ces questionnaires permettent également d'évaluer l'appréciation globale de l'expérimentation, ainsi que son influence sur l'état émotionnel des participants, notamment en termes de valence, d'excitation et d'émotion dominante.

6.2.1 Classement des stimuli par ordre de préférence

L'étude des données de classement des stimuli par ordre de préférence (cf. section 4.5.1.4) repose sur le modèle de Plackett–Luce [Turner et al., 2020]. Spécifiquement conçu pour l'analyse de données ordinales, ce modèle probabiliste permet d'estimer un classement global à partir des classements individuels, effectués par les participants, en tenant compte de l'ensemble des positions occupées par chaque stimulus dans les différents classements. Plus précisément, il attribue à chaque stimulus une probabilité d'être préféré relativement aux autres stimuli présentés. Ainsi, un stimulus systématiquement classé en deuxième position peut être considéré comme globalement plus apprécié qu'un stimulus n'occupant la première position qu'occasionnellement, mais étant par ailleurs fréquemment classé parmi les dernières positions.

Les résultats varient en fonction du niveau sonore (figure 6.1).

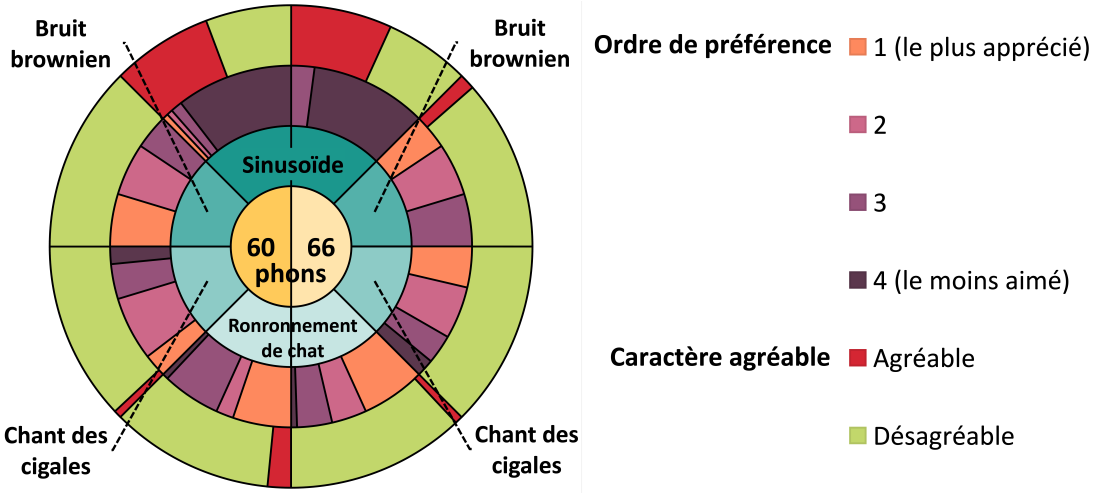


FIGURE 6.1 – Retours subjectifs des participants relatifs à la perception des stimuli auditifs, incluant leur ordre de préférence et leur acceptabilité.

A 60 phones, le bruit Brownien présente la probabilité la plus élevée d'être classé comme stimulus préféré (tableau 6.1), suivi du ronronnement de chat. À l'inverse, à 66 phones, le ronronnement de chat devient le stimulus ayant la plus forte probabilité d'être classé en première position, devant le bruit Brownien. Pour les deux niveaux sonores, le chant des cigales occupe la troisième place, tandis que le son pur présente systématiquement la probabilité la plus faible d'être préféré. Lorsque les données des deux niveaux sonores sont considérées conjointement, le modèle de Plackett–Luce classe le bruit Brownien en première position, devant le ronronnement de chat, le chant des cigales et, enfin, le son pur.

Niveau sonore (phones)	Sinusoïde	Bruit Brownien	Chant des cigales	Ronronnement de chat
60	0,031	0,429	0,256	0,284
66	0,031	0,306	0,272	0,391
60 et 66	0,032	0,365	0,266	0,337

TABLE 6.1 – Probabilités, pour chaque stimulus, d'être classé comme stimulus préféré, à chaque niveau de sonie, ainsi que pour les deux niveaux de sonie confondus, selon le modèle de Plackett–Luce.

6.2.2 Acceptabilité des stimuli

La proportion de participants rapportant une gêne associée aux stimuli varie significativement selon le stimulus considéré, et ce pour les deux niveaux sonores (test de Friedman, $p < 0{,}001$). Dans les deux conditions, plus de la moitié des sujets (54,2%) jugent la sinusoïde désagréable (figure 6.1), soit une proportion significativement plus élevée que pour les autres stimuli (tableau 6.2). À 60 phones, le ronronnement de chat est jugé désagréable par 12,5% des participants et le chant des cigales par 4,17%, tandis qu'aucun participant ne considère le bruit Brownien comme désagréable. Cette distribution diffère à 66 phones : le bruit Brownien et le ronronnement de chat sont jugés désagréables par 8,33% et 4,17% des participants, respectivement, alors qu'aucune gêne n'est rapportée pour le chant des cigales.

Niveau sonore (phones)	Bruit Brownien	Chant des cigales	Ronronnement de chat
60	0,001	0,001	0,023
66	0,002	0,002	0,002

TABLE 6.2 – Test des rangs signés de Wilcoxon : *p-values* comparant, pour chaque niveau sonore, la proportion de participants jugeant la sinusoïde désagréable à celle observée pour les autres stimuli.

6.2.3 Perception de la sonie des stimuli

Bien que les stimuli soient égalisés en sonie, leur intensité n'est pas perçue de manière équivalente par les participants (test de Friedman, $p = 0.033$ et $p = 0.006$, respectivement à 60 phones et à 66 phones). Cette différence peut notamment s'expliquer par la variabilité interindividuelle de la perception de la sonie, les courbes d'isosonie étant établies à partir de réponses moyennes issues d'un large échantillon de population [ISO, 2023; Edjekouane, 2016].

La sinusoïde est le stimulus le plus fréquemment perçu comme plus fort que les autres (figure 6.2), avec des proportions significativement plus élevées que celles observées pour le chant des cigales. Le ronronnement de chat est également perçu comme relativement fort et, à 66 phones, il est jugé significativement plus fort que le chant des cigales. À l'inverse, le bruit Brownien est moins souvent perçu comme plus fort, tandis que le chant des cigales est le stimulus le plus rarement considéré comme tel. Globalement, la sinusoïde et le ronronnement de chat sont significativement plus fréquemment perçus comme plus forts que le bruit Brownien et le chant des cigales.

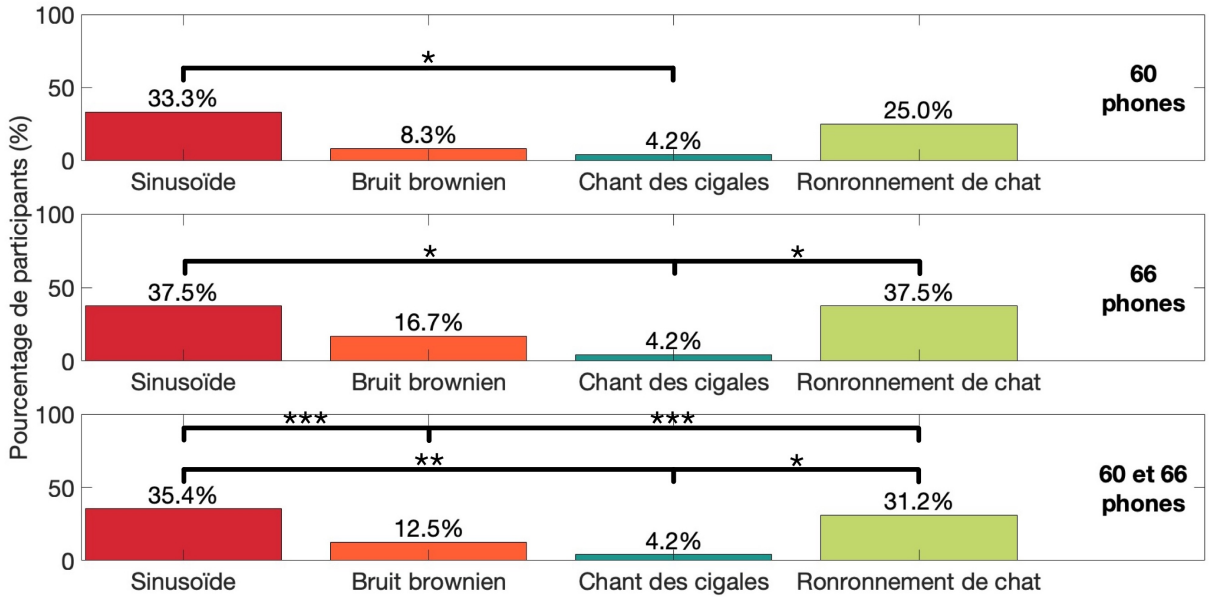


FIGURE 6.2 – Proportion de participants ayant indiqué, pour chaque stimulus, l'avoir perçu comme plus fort que les autres. * indique $p < 0.05$, ** indique $p < 0.01$ et *** indique $p < 0.001$ (test des rangs signés de Wilcoxon).

La perception d'une intensité plus élevée peut être influencée par le caractère potentiellement désagréable de la sinusoïde, ainsi que par des effets de reconnaissance des stimuli : certains

participants n'ont pas identifié le ronronnement de chat et l'ont parfois décrit comme menaçant dans leurs retours informels.

6.2.4 Acceptabilité et impact émotionnel de l'expérimentation

L'expérimentation est jugée agréable, aux deux niveaux sonores, par 87,5% des participants.

Concernant la valence et l'excitation, la différence entre les scores pré- et post-test (évalués sur une échelle de 1 à 9, cf. section 4.5.1.3) est calculée pour chaque sujet, afin de limiter une éventuelle influence de l'état initial des participants sur les données, puis moyennée sur l'ensemble des sujets. Pour l'expérimentation menée à 60 phones, la variation moyenne entre les scores pré- et post-test est de $-0{,}17$ pour la valence et de -1 pour l'excitation. À 66 phones, elle est respectivement de 0,25 et de $-0{,}54$, pour la valence et l'excitation.

Dans les deux conditions, une diminution de l'excitation moyenne est observée, ce qui concorde avec les résultats relatifs à l'émotion dominante (figure 6.3). L'émotion dominante la plus fréquemment rapportée avant le test est "enthousiaste", tandis qu'après le test, il s'agit principalement de l'état "relaxant". En effet, à 60 phones, un seul participant déclare un état "relaxant" avant la phase de test, contre 11 après celle-ci, tandis que le nombre de participants se disant "enthousiastes" passe de 9 à 3. Une tendance similaire est observée à 66 phones : avant la phase de test, 1 participant se déclare "relaxant" et 1 autre "somnolent", alors qu'après celle-ci, 8 participants rapportent un état "relaxant" et 5 un état "somnolent".

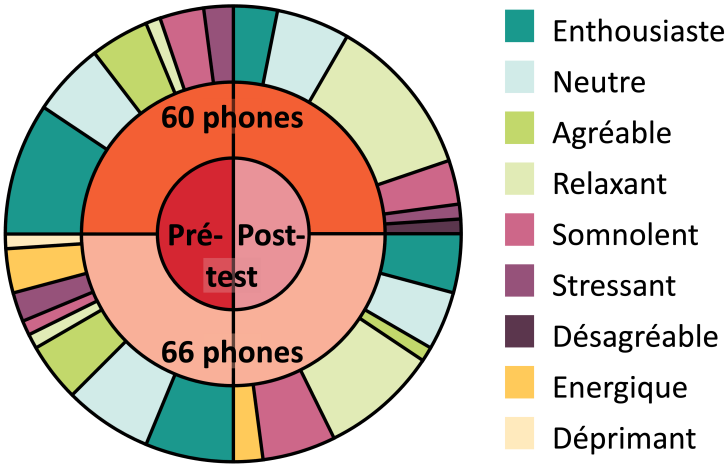


FIGURE 6.3 – Emotion dominante rapportée par les participants avant et après la phase de test, pour chaque niveau de sonie (i.e. 60 et 66 phones).

6.3 Profils des participants

Tous les participants retenus pour cette expérimentation présentent, selon la procédure audiométrique développée, une audition normale pour leur âge, en tenant compte des effets de la presbyacousie, susceptibles d'apparaître à partir de 40 ans (cf. section 4.4.2.2). Toutefois, 5 participants du premier groupe (60 phones) et 3 du second (66 phones) déclarent des acouphènes. La répartition hommes–femmes, ainsi que la distribution des âges, sont similaires entre les deux groupes (figure 6.4). De manière générale, la répartition des profils, en termes de latéralité et de musicalité, est comparable entre les deux conditions : les droitiers et les musiciens y sont majoritaires. Ainsi, dans le premier groupe, un participant est gaucher (4,2%), quatre sont ambidextres

(16,7%) et dix-neuf (79,0%) sont droitiers. Le second groupe comprend également dix-neuf droitiers, ainsi que deux gauchers (8,3%) et trois ambidextres (12,5%). Concernant la musicalité, une différence apparaît toutefois entre les groupes : quinze participants sont musiciens dans le premier (62,5%), contre dix-neuf dans le second (79,0%).

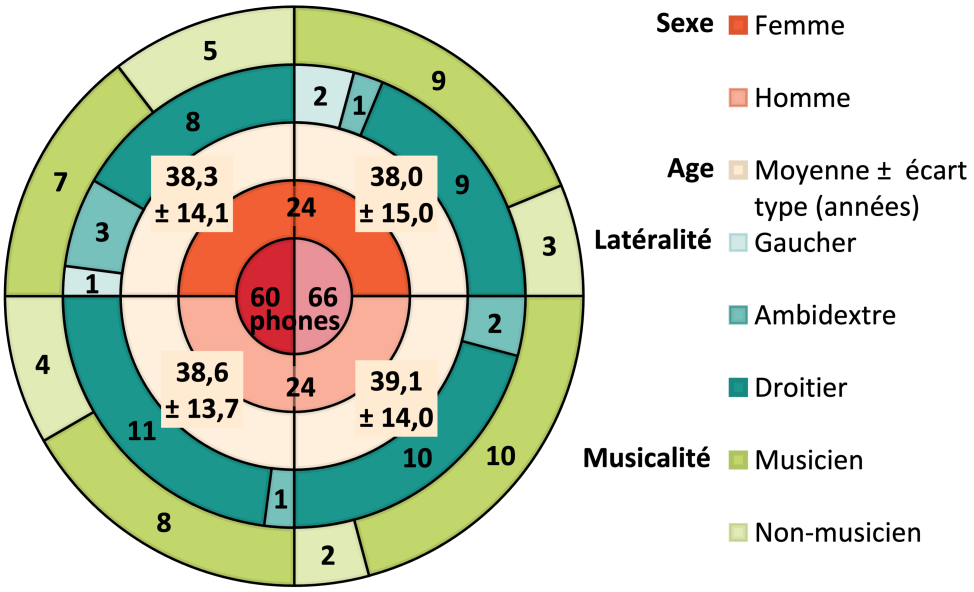


FIGURE 6.4 – Description quantitative et qualitative des participants à l'expérimentation.

6.4 Conclusion

Les résultats perceptifs mettent en évidence un décalage marqué entre performance et acceptabilité des stimuli. Le son pur, bien qu'il demeure le signal porteur le plus efficace, est jugé désagréable par plus de la moitié des participants. Cette faible acceptabilité constitue une limite importante pour son utilisation dans des BCI auditives, en particulier dans le cadre d'usages prolongés. À l'inverse, les stimulations alternatives, telles que le bruit Brownien, le chant des cigales et le ronronnement de chat, sont globalement perçues comme agréables et figurent parmi les préférées des participants. Toutefois, cette préférence ne s'accompagne pas d'une détectabilité satisfaisante de la fréquence de modulation, ce qui limite, à ce stade, leur utilisation directe en tant que signaux porteurs. Ces résultats indiquent néanmoins que ces stimuli pourraient constituer des alternatives pertinentes, à condition d'adapter leurs propriétés, afin d'améliorer leur efficacité, tout en préservant leur confort auditif. Par ailleurs, l'expérimentation est jugée agréable par une large majorité des participants et est même associée, à l'issue de celle-ci, à un état émotionnel "relaxant". Ces éléments suggèrent que ce type de paradigme ne constitue pas, en l'état, un frein à son acceptation par le grand public.

L'analyse de la perception de l'intensité sonore révèle également que, malgré l'égalisation en sonie, le son pur est plus souvent perçu comme plus intense que les autres stimuli : environ un tiers des participants le jugent plus fort, et ce dans les deux conditions expérimentales. Cette différence perceptive pourrait contribuer à expliquer son efficacité en tant que signal porteur, sans toutefois remettre en cause sa supériorité relative, observée de manière cohérente aux deux niveaux de sonie. Dans cette perspective, il serait pertinent d'examiner plus spécifiquement les

participants rapportant cette perception accrue de l'intensité de la sinusoïde, afin de déterminer si celle-ci est associée à des réponses cérébrales de plus grande amplitude.

Enfin, la répartition équilibrée entre femmes et hommes constitue un point fort de cette expérimentation au regard de la littérature, encore souvent marquée par des biais de genre. Certaines catégories restent toutefois sous-représentées, en particulier les participants gauchers et les non-musiciens, ce qui devra être pris en compte dans de futurs travaux.

Dans l'ensemble, ces résultats soulignent que l'optimisation des BCI basées sur les SSAEP ne peut reposer uniquement sur les performances des stimuli, mais doit également intégrer leur acceptabilité.

7

Conclusion et perspectives

Sommaire

7.1 Synthèse générale 137
7.2 Perspectives 139
7.2.1 Conception de nouveaux signaux porteurs 139
7.2.2 Influence de la fréquence de modulation, du niveau sonore et de la durée de stimulation sur les performances 139
7.2.3 Analyse de la variabilité interindividuelle 140

7.1 Synthèse générale

Ces travaux ont conduit au développement d'un paradigme expérimental dédié à l'évaluation de différents signaux porteurs dans le cadre de BCI basées sur les SSAEP. Ce paradigme intègre à la fois des considérations acoustiques, perceptives et humaines, afin de contrôler précisément la reproduction des stimulations sonores, la perception de celles-ci et d'interpréter les réponses cérébrales en tenant compte des spécificités individuelles.

Ainsi, l'expérimentation a permis de constituer une base de données, comprenant 48 participants, qui sera prochainement mise à disposition de la communauté scientifique. Les signaux EEG ont été enregistrés en réponse à quatre stimuli auditifs, ainsi qu'à une condition de silence. Une unique fréquence de modulation a été utilisée, l'objectif principal étant d'évaluer l'influence des signaux porteurs sur l'activité cérébrale. Les stimulations sonores ont été soigneusement conçues et traitées, de manière à garantir leur restitution fidèle, indépendamment du matériel de reproduction. Elles ont également été égalisées en sonie pour limiter l'influence du contenu fréquentiel sur l'intensité perçue. L'étude a par ailleurs été menée à deux niveaux de sonie, afin d'examiner l'effet de ce paramètre sur la détectabilité de la fréquence de modulation. Les enregistrements EEG sont accompagnés de métadonnées incluant des variables individuelles, telles que le seuil d'audition, la latéralité, la musicalité, la consommation de substances psychoactives et l'état d'esprit.

Afin de mesurer le seuil d'audition de chaque participant de manière à garantir une perception correcte des stimuli, un test audiométrique simplifié a été implémenté, ne nécessitant pas de matériel normalisé. Cette procédure est adaptable et pourra être réutilisée dans de futures études, sous réserve de disposer d'un HATS pour la calibration ou de développer une méthode alternative. Les questionnaires élaborés pour collecter des données individuelles ont également été conçus dans une perspective de réutilisation et d'adaptation à d'autres protocoles expérimentaux.

Les signaux EEG ont fait l'objet de divers traitements, parmi lesquels le filtrage passe-bande et le fenêtrage temporel ont été étudiés plus spécifiquement. Les résultats mettent en évidence que le filtrage passe-bande centré sur la fréquence de modulation constitue le facteur le plus déterminant. En particulier, l'utilisation d'un filtre de Bessel à bande étroite améliore significativement les performances de classification. L'effet du fenêtrage temporel, quant à lui, apparaît plus modéré lorsque le filtrage est optimisé, bien qu'il puisse compenser l'utilisation de bandes passantes plus larges. Toutefois, ces résultats ont été obtenus à l'aide d'un unique classifieur (MDM) et demeurent donc susceptibles de dépendre de la méthode de classification considérée. L'optimisation des performances du MDM peut être interprétée comme une augmentation de la séparabilité entre les classes, aussi bien sur la variété riemannienne que dans l'espace tangent. Dans cette perspective, les paramètres optimisés pourraient donc ne pas bénéficier uniquement au classifieur MDM, mais également améliorer les performances d'autres méthodes basées sur la géométrie riemannienne, dans la mesure où celles-ci exploitent une structure géométrique similaire des données. Cette hypothèse reste néanmoins à confirmer expérimentalement, certaines méthodes pouvant exploiter des propriétés ne se réduisant pas à la seule séparabilité globale. En outre, les résultats présentés ont été obtenus uniquement à partir des données acquises lors de l'expérimentation menée à 60 phones, les performances de classification étant similaires pour les deux niveaux de sonie. Ce choix ne permet néanmoins pas d'exclure que l'analyse de données acquises à un autre niveau sonore puisse conduire à l'identification de paramètres optimaux différents. Une étude comparable devrait donc être réalisée dans le cadre de toute nouvelle expérimentation, afin d'identifier les paramètres de prétraitement les plus adaptés aux données considérées.

La détectabilité de la fréquence de modulation a été évaluée à l'aide du RSB, puis par l'application de différentes méthodes de classification issues de l'apprentissage automatique, incluant des approches classiques, des modèles riemanniens, ainsi que des algorithmes d'apprentissage profond. Dix chaînes de traitement ont ainsi été comparées. Les alternatives au son pur testées ne parviennent pas à égaler le niveau de détectabilité de la référence à 1 kHz, dont les précisions intra-session dépassent 88%. Le chant des cigales atteint des performances d'environ 60%, tandis que le bruit Brownien et le ronronnement de chat ne permettent pas d'obtenir une classification fiable avec les méthodes considérées. Les approches basées sur la géométrie riemannienne, en particulier celles reposant sur l'espace tangent, ainsi que les méthodes GREEN et MDM, se révèlent les plus performantes. Par ailleurs, bien que l'écart de sonie considéré demeure relativement faible, aucun effet significatif du niveau sonore sur les performances de classification n'a été observé.

L'identification des régions cérébrales associées au traitement des stimuli est limitée par la robustesse des critères d'évaluation utilisés (cf. section 5.2.2). Néanmoins, l'écoute du son pur met en évidence l'activation de régions situées à proximité des oreilles, ainsi que de zones situées dans les lobes occipitaux et pariétaux. Ces observations sont cohérentes avec l'état de l'art sur les aires cérébrales impliquées dans le traitement auditif (chapitre 1), qui indique que ce processus sollicite le cortex auditif primaire, situé sur la face supérieure du lobe temporal, ainsi que des réseaux plus étendus. Ces derniers incluent notamment des aires associatives réparties dans les lobes temporaux, frontaux, pariétaux et occipitaux des deux hémisphères, intervenant dans la discrimination des catégories de sons et leur localisation.

Enfin, les résultats relatifs à la perception des stimuli par les participants mettent en évidence une faible acceptabilité de la sinusoïde, jugée peu agréable par plus de la moitié d'entre eux. À l'inverse, les stimuli naturels et artificiels sont globalement préférés, bien que leur détectabilité demeure limitée.

7.2 Perspectives

Ces travaux ouvrent plusieurs perspectives de recherche. La première concerne la conception de nouveaux signaux porteurs, qui pourraient être définis à partir des caractéristiques, spectrales et temporelles, des signaux porteurs les plus efficaces (cf. section 5.5). Une seconde perspective porte sur l'influence de paramètres expérimentaux autres que le signal porteur, tels que la fréquence de modulation, le niveau sonore ou la durée de stimulation. Par ailleurs, une analyse plus approfondie des performances et des régions cérébrales impliquées, en fonction des profils individuels et de la nature des stimuli, permettrait de mieux interpréter les résultats et d'envisager, à terme, une adaptation des systèmes BCI aux utilisateurs.

7.2.1 Conception de nouveaux signaux porteurs

Les analyses réalisées sur les différents stimuli dans la section 5.5 mettent en évidence plusieurs caractéristiques acoustiques susceptibles de favoriser la détectabilité de la fréquence de modulation, parmi lesquelles un centroïde spectral élevé par rapport à la fréquence de modulation, une largeur spectrale relativement étroite et un caractère stationnaire du signal.

Plutôt que de rechercher ce profil idéal dans des sources audio réelles, des techniques de *morphing* audio peuvent être envisagées, afin de transformer des sons naturels (tels que le chant des cigales ou le ronronnement de chat) en sons "hybrides" dont les propriétés se rapprochent progressivement de celles d'un son pur [Kazazis et al., 2016]. L'objectif est de trouver un compromis optimal entre la détectabilité et le confort d'écoute. Des approches de synthèse sonore pourraient également être explorées, de manière à générer des stimuli inspirés de sons naturels, tout en contrôlant explicitement leurs paramètres pour obtenir les propriétés recherchées. Par ailleurs, afin d'améliorer la stationnarité des stimuli, il est possible d'extraire un court segment du signal, puis de le répéter, de manière à produire un motif auditif plus stable. Ces approches nécessitent cependant des travaux complémentaires pour mieux caractériser les propriétés des signaux porteurs et comprendre les facteurs influençant la détectabilité de la fréquence de modulation.

7.2.2 Influence de la fréquence de modulation, du niveau sonore et de la durée de stimulation sur les performances

La détectabilité des stimuli auditifs dans les signaux cérébraux dépend également de la fréquence de modulation utilisée. Il serait donc pertinent d'évaluer différentes fréquences, notamment dans la plage comprise entre 1 et 200 Hz [Picton et al., 2003], afin d'identifier celles offrant les meilleures performances. À terme, une personnalisation des fréquences de modulation pourrait être envisagée, en sélectionnant pour chaque individu les valeurs les plus détectables, à la manière de [Harada et al., 2022].

Concernant le niveau sonore, les résultats obtenus laissent apparaître qu'une augmentation de 6 phones n'a pas d'impact significatif sur les performances. L'exploration de niveaux de sonie plus faibles apparaît ainsi comme une perspective intéressante, par exemple en diminuant progressivement le niveau par paliers de 6 phones à partir de 60 phones, afin d'évaluer l'existence éventuelle d'un effet tout en préservant le confort des participants.

La durée des stimuli constitue également un paramètre d'intérêt. Des stimulations plus courtes pourraient en effet réduire le temps d'expérimentation et améliorer la réactivité des BCI. Une étude suggère que des durées de l'ordre de quelques dizaines à quelques centaines de millisecondes pourraient être suffisantes [Kaongoen et Jo, 2017], ce qui mérite d'être exploré plus en détail.

7.2.3 Analyse de la variabilité interindividuelle

L'analyse des données individuelles constitue aussi un axe de recherche important. Les questionnaires pré-test offrent notamment la possibilité d'évaluer les différences interindividuelles observées, en les mettant en relation avec les caractéristiques des participants (sexe, substances consommées, latéralité, etc.). En particulier, l'étude des liens entre les zones cérébrales impliquées, les performances de classification, le profil des participants et le type de stimulation pourrait apporter une meilleure compréhension des mécanismes sous-jacents aux variations observées.

Bien que tous les participants présentent une audition normale pour leur âge, ce critère n'exclut pas une diminution des capacités auditives dans les hautes fréquences, phénomène lié à la presbyacousie et dont l'ampleur varie selon l'âge (cf. section 4.4.2.2). Dans ce contexte, une analyse des réponses cérébrales en fonction de l'âge, en particulier en comparant les participants de moins de 40 ans aux plus âgés, pourrait permettre d'étudier l'influence de ces variations auditives sur la détectabilité de la fréquence de modulation. Plus généralement, les différences de sensibilité auditive entre participants mériteraient d'être prises en compte dans l'interprétation des réponses cérébrales.

La répartition équilibrée entre femmes et hommes constitue par ailleurs une opportunité d'explorer d'éventuelles différences liées au sexe, en particulier en termes de distribution spatiale des réponses cérébrales et d'asymétries interhémisphériques (cf. section 1.7.3). Une démarche similaire pourrait être envisagée pour la latéralité (cf. section 1.7.2) et la musicalité (cf. section 1.7.1). Toutefois, la faible représentation des participants gauchers et non musiciens dans l'échantillon ne permet pas, en l'état, de mener des analyses robustes. Avec des effectifs plus équilibrés, il serait notamment possible d'examiner si les participants pratiquant la musique présentent des activations renforcées, en particulier au niveau de l'aire motrice supplémentaire gauche, du gyrus temporal supérieur ou de l'insula droite. De même, la latéralité pourrait être associée à des asymétries interhémisphériques, ainsi qu'à des différences dans la perception des stimuli. L'influence des substances consommées (cf. section 1.8) mérite aussi d'être étudiée, celles-ci étant susceptibles de moduler les signaux cérébraux. Ces pistes d'analyse s'inscrivent dans un cadre plus large, appuyé par des travaux suggérant que les sons désagréables recrutent des réseaux cérébraux plus étendus, tandis que les sons naturels peuvent induire des réponses d'amplitude plus élevée [Hong et Santosa, 2016]. À terme, une meilleure compréhension de ces facteurs de variabilité pourrait permettre d'adapter les systèmes BCI, notamment en optimisant le nombre et le positionnement des électrodes en fonction des profils individuels et des caractéristiques du stimulus utilisé.

Contributions

Journaux

1. Guého L., Lefundes da Silva, H., Plapous, C., Bougrain, L., Henaff, P., et Nicol R. (2026). Alternatives to sine carrier in auditory BCI : exploring machine learning strategies for assessing modulation detectability in EEG. *IEEE Open Journal of Signal Processing.* ⟨https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=11395630⟩

Conférences

1. Guého, L., Plapous, C., Bougrain, L., Henaff, P., et Nicol, R. (2023). Conception et évaluation d'une interface cerveau-machine exploitant des stimuli auditifs spatialisés. *Journées Perception Sonore, Nov. 2023, Paris, France.* ⟨https://hal.science/hal-04287488v1⟩

2. Guého, L., Plapous, C., Nicol, R., Henaff, P., et Bougrain, L. (2024). Evaluation de stimulations sonores pour améliorer le confort utilisateur et leur détection par une ICO auditive. *Journées Cortico, Mai 2024, Nancy, France.* ⟨https://hal.science/hal-04631298v1⟩

3. Guého, L., Plapous, C., Nicol, R., Henaff, P., et Bougrain, L. (2025). Assessing detectability and pleasantness of stimuli for auditory BCIs. *Journées Cortico, Mai 2025, Lyon, France.* ⟨https://hal.science/hal-05413844v1⟩

4. Guého, L., Bougrain, L., Plapous, C., Henaff, P., et Nicol, R. (2026). Assessing Stimuli Detectability and Pleasantness for Auditory BCI. *19th International Joint Conference on Biomedical Engineering Systems and Technologies.* ⟨https://hal.science/hal-05471967v1⟩

5. Guého, L., Bougrain, L., Plapous, C., Henaff, P., et Nicol, R. (2026). An enhanced experimental paradigm for auditory BCIs by addressing both acoustic and human factors. *14th International Winter Conference on Brain-Computer Interface (BCI).* ⟨https://hal.science/hal-05471975v1⟩

6. Guého, L., Lefundes da Silva, H., Plapous, C., Henaff, P., Nicol, R., et Bougrain, L. (2026). Toward More Comfortable Auditory BCIs : Can Natural Sounds Replace Pure Tones ? *Journées Cortico, Mai 2026, Lille, France.* ⟨https://inria.hal.science/hal-05641025v1⟩

7. Lefundes da Silva, H., Guého, L., Plapous, C., Bougrain, L., Henaff, P., et Nicol, R. (2026). Benchmarking Machine Learning strategies to classify Steady-State Auditory Evoked Potentials. *10th Graz Brain-Computer Interface Conference 2026.*

Brevets

1. De Souza, N., Gautier, N., et Guého, L. (2023). Contrôle d'un système multi-robots par stimuli auditifs via l'analyse de signaux cérébraux.
2. Guého, L., Nicol, R., et Plapous, C. (2023). Interface cerveau-ordinateur pour le contrôle augmenté de la signalétique des bicyclettes.

Base de données

1. Base de données EEG acquise sur 24 canaux, auprès de 48 participants exposés à quatre stimuli auditifs modulés en amplitude par une sinusoïde de 40 Hz, ainsi qu'à une condition sans stimulation auditive. Les stimuli, d'une durée de 10 secondes, sont présentés 20 fois chacun, au cours de deux *runs*, dans un ordre pseudo-aléatoire. L'expérimentation est réalisée à deux niveaux de sonie (60 et 66 phones, présentation diotique), avec 24 participants par niveau. Publiée dans MOABB [Chevallier et al., 2024] ⟨`https://moabb.neurotechx.com/docs/generated/moabb.datasets.Lenaig2026.html`⟩.

Bibliographie

Abdi, H. et Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews : computational statistics*, 2(4) :433–459.

Acampora, G., Trinchese, P., et Vitiello, A. (2019). Applying logistic regression for classification in single-channel ssvep-based bcis. *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pages 33–38. IEEE.

Aggarwal, S. et Chugh, N. (2022). Review of machine learning techniques for eeg based brain computer interface. *Archives of Computational Methods in Engineering*, 29(5) :3001–3020.

Aguilar, J. M., Castillo, J., et Elias, D. (2015). Eeg signals processing based on fractal dimension features and classified by neural network and support vector machine in motor imagery for a bci. *VI Latin American Congress on Biomedical Engineering CLAIB 2014, Paraná, Argentina 29, 30 & 31 October 2014*, pages 615–618. Springer.

Alberti, M., Seuret, M., Pondenkandath, V., Ingold, R., et Liwicki, M. (2017). Historical document image segmentation with lda-initialized deep neural networks. *Proceedings of the 4th international workshop on historical document imaging and processing*, pages 95–100.

Allison, B. Z., McFarland, D. J., Schalk, G., Zheng, S. D., Jackson, M. M., et Wolpaw, J. R. (2008). Towards an independent brain–computer interface using steady state visual evoked potentials. *Clinical neurophysiology*, 119(2) :399–408.

Amiri, S., Fazel-Rezai, R., et Asadpour, V. (2013). A review of hybrid brain-computer interface systems. *Advances in Human-Computer Interaction*, 2013(1) :187024.

André, N. (2025). *Représentations et fonctions d'activation réelles et hyper-complexes dans les réseaux de neurones pour le traitement du signal*. PhD thesis, Université d'Avignon.

Ang, K. K., Chin, Z. Y., Wang, C., Guan, C., et Zhang, H. (2012). Filter bank common spatial pattern algorithm on bci competition iv datasets 2a and 2b. *Frontiers in neuroscience*, 6 :39.

Aras, V. P. (2003). Audiometry techniques, circuits, and systems. *M. Tech. Credit Seminar Report, Electronic Systems Group, EE Department, Bombay*.

Arlinger, S. (1979). Comparison of ascending and bracketing methods in pure tone audiometry a multi-laboratory study. *Scandinavian Audiology*, 8(4) :247–251.

Arsigny, V., Fillard, P., Pennec, X., et Ayache, N. (2005). Fast and simple calculus on tensors in the log-euclidean framework. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 115–122. Springer.

Artières, F., Bizaguet, D. B., Dauman, R., Meyer, B., Meyer-Bisch, C., Ohresser, M., Renard, X., et Collette, J. (2006). Guide des bonnes pratiques en audiométrie de l'adulte. *Rap. tech., Société Française d'Audiologie*.

Banks, D. L. et Fienberg, S. E. (2003). Data mining, statistics. Meyers, R. A., editor, *Encyclopedia of Physical Science and Technology (Third Edition)*, pages 247–261. Academic Press, New York, third edition edition.

Banoczi, W. R. (2005). How some drugs affect the electroencephalogram (eeg). *American journal of electroneurodiagnostic technology*, 45(2) :118–129.

Barachant, A., Andreev, A., et Congedo, M. (2013a). The riemannian potato : an automatic and adaptive artifact detection method for online experiments using riemannian geometry. *TOBI Workshop lV*, pages 19–20.

Barachant, A., Barthélemy, Q., vom Berg, G. W., Gramfort, A., King, J.-R., Rodrigues, P. L. C., Lopes, B. J., gcattan, Dave, Olivetti, E., Goncharenko, V., Dolle, M., toncho11, Näher, T., Yamamoto, M. S., GhilesReguig, Mellot, A., Mian, A., stonebig, Mortier, S., mhurte, mccorsi, Sassenhagen, J., jliersch, de Surrel, T., Höchenberger, R., Clisson, P., JJ, Carrara, I., et Heskebeck, F. (n.d.). pyriemann/pyriemann : v0.10. `https://doi.org/10.5281/zenodo.18160196`. Dernière consultation le 3 Avril 2026.

Barachant, A., Bonnet, S., Congedo, M., et Jutten, C. (2010a). Common spatial pattern revisited by riemannian geometry. *2010 IEEE int. workshop on multimedia signal processing*, pages 472–476. IEEE.

Barachant, A., Bonnet, S., Congedo, M., et Jutten, C. (2010b). Riemannian geometry applied to bci classification. *International conference on latent variable analysis and signal separation*, pages 629–636. Springer.

Barachant, A., Bonnet, S., Congedo, M., et Jutten, C. (2011). Multiclass brain–computer interface classification by riemannian geometry. *IEEE Transactions on Biomedical Engineering*, 59(4) :920–928.

Barachant, A., Bonnet, S., Congedo, M., et Jutten, C. (2013b). Classification of covariance matrices using a riemannian-based kernel for bci applications. *Neurocomputing*, 112 :172–178.

Barthélemy, Q., Mayaud, L., Ojeda, D., et Congedo, M. (2019). The riemannian potato field : a tool for online signal quality index of eeg. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(2) :244–255.

Beauchamp, M. S., Lee, K. E., Argall, B. D., et Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, 41(5) :809–823.

Behtani, L. (2023). La repondération sensorielle chez les adultes ayant une presbyacousie, porteurs ou non-porteurs d'appareils auditifs.

Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, Anne, Masure, M.-C., et Samson, Y. (1998). Lateralization of speech and auditory temporal processing. *Journal of cognitive neuroscience*, 10(4) :536–540.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., et Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral cortex*, 10(5) :512–528.

Bizaguet, P. É. (2011). L'hyperacousie et sa prise en charge. *Comité scientifique*.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., et Muller, K.-R. (2007). Optimizing spatial filters for robust eeg single-trial analysis. *IEEE Signal processing magazine*, 25(1) :41–56.

Bleuzé, A., Mattout, J., et Congedo, M. (2021). Transfer learning for the riemannian tangent space : applications to brain-computer interfaces. *2021 International Conference on Engineering and Emerging Technologies (ICEET)*, pages 1–6. IEEE.

Blume, W. T. (2006). Drug effects on eeg. *Journal of Clinical Neurophysiology*, 23(4) :306–311.

Boatman, D., Freeman, J., Vining, E., Pulsifer, M., Miglioretti, D., Minahan, R., Carson, B., Brandt, J., et McKhann, G. (1999). Language recovery after left hemispherectomy in children with late-onset seizures. *Annals of Neurology : Official Journal of the American Neurological Association and the Child Neurology Society*, 46(4) :579–586.

Bonetti, L., Brattico, E., Carlomagno, F., Donati, G., Cabral, J., Haumann, N. T., Deco, G., Vuust, P., et Kringelbach, M. L. (2021). Rapid encoding of musical tones discovered in whole-brain connectivity. *NeuroImage*, 245 :118735.

Bradley, M. M. et Lang, P. J. (1994). Measuring emotion : the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1) :49–59.

Briscik, M., Dillies, M.-A., et Déjean, S. (2023). Improvement of variables interpretability in kernel pca. *BMC bioinformatics*, 24(1) :282.

Brüel & Kjaer (2019). *Wideband Ear Simulator Type 4987 Family (product data)*. `https://www.bksv.com/-/media/literature/Product-Data/bp2605.ashx`.

Brüel & Kjaer (2021). *Head and Torso Simulator Types 4128-C and 4128-D Handset Positioner for HATS Type 4606*. `https://www.bksv.com/media/doc/bp0521.pdf`.

Brüel & Kjaer (2022). *Sound Calibrator Type 4231*. `https://www.bksv.com/media/doc/bp1311.pdf`.

Brüel & Kjaer (n.d.). *Wide Range Measuring Amplifiers — Types 2610 and 2636*. `https://pearl-hifi.com/06_Lit_Archive/15_Mfrs_Publications/10_Bruel_Kjaer/02_Product_Data/2610_2636_Product%20Data.pdf`.

Burle, B., Spieser, L., Roger, C., Casini, L., Hasbroucq, T., et Vidal, F. (2015). Spatial and temporal resolutions of eeg : Is it really black and white ? a scalp current density view. *International Journal of Psychophysiology*, 97(3) :210–220.

Cabestaing, F. et Derambure, P. (2016). Marqueurs physiologiques pour le contrôle de bci actives/réactives. *Les interfaces cerveau-ordinateur*, 1.

Carhart, R. et Jerger, J. F. (1959). Preferred method for clinical determination of pure-tone thresholds. *Journal of speech and hearing disorders*, 24(4) :330–345.

Carlach, J.-C., Crom, J.-M., et Javaudin, J.-P. (2017). Brain Computer Interfaces - A state of the art and perspectives for Orange.

Carlach, J.-C. et Javaudin, J.-P. (2020). Note technique "avancement du projet de recherches sur les Interfaces Cerveau-Machine".

Carrara, I. et Papadopoulo, T. (2024). Pseudo-online framework for bci evaluation : a moabb perspective using various mi and ssvep datasets. *Journal of Neural Engineering*, 21(1) :016003.

Carvalho, S. N., Costa, T. B., Uribe, L. F., Soriano, D. C., Yared, G. F., Coradine, L. C., et Attux, R. (2015). Comparative analysis of strategies for feature extraction and classification in ssvep bcis. *Biomedical Signal Processing and Control*, 21 :34–42.

Cheng, M., Gao, X., Gao, S., et Xu, D. (2002). Design and implementation of a brain-computer interface with high transfer rates. *IEEE transactions on biomedical engineering*, 49(10) :1181–1186.

Chevallier, S., Carrara, I., Aristimunha, B., Guetschel, P., Sedlar, S., Lopes, B., Velut, S., Khazem, S., et Moreau, T. (2024). The largest eeg-based bci reproducibility study for open science : the moabb benchmark. *arXiv preprint arXiv :2404.15319*.

Chevallier, S., Kalunga, E. K., Barthélemy, Q., et Monacelli, E. (2021). Review of riemannian distances and divergences, applied to ssvep-based bci. *Neuroinformatics*, 19(1) :93–106.

Chevallier, S., Kalunga, E. K., Barthélemy, Q., et Yger, F. (2018). Riemannian classification for ssvep-based bci : Offline versus online implementations. *Brain–Computer Interfaces Handbook*, pages 371–396. CRC Press.

Chollet, F. (2017). Xception : Deep learning with depthwise separable convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258.

Clerc, M., Bougrain, L., et Lotte, F. (2016). *Les interfaces cerveau-ordinateur 1 : technologie et applications*, volume 1. ISTE Group.

Corballis, M. C., Badzakova-Trajkov, G., et Häberling, I. S. (2012). Right hand, left brain : genetic and evolutionary bases of cerebral asymmetries for language and manual action. *Wiley Interdisciplinary Reviews : Cognitive Science*, 3(1) :1–17.

Cui, X., Bray, S., Bryant, D. M., Glover, G. H., et Reiss, A. L. (2011). A quantitative comparison of nirs and fmri across multiple cognitive tasks. *Neuroimage*, 54(4) :2808–2821.

Damien, B., Quentin, D., Thomas, F., et Stéphanie, H. (2023). Les réseaux de neurones appliqués à la statistique publique : méthodes et cas d'usages.

De Dreuille, O., Maszelin, P., Foehrenbach, H., Bonardel, G., et Gaillard, J. (2002). Principe et technique de la tomographie par émission de positons (tep). *Encycl Méd Chir*, 35.

Defrise, M. et Trebossen, R. (2002). La tomographie d'émission de positons. *La tomographie médicale. Paris : Hermes-Lavoisier*.

Degrave, P. et Dedonder, J. (2019). A french translation of the goldsmiths musical sophistication index, an instrument to assess self-reported musical skills, abilities and behaviours. *Journal of New Music Research*, 48(2) :138–144.

Dehais, F., Rida, I., Roy, R. N., Iversen, J., Mullen, T., et Callan, D. (2019). A pbci to predict attentional error before it happens in real flight conditions. *2019 ieee international conference on systems, man and cybernetics (smc)*, pages 4155–4160. IEEE.

Delorme, A. (2011). Eeglab.

Delorme, A. et Makeig, S. (2004). Eeglab : an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1) :9–21.

Deng, X., Zhang, B., Yu, N., Liu, K., et Sun, K. (2021). Advanced tsgl-eegnet for motor imagery eeg-based brain-computer interfaces. *IEEE access*, 9 :25118–25130.

Deutsch, D. (1983). The octave illusion in relation to handedness and familial handedness background. *Neuropsychologia*, 21(3) :289–293.

Deutsch, D., Hamaoui, K., et Henthorn, T. (2007). The glissando illusion and handedness. *Neuropsychologia*, 45(13) :2981–2988.

EASYCAP (n.d.). Cap for real-behavior-EEG, 24ch neutral layout, for smarting. `https://mbraintrain.com/wp-content/uploads/2024/01/RBE-24-STD.pdf`. Dernière consultation le 24 Avril 2026.

Edjekouane, I. (2016). *La sonie des signaux non-stationnaires et la sonie binaurale : application à la téléphonométrie*. PhD thesis, Aix-Marseille.

Edlin, J. M., Leppanen, M. L., Fain, R. J., Hackländer, R. P., Hanaver-Torrez, S. D., et Lyle, K. B. (2015). On the use (and misuse ?) of the edinburgh handedness inventory. *Brain and cognition*, 94 :44–51.

Ehlers, C. L., Wall, T. L., et Schuckit, M. A. (1989). Eeg spectral characteristics following ethanol administration in young men. *Electroencephalography and clinical neurophysiology*, 73(3) :179–187.

Erhart, N., Bannier, E., Corouge, I., Abaakil, I., et Ferré, J.-C. (2023). 15 minutes pour comprendre la spectroscopie du proche infrarouge (nirs) en imagerie cérébrale fonctionnelle. *Journal d'imagerie diagnostique et interventionnelle*, 6(2) :77–81.

Ferracuti, F., Freddi, A., Iarlori, S., Longhi, S., et Peretti, P. (2013). Auditory paradigm for a p300 bci system using spatial hearing. *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 871–876. IEEE.

Fillard, P. (2004). Un cadre riemannien pour l'imagerie du tenseur. *Rapport de stage de Master Optique-Image-Vision, Spécialité recherche Image-Vision-Signal 2ème année, Institut National de Recherche en Informatique et Automatique*.

Fillon, T. (2004). *Traitement numérique du signal acoustique pour une aide aux malentendants*. PhD thesis, Télécom ParisTech.

Fletcher, H. et Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *Bell System Technical Journal*, 12(4) :377–430.

Fox, M. D. et Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature reviews neuroscience*, 8(9) :700–711.

Galambos, R., Makeig, S., et Talmachoff, P. J. (1981). A 40-hz auditory potential recorded from the human scalp. *Proceedings of the national academy of sciences*, 78(4) :2643–2647.

Gaume, A., Vialatte, F., et Dreyfus, G. (2014). Transient brain activity explains the spectral content of steady-state visual evoked potentials. *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 688–692. IEEE.

Gaur, P., Chowdhury, A., McCreadie, K., Pachori, R. B., et Wang, H. (2021). Logistic regression with tangent space-based cross-subject learning for enhancing motor imagery classification. *IEEE Transactions on Cognitive and Developmental Systems*, 14(3) :1188–1197.

Gevensleben, H., Albrecht, B., Lütcke, H., Auer, T., Dewiputri, W. I., Schweizer, R., Moll, G., Heinrich, H., et Rothenberger, A. (2014). Neurofeedback of slow cortical potentials : neural mechanisms and feasibility of a placebo-controlled design in healthy adults. *Frontiers in human neuroscience*, 8 :990.

Gilbert, D. G., Dibb, W. D., Plath, L. C., et Hiyane, S. G. (2000). Effects of nicotine and caffeine, separately and in combination, on eeg topography, mood, heart rate, cortisol, and vigilance. *Psychophysiology*, 37(5) :583–595.

Golestani, N. (2014). Brain structural correlates of individual differences at low-to high-levels of the language processing hierarchy : A review of new approaches to imaging research. *International Journal of Bilingualism*, 18(1) :6–34.

Gosseries, O., Demertzi, A., Noirhomme, Q., Tshibanda, J., Boly, M., de Beeck, M. O., Hustinx, R., Maquet, P., Salmon, E., Moonen, G., et al. (2008). Que mesure la neuro-imagerie fonctionnelle : Irmf, tep & meg ? *Revue Médicale de Liège*, 63(5-6).

Griffiths, T., Bench, C., et Frackowiak, R. (1994). Human cortical areas selectively activated by apparent sound movement. *Current Biology*, 4(10) :892–895.

Grosse-Wentrup, M. et Buss, M. (2008). Multiclass common spatial patterns and information theoretic feature extraction. *IEEE transactions on Biomedical Engineering*, 55(8) :1991–2000.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., et al. (2018). Recent advances in convolutional neural networks. *Pattern recognition*, 77 :354–377.

Guého, L. (2022). Brain-computer interface based on auditory stimuli. *Rapport de stage Master 2 Acoustique et Musicologie, Aix-Marseille Université, Orange Labs.*

Guo, Y., Li, Y., Wang, L., et Rosing, T. (2019). Depthwise convolution is all you need for learning multiple visual domains. *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 8368–8375.

Harada, K., Matsubara, A., Nakashima, S., et Nishifuji, S. (2022). Classification of three intentions based on auditory steady state response to natural sound. *2022 IEEE 11th Global Conference on Consumer Electronics (GCCE)*, pages 335–338. IEEE.

Hashimoto, T., Usui, N., Taira, M., Nose, I., Haji, T., et Kojima, S. (2006). The neural mechanism associated with the processing of onomatopoeic sounds. *Neuroimage*, 31(4) :1762–1770.

Hastie, T., Tibshirani, R., et Friedman, J. (2008). Linear methods for classification. *The elements of statistical learning : data mining, inference, and prediction*, pages 101–137. Springer.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., et Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage*, 87 :96–110.

Heo, J., Baek, H. J., Hong, S., Chang, M. H., Lee, J. S., et Park, K. S. (2017). Music and natural sounds in an auditory steady-state response based brain–computer interface to increase user acceptance. *Computers in Biology and Medicine*, 84 :45–52.

Herrera, A. G. (2024). *Vibrotactile feedback to support kinesthetic motor imagery in a brain-computer interface for post-stroke motor rehabilitation*. PhD thesis, Université de Lorraine.

Hickok, G. et Poeppel, D. (2004). Dorsal and ventral streams : a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1-2) :67–99.

Hickok, G. et Poeppel, D. (2007). The cortical organization of speech processing. *Nature reviews neuroscience*, 8(5) :393–402.

Higashi, H., Rutkowski, T. M., Washizawa, Y., Cichocki, A., et Tanaka, T. (2011). Eeg auditory steady state responses classification for the novel bci. *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4576–4579. IEEE.

Hinterberger, T., Schmidt, S., Neumann, N., Mellinger, J., Blankertz, B., Curio, G., et Birbaumer, N. (2004). Brain-computer communication and slow cortical potentials. *IEEE Transactions on Biomedical Engineering*, 51(6) :1011–1018.

Hodgetts, S. et Hausmann, M. (2022). Sex/gender differences in brain lateralisation and connectivity. *Sex Differences in Brain Function and Dysfunction*, pages 71–99. Springer.

Hong, K.-S. et Santosa, H. (2016). Decoding four different sound-categories in the auditory cortex using functional near-infrared spectroscopy. *Hearing research*, 333 :157–166.

Hosmer Jr, D. W., Lemeshow, S., et Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons.

Huang, Z. et Van Gool, L. (2017). A riemannian network for spd matrix learning. *Proceedings of the AAAI conference on artificial intelligence*, volume 31.

Hughson, W. et Westlake, H. (1944). Manual for program outline for rehabilitation of aural casualties both military and civilian. *Trans Am Acad Ophthalmol Otolaryngol*, 48(Suppl) :1–15.

Hull, R. H. et Kerschen, S. R. (2010). The influence of cardiovascular health on peripheral and central auditory function in adults : a research review. *Clinical Focus*.

Islam, M., Chen, G., et Jin, S. (2019). An overview of neural network. *American Journal of Neural Networks and Applications*, 5(1) :7–11.

ISO (2023). Acoustics : Normal equal-loudness-level contours. *International Organization for Standardization*.

ITU (1996). Methods for subjective determination of transmission quality. *International Telecommunication Union*.

ITU (2013). Head and torso simulator for telephonometry. *Genève : ITU Telecommunication Standardization Sector*, page 58.

Izumi, S., Itoh, K., Matsuzawa, H., Takahashi, S., Kwee, I. L., et Nakada, T. (2011). Functional asymmetry in primary auditory cortex for processing musical sounds : temporal pattern analysis of fmri time series. *Neuroreport*, 22(10) :470–473.

Jakkula, V. (2006). Tutorial on support vector machine (svm). *School of EECS, Washington State University*, 37(2.5) :3.

Jeon, Y., Nam, C. S., Kim, Y.-J., et Whang, M. C. (2011). Event-related (de) synchronization (erd/ers) during motor imagery tasks : Implications for brain–computer interfaces. *International Journal of Industrial Ergonomics*, 41(5) :428–436.

Jolliffe, I. T. (2002). *Introduction*, pages 1–9. Springer New York, New York, NY.

Jouvie, C. (2013). *Estimation de la fonction d'entrée en tomographie par émission de positons dynamique : application au fluorodesoxyglucose*. PhD thesis, Université Paris Sud-Paris XI.

Kalunga, E. K., Chevallier, S., Barthélemy, Q., Djouani, K., Monacelli, E., et Hamam, Y. (2016). Online ssvep-based bci using riemannian geometry. *Neurocomputing*, 191 :55–68.

Kanagaluru, V. et Sasikala, M. (2024). Two class motor imagery eeg signal classification for bci using lda and svm. *Traitement du Signal*, 41(5) :2743.

Kaongoen, N. et Jo, S. (2017). A novel hybrid auditory bci paradigm combining assr and p300. *Journal of neuroscience methods*, 279 :44–51.

Kazazis, S., Depalle, P., et McAdams, S. (2016). Sound morphing by audio descriptors and parameter interpolation. *Proceedings of the 19th International Conference on Digital Audio Effects (DAFx-16). Brno, Czech Republic*.

Khan, R. A., Rashid, N., Shahzaib, M., Malik, U. F., Arif, A., Iqbal, J., Saleem, M., Khan, U. S., et Tiwana, M. (2023). A novel framework for classification of two-class motor imagery eeg signals using logistic regression classification algorithm. *Plos one*, 18(9) :e0276133.

Khoa, T. Q. D. et Nakagawa, M. (2008). Recognizing brain activities by functional near-infrared spectroscope signal analysis. *Nonlinear biomedical physics*, 2(1) :3.

Kim, D.-W., Cho, J.-H., Hwang, H.-J., Lim, J.-H., et Im, C.-H. (2011a). A vision-free brain-computer interface (bci) paradigm based on auditory selective attention. *2011 Int. Conf. of the IEEE Eng. in Medicine and Biology Society*, pages 3684–3687. IEEE.

Kim, D.-W., Hwang, H.-J., Lim, J.-H., Lee, Y.-H., Jung, K.-Y., et Im, C.-H. (2011b). Classification of selective attention to auditory stimuli : toward vision-free brain–computer interfacing. *Journal of neuroscience methods*, 197(1) :180–185.

Kim, H., Drake, B. L., et Park, H. (2007). Multiclass classifiers based on dimension reduction with generalized lda. *Pattern recognition*, 40(11) :2939–2945.

Kim, S., Lim, E. J., Kim, H. S., Park, J. H., Jarng, S. S., et Lee, S. H. (2010). Sex differences in a cross sectional study of age-related hearing loss in korean. *Clinical and experimental otorhinolaryngology*, 3(1) :27–31.

Koles, Z. J., Lazar, M. S., et Zhou, S. Z. (1990). Spatial patterns underlying population differences in the background eeg. *Brain topography*, 2(4) :275–284.

Krainik, A., Rubin, C., Grand, S., David, O., Baciu, M., Jaillard, A., Troprès, I., Lamalle, L., Duffau, H., Le Bas, J., et al. (2006). L'imagerie par résonance magnétique cérébrale fonctionnelle en pratique clinique. *Journal de radiologie*, 87(6) :607–617.

Krishnan, G. P., Hetrick, W. P., Brenner, C. A., Shekhar, A., Steffen, A. N., et O'Donnell, B. F. (2009). Steady state and induced auditory gamma deficits in schizophrenia. *Neuroimage*, 47(4) :1711–1719.

Kulynych, J. J., Vladar, K., Jones, D. W., et Weinberger, D. R. (1994). Gender differences in the normal lateralization of the supratemporal cortex : Mri surface-rendering morphometry of heschl's gyrus and the planum temporale. *Cerebral cortex*, 4(2) :107–118.

Kumar, S., Yger, F., et Lotte, F. (2019). Towards adaptive classification using riemannian geometry approaches in brain-computer interfaces. *2019 7th International Winter Conference on Brain-Computer Interface (BCI)*, pages 1–6. IEEE.

La Marque et La Micronutrition (n.d.). Caféine : une molécule aux propriétés psychoactives. *Santé*, 24 :50.

Lal, T. N., Schroder, M., Hinterberger, T., Weston, J., Bogdan, M., Birbaumer, N., et Scholkopf, B. (2004). Support vector channel selection in bci. *IEEE transactions on biomedical engineering*, 51(6) :1003–1010.

Lashgari, E., Liang, D., et Maoz, U. (2020). Data augmentation for deep-learning-based electroencephalography. *Journal of Neuroscience Methods*, 346 :108885.

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., et Lance, B. J. (2018). Eegnet : a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5) :056013.

Le cerveau à tous les niveaux (n.d.). Broca, wernicke et les autres aires du langage. `https://www.lecerveau.ca/flash/d/d_10/d_10_cr/d_10_cr_lan/d_10_cr_lan.html`. Dernière consultation le 3 Avril 2026.

LeCun, Y., Bengio, Y., et Hinton, G. (2015). Deep learning. *nature*, 521(7553) :436–444.

Ledoit, O. et Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2) :365–411.

Lewald, J., Riederer, K. A., Lentz, T., et Meister, I. G. (2008). Processing of sound location in human cortex. *European Journal of Neuroscience*, 27(5) :1261–1270.

Lewis, J. W., Wightman, F. L., Brefczynski, J. A., Phinney, R. E., Binder, J. R., et DeYoe, E. A. (2004). Human brain regions involved in recognizing environmental sounds. *Cerebral cortex*, 14(9) :1008–1021.

Li, G., Wang, S., et Duan, Y. Y. (2018). Towards conductive-gel-free electrodes : Understanding the wet electrode, semi-dry electrode and dry electrode-skin interface impedance using electrochemical impedance spectroscopy fitting. *Sensors and Actuators B : Chemical*, 277 :250–260.

Li, G.-L., Wu, J.-T., Xia, Y.-H., He, Q.-G., et Jin, H.-G. (2020). Review of semi-dry electrodes for eeg recording. *Journal of Neural Engineering*, 17(5) :051004.

Li, P., Xu, P., Zhang, R., Guo, L., et Yao, D. (2013). L1 norm based common spatial patterns decomposition for scalp eeg bci. *Biomed. engin. online*, 12 :1–12.

Liu, H., Du, B., Liu, B., Fu, X., et Wang, Y. (2022). Clinical comparison of two automated audiometry procedures. *Frontiers in Neuroscience*, 16 :1011016.

Lopez, M.-A., Pomares, H., Pelayo, F., Urquiza, J., et Perez, J. (2009). Evidences of cognitive effects over auditory steady-state responses by means of artificial neural networks and its use in brain–computer interfaces. *Neurocomputing*, 72(16-18) :3617–3623.

Lotte, F. (2012). Les interfaces cerveau-ordinateur. *Technique et science informatiques*, 3 :1–22.

Lotte, F. (2021). Les interfaces-cerveau ordinateur. *Médecine du Sommeil*, 18(4) :189.

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., et Yger, F. (2018). A review of classification algorithms for eeg-based brain–computer interfaces : a 10 year update. *Journal of neural engineering*, 15(3) :031005.

Lotte, F. et Guan, C. (2010). Regularizing common spatial patterns to improve bci designs : unified theory and new algorithms. *IEEE Transactions on biomedical Engineering*, 58(2) :355–362.

Lu, X. et Hu, L. (2019). Electroencephalography, evoked potentials, and event-related potentials. *EEG signal processing and feature extraction*, pages 23–42. Springer.

Lécuyer, A. et Si-Mohammed, H. (2019). Livrable lot 2 : Rapport sur les algorithmes.

Mahmoodi, M., Abadi, B. M., Khajepur, H., et Harirchian, M. H. (2017). A robust beamforming approach for early detection of readiness potential with application to brain-computer interface systems. *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2980–2983. IEEE.

Marissens Cueva, V., Bougrain, L., Lotte, F., et Rimbert, S. (2025). Reliable predictor of bci motor imagery performance using median nerve stimulation. *Journal of Neural Engineering*, 22(2) :026039.

McNamara, D. et Ziarani, A. (2004). A new adaptive technique of estimation of steady state auditory evoked potentials. *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 2, pages 4544–4547. IEEE.

Millán, J. d. R. et Carmena, J. (2010). Invasive or noninvasive : understanding brain-machine interface technology. Institute of Electrical and Electronics Engineers.

MucLeun, P. D. (1996). Women : A more balanced brain ? *Zygon®*, 31(3) :421–439.

Müllensiefen, D., Gingras, B., Stewart, L., et Musil, J. J. (2013). Goldsmiths musical sophistication index (gold-msi) v1.0 : tech. report and doc. revision 0.3.

Nambu, I., Ebisawa, M., Kogure, M., Yano, S., Hokari, H., et Wada, Y. (2013). Estimating the intended sound direction of the user : toward an auditory brain-computer interface using out-of-head sound localization. *PloS one*, 8(2) :e57174.

neuromedia (n.d.). Cerveau : anatomie et divisions. `https://www.neuromedia.ca/lorganisation-du-cerveau/`. Dernière consultation le 3 Avril 2026.

Nguyen, T., Hettiarachchi, I., Khatami, A., Gordon-Brown, L., Lim, C. P., et Nahavandi, S. (2018). Classification of multi-class bci data by common spatial pattern and fuzzy system. *IEEE access*, 6 :27873–27884.

Noreña, A., Lacher-Fougère, S., Fraysse, M., Bizaguet, E., Grevin, P., Thai-Van H, M. L., Le Pajolec, C., et Ohresser, M. (2020). Qu'est-ce que l'acouphène ? enjeux et justifications d'une définition. *Audiologie Demain*.

Odom, J. V., Bach, M., Barber, C., Brigell, M., Marmor, M. F., Tormene, A. P., Holder, G. E., et Vaegan (2004). Visual evoked potentials standard (2004). *Documenta ophthalmologica*, 108(2) :115–123.

Oldfield, R. C. (1971). The assessment and analysis of handedness : the edinburgh inventory. *Neuropsychologia*, 9(1) :97–113.

Olive, S., Welti, T., et McMullin, E. (2013). Listener preference for different headphone target response curves. *134th Convention of the Audio Engineering Society*.

O'shea, K. et Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv :1511.08458*.

Padfield, N., Camilleri, K., Camilleri, T., Fabri, S., et Bugeja, M. (2022). A comprehensive review of endogenous eeg-based bcis for dynamic device control. *Sensors*, 22(15) :5802.

Paillard, J., Hipp, J. F., et Engemann, D. A. (2024). Green : a lightweight architecture using learnable wavelets and riemannian geometry for biomarker exploration. *bioRxiv*, pages 2024–05.

Paillard, J., Hipp, J. F., et Engemann, D. A. (2025). Green : A lightweight architecture using learnable wavelets and riemannian geometry for biomarker exploration with eeg signals. *Patterns*, 6(3).

Pan, H., Gao, H., Zhang, Y., Yu, X., Li, Z., Lei, X., et Mi, W. (2025). Design and implementation of a writing-stroke motor imagery paradigm for multi-character eeg classification. *Neuroscience*.

Pan, J., Yang, F., Qiu, L., et Huang, H. (2022). Fusion of eeg-based activation, spatial, and connection patterns for fear emotion recognition. *Computational Intelligence and Neuroscience*, 2022(1) :3854513.

Parvandeh, S., Yeh, H.-W., Paulus, M. P., et McKinney, B. A. (2020). Consensus features nested cross-validation. *Bioinformatics*, 36(10) :3093–3098.

Pascual-Leone, A., Brasil-Neto, J. P., Valls-Solé, J., Cohen, L. G., et Halett, M. (1992). Simple reaction time to focal transcranial magnetic stimulation : comparison with reaction time to acoustic, visual and somatosensory stimuli. *Brain*, 115(1) :109–122.

Patterson, R. D. et Johnsrude, I. S. (2008). Functional imaging of the auditory processing applied to speech sounds. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 363(1493) :1023–1035.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn : Machine learning in python. *the Journal of machine Learning research*, 12 :2825–2830.

Peeters, G. et al. (2004). A large set of audio features for sound description (similarity and classification) in the cuidado project. *CUIDADO Ist Project Report*, 54(0) :1–25.

Pfurtscheller, G., Allison, B. Z., Bauernfeind, G., Brunner, C., Solis Escalante, T., Scherer, R., Zander, T. O., Mueller-Putz, G., Neuper, C., et Birbaumer, N. (2010). The hybrid bci. *Frontiers in neuroscience*, 4 :1283.

Pham, M., Hinterberger, T., Neumann, N., Kübler, A., Hofmayer, N., Grether, A., Wilhelm, B., Vatine, J.-J., et Birbaumer, N. (2005). An auditory brain-computer interface based on the self-regulation of slow cortical potentials. *Neurorehabilitation and Neural Repair*, 19(3) :206–218.

Picton, T. W. (1992). The p300 wave of the human event-related potential. *Journal of clinical neurophysiology*, 9(4) :456–479.

Picton, T. W., John, M. S., Dimitrijevic, A., et Purcell, D. (2003). Human auditory steady-state responses : Respuestas auditivas de estado estable en humanos. *International journal of audiology*, 42(4) :177–219.

Picton, T. W., Lins, O. G., et Scherg, M. (1995). The recording and analysis of event-related potentials. *Handbook of neuropsychology*, 10 :3–3.

Pineda, J. A., Allison, B. Z., et Vankov, A. (2000). The effects of self-movement, observation, and imagination on/spl mu/rhythms and readiness potentials (rp's) : toward a brain-computer interface (bci). *IEEE Transactions on Rehabilitation Engineering*, 8(2) :219–222.

Pinel, J. P. (2007). *Basics of biopsychology*. Pearson Allyn and Bacon Boston, MA, USA :.

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., et Burgess, P. W. (2020). The present and future use of functional near-infrared spectroscopy (fnirs) for cognitive neuroscience. *Annals of the new York Academy of Sciences*, 1464(1) :5–29.

Pouchain, D., Dupuy, C., San Jullian, M., Dumas, S., Vogel, M.-F., Hamdaoui, J., et Vergnon, L. (2007). La presbyacousie est-elle un facteur de risque de démence ? etude acoudem. *La Revue de gériatrie*, 32(6) :439–445.

Prasad, P. D., Halahalli, H. N., John, J. P., et Majumdar, K. K. (2013). Single-trial eeg classification using logistic regression based on ensemble synchronization. *IEEE journal of biomedical and health informatics*, 18(3) :1074–1080.

Punsawad, Y. et Wongsawat, Y. (2011). Enhancement of ssvep-based bci system via ssaep during eye fatigue period. *Apsipa Asc*.

Quitadamo, L., Cavrini, F., Sbernini, L., Riillo, F., Bianchi, L., Seri, S., et Saggio, G. (2017). Support vector machines to detect physiological patterns for eeg and emg-based human–computer interaction : a review. *Journal of neural engineering*, 14(1) :011001.

Renard, Y., Lotte, F., Gibert, G., Congedo, M., Maby, E., Delannoy, V., Bertrand, O., et Lécuyer, A. (2010). Openvibe : An open-source software platform to design, test, and use brain–computer interfaces in real and virtual environments. *Presence*, 19(1) :35–53.

Roy, R. et Frey, J. (2016). Marqueurs neurophysiologiques pour les interfaces cerveau-ordinateur passives. *Les interfaces cerveau-ordinateur 1*.

Saideepthi, P., Chowdhury, A., Gaur, P., et Pachori, R. B. (2023). Sliding window along with eegnet-based prediction of eeg motor imagery. *IEEE Sensors Journal*, 23(15) :17703–17713.

Sasatake, Y. et Matsushita, K. (2025). Eeg baseline analysis for effective extraction of p300 event-related potentials for welfare interfaces. *Sensors*, 25(10) :3102.

Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., et Ball, T. (2017). Deep learning with convolutional neural networks for eeg decoding and visualization. *Human brain mapping*, 38(11) :5391–5420.

Schölkopf, B., Smola, A., et Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, 10(5) :1299–1319.

Sen, D., Mishra, B. B., et Pattnaik, P. K. (2023). A review of the filtering techniques used in eeg signal processing. *2023 7th International Conference on Trends in Electronics and Informatics (ICOEI)*, pages 270–277. IEEE.

Shahbakhti, M., Rodrigues, A. S., Augustyniak, P., Broniec-Wójcik, A., Sološenko, A., Beiramvand, M., et Marozas, V. (2021). Swt-kurtosis based algorithm for elimination of electrical shift and linear trend from eeg signals. *Biomedical Signal Processing and Control*, 65 :102373.

Shelton, J. et Kumar, G. P. (2010). Comparison between auditory and visual simple reaction times. *Neuroscience and medicine*, 1(01) :30–32.

Shenoy, P., Krauledat, M., Blankertz, B., Rao, R. P., et Müller, K.-R. (2006). Towards adaptive classification for bci. *Journal of neural engineering*, 3(1) :R13–R23.

Shervine Amidi (n.d.). Pense-bête de réseaux de neurones convolutionnels. `https://stanford.edu/~shervine/l/fr/teaching/cs-230/pense-bete-reseaux-neurones-convolutionnels/`. Dernière consultation le 14 Mai 2026.

SHURE (2023). SE215 ecouteurs sound isolating. `https://www.shure.com/en-US/docs/guide/SE215`.

Si-Mohammed, H., Haumont, C., Sanchez, A., Plapous, C., Bouchnak, F., Javaudin, J.-P., et Lécuyer, A. (2022). Designing functional prototypes combining bci and ar for home automation. *International Conference on Virtual Reality and Mixed Reality*, pages 3–21. Springer.

Si-Mohammed, H., Holz, C., Wilson, A., Gamper, H., Lee, A. K., Emmanouilidou, D., Cutrell, E., et Tashev, I. (2023). On the effect of size and contrast of the ssvep visual stimuations on classification accuracy and user-friendliness in virtual reality. *2023 11th International Winter Conference on Brain-Computer Interface (BCI)*, pages 1–6. IEEE.

Siepmann, M. et Kirch, W. (2002). Effects of caffeine on topographic quantitative eeg. *Neuropsychobiology*, 45(3) :161–166.

Smith, D. J., Varghese, L. A., Stepp, C. E., et Guenther, F. H. (2014). Comparison of steady-state visual and somatosensory evoked potentials for brain-computer interface control. *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1234–1237. IEEE.

Specht, K. et Reul, J. (2003). Functional segregation of the temporal lobes into highly differentiated subsystems for auditory perception : an auditory rapid event-related fmri-task. *Neuroimage*, 20(4) :1944–1954.

Staeren, N., Renvall, H., De Martino, F., Goebel, R., et Formisano, E. (2009). Sound categories are represented as distributed patterns in the human auditory cortex. *Current biology*, 19(6) :498–502.

Stapells, D. R., Linden, D., Suffield, J. B., Hamel, G., et Picton, T. W. (1984). Human auditory steady state potentials. *Ear and hearing*, 5(2) :105–113.

Sterkers-Artières, F. et Vincent, V. (2014). Consensus formalise d'experts concernant l'audiometrie de l'adulte et de l'enfant. sforl. 2014.

Strotzer, M. (2009). One century of brain mapping using brodmann areas. *Clinical Neuroradiology*, 19(3) :179–186.

Talevi, L., Mondini, V., Mangia, A. L., Lannocca, M., et Cappello, A. (2017). A ssvep bci based on canonical correlation analysis. *GBCIC*.

Tang, L., Peng, S., Bi, Y., Shan, P., et Hu, X. (2014). A new method combining lda and pls for dimension reduction. *PloS one*, 9(5) :e96944.

Tang, Z., Sun, S., Zhang, S., Chen, Y., Li, C., et Chen, S. (2016). A brain-machine interface based on erd/ers for an upper-limb exoskeleton control. *Sensors*, 16(12) :2050.

Tayeb, S., Mahmoudi, A., Regragui, F., et Himmi, M. M. (2014). Efficient detection of p300 using kernel pca and support vector machine. *2014 second world conference on complex systems (WCCS)*, pages 17–22. IEEE.

Thanwerdas, Y. et Pennec, X. (2019). Is affine-invariance well defined on spd matrices ? a principled continuum of metrics. *International Conference on Geometric Science of Information*, pages 502–510. Springer.

Tibermacine, I. E., Russo, S., Tibermacine, A., Rabehi, A., Nail, B., Kadri, K., et Napoli, C. (2024). Riemannian geometry-based eeg approaches : A literature review. *arXiv preprint arXiv :2407.20250*.

Turner, H. L., van Etten, J., Firth, D., et Kosmidis, I. (2020). Modelling rankings in r : the plackettluce package. *Computational Statistics*, 35(3) :1027–1057.

Vaid, S., Singh, P., et Kaur, C. (2015). Eeg signal analysis for bci interface : A review. *2015 fifth international conference on advanced computing & communication technologies*, pages 143–147. IEEE.

Van Hulle, M. M. (2012). Self-organizing maps. *Handbook of natural computing*, 1 :585–622.

Vialatte, F.-B., Maurice, M., Dauwels, J., et Cichocki, A. (2010). Steady-state visually evoked potentials : focus on essential paradigms and future perspectives. *Progress in neurobiology*, 90(4) :418–438.

Viceic, D., Fornari, E., Thiran, J.-P., Maeder, P. P., Meuli, R., Adriani, M., et Clarke, S. (2006). Human auditory belt areas specialized in sound recognition : a functional magnetic resonance imaging study. *Neuroreport*, 17(16) :1659–1662.

Vidaurre, C. et Blankertz, B. (2010). Towards a cure for bci illiteracy. *Brain topography*, 23(2) :194–198.

Vishwakarma, R., Khwaja, H., Samant, V., Gaude, P., Gambhir, M., et Aswale, S. (2020). Eeg signals analysis and classification for bci systems : a review. *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, pages 1–6. IEEE.

Wainer, J. et Cawley, G. (2021). Nested cross-validation when selecting classifiers is overzealous for most practical applications. *Expert Systems with Applications*, 182 :115222.

Waldert, S. (2016). Invasive vs. non-invasive neuronal signals for brain-machine interfaces : will one prevail ? *Frontiers in neuroscience*, 10 :295.

Wang, X., Hersche, M., Tömekce, B., Kaya, B., Magno, M., et Benini, L. (2020). An accurate eegnet-based motor-imagery brain–computer interface for low-power edge computing. *2020 IEEE international symposium on medical measurements and applications (MeMeA)*, pages 1–6. IEEE.

Wang, Y., Berg, P., et Scherg, M. (1999). Common spatial subspace decomposition applied to analysis of brain responses under multiple task conditions : a simulation study. *Clinical neurophysiology*, 110(4) :604–614.

Wang, Y., Gao, S., et Gao, X. (2006). Common spatial pattern method for channel selection in motor imagery based brain-computer interface. *2005 IEEE engineering in medicine and biology 27th annual conference*, pages 5392–5395. IEEE.

Warren, J. D., Jennings, A., et Griffiths, T. D. (2005). Analysis of the spectral envelope of sounds by the human brain. *Neuroimage*, 24(4) :1052–1057.

Whilby, S., Florentine, M., Wagner, E., et Marozeau, J. (2006). Monaural and binaural loudness of 5-and 200-ms tones in normal and impaired hearing. *Journal of the Acoustical Society of America*, 119(6) :3931–3939.

Wikimedia commons (2017). International 10-20 system for eeg-mcn. `https://commons.wikimedia.org/wiki/File:International_10-20_system_for_EEG-MCN.svg`. Dernière consultation le 3 Avril 2026.

Wikipédia (2006a). Lobe (cerveau). `https://fr.wikipedia.org/wiki/Lobe_%28cerveau%29`. Dernière consultation le 3 Avril 2026.

Wikipédia (2006b). A simplified view of an artificial neural network. `https://fr.wikipedia.org/wiki/R%C3%A9seau_de_neurones_artificiels#/media/Fichier:Neural_network.svg`. Dernière consultation le 6 Mai 2026.

Wikipédia (2007). Representation du plan tangent à la sphère en un point. `https://fr.wikipedia.org/wiki/Espace_tangent#/media/Fichier:Image_Tangent-plane.svg`. Dernière consultation le 6 Mai 2026.

Wolpaw, J. R. (2013). Brain–computer interfaces. *Handbook of clinical neurology*, volume 110, pages 67–74. Elsevier.

Wu, X. et Liu, X. (1995). Study of the alpha frequency band of healthy adults in quantitative eeg. *Clinical Electroencephalography*, 26(2) :131–136.

Xie, J., Ren, Z., Liu, Y., Fang, P., Li, G., et Zhang, M. (2021). A novel synchronous hybrid steady-state brain-computer interface based on visual and auditory integration. *2021 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4. 0&IoT)*, pages 615–619. IEEE.

Yger, F., Berar, M., et Lotte, F. (2016). Riemannian approaches in brain-computer interfaces : a review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(10) :1753–1762.

Yijun, W., Ruiping, W., Xiaorong, G., et Shangkai, G. (2005). Brain-computer interface based on the high-frequency steady-state visual evoked potential. *Proceedings. 2005 First International Conference on Neural Interface and Control, 2005.*, pages 37–39. IEEE.

Yuan, M. et Cai, T. T. (2010). A reproducing kernel hilbert space approach to functional linear regression.

Yun, Y., Gu, I. Y.-H., et Aghajan, H. (2013). Riemannian manifold-based support vector machine for human activity classification in images. *2013 IEEE International Conference on Image Processing*, pages 3466–3469. IEEE.

Zatorre, R. J. et Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral cortex*, 11(10) :946–953.

Zatorre, R. J., Belin, P., et Penhune, V. B. (2002). Structure and function of auditory cortex : music and speech. *Trends in cognitive sciences*, 6(1) :37–46.

Zhang, J., Zhang, G., Li, X., Wang, P., Wang, B., et Liu, B. (2020). Decoding sound categories based on whole-brain functional connectivity patterns. *Brain imaging and behavior*, 14(1) :100–109.

Zhang, Y., Zhou, G., Jin, J., Wang, X., et Cichocki, A. (2014). Frequency recognition in ssvep-based bci using multiset canonical correlation analysis. *International journal of neural systems*, 24(04) :1450013.

Zhang, Y.-t., Zhang, Q., Zhang, J., et al. (2006). Auditory cortical responses evoked by pure tones in healthy and sensorineural hearing loss subjects : functional mri and magnetoencephalography. *Chinese Medical Journal*, 119(18) :1548–1554.

Zhivomirov, H. (2024). An empirical method for signal stationarity estimation. *The Scientific Bulletin, Series C : Electrical Engineering and Computer Science*, 86 :243–256.

Zhivomirov, H. et Nedelchev, I. (2020). A method for signal stationarity estimation. *Romanian Journal of Acoustics and Vibration*, 17(2) :149–155.

Zhou, S., Allison, B. Z., Kübler, A., Cichocki, A., Wang, X., et Jin, J. (2016). Effects of background music on objective and subjective performance measures in an auditory bci. *Frontiers in computational neuroscience*, 10 :105.

Zhu, D., Bieger, J., Garcia Molina, G., et Aarts, R. M. (2010). A survey of stimulation methods used in ssvep-based bcis. *Computational intelligence and neuroscience*, 2010(1) :702357.

Zwicker, E. (1958). Über psychologische und methodische grundlagen der lautheit. *Acta Acustica united with Acustica*, 8(4) :237–258.

Résumé

Cette thèse porte sur les interfaces cerveau-ordinateur (*Brain-Computer Interface* ou BCI), et plus précisément sur l'optimisation des stimulations sonores utilisées pour les BCI réactives basées sur les potentiels évoqués stationnaires auditifs. Les BCI réactives permettent à un utilisateur de contrôler des dispositifs en exploitant ses réponses cérébrales à des stimulations sensorielles. Dans les BCI visuelles, la fréquence de clignotement d'une cible observée est détectée dans l'activité cérébrale de l'utilisateur, permettant de déclencher l'action associée. Bien que largement utilisées, ces stimulations présentent des limites : elles ne conviennent pas à tous et nécessitent des équipements ne garantissant pas la confidentialité, ou coûteux. Les stimulations auditives constituent alors une alternative prometteuse. Elles reposent généralement sur un signal porteur (souvent sinusoïdal) modulé en amplitude par une sinusoïde dont la fréquence est identifiable dans les signaux cérébraux. Toutefois, leur usage prolongé pose des questions de confort et d'acceptabilité. Cette thèse vise donc à identifier des stimuli à la fois détectables et confortables.

Une expérimentation a été menée afin d'évaluer différents signaux porteurs modulés à 40 Hz : sinusoïde, bruit Brownien, chant des cigales et ronronnement de chat. Egalisés en sonie pour homogénéiser l'intensité perçue, indépendamment de leur contenu fréquentiel et du matériel de reproduction, ils ont été présentés à deux niveaux, afin d'analyser l'effet du niveau sonore sur la détectabilité de la fréquence de modulation.

Les analyses des signaux cérébraux montrent que la sinusoïde offre le meilleur taux de détection de la fréquence de modulation. Le chant des cigales constitue une alternative intéressante, bien que moins efficace, tandis que le bruit Brownien et le ronronnement de chat ne permettent pas une détection fiable. Le niveau sonore n'a pas d'effet significatif. Ces résultats suggèrent que certaines propriétés acoustiques favorisent la détectabilité, notamment un centroïde spectral élevé, une largeur spectrale étroite et un caractère stationnaire. La détection de la fréquence de modulation a été évaluée via le rapport signal sur bruit et dix classifieurs, incluant des approches classiques, d'apprentissage profond et riemanniennes, ces dernières étant les plus performantes. Un filtrage adapté en amont améliore les performances de classification jusqu'à 15%, tandis qu'un fenêtrage temporel de 2 secondes sans recouvrement offre un bon compromis entre nombre d'échantillons et détectabilité des classes. L'analyse des zones cérébrales activées met en évidence une implication des régions proches des oreilles, ainsi que des régions pariétales et occipitales en réponse à la sinusoïde. L'étendue de ces activations augmente avec le niveau sonore. Le profil des participants et leur perception des stimuli ont été étudiés par des questionnaires. Ainsi, bien que la sinusoïde soit la plus efficace, elle est aussi la moins appréciée et est jugée désagréable par plus de la moitié des participants. Parmi les 48 participants, la répartition hommes/femmes est équilibrée, la majorité est droitière et musicienne. Tous présentent une audition normale, évaluée à l'aide d'un test audiométrique développé spécifiquement pour l'étude.

En conclusion, cette thèse propose une démarche complète pour identifier des stimuli à la fois confortables et détectables pour les BCI. Elle combine l'évaluation de nouveaux sons porteurs, le développement d'outils d'évaluation de l'audition, l'analyse des différences individuelles et l'optimisation de la chaîne de détection.

Mots-clés: Interfaces cerveau-ordinateur, Potentiels évoqués stationnaires auditifs, Electroencéphalographie, Classification.

Abstract

This thesis focuses on Brain-Computer Interfaces (BCIs), and more specifically on improving auditory stimuli used in reactive BCIs based on Steady-State Auditory Evoked Potentials (SSAEP). Reactive BCIs enable users to control devices by exploiting their brain responses to sensory stimulation. In visual BCIs, the flickering frequency of an observed target is detected in the user's brain activity, enabling the associated action to be triggered. Although widely used, these stimuli have limitations : they are not suitable for everyone and require equipment that may not guarantee privacy or can be costly. Auditory stimuli therefore represent a promising alternative. They generally rely on a carrier signal (often sinusoidal) amplitude-modulated by a sinusoid whose frequency can be identified in brain signals. However, prolonged use raises concerns regarding comfort and acceptability. This thesis thus aims to identify stimuli that are both detectable and comfortable.

An experiment was conducted to evaluate different carrier signals modulated at 40 Hz : sinusoid, Brownian noise, cicada song and cat's purr. Loudness-equalized to ensure consistent perceived intensity regardless of spectral content and playback equipment, they were presented at two levels to analyze the effect of sound level on modulation frequency detectability.

Analysis of brain signals shows that the sinusoid provides the highest detection rate of the modulation frequency. The cicada sound is an interesting alternative, although less effective, while Brownian noise and cat's purr do not allow reliable detection. Sound level has no significant effect. These results suggest that certain acoustic properties favor detectability, particularly a high spectral centroid, narrow bandwidth, and stationary characteristics. Detection of the modulation frequency was evaluated using signal-to-noise ratio and ten classifiers, including classical, deep learning, and Riemannian approaches, the latter being the most effective. Appropriate upstream filtering improves classification performance by up to 15%, while a 2-second non-overlapping time window offers a good compromise between sample size and class detectability. Analysis of activated brain regions highlights involvement of areas near the ears, as well as parietal regions and occipital regions in response to the sinusoid. The extent of these activations increases with sound level. The participants' profiles and their perceptions of the stimuli were assessed using questionnaires. Although the sinusoid is the most effective, it is also the least appreciated and is considered unpleasant by more than half of the participants. Among the 48 participants, the gender distribution is balanced, most are right-handed and musicians, and all have normal hearing, assessed using an audiometric test specifically developed for the study.

In conclusion, this thesis proposes a comprehensive approach to identify stimuli that are both comfortable and detectable for BCIs. It combines the evaluation of new carrier sounds, the development of auditory assessment tools, the analysis of individual differences, and optimization of the detection pipeline.

Keywords: Brain-computer interfaces, Steady-state auditory evoked potentials, Electroencephalography, Classification.